CRC Press
Taylor & Francis Group

People and Rail Systems

Human Factors at the Heart of the Railway

Edited by John R. Wilson, Beverley Norris,
Theresa Clarke and Ann Mills

HUMAN FACTORS IN ROAD AND RAIL TRANSPORT

PEOPLE AND RAIL SYSTEMS

This page intentionally left blank

People and Rail Systems

Human Factors at the Heart of the Railway

Edited by

JOHN WILSON
University of Nottingham, UK

BEVERLEY NORRIS
University of Nottingham, UK

THERESA CLARKE
Network Rail, UK

and

ANN MILLS
Rail Safety and Standards Board, UK

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2007 by John R. Wilson, Beverly Norris, Theresa Clarke and Ann Mills
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper
Version Date: 20160226

International Standard Book Number-13: 978-0-7546-7184-8 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

List of Figures *xi*
List of Tables *xvii*
List of Contributors *xxi*

PART I: INTRODUCTION

1. Introduction 3
John R. Wilson, Ann Mills, Theresa Clarke and Beverley Norris

2. Real Prediction of Real Performance 9
Neville Moray

PART II: PASSENGERS AND PUBLIC

3. Wayfinding, Accessibility, Inclusive Design and Passenger Information Systems: 'Euston Station and Beyond' 25
Matt Pattison, John R. Wilson and Theresa Clarke

4. Improving the Content and Placement of Anti-Trespass Signs 37
Nicola Forsdike, Claire Turner, Fiona Bellerby, Sue Nelson and Paul McGuire

5. Travel Information vs Commercial Signboards: The Battle between Travel and Shopping 49
Answina D.P. Zwaagstra

PART III: DRIVER PERFORMANCE AND WORKLOAD

6. Assessing the Impact of Increased Numbers of CCTV Images on Driver Only Operation of a Train 63
M. Wood, M. Freer, E. Grimes and N. Brook-Carter

7. Understanding Driver Route Knowledge 71
R. Luther, H. Livingstone, T. Gipson and E. Grimes

8. Robot Trains: Results of Focus Groups with Remote Control Operators in the United States and Canada 79
Stephen Reinach

PART IV: DRIVING AND COGNITION

9. Cognitive Workload of Train Drivers 91
Ilse Gillis

10. Train Drivers and Fatal Accidents on the Rails: Psychological Aspects and Safety 103
Valdimar Briem, Sonia de Lima and Camilla Siotis

11. The Cognitive Tasks of the Driver: The Approach and Passage Through Diverging Junctions 115
Amanda C. Elliott, Sarah D. Garner and Elaine Grimes

12. A Train Driving Simulator Experiment to Investigate Driver Fault Diagnosis 125
W.H. Gibson, M.W. Halliday, L. Sutton, J. Shelton, and M. Bond

PART V: TRAIN CAB AND INTERFACES: SIMULATION AND DESIGN

13. An Ergonomics Methodology for Retrofitting Equipment in Train Cabs 137
Katie Buchanan, Gary Davis and Richard Roels

14. The Waterfall Train Accident: Implications and Lessons Learnt 147
Andrew S. McIntosh and Graham Edkins

15. Determining User Requirements for a Human Factors Research Train Driver Simulator 155
Thomas K. Yates, Sarah C. Sharples, Ged Morrisroe and Theresa Clarke

16. A Facility for Testing ERTMS/ETCS Conformity and Human Factors 167
Klaus P. Jaschke, Katrin Hartwig, Michael Meyer zu Hörste and Karsten Lemmer

17. Cardboard to Computers: An Evolution of Design Visualisation 175
Karen Wright

PART VI: ROUTES, SIGNAGE, SIGNALS AND DRIVABILITY

18. Early Route Drivability Assessment in Support of Railway Investment 185
W. Ian Hamilton, Emma Lowe and Charlotte Hill

19. Development of a Route Assessment Checklist Tool for Train Operators 197
Mark Newman, Claire Turner, Ann Mills and John Peters

20. Informing Signage Positioning Rules Through a Human Factors Assessment of Signal/Sign Co-location 203
Charlotte Hill, Harry Blanchard and Mike Carey

PART VII: SIGNALLING AND CONTROL OF THE RAILWAY

21. Mental Workload Assessment and the Development of the Operational Demand Evaluation Checklist (ODEC) for Signallers 215
Laura Pickup and John R. Wilson

22. Supervision in Signal Boxes 225
Stuart Dickinson and Emma Lowe

23. Prediction of Signaller Workload 237
Toby Garner, Mark Newman, Chris Lowe and W. Ian Hamilton

24. Collecting Human Factors Attitudes and Opinions from Signallers: Development and Use of REQUEST (The Railway Ergonomics Questionnaire) 247
Brendan Ryan, John R. Wilson, Sarah Sharples and Eleanor Marshall

25. Ergonomics Assessment of Lever Operation in Mechanical Signalling 257
R.J. Muffett

26. From the Horse's Mouth: The Contribution of Subject Matter Experts (SMEs) to Study of Rail Work Systems 267
Gemma Cox, Trudi Farrington-Darby and Richard Bye

27. Work Analysis and Distributed Cognition Representation of Integrated Rail Operations 275
R. Bye, T. Farrington-Darby, G. Cox, G.R.J. Hockey, J.R. Wilson and T. Clarke

PART VIII: PLANNING FOR THE RAILWAY

28. Planning, Reasoning and Patterns of Inferences: An Empirical Study into the Reasoning of Staff Planners in The Netherlands Railways 287
René Jorna and Derk Jan Kiewiet

29. Implementation of a Test System for Evaluation of New Concepts in Rail Traffic Planning and Control 301
Bengt Sandblad, Arne W. Andersson, Arvid Kauppi and Johan Wikström

30. Task Oriented Support for Train Shunting Planning 309
Wout Van Wezel

PART IX: ENGINEERING WORK AND MAINTENANCE

31. Understanding the Underlying Causes of Procedures Violations and Developing Effective Preventative Strategies 321
David Embrey

32. Human Factors in the Management of Engineering Possessions: The Roles of the Engineering Supervisor and PICOP 331
Brendan Ryan, John R. Wilson, Alex Schock, Emma Lowe and Fiona Kenvyn

33. Human Factors Integration for the Computerised Track Access Control System 343
Chris Lowe, Dan Lock, Barnaby Annan, Paul Thompson and Paul Raistrick

PART X: LEVEL CROSSINGS

34. The Impact of Aggressive Driving on the Design of Level Crossing Safety Measures 355
Andrej Godec and Zdravko Toš

35. Human Factors Issues at Level Crossings: A Reference Tool for Inspectors 363
Hayley Dixon, Andy Baker and Claire Dickinson

36. Human Factors in the Upgrading of Railway Control Equipment 371
John Wood, Mark Brunt, Claire Fix, Andy Harding and Theresa Clarke

37. The Effect of a Level Crossing Upgrade on Signaller Workload and Staffing Requirements: A Case Study 379
Amanda J. Widdowson

PART XI: ACCIDENTS AND SAFETY

38. Maximising Information: An Effective Interview Technique for Investigators in the Rail Industry 387
Emma Lowe and Claire Turner

39. Changing Safety Critical Communications Behaviour 399
Emma Lowe and Peter Nock

40. Safety Climate in Australian Railways 409
A. Ian Glendon and Bronwyn Evans

41. The Waterfall Train Accident: The Critical Role of Human Factors 419
Andrew S. McIntosh and Graham Edkins

42. The Role of Communication Errors in Railway Incident Causation 427
Dr Paul Shanahan, Dik Gregory, Mike Shannon and Huw Gibson

PART XII: HUMAN ERROR AND HUMAN RELIABILITY

43. Assessing the Potential for Human Error in Level 4 Rail Vehicle Maintenance 439
Steve Mason

44. Generic Human Reliability Assessment for Railways: Results 477
Ned Hickling, Liz Gaskell and Theresa Clarke

45. The Development and Application of a Rail Human Reliability Assessment Tool 455
Jenny Gilroy and Elaine Grimes

46. Improving the Predictive Power of an Existing Human Reliability Assessment Method 465
J.L. Bell, J.C. Williams, C. Daniels and N.D. Warren

PART XIII: SPADS – SIGNALS PASSED AT DANGER

47. The Impact of Returning from Rest Days on SPAD Incidents 475
W. Huw Gibson, John Shelton and Ann Mills

48. Driver Assessment of the Effectiveness of Spad Mitigation Measures 483
Tidi Wisawayodhin, Suzanne Heape, Guangyan Li, Claire Turner, Ann Mills and John Peters

49. Predicting the Causes of Spad Incidents 491
Linda Wright, Mark Dabekaussen, Tjerk van der Schaaf

PART XIV: HUMAN FACTORS INTEGRATION AND STANDARDS

50. Integrating Ergonomics into Engineering and Engineering into Ergonomics 501
Mike Carey

51. Focussed and Effective Human Factors Integration Programmes for Railways 511
Martin Dooley, Ned Hickling and Paul Stephens

52. Experience of Rail Human Factors in Hong Kong 517
Karen Priestley

PART XV: IMPAIRMENTS TO PERFORMANCE

53. The Prevalence of Obstructive Sleep Apnoea Syndrome (OSAS) in Train Drivers 529
Melissa Hack, Elaine Grimes, Toyin Davies and John Shelton

54. Fatigue Management on the Operational Railway: Developing a Strategic Approach 535
Fiona Kenvyn

55. Alcohol and Drugs as Risk Factors in Railway Traffic 545
O. Ervasti, L. Pitkämäki, H.K. Otterstad, M. Sternudd and J. Worm-Petersen

56. Research on Fatigue and Mental Workload of Railway Drivers and Traffic Controllers 553
Anabela Simoes, José Carvalhais, Pedro Ferreira, Júlia Correia and Miguel Lourenco

57. Fatigue and Shift Work in UK Train Drivers 565
Barbara Stone, Alison McGuffog, Mick Spencer, Claire Turner and Ann Mills

PART XVI: COMPETENCIES, TRAINING AND PROCEDURES

58. Educational Activities in Relation to Affective Commitment, Voluntary Turnover, Perceived Inner Motivation and Work Effort 577
Cathrine Brustad, Anna Gunningberg and Marie Svensli

59. Management Toolkits: Solutions for Rule Compliance 583
Jonathan Berman, Peter Ackroyd, Ann Mills and Toyin Davies

60. Rules Management: How to Achieve Flexible Guidance for Safe Railway Operations 593
Gudela Grote

Index *601*

List of Figures

2.1	The general organization of human factors and ergonomics	10
2.2	An example of prediction of the ability of Optimal Control Theory to predict human behaviour	13
2.3	Cumulative probability of response related to log (time) since incident	16
2.4	A hypothetical case of a train approaching an obscured sign	17
3.1	Decision tree for wayfinding	29
3.2	Main Station sites: Watford Junction (a), Paddington (b) and Euston (c)	30
3.3	Where to get information? – Varied information points across the stations	31
3.4	Car stop signage can be confused with platform numbering	32
4.1	Anti-trespass signs	40
4.2	Anti-trespass signage from outside the UK	41
4.3	Existing UK anti-trespass signs	42
4.4	New text-only signs	42
4.5	New colours and shapes	42
4.6	New graphics	43
4.7	New graphics and text	43
4.8	Existing UK posters and youth campaigns	44
5.1	A big station building the historic Amsterdam Central Station	50
5.2	The smaller Hengelo Station	50
5.3	Travel information and commercial information are in the same line of sight	52
5.4	Travel information is square to the passenger flow, advertising is parallel to the passenger flow	52
5.5	Confusion: The bookstore advertises right above ticket vending machines	53
5.6	Travel information and advertising have the same orientation with the advertisement dwarfing travel information	53
5.7	Illustration of domains in the area ‘Platform and tunnels’	54
6.1	Arrangement for DOO(P) CCTV on Rolling Stock A) arrangement in cab B) viewing cone for externally mounted camera	63
6.2	Range of ‘Targets’ used during recording of test material: From top and clockwise ‘trapped waving’, trapped not moving’, ‘child’, ‘grey board’, ‘buggy’	65
6.3	Layout of DOO(P) images – left to right – 4, 6, 8 and 12 carriages	65
6.4	Experimental set-up for DOO(P) trials	67
6.5	Experimental protocol for DOO(P) trials	68
7.1	Model of Situation Awareness in Dynamic Decision Making (Endsley, 1995)	72
7.2	Train Driver Route Knowledge Model	73
7.3	Factors contained within Train Driver Route Knowledge Model components	74
7.4	Screen shot of relational database	74

8.1 OCU top (left) and front (right) views (Courtesy of Cattron-Theimeg, Inc. 2004. Reprinted with permission) 79
8.2 RCO coupling rail cars 80

9.1 Overview of the various phases of the investigation 93
9.2 Timeline and total mental load – extract from the application to an actual situation 97
9.3 Timeline and task interference – Extract from application to an actual situation 98

10.1 Associations between variables and test scores 107

11.1 The high-level tasks of approaching and driving through junctions, and their task relationships to each other using the demarcations within Endsley's model of situational awareness (SA) 117
11.2 A detailed activity-flow analysis for the identification tasks of the train driver 118
11.3 A detailed activity-flow analysis for the prediction tasks of the train driver 120

12.1 Interaction plot for interface type for the different brake demands 129
12.2 Average workload rating by variant 131

13.1 The three possible survey methods 139
13.2 An anthropometric rig in use 140
13.3 A group of users 141
13.4 The use of computer-aided drawings and software human manikins 143
13.5 An example of photographs taken of possible fitment locations for a mock-up of GSM-R CDU and handset 145

14.1 Comparison of measured range of foot forces (with shoes) to activation and deactivation forces on Melbourne trains. 149

15.1 Three level node structure of raw interview data. 159

16.1 RailSiTe® concept 168
16.2 Basic simulation concept of the RailSiTe® 170

17.1 Example light 176
17.2 Full scale mock-up 179
17.3 3D CAD mock-up 180

18.1 GRIP process project lifecycle 185
18.2 Engineering lifecycle with ergonomics milestones and output 186
18.3 A task-group anchored to an event 188
18.4 Speed profiles 191
18.5 Workload profiles for demand values 192
18.6 The basic RDT screen 193
18.7 Inputs from the RDT to the GRIP stages. 194

19.1 Guidance example 199
19.2 Overview of the process for scoring questions 200

19.3 Checklist scoring example 201

20.1 Co-location scenario used as worked example of preliminary modelling 204
20.2 Sign positions included in the ATLAS modelling 207
20.3 Alternative strategies for rescheduling tasks 209
20.4 Example of a co-location scenario using the Excel modelling method 210

21.1 Mental workload Framework 216

22.1 Overview of the approach 226
22.2 Structure of the signal box assessment proforma 233

23.1 Extract from the Signaller Hierarchical Task Analysis 240
23.2 Comparison of actual and predicted workload, ATLAS (T2) model 242
23.3 Predicted vs. actual workload, high load period 243

24.1 Mean ratings of job satisfaction by route 251
24.2 Percentage of signallers rating communications with other groups as "good", by general manager areas 252
24.3 Relationship between mean ratings for general and safety culture and mean ratings for job satisfaction 254
24.4 Factors affecting ratings of culture, identified through multiple regression analysis 255

25.1 Lever boxes required heavy work many years ago and still do so today 258
25.2 Anthropometric measurements (not to scale) 261
25.3 Melton station with British Rail Standard Midland frame 262
25.4 Maximum force of pulling (left) and pushing (right) for a man, feet 300 mm apart 262

27.1 Levels of domain decomposition 277
27.2 Further descriptions within levels of the domain decomposition 279

28.1 Model suite in the CommonKADS methodology 289
28.2 Example of a part of the inference structure for planning 290
28.3 Assignment 2, new task to be inserted into three shifts 293
28.4 Percentage of planners that uses inferences 296
28.5 Percentage of planners that uses the sequence of inferences 298

30.1 Scheduling system architecture 310
30.2 Example of a train shunting plan 311
30.3 Shunting plan 316
30.4 Graphical overview of the shunting yard 316
30.5 Routing alternatives 316

31.1 Reasons for non-usage of procedures 324
31.2 Stage 1 of CARMAN 327
31.3 Stage 2 of CARMAN 328
31.4 Decision aid for choosing level of Job Aid Support 328

32.1 Part of the hierarchical task analysis and a relevant extract from the tabular task analysis for the work of the Engineering Supervisor 333
32.2 Timeline showing communication to and from the ES 334
32.3 Activity analysis of the PICOP 339

33.1 TAC department roles, equipment and interactions 343
33.2 Top level goals of Track Access Control operator 344
33.3 Alignment of HFI work packages with project lifecycle 347
33.4 Human Factors Integration Manager Issues Log 348

34.1 The level crossing (LX) problem in the system approach 356
34.2 A diagram of risk indicator calculation 357
34.3 An example of comparison of three trial scenarios 358
34.4 An example of the calculation of queues 360

35.1 Drop-off locations 364
35.2 Queues over the crossing 364
35.3 A database entry with all of the data fields 365

36.1 Targets used for trials 372
36.2 Mobile Camera Van 373
36.3 Locations used on crossings for test targets 374
36.4 Experimental protocol used during trials 374
36.5 Targets missed v display type and image 375
36.6 Reaction time v display type and image 375
36.7 View of 'Maxey' level crossing 376

37.1 Panel 4 (Left) and 5 (Right) 380
37.2 Timeline histogram (23.01-00.59) 381
37.3 ODEC scores 383

38.1 Incident memory and interview process (adapted from Milne and Bull, 2003) 390

39.1 Overview of the Safety Critical Communications Competence Model (from Competence Assurance Solutions) 400
39.2 Burke-Litwin Model of Organisational Performance and Change (as cited in French and Bell, 1999) 404

41.1 Schematic diagram of the deadman system 421
41.2 Estimate of proportion of SRA Tangara drivers able to hold the deadman's pedal in the activated position while incapacitated 422

42.1 The process for analysing investigation reports 428
42.2 A layered model of the communication domain 432

45.1 Representation of integrated use of HEI and HEQ tools 458

46.1 Probability of failure data taken from each relevant study 467

46.2 Correlation between observed and predicted experimental results 469

47.1 SPAD rates for different rest break durations 479

48.1 Examples of before (a) and after (b) picture cards illustrating 'backplate enlargement' mitigation measure 486
48.2 Examples of the implementation of banner repeaters 487
48.3 An example of a mitigation measure rating card 487

49.1 The PRISMA incident causation model (van der Schaaf, 1992) 491
49.2 The main components of the SPAD Bow Tie Model (IVW 2005) 492
49.3 The sub-categories contained in the main category "braking operations". 493
49.4 Percentage of causal factors per category 495

50.1 Network Rail engineering executive team 503

52.1 Representation of human factors input to projects 520

57.1 Mental tiredness related to shift type 567
57.2 Mental tiredness related to shift length 567
57.3 Relative SPAD risk related to time on task 568
57.4 SPAD risk over consecutive days 569

58.1 An illustration of the hypotheses 576

59.1 Compliance solutions 584

60.1 Two approaches to managing uncertainties in organizations (from Grote, 2004a) 592

This page intentionally left blank

List of Tables

2.1a	Transition probabilities of a radar controller's eye movements.	14
2.1b	Mean First Passage Times, MFPT, derived from Table 1a	14
2.1c	Mean fixation times derived from Table 1a. Times in seconds	14
2.1d	Proportion of time spent looking at each locus derived from Table 1a	14
2.2	Means and sigmas of FPT for looking at F repeatedly. P shows the probability of a FPT of each value.	15
2.3	Categories needed to classify drivers' eye movements. From Groeger et al. (2004)	18
2.4	Relative efficiency of colour vision with visual eccentricity	19
2.5	Probability that at least one fixation will be on the signal as a function of the number of fixations	21
3.1	Best practice appraisal	33
4.1	Profile of a young trespasser	39
4.2	Focus groups	45
4.3	Focus group structure	46
6.1	Target detection reliability scores excluding 'grey board' and 'unaware' targets for different numbers of images	68
6.2	Response times for different car sets and numbers of images	69
8.1	Participating RCO railroad experience	82
10.1	Number of accidents on the Swedish railways, including suicide or attempted suicide, from 1997 to 2003	103
10.2	Descriptive data for the train drivers	106
12.1	Red brake light illumination state for the design variants following acknowledgement	126
12.2	Confusion Matrix for variant 2 incorrect diagnoses	130
12.3	Confusion Matrix for variant 2 incorrect diagnoses	130
13.1	The advantages and disadvantages of using an anthropometric rig	141
13.2	The advantages and disadvantages of using user participants	142
13.3	The advantages and disadvantages of using software manikins	143
14.1	The foot pedal force was measured with participants wearing their normal shoes.	148
15.1	Summary of simulator requirements (numbers in parentheses indicate frequency of comments)	162
15.2	Summary of simulator requirements (cont'd)	163
18.1	VACP Description of the "AWS Horn Response" Behavioural Building Block	190

18.2	Comparing predicted (fast, slow and median) and observed (min, max and mean) timings	192
19.1	Route assessment checklist categories	198
20.1	Performance times for CTAs used to model driver strategy	206
20.2	Unacceptable behavioural conflicts relating to co-location	208
20.3	Examples of minimum distances for speed sign location relative to signal	211
21.1	The workload toolkit for field based workload assessments of the railway signaller	217
21.2	Constructs most associated to each group of elements	220
22.1	Checklist of roles and activities the supervisor may perform in each signal box visited	228
22.2	Perception of the supervisory role	228
23.1	Correlation between D-VACP dimensions and IWS, high workload period	243
24.1	Main scales from the REQUEST questionnaire and other key question areas which are used in comparisons in the current report	248
24.2	Summary of the analyses from the REQUEST survey	251
24.3	Differences in ratings between respondents operating different system types	253
24.4	Factors with multiple relationships with other factors	253
27.1	Questions reflecting the information sought at each level of decomposition	278
27.2	Methods used for data capture	280
28.1	Solution times	294
28.2	Solution choice	294
28.3	Frequency of planners that uses inferences	295
28.4	Frequency of planners that uses the sequence of inferences	297
32.1	Themes on which information has been collected in fieldwork, interviews and review of documentation	335
33.1	High-Level CTAC Human Factors Goals	346
40.1	Respondents' occupational group	411
40.2	Mean scores for seven occupational groups on two safety climate factors	412
40.3	Mean scores for seven occupational groups on four safety climate factors	413
42.1	Percentage of incidents in which communication problems played a causal role	429
47.1	SPAD categories and the revised classification	476
47.2	SPADs and preceding days work/not worked	477
47.3(a)	Days working and not working	478
47.3(b)	SPAD rates following a rest day and following a day worked	479

48.1 List of testable mitigation measures 485
48.2 Comparison of the mitigation measures rated effective by the drivers and those suggested by the application of the framework 489

49.1 Number of incidents and mean number of causal factors 494
49.2 Results of chi-square for proportions 495

52.1 Major projects on MTRCL and KCRC since 1992 517
52.2 Summary of system characteristics for Hong Kong Railways 517
52.3 Contractual and cultural influences on the success of human factors implementation (based on Courtney, 1996) 519

53.1 Prevalence of smoking and medical conditions in the study population 533

55.1 Results of alcohol and drug testing in the transport industry in Stockholm 1998–2004 547
55.2 Different level of risks in road-traffic and driving 548

56.1 Categories of measures for intervention 556
56.2 Mental workload indexes obtained with DALI 558
56.3 Mental workload indexes obtained with DALI 561

60.1 Results of rules analysis 596

This page intentionally left blank

Contributors

Peter Ackroyd	Greenstreet Berman Limited, UK
Arne W. Andersson	Dept of Information Technology, Uppsala University, Sweden
Barnaby Annan	Human Engineering Ltd, Bristol, UK
Andrew Baker	Davis Associates Ltd, UKJ
Hayley Dixon	Davis Associates Ltd, UK
Julie Bell	Health & Safety Laboratory, UK
Fiona Bellerby	Rail Safety and Standards Board, UK
Jonathan Berman	Greenstreet Berman Limited, UK
Harry Blanchard	Human Engineering Ltd, London, UK
Mike Bond	Thales Ltd, UK
Valdimar Briem	Kristianstad University, Sweden
N. Brook-Carter	Rail Safety and Standards Board, UK
Mark Brunt	CCD Design and Ergonomics Ltd, UK
Cathrine Brustad	Jernbaneverket, Norway
Katie Buchanan	Davis Associates Ltd, UK
Richard Bye	University of Sheffield, UK
Mike Carey	Network Rail, UK
José Carvalhais	Technical University of Lisbon, Portugal
Theresa Clarke	Network Rail, UK
Júlia Correia	Technical University of Lisbon, Portugal
Gemma Cox	Human Engineering Ltd, UK
Mark Dabekaussen	Tebodin Consultants and Engineers, The Netherlands
Christine Daniels	Health and Safety Laboratory, UK
Toyin Davies	Rail Safety and Standards Board, UK
Gary Davis	Davis Associates Ltd, UK
Sonia de Lima	Kristianstad University, Sweden
Claire Dickinson	HMRI, Health and Safety Executive, UK
Stuart Dickinson	ERM Risk, UK
Martin Dooley	Vectra Group Ltd, UK
Graham Edkins	Public Transport Safety Victoria, Dept of Infrastructure, Victoria, Australia

Amanda C. Elliott Lloyd's Register Rail Ltd, UK
David Embrey Human Reliability Associates, UK
O. Ervasti Finish Railways, Finland
Bronwyn Evans Griffith University, Australia
Trudi Farrington-Darby Network Rail, UK
Pedro Ferreira Technical University of Lisbon, Portugal
Claire Fix CCD Design and Ergonomics Ltd, UK
Nicola Forsdike Halcrow Group, UK
M. Freer CCD Design and Ergonomics Ltd, UK
Sarah Garner Lloyd's Register Rail (Asia) Ltd, UK
Toby Garner Human Engineering Ltd, Bristol, UK
Liz Gaskell Vectra Group Ltd, UK
Huw Gibson University of Birmingham, UK
Ilse Gillis Corporate Prevention Services, Belgian Railways, Belgium
Jenny Gilroy Atkins Rail, UK
T. Gipson Air Affairs (UK) Ltd, UK
Ian Glendon Griffith University, Australia
Andrej Godec SITEP, Ljubljana, Slovenia
Dik Gregory Gregory Harland Ltd, UK
Elaine Grimes Rail Safety & Standards Board, UK
Gudela Grote Swiss Federal Institute of Technology, Switzerland
Anna Gunningberg Jernbaneverket, Norway
Melissa Hack The Sleep Centre Ltd, UK
Mark Halliday M W Halliday Associates Ltd, UK
W Ian Hamilton Human Engineering Ltd, Bristol, UK
Andy Harding Southern Rail, UK
Katrin Hartwig German Aerospace Centre, Braunschweig, Germany
Suzanne Heape Human Engineering, Bristol, UK
Ned Hickling Vectra Group Ltd, UK
Charlotte Hill Human Engineering Ltd, UK
G R J Hockey University of Sheffield, UK
Klaus P Jaschke German Aerospace Centre, Braunschweig, Germany
Rene Jorna University of Groningen, The Netherlands
Arvid Kauppi Dept of Information Technology, Uppsala University, Sweden
Fiona Kenvyn Network Rail, UK
Derk Jan Kiewiet University of Groningen

Karsten Lemmer	German Aerospace Centre, Braunschweig, Germany
Guangyan Li	Human Engineering Ltd, Bristol, UK
Harriett Livingstone	Air Affairs (UK) Ltd, UK
Dan Lock	Human Engineering Ltd, Bristol, UK
Miguel Lourenço	Technical University of Lisbon, Portugal
Chris Lowe	Human Engineering Ltd, Bristol. UK
Emma Lowe	Network Rail, UK
Rebecca Luther	Air Affairs (UK) Ltd, UK
Alison McGuffog	QinetiQ Ltd , UK
Paul McGuire	Rail Safety and Standards Board, UK
Andrew McIntosh	School of Safety Science, UNSW, Australia
Stephen Mason	Health, Safety & Engineering Consultants Ltd, UK
Eleanor Marshall	University of Nottingham, UK
Michael Meyer zu Hörste	German Aerospace Centre, Braunschweig, Germany
Ann Mills	Rail Safety and Standards Board, UK
Neville Moray	University of Surrey, UK
Ged Morrisroe	Network Rail, UK
Bob Muffett	Network Rail, UK
Sue Nelson	Rail Industry Community Safety Support Unit, UK
Mark Newman	Human Engineering Ltd, Bristol, UK
Peter Nock	Network Rail, UK
Beverley Norris	University of Nottingham, UK
Hans Knut Otterstad	Norwegian Rail Agency, Norway
John Peters	Rail Safety and Standards Board, UK
Matt Pattison	Centre for Rail Human Factors, University of Nottingham, UK
Laura Pickup	University of Nottingham, UK
L Pitkämäki	Finish Railways, Finland
Karen Priestley	Lloyd's Register Rail (Asia) Ltd, Hong Kong
Paul Raistrick	Data Systems and Solutions, UK
Stephen Reinach	Foster-Miller Inc, Waltham, MA. USA
Richard Roels	Rail Safety and Standards Board, UK
Brendan Ryan	University of Nottingham, UK
Bengt Sandblad	Dept of Information Technology, Uppsala University, Sweden
Alex Schock	University of Nottingham, UK
Paul Shanahan	Gregory Harland Ltd, UK
M Shannon	Detica Ltd, UK

Sarah Sharples	University of Nottingham, UK
John Shelton	Rail Safety and Standards Board, UK
Camilla Siotis	Kristianstad University, Sweden
Anabela Simões	Technical University of Lisbon, Portugal
Mick Spencer	QinetiQ Ltd, UK
Paul Stephens	Vectra Group Ltd, UK
M Sternudd	National Rail Authority, Sweden
Barbara Stone	QinetiQ Ltd, UK
Laura Sutton	Rail Safety and Standards Board, UK
Marie Svensli	JBV, Norwegian Rail, Norway
Paul Thompson	Northern Rail, UK
Zdravko Tos	Faculty of Traffic and Transport Safety, Zagreb, Croatia
Claire Turner	Human Engineering Limited, UK
Tjerk van der Schaaf	Eindhoven University of Technology, The Netherlands
Wout Van Wezel	University of Groningen, The Netherlands
Nicholas Warren	Health and Safety Laboratory, UK
Johan Wikström	Dept of Information Technology, Uppsala University, Sweden
Amanda Widdowson	Vectra Group Ltd, UK
Jeremy Williams	Health & Safety Laboratory, UK
John R. Wilson	University of Nottingham , UK
Tidi Wisawayodhin	Human Engineering Ltd, UK
J. Worm-Petersen	National Rail Authority, Denmark
Karen Wright	Mott MacDonald, UK
Linda Wright	Prorail, The Netherlands
John Wood	CCD Design and Ergonomics Ltd., UK
Thomas Yates	University of Nottingham, UK
Answina Zwaagstra	Intergo, Utrecht, The Netherlands

PART I
Introduction

This page intentionally left blank

Chapter 1

Introduction

John R. Wilson, Ann Mills, Theresa Clarke and Beverley Norris

It seems only a handful of years since those of us working in rail human factors were almost apologetic about its 'little brother' status as compared to the well researched and published areas such as aviation or nuclear human factors. We perhaps felt too much the need to apologise for the barren years where little research was carried out apart from amongst a few brave souls in places such as The Netherlands and for the consequent dearth of specifically rail human factors oriented tools, methods and theory-led developments. This of course is not to say that we felt that we had to reinvent the wheel just to be specific to rail – we have always known that there was much that we could adapt from other fields and domains which was quite serviceable within the rail domain. However, we certainly felt that the real strength of activity in this field would only be evident when it began to have its own literature reporting on its own original investigations and method and theory development.

Now in just a few short years we have held the first European Rail Human Factors Conference, published the first full book in the area (Wilson et al. 2005), and published two full special issues of highly reputable international scientific journals (*Applied Ergonomics*, 36, 6, 2005 and Cognition Technology and Work, 8, 1, 2006). These publications describe rail human factors work carried out between, approximately, 2001 and 2004, and represent well the breadth and depth of the field in that period and the basis on which we are now going forward. A good, if somewhat self-referential, way of summarising current rail human factors work is to reflect on the activities of our own groups, the Rail Safety and Standards Board (RSSB) Human Factors Group, Network Rail's Ergonomics National Specialist Team (NST), and the University of Nottingham Centre for Rail Human Factors (CRHF) within the Institute for Occupational Ergonomics (IOE). These three groups are, in combination, perhaps typical of, and we like to think to some extent lead, rail human factors around the world in terms of their priorities, concerns and approach. The three groups and their parent bodies reflect and represent different business and operational needs and support, and have somewhat different (if overlapping and complementary) requirements for human factors knowledge, research, methods, tools and applications; all have a commitment to the strengthening of rail human factors as a branch of the discipline and as a major contributor to a strong successful rail business and network.

RSSB's human factors team recently published 'Understanding Human Factors – a Guide for the Rail Industry' as part of the team's strategy to assist the UK railway industry in integrating human factors into all of its activities. The guide covers the full breadth of human factors issues and focuses around the frequently asked questions that the industry has about our discipline. The vision of the team is that human factors will be adequately identified, managed and mitigated, to improve the safety of all users of the railway system, while improving industry and company performance. A significant programme of work is being carried out to assist the industry to understand the causal factors of human failure following an incident, and to support the identification of solutions to reduce the likelihood of the incident occurring again.

This work will include further accident investigation training, human factors input to incident investigation and continued support to the development of SMIS (Safety Management Information System). In the short-term the team is retrospectively analysing a significant number of incidents which will be used to prioritise work activity and identify the most salient human factors issues as well as provide tools for front line staff in error prevention – such as rule compliance, CCTV, alarm and alerts and safety culture toolkits. The team is currently developing a train cab design standard and guidance note along with the development of CAD models and an anthropometric database of UK train drivers, allowing us to support industry with emerging cab design data and factors. The team is also supporting a variety of European projects aimed at developing optimal driving cabs in the future.

The team will be continuing to support the RSSB research programme but will place significant emphasis on helping the industry to make best use of the knowledge that has been generated in the past few years from the increase in human factors research. In this we will maintain a balance between exploration and exploitation. Finally the team will increase the number of in-house projects it undertakes such as the recent reviews into TPWS 'reset and continue', depot incidents, evaluation of SPAD initiatives, development of new train driver selection systems, guidance on fatigue management systems and further supporting the development of new systems such as GSM-R and ERTMS.

For Network Rail, recognising the human factor in railway systems remains integral to the delivery of reliable human performance. The Ergonomics NST views the railway network as a whole, taking a systems approach to identify and address human factors associated with the design, installation, operation and maintenance of the entire infrastructure. This activity extends throughout the business: the driver's interaction with the network to ensure that infrastructure is drivable; all operational elements, from our signallers using signalling equipment to the information required to direct and control train movements; Network Rail's own vehicles and their design; designing for installation and maintainability; and understanding safety behaviour and ideals of error free systems!

Projects are major and minor: new control centres technology, changes in existing boxes, changes to stations, new installation processes for switches and crossings, new business systems and integration of functions. This range of projects provides a diverse set of integration requirements and opportunities for involvement from initial idea to implementation. The delivery of GSM-R and ETCS, next generation of communications technologies and signalling-train control systems, requires a whole systems view for the design and development activity. Eliciting and refining cross-industry ergonomic requirements and then articulating those to other disciplines requires software systems engineering, behavioural and communications skills. This ability to interpret information between diverse but interested parties is a challenge and makes ergonomics an exciting place to be right at the heart of engineering change!

Understanding the Network Rail workforce, whether that is operator, maintainer or deliverer, working to shape and change underlying attitudes and address root causes to problems and systemic contributors to performance loss, is a core activity. Work continues on diverse topics of signaller competency and selection arrangements, investigation, fatigue management, revised working arrangements for access to the infrastructure, better comprehension of the capabilities and motivations of our workforce, and safety behavioural programmes. Amongst all the huge opportunities for ergonomics to make a difference it is often the simplest of things that may have the greatest effects for the guy on the ground – like making forms easier to fill in or key information clearer to understand.

Growing Network Rail's ergonomics capability, in terms of bringing on the next generation of practitioners to the discipline and to the rail industry, adding to the ergonomics body of understanding, and spreading this through the organisation has proved to be hugely satisfying.

At University of Nottingham the Centre for Rail Human Factors has continued to make a bridge between fundamental and applied research of relevance to the rail network. Some of the fundamental work has been funded by our partners in producing this book, Network Rail and RSSB, but also through our membership of the EPSRC funded network of several UK universities, Rail Research UK. In addition our fundamental human factors work is informed by our membership of the European network, EURNEX and our leadership of Pole 9 Human Factors within this.

In terms of particular strands of research Nottingham has carried out a substantial effort which has involved understanding the fundamental nature of work on the railways, examining strategies of working, activities and experience and competences within signalling and control for instance. In parallel with this we have been working to produce a number of methods and tools for particular studies, for instance of situation awareness, mental workload in signallers and team mental models. These programmes of research have now extended into studies of appropriate technical support, including levels of automation, for signalling and control.

One of the under-researched areas of rail human factors has been to do with engineering work – maintenance, enhancements and renewals. We have carried out a number of basic studies of the work of individual functions (engineering supervisors, PICOPs etc.) and this has now extended into full function analysis and HAZOP (hazard and operability study) for engineering work. Another strand of research is to do with display systems and this includes the development of virtual reality based simulators for train driving (and by extension for siting of signals and signs), development of interaction models and interfaces for mobile or personal information and communication devices for people out on the network, and fundamental investigations of monitoring and decision making when using CCTV. At an overall level our work continues into development and administration of ergonomics audit instruments and the detailed complex analyses required in order to interpret the large amounts of data emerging and to utilise these in establishing the direction forward for rail human factors.

A few short years ago it was not certain that even rail research generally, never mind the part that is rail human factors, would outlive an initial growth spurt. That it has done so is evidenced by such events as the second awarding by the Engineering Physical Sciences Research Council in the UK of a programme grant (RRUK2) across all rail research to nine universities, the central role being played by the Rail Research network, EURNEX, in establishing a business case for itself and in informing the next European Commission framework programmes for research, and continuing research with universities and consultancies contracted by Network Rail and by Rail Safety and Standards Board. Any familiarity with these networks makes it clear that there is substantial and vibrant research in such topics as improving the wheel rail interface, developing lighter and higher performance trains, exploring intelligent mobility and responsive transport networks and better understanding of the business, economic and social cases for rail transport. However it is at least arguable that per researcher involved no rail research field has grown so rapidly and so vibrantly as rail human factors. It is likely that this is because there is a widespread acceptance that in order to obtain a railway which is reliable, safe, efficient and makes maximum use of capacity, changes and improvements must be made in organisational and human network systems.

Moreover where technical systems change is considered, apart from in such areas as the wheel rail interface, these will only take best effect if there is a substantial human factors effort in parallel. For instance there may be major gains to be had from a new generation of signalling and control centres, using the best in automation, intelligent decision support systems and communication networks, but the benefits will only be realised if these centres are developed on the basis of the

current skills, knowledge and attitudes of staff, and on what these are likely to be in the future. Equally more sophisticated warning, reminder and protection devices in the train cab may have benefits of improved safety and also efficiency in train paths and train movements, but their design and implementation must reflect clear understanding of driver information processing, driver behaviour, driver expertise etc. As a third example one of the greatest gains in efficiency in running the whole network currently would be in improving the way that engineering work is planned and managed. However since this involves close understanding of good practice in command and control (for instance moving engineering trains on and off the track), communication and collaboration (for instance in briefing) and coordination (for instance in self-organising teamwork distributed in time and space) then considerable part of the improvement will come about through application of human factors expertise.

In all the above examples, and in countless others, what is particularly clear apart from the need for a human factors contribution is that this contribution must be based upon good practice and good knowledge. This means that we need to constantly renew, refresh and report the rail human factors research base and at the same time develop better techniques, standards and programmes for implementation of that human factors knowledge in practice. Moreover this must be done in the clear understanding that the railway is a complex socio technical system, that it is rare that examination and design for just one particular facet of operation (for instance a single type of computer display interface) will have the intended benefits if this is in isolation of understanding of the wider physical and social environment, and that holistic approach and methods are needed to achieve success.

This book has grown out of the second European Rail Human Factors Conference, held in London in November 2005. The chapters in the book are based upon the papers presented at the conference, which attracted a world wide audience from nearly 20 countries, and a full spread of interested people from academia, consultancies and the rail industry itself. Indeed one of the strengths of rail human factors at the moment is the close relationship many within the industry have with the work that is going on to understand the rail socio-technical system and to provide related understanding, guidance and design and implementation recommendations. The presentations at the conference covered the range of human and organisational issues on the railway from driving to signalling and control to maintenance and engineering work to passengers and security issues such as trespass. Therefore this book represents the best of recent work in rail human factors, and in the recommendations for future work and descriptions of next stages of work contained in many of the chapters it starts to define the framework for the next few years.

References

Wilson, J.R., Norris, B.J., Clarke, T. & Mills, A. (eds) *Rail Human Factors: supporting the Integrated Railway.* London: Ashgate, 2005).

Acknowledgements

Conferences and books are always a collective effort, and the second European Rail Human Factors Conference and subsequently this book are no exceptions.

First we thank all our authors and delegates – without them and their contribution then there would be no conference or book, and indeed no thriving community in rail human factors. The editors are grateful for the help of their colleagues at Network Rail, RSSB and University of Nottingham for their work on the conference and their own research and project management activities which

inform a part of the book. We are indebted to the assistance of Anne Floyde in collecting and collating the original papers which formed the conference proceedings and subsequently the book, and especially to our conference organiser Kerry Taylor who took responsibility for all site and delegate arrangements. After the conference most of the work of reformatting chapters and chasing recalcitrant authors (and editors!) fell to Lynne Mills and we are very grateful as usual for her vital contribution. At an institutional level we are also grateful for the support of our employers in putting the book together, and the staff at University of Nottingham also acknowledge the contribution made by RRUK and EURNEX in funding some of the time spent in editing and producing the book. Finally thanks are due also to our publishers and editors at Ashgate, and particularly Guy Loft.

This page intentionally left blank

Chapter 2

Real Prediction of Real Performance

Neville Moray

Editor's note: This opening chapter is taken directly from the opening address by Professor Neville Moray to the 2nd European Rail Human Factors Conference. In it he described, to a mixed audience of human factors researchers, practitioners and clients, how ergonomics is better placed today than ever before to make exact predictions about human performance. He suggests, however, that this will require a change in the way in which empirical data are used, in how statistics are reported and used, and a return to the use of mathematical models in co-operation with engineers.

Introduction

Ergonomists and human factors experts (I take them to be identical) have long claimed to be able to improve the efficiency of work, reduce errors and accidents, and ameliorate the working conditions of those who operate complex technical systems such as railways. In recent years indeed there has been a very great increase in the acceptance of ergonomics in the design and operation of such systems. It seems as though ergonomics has become accepted by the design and engineering community, by management and politicians, and even by lawyers, although the members of some of those categories have had to be dragged by main force into admitting the fact. We have seen ergonomists have major impacts on public inquiries in the nuclear industry, in accident investigations, in aviation, in the military and now finally even in the railways, long a privileged domain of civil and mechanical engineers. At the end of a career of in which I have spent many years working in the human factors of high technology systems this is gratifying. Furthermore it offers great promise for the next generation of ergonomists, and a chance to have a major impact on the well-being of society.

I do not have to tell this audience of the many ways in which ergonomics can be applied to complex technological systems. Not the least remarkable characteristic of the history of ergonomics in recent decades has been they way in which it has integrated social science and organisational factors into its approach and it is perhaps that aspect that has dominated recent developments. In my own case it was when defining the research needs of the American nuclear industry for the Nuclear Regulatory Commission that I came to realize that organisational factors were as important as the design of instruments and the ergonomics of the work place, with which I was most familiar. Traditionally also ergonomics can be used to evaluate and predict, at least ordinally, alternative systems designs, by using our empirical knowledge of the desirable qualities of displays and controls. But today I want to concentrate on a different way to use ergonomics, one which seems to have become less common in recent years, namely to make detailed quantitative predictions of moment to moment behaviour, and to use such methods to foresee both system success and system failure, and to advise on system design and modification.

Quantitative and engineering models

We have of course traditionally predicted performance on the basis of a statistical analysis of empirical data. This is what the vast majority of ergonomics relies on. Either by reference to data from previous field or laboratory research, or by empirical investigation of a particular system, we measure the mean and standard deviation (sigma) of performance measures or properties of the system, and use these, often with analysis of variance, regression, or other statistics, to show or predict that one system is "*significantly better*" than another, or at least "*significantly different*". This is a laudable and often cost-beneficial achievement. But it does not often let us make quantitative predictions, in the sense that I am proposing. Moreover, a point to which I will return, the traditional concentration on the *mean* as the measure of performance is not the most useful for ergonomics. Rather, as engineers know, prediction requires a model of the system whose performance is to be predicted, and it is because engineering has many models that it is so successful. Why, in ergonomics, is so little use made of models for detailed prediction? Do we not possess models that are sufficiently powerful for our purposes?

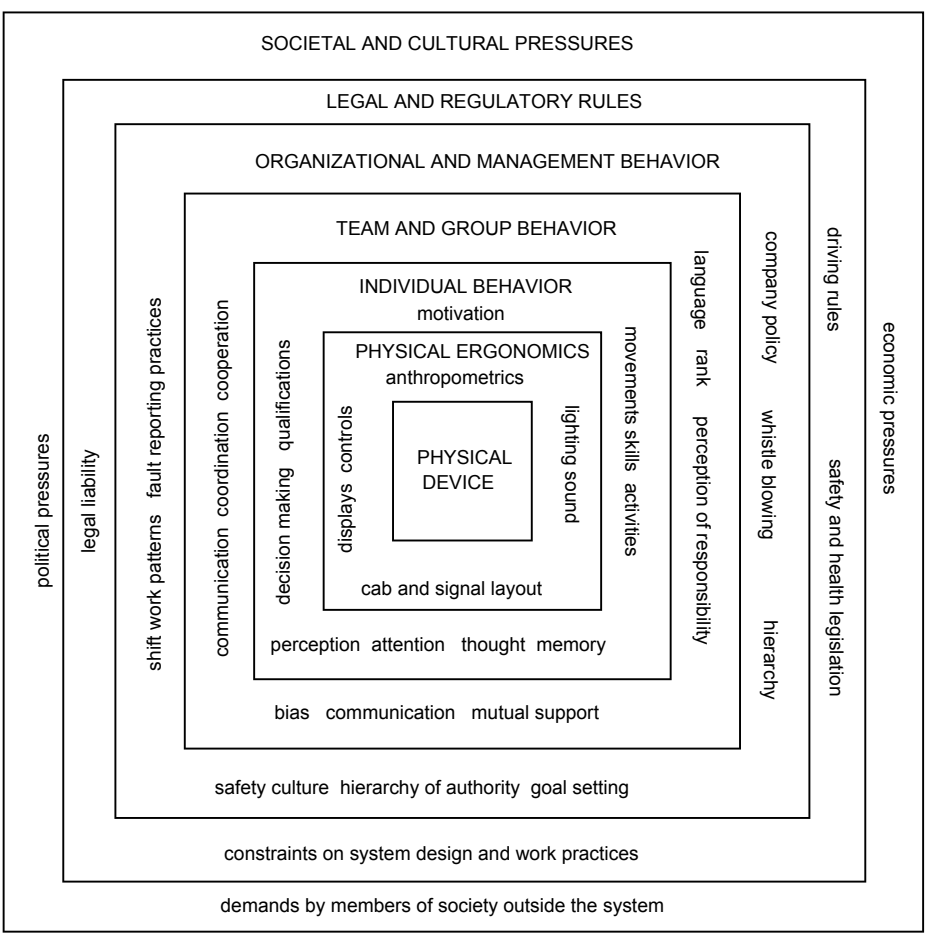


Figure 2.1 The general organization of human factors and ergonomics. The terms in upper case define levels of application. The terms in lower case describe typical variables relevant to each level of description

I have been worried in recent years by a change in the relation between ergonomists and engineers. Some people will feel that there is less reason to be worried than there used to be: at least we seem to be increasingly welcomed as colleagues by engineers. But 20 years ago there was a considerable international cadre of engineers who not merely welcomed ergonomists as colleagues, but were themselves very active in the study and prediction of human performance. As examples, one can cite Henk Stassen (a control engineer) from the Netherlands, Thomas Sheridan and Duane McRuer (mechanical engineers) from the USA., and Jens Rasmussen (an electronics engineer) from Denmark. All these have made enormous contributions to our ability to understand and to predict human performance. But all of them have now retired, and while there is a great increase in young practitioners entering ergonomics from psychological backgrounds, there are very few programmes in which new engineers are entering the profession. For thirty years there was an annual meeting in the USA., the famous "Annual Manual", to which fifty or more engineers, psychologists and ergonomists used to come to discuss human performance models. That no longer exists. Now there seems to be a very disturbing tendency to say that engineering models, and quantitative mathematical models in general, are "too difficult" to teach or use. I have recently come across two cases where authors, writing on the human factors of high technology systems, have been asked by editors to delete mathematical models because they would not be understood by readers and might put readers off. (One was myself.)

Now I am not an engineer, and have not had a formal training in mathematics beyond what used to be called "A Level". But if behavioural scientists are capable of understanding statistics and probability theory they are capable of learning enough mathematics to understand quantitative models, and even sometimes putting them to use. The precision with which performance can be predicted by such models is in some cases very great, and failing to use such models is a grave defect in our discipline. Furthermore, if we are to co-operate with engineers we must show that even if we cannot manipulate their models ourselves, we can talk with them and both understand the conversation and direct them as to where we want them to apply their models. Further, it is common for one to be asked for examples of where ergonomics is cost effective, and a major (although not the only) way to show this is by showing how predictions can be made.

I have written elsewhere about the decrease in the use of predictive quantitative models, and particularly engineering models:

- "The loss of contact with engineering has led people to avoid models tightly coupled to the dynamics of the systems and environments.
- A zeitgeist developed in the 1980s that rejected mathematical models in favour of a deeper psychological understanding of the operators' cognitive processes. But almost no predictive models have resulted.
- In modern systems the coupling between human and machine is much looser. (It was loose even for very slow manual control processes and was never developed for discrete manufacturing. See papers in Edwards and Lees (1974) for discrete intervention in the case of slow process control.)
- The coupling changed from physical causality to logical and semantic relationships (Rasmussen, 1986).
- The operators' role changed from controller to supervisor and intervener.
- Operators' actions afforded by the systems became ever less analog, and less depended on physical strength and skill. (Interfaces intervened between operator actions and provided digital control of the systems.)"
- As human-machine systems (HMS) become ever more complex and larger, teams, rather

> than individuals control them, and modelling team behaviour seems at first sight very different from modelling individual behaviour. Moray, (2005a)

Those not familiar with engineering models may find a collection of several of the most important, in the context of the history of ergonomics, in Moray (2005b).

Quantitative prediction

To encourage interest in quantitative prediction I will present some examples for several kinds of performance, and then give a "worked example" of how modelling may be applied in a railway context.

What should be a famous example is not from truly an engineering model of the human, but from the "brute force" model called *The Model Human Processor* or *GOMS* (Card, Moran and Newell (1983). This model that relies simply on best estimates of mean and variance of empirical times for perception, attention, decision, movement, etc., allowed Gray, John and Atwood (1993) to predict that a proposed new work station for answering telephone queries would be slightly slower to use than an alternative. The authors state,

> By this benchmark-based, non-cognitive procedure, we would predict an average advantage for the proposed work-station of 5.2%. When call categories are weighted by their frequency of occurrence, the predicted advantage becomes 18.6% (4.1 sec), for an estimated savings in annual operating costs of $12.2 million (my emphasis.)

Note that the prediction was accurate to the order of a few seconds, and that for once the Holy Grail was actually achieved: ergonomic intervention *is* cost effective! This kind of model could be used to predict response times of skilled operators in a signal box or the driving cab of a train.

GOMS is mathematically very simple, and does not involve the mathematics of continuous temporal processes. Engineering models are often more complex, and typically deal with continuous dynamic processes which are not well served by the kind of statistical models typically taught and used in behavioural science. Classical and Optimal Control Theory, incorporating Fourier Analysis and Power Spectral Analysis, can make very exact predictions about whether and how well vehicles can be controlled by human operators. Classical control theory can predict whether a vehicle will be stable under manual control, and can indicate when some kind of machine (computer) aiding will be needed. Optimal control can predict with precision not just mean performance but the moment-to-moment movements in real time of a human controller. An example, from Stassen's work, is shown in Figure 2.2. The actual movements of the wheel by a helmsman, not just the mean course, are accurately predicted. This kind of model can be applied to a wide variety of systems, including process control, vehicular control, the design of prostheses, and even systems in which only estimation of noisy data – that is, dynamic pattern recognition rather than movement – is to be predicted. These models require a knowledge of calculus, Laplace transforms, and matrix algebra, but that level of mathematics can be acquired by anyone capable of graduate level statistics, at least to a level at which discussions about their application can be held with engineers. Control theory can be used to model the manual control of train dynamics.

Next let us look at a mathematical model of eye movement dynamics used to predict operator workload (Moray, Richards and Low, 1980; Moray, Neil, and Brophy, 1983). It is becoming common to use eye movement recording to describe how people allocate their attention over time. Usually the data are reported in terms of the mean and variance of the proportion of time spent looking at each locus in the environment; but there are far more powerful ways to model the dynamics. By recording the transitions of glances from one locus to the next, we can apply the mathematics of

Markov processes to derive measures such as the Mean First Passage Time (MFPT), the proportion of time on each locus, the duration of each glance, and the variances of these measures. Table 2.1 shows the result of analyzing the eye movements of an individual radar controller directing fighters during an interception.

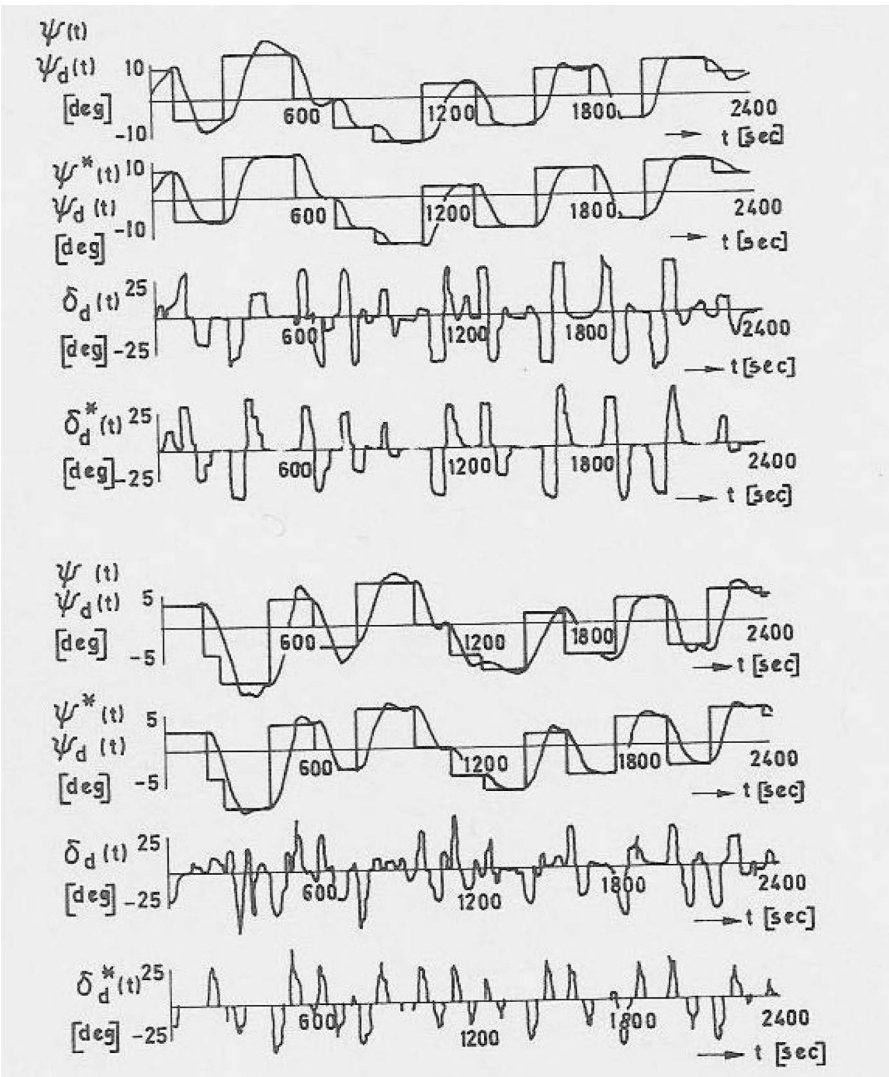


Figure 2.2 **An example of prediction of the ability of Optimal Control Theory to predict human behaviour. In each pair of lines the upper (for examplę ψ is the behaviour of the human, and the lower starred line for example ψ* is the behaviour of the model. ψ shows the course of the ship, δ the movements of the helm. From Veldhuyzen and Stassen (1976)**

Table 2.1a Transition probabilities of a radar controller's eye movements. (The table shows, e.g., that after looking at Str the probability that the next locus looked at is T is equal to 0.46.)

NEW LOOK

LAST LOOK

(P)	F	T	Str	E	To	C	Se
F	0.40	0.53	0.00	0.04	0.01	0.00	0.02
T	0.36	0.49	0.02	0.07	0.01	0.03	0.02
Str	0.15	0.46	0.31	0.00	0.00	0.08	0.00
E	0.27	0.54	0.01	0.08	0.02	0.02	0.06
To	0.23	0.08	0.00	0.00	0.69	0.00	0.00
C	0.05	0.28	0.05	0.03	0.00	0.56	0.03
Se	0.28	0.40	0.04	0.04	0.00	0.04	0.20

Table 2.1b Mean First Passage Times, MFPT, derived from Table 1a. Times in seconds. (If currently looking at Str, the mean time elapsed before looking at T is 2.22 seconds. Sigmas are approximately the same magnitude as the means.)

NEW LOOK

LAST LOOK

	T	Str	E	To	C	Se	
F	2.88	1.95	75.41	19.03	103.14	54.30	46.14
T	3.09	2.06	73.70	18.52	103.22	52.88	46.08
Str	4.11	2.22	51.63	20.21	104.94	48.29	47.44
E	3.37	1.99	74.31	18.32	102.20	53.00	44.24
To	4.01	4.66	78.18	22.11	32.98	57.04	49.34
C	5.18	3.04	67.70	19.70	105.72	23.99	45.24
Se	3.43	2.29	71.59	19.16	104.60	51.60	37.83

Table 2.1c Mean fixation times derived from Table 1a. Times in seconds. (The mean duration of a fixation on F over all the occasions it is fixated is 1.67 seconds.)

F	T	Str	E	To	C	Se
1.67	1.85	1.45	1.09	3.22	2.27	1.25

Table 2.1d Proportion of time spent looking at each locus derived from Table 1a. (48% of the operator's time is spent looking at T.

F	T	Str	E	To	C	Se
0.35	0.48	0.02	0.05	0.03	0.04	0.03

Although not shown, it is important to note that the standard deviation of the MFPT entries is about the same magnitude as the entries in the table. If the time between successive looks at an aircraft being controlled becomes too long, its trajectory will not be able to be accurately controlled hence the authors were able to use the MFPTs to estimate workload. They were able to predict when overload would occur for each operator by relating growing MFPTs to forgetting rates as the number of aircraft increased.

These data also allow us to see a second point about methodology. Although the mean is by definition the value most typical of the data, *in many real situations it is not the most important value for predicting performance.* Suppose that the controller is looking at the friendly fighter "F". The MFPT shows that on the average it will only be about 3 seconds before he looks at it again (actually 2.88 secs.). But if the sigma is also about 3 seconds, then we can draw up table 2.2.

Table 2.2 Means and sigmas of FPT for looking at F repeatedly. P shows the probability of a FPT of each value.

F|F mean	F|F mean+1.0 sigma	F|F mean+2.0 sigma
Time 3.0sec	6.0sec	9.0sec
P ≤ 0.5	≤ 0.16	≤0.01

Fighter aircraft travel very rapidly, and as they near their target there may be very little time left for the controller to affect the course of the fighter. One might think it enough that the controller look at the fighter every 3.0 seconds. But given the critical nature of the task, it is perhaps more important to notice that the controller will look back only after 6.0 seconds on 16% of the occasions, and will look back *only after nearly 9 seconds on at least 1% of occasions*. Given that fighters may fly at least as fast as 600 mph, that interval represents a distance of about 1.5 miles, sufficient to cause either a collision or a complete miss of the target. At supersonic speeds the distances are even greater, and it was reported that many radar operators could control subsonic fighters but lost control at supersonic speeds. As with many dangerous situations, performance at 3 sigma does not happen very often, but once is enough: and given the number of eye movements made per sortie, and the number of sorties, how many occasions are represented by 1%?

The important conclusion is that *for many real tasks, it is not important that the mean value of performance is acceptable, but rather that the variance implies that an unacceptable performance may occur occasionally.* Furthermore, by citing the variance, we can predict how often the extreme value will arise and what its magnitude will be. *Results should be presented to clients in terms of the extreme values and their probabilities, not in terms of means.* The probability level of interest must be chosen in terms of cost-benefit analysis, for good or evil outcomes.

Closely related to the last example is the use of time-reliability curves to emphasise the implications of extreme values, rather than merely giving means and sigmas. A classical example is the work of Wohl (1982) on the time to repair radar sets, but almost any measures of response time, completion time, etc. can be plotted in a similar way. One can either use the mathematics of the Weibull distribution, or the cumulative probability-versus-log(time) plot. An example, from the nuclear industry, is shown in Figure 2.3.

Beare et al. (1984) investigated the time it took teams of operators to respond to a nuclear power plant incident. They could have presented their results as means and sigmas, but the time reliability plot is much more revealing. Suppose we knew that the process would become uncontrollable after about 600 seconds. The median performance (cumulative probability = 50%) is well under that value. But one can easily see from this model that we can expect that somewhere between 5% and 10% of interventions will not be completed by 600 seconds, and can then ask what proportion is an acceptable risk. By fitting an equation to the curve, we can therefore predict future performance well beyond our data, and find critical values that call for system redesign, increase in manning levels, improved training, better operating procedures, etc.

A worked example: Predicting a SPAD

Finally, to show how one might approach quantitative modelling in a specific case, let us develop a quantitative model; and since this is a conference on rail human factors, let us consider how we might model the cognitive processes that underlie a SPAD. Since this is a hypothetical example, and is not based entirely on real data, we make some assumptions, (see Figure 2.4).

We assume:

1. A train of three carriages is approaching a signal that is obscured by a building and by a bridge, so that the signal first comes into view when the train passes under the bridge, about 100 metres from the signal.
2. Trains normally run at this point at a speed of 40 mph (18 m.sec^{-1}).
3. That train speed implies that there will be about 6 seconds from the time the signal first becomes visible to the moment when the train passes the signal (if the driver takes no

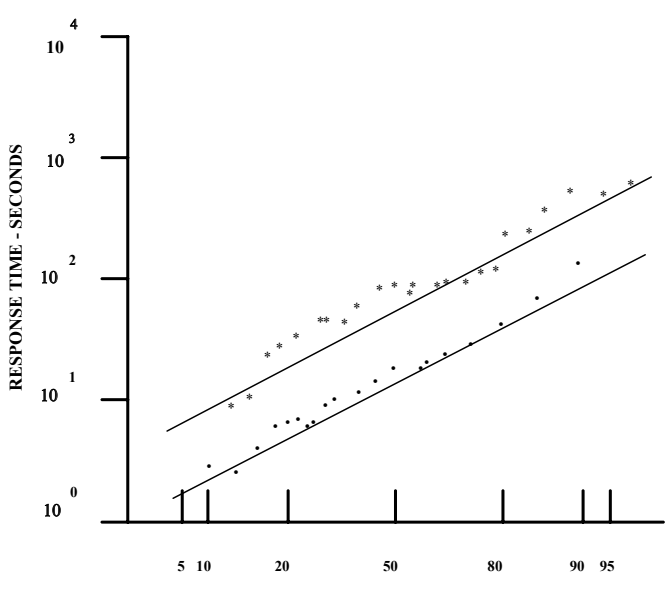


Figure 2.3 Cumulative probability of response related to log(time) since incident

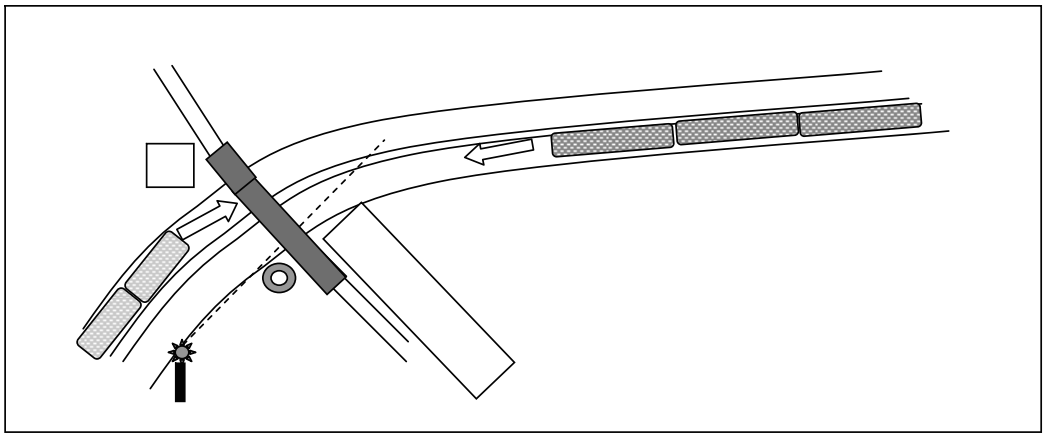


Figure 2.4 A hypothetical case of a train approaching an obscured sign

action).

4. There is a train of two carriages approaching from the other direction.
5. The signal is showing a red aspect and is correctly aligned according to engineering standards
6. There is a posted speed sign just after the bridge.
7. The driver is well trained and well motivated.
8. He has driven the route 20 times in the past 10 days and the signal has always shown a green or double yellow aspect.
9. Since starting his journey today he has passed 10 signals all showing a green aspect.

Can we predict quantitatively whether he will commit a SPAD?

We can rephrase the question in a way that suggests how to develop an appropriate model. "Can we predict the pattern of attention that will be shown by the driver, and does the latter suggest that there will be an unacceptable probability of failing to see the Red aspect in the time available and acting appropriately?"

To predict this performance we need to predict both the probability that the driver will pay attention to (that is, fixate) the signal, and that if so he will have time to detect the colour of the light. As we saw in the case of the radar operator a Markov process is a good model for visual attention in a dynamic environment. Unfortunately we do not have such data for train drivers. We do however have at least some eye movement data, recently collected and analysed by Groeger, Bradshaw, Everatt, Merat, and Field, (2004).[1] Those authors found that in analyzing train drivers' eye movements they needed to sort fixations into nearly 25 categories, as shown in Table 2.3. Note that in fact there are even more categories, since a driver cannot simultaneously view all the different instruments, controls etc. "In Cab".

Groeger et al. pooled their data into two categories, "signals" (when the driver is looking at signals relevant to his train) and "non-signals". They found that in the last 15 seconds approaching signals drivers typically spent only about 22% of the time looking at their signals (mean proportion = 0.22, sigma = 0.15). Given that in our example the signal is only visible for 6 seconds, this means that a driver will typically look at the signal for (0.22*6.0 = 1.32) seconds. On the assumption that the data of Groeger et al. are normally distributed, there is therefore a

1 The quantitative values in the following model are also supported by Luke et al., 2006.

Table 2.3 Categories needed to classify drivers' eye movements. From Groeger et al. (2004)

Category	Individual objects and object locations
Own Signal	Own signal (separately coded for 'looks' within 1, 2 and 4 degrees of signal) Own direction indicator
Other Signal	Signals to the left and right of driver's own signal (on the same gantry) Upcoming signal on the same line as driver's own signal Upcoming signal to the left or right of driver's own signal Direction indicators to the left or right of the driver's own signal
Own Gantry	Any position on driver's own gantry, except the signal
Other Gantry	Any signals not facing the driver Any gantries to the right or left of driver's own gantry
Signage	All speed restriction signs AWS magnet TPWS grid CDRA sign Station TV monitors Station clock Station electronic train information screen Platforms to the right and left
Moving Objects	Travelling trains to the right and left of the driver Station or track crew
Off Track	Fields to the right and left Verge between track and fields Trees Oncoming bridges and city buildings Tunnel wall and ceiling
Track Ahead	Driver's own track Track to the left or right of the driver
Sky	The sky
In Cab	All objects and controls inside the cab Driver's own window Side windows Instructor

50% chance that a driver will spend not more than 1.32 seconds looking at the signal. That may be thought sufficient to identify the aspect of the signal, *provided that the fixation occurs as the driver passes through the centre of the aligned beam.* There is of course no guarantee that such will be the case. Moreover, the value of the sigma gives cause for greater concern, since it implies that there is a probability of at least $p = 0.16$ that the time spent looking will be as short as $(0.22 – 0.15)* 6.0 = 0.42$ secs.; and there is, statistically, a substantially non-zero probability that the time spent fixating will be zero, since the value of sigma means that zero is to be found less than 2 sigma from the mean.

While the statistical model of subtracting standard deviations obviously fails when the result is a time less than zero, the conclusion must be that at the speed typical of trains on this section of track we may well expect an occasional "near miss" or even SPAD if there is a red aspect showing. The situation is even worse, since we have assumed that the signal is visible for 6 seconds (100 metres) if looked at. But Groeger et al. also found that the mean proportion of that distance that typically elapsed before first fixation was about 0.25 (sd = 0.28) so that the actual time available is more likely to be less than 4.5 seconds rather than 6.0, and hence all the above values should be reduced.

Groeger et al. also found that the mean frequency of eye movements was about 0.46 per second (sigma = 0.3), that is approximately one new fixation every 2.5 seconds. Let us conservatively assume this value to be one every 2.0 seconds. There will therefore only be time for two fixations after the signal becomes visible to the driver, of which, even if we use the mean, only 0.33*2.0, or at the most *one*, will be directed to the signal. If we include the sigma of the mean, there will be a high probability that neither will be directed to the signal.

It may be thought that two other features of the situation will save the day. The first is that the driver may be looking at another feature of the environment as he approaches the signal, (in particular his attention will tend to be caught by sudden changes in the visual field such as the oncoming train entering the visual field,) but he may see the signal in his peripheral vision. Will this not attract his attention in time, or at least let him recognize the colour?

It is important to realize that the reason he is distributing his attention over a large number of non-signal loci is *not* because he is driving badly. Although there used to be a myth current among railway management to the effect that drivers always look at upcoming signals before they appear (at least according to statements by train operating companies in public inquiries) it is not true. Groeger et al. found that on as many as 33% of approaches the signal was only fixated after the AWS horn sounded. The reason is that to drive well a driver must look at many things – the track side to look for working parties, vandals or others who may stray onto the lines, the control panel in the cab to monitor speed, signals on gantries to perform "counting across" to identify their signal, and so on. The fact that attention is not paid exclusively to signals is precisely a sign that the driver is performing his task *well*.

What then is the probability that if his gaze is elsewhere as he passes through the beam he will identify the colour of the aspect in peripheral vision? This is difficult to calculate without knowing the exact sequence of eye movements, which will of course differ on different occasions. However, we can put some boundaries on the probabilities. Let us assume that the visibility is good[2]. From the work of Van Esch, Koldenhof, van Doorn, and Koenderink, (1984) and Abramov, Gordon, and Chan (1991) we can predict that if the colour is identified perfectly when fixated on the fovea, the efficiency of colour detection with retinal eccentricity will fall as shown in Table 2.4.

Table 2.4 Relative efficiency of colour vision with visual eccentricity

Distance from the fovea in degrees visual angle	0	8	10	25	40	80
Relative ability to detect colour if size of signal is constant	100%	25%	12.5%	10%	6%	3%

2 If there is fog or rain it can only make things worse, and if needed there are nomograms to predict quantitatively the deterioration of signal visibility, such as that given in *7.602 Nomographic charts for daylight and overcast sky conditions.*(Boff and Lincoln, 1988).

Clearly we must predict that if the driver fixates more than about 20^0 away from the light his ability to identify the colour will be greatly reduced. In fact in many commuter trains the driver must look down at an angle of at least 45^0 to see instruments in the cab. Furthermore, it is known that colours presented in the peripheral field appear desaturated, and progressively achromatic. We may expect (although here I have not found a model to quantify the prediction) that a red light appearing weak, desaturated and achromatic in the periphery of vision will appear more like a Yellow aspect.

The detection of colour at high retinal eccentricities then becomes a problem in Signal Detection Theory. You will remember that in the latter the subjective criterion for the identification of a signal depends on the payoff matrix and the subjective probability of the signal. Now we have to assume that the relative payoffs associated with red, yellow and green aspects stay constant, and are functions of training. On the other hand, the subjective probability of a particular colour aspect will depend on the *a priori* probability of that aspect appearing, and this will be dynamic. Given that a Double Yellow aspect has just been passed, the subjective probability (expectation) that the next signal will be Single Yellow will be much higher than if the previous signal was a Green aspect. But we know from our assumptions that the driver in this case has driven over this route 20 times and has never seen anything but Green or Double Yellow aspects at this point. His decision criterion (beta) will therefore be particularly low for detecting such an aspect, and high for red, and the desaturated red in the periphery of vision will strongly tend to be judged to be Single Yellow. (Indeed, if we knew the total history of his driving over this route, we might be able to make a reasonable, conservative, quantitative estimate of his decision criterion.) There are hints of different criteria in the behaviour of Groeger et al.'s drivers when approaching signals showing different aspects, but I will not discuss them here. We may conclude from this (semi-quantitative) model that it is unlikely that peripheral vision will provide the necessary information if the signal is not fixated. It certainly would not be a conservative conclusion to assume otherwise, and we could probably make a reasonable conservative quantitative estimate.

There remains the AWS which will sound as the train approaches the signal. We have seen that Groeger et al. found that on about a third of occasions the driver was looking elsewhere when the AWS sounded. Since an eye movement takes only a fraction of a second, surely the AWS will direct the driver's attention to the signal. Although we cannot put a quantitative value on this effect, the result is uncertain. AWS horns are used not only to direct attention to signals, but also to direct attention to posted speed limits. We have seen in the Signal Detection analysis that there are reasons for the driver to be expecting a Green aspect at this point, and being a good driver his route knowledge tells him that there is a posted speed limit coming up. He may therefore assume that the AWS horn is relevant to the speed limit, not to the (highly improbable $p << 0.05$ in his experience) occurrence of a red aspect.

A quantitative analysis on these lines clearly would predict that the combination of signal position and train speed means that the probability of missing or mis-reading the signal is of the order of $p \geq 0.01$. Given that many passes of this point are made each week, that probability is far too high to be acceptable, which indicates that there is a need to redesign the system at this point. Either the signal must be moved or the speed reduced. By how much? We can work the quantitative analysis in reverse to make a recommendation.

We know that as a driver approaches a signal he spends a mean of 0.22 of the last 15 seconds looking at the signal, and that the sigma of this value is 0.15. Let us interpret this to mean that the mean probability that the first fixation after the bridge will be on the signal is 0.22. Therefore the probability that the first fixation will *not* be on the signal is 0.78. Therefore, the probability of n successive looks all not being on the signal is equal to $(1 - 0.78^n)$. This allows us to construct Table 2.5.

Table 2.5 Probability that at least one fixation will be on the signal as a function of the number of fixations

Number of fixations	2	4	6	8	10	12
Probability that one will be on the signal	0.39	0.63	0.78	0.87	0.92	0.95

Let us say that we can accept a probability of 0.92 as safe. (This may be too conservative, because some of the other fixations may well remind the driver of where he is, and thus direct attention towards the signal.) Since we know that the driver makes a fixation about every 2 seconds, this means we need to allow 20 seconds for 10 fixations. That in turn means the speed must not exceed 5 metres a second during the 100 metres prior to the signal. This is equivalent to about 11 mph! It also follows, roughly speaking, that if the signal is visible for 300 metres a speed of 33 mph would be appropriate, for 500 metres a speed of 55 mph, and so on. We can therefore quantitatively define acceptable changes in design.

Let us stop here (no pun intended). There may well be many other features of the situation that would lead to these figures being revised up or down. Many of you may think the scenario unrealistic. That is not important. My intention is not to develop a new standard for signal sighting in this chapter, nor to consider all possible details of such a situation, since it is hypothetical. I wish merely to show how by applying a combination of psychological and ergonomic empirical research and mathematical modelling we can go far beyond simply asserting that by making this or that change in accordance with ergonomic principles "things will be improved". As I have said, I would like to have the data to apply Markov modelling as adding a powerful extra mathematical tool. I would also want to add an engineering analysis. Suppose that it were to be suggested that automatic braking on Red aspect should be installed. There remains the problem of rare failures of the automation, and the intervention of the human driver in such situations. By modelling the braking dynamics of the train, (a purely mechanical engineering problem) we could understand how much time would be available for the driver to intervene if he thought that the automatic braking had failed. We could then apply the kind of analysis developed by Inagaki (1993, 1995) to model human intervention in automatic control for aircraft takeoff, and decide how to balance automation with human intervention, using the ideas of Moray (2003) and Moray and Inagaki (2001) on avoiding complacency with automation.

Conclusion

I want to leave you with a *cri-de-coeur*. It is time to develop again quantitative methods, even though in recent years the dominant emphasis on the ergonomics of complex systems has been, probably correctly, on organizational factors (Reason, 1997). Rediscover quantitative methods such as signal detection theory, control theory, Markov modelling, scheduling theory, and mathematical modelling in general. We have today an enormously greater data base of empirical facts about human performance than we had a generation ago, much greater even than when Boff and Lincoln (1988) published the Engineering Data Compendium. We have better equipment and apparatus. We have computers to perform advanced mathematical calculations. We even at this moment have the good will of society and other disciplines towards ergonomics. Let us make the most of all this to provide a truly predictive applied science of human behaviour wherever that is possible, with all

it can offer, including cost-benefit justification of ergonomic intervention, improved safety for the work-force and society, and greater efficiency of complex human-machine systems.

References

Abramov, I., Gordon, J. and Chan, H. 1991. Color appearance in the peripheral retina: effects of stimulus size. *Journal of the Optical Society of America, A, 8,* 404-413.

Beare, A. N., Dorris, R. E., Bovell, C. R., Crowe, D. S., and Kozinsky, E. J. 1984. *A Simulator-Based Study of Human Errors in Nuclear Power Plant Control Room Tasks. NUREG/CR-3309.* U.S. Nuclear Regulatory Commission, Washington, D.C.

Boff, K. and Lincoln, J. E. 1988. *Engineering Data Compendium: Human Perception and Performance.* AAMRL, Wright Paterson AFB, Ohio.

Card, S.K., Moran, T.P., and Newell, A. 1983. *The Psychology of Human-Computer Interaction.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Gray, W. D., John, B. E. and Atwood, M. E. 1993. Project Ernestine: validating a GOMS analysis for predicting and explaining real-world task performance. *Human–Computer Interaction, 8,* 237-309.

Groeger, J. A., Bradshaw, M. F., Everatt, J., Merat, N., and Field, D., 2004. "Pilot study of Train Drivers' Eye Movements." Rail Safety and Standards Board.

Inagaki, T. 1993. Situation adaptive degree of automation for system safety. *Proceedings of IEEE International Workshop on Robot and Human Communication.* 231-236.

Inagaki, T. 1995. Situation adaptive responsibility allocation for human-centered automation. *Transactions of the SICE of Japan, 31(3)*, 292-298.

Luke, T., Brook-Carter, N., Parkes, A. M., Grimes, E., and Mills, A. 2006. An investigation of train driver visual strategies. *Cognition, Technology and Work, 8(1),* 15-29.

Moray, N. 2003. Monitoring, complacency, scepticism and eutactic behaviour. *International Journal of Industrial Ergonomics, 31,* 175-178.

Moray, N. 2005a. Où sont les neiges d'antan? *Keynote paper at HPSAA/ERAU Conference, Orlando, Florida.*

Moray, N. (Ed.) 2005b. Ergonomics: Major Writings. Vols I – IV. London: Routledge/Taylor and Francis.

Moray, N., and Inagaki, T. 2001. Attention and Complacency. Theoretical Issues in Ergonomic Science, 1(4), 354-365.

Moray, N., Neil, G. and Brophy, C. M. 1983. Selection and Behaviour of Fighter Controllers, Contract Report, Ministry of Defence, London.

Moray, N., Richards, M. and Low, J. 1980.*The Behaviour of Fighter Controllers*, Contract Report, Ministry of Defense, London.

Reason, J. 1997. Managing Risks of Organizational Accidents. Ashgate Gower.

Van Esch, J. A.,Koldenhof, E. E., van Doorn, A. J. and Koenderink, J. J. 1984. Spectral sensitivity and wavelength discrimination of the human peripheral visual field. *Journal of the Optical Society of America, A, 1,* 443-450.

Veldhuyzen, W. and Stassen, H. G. 1976. The internal model: what does it mean in human control? In T. B. Sheridan and G. Johannsen (Eds.). *Monitoring Behavior and Supervisory Control*. New York: Plenum Press. 157-171.

Wohl, J.G. 1982. Maintainability prediction revisited: diagnostic behavior, system complexity, and repair time. *IEEE Transactions on Systems, Man and Cybernetics, SMC-12,* 241-250.

7.602 Nomographic charts for daylight and overcast sky conditions. In Boff, K. and Lincoln, J. E. 1988. *Engineering data Compendium: Human Perception and Performance. AAMRL*, Wright Paterson AFB, Ohio. 1606-1607.

PART II
Passengers and Public

This page intentionally left blank

Chapter 3

Wayfinding, Accessibility, Inclusive Design and Passenger Information Systems: 'Euston Station and Beyond'

Matt Pattison, John R. Wilson and Theresa Clarke

Introduction

The social goal of increasingly accessible environments is imperative for an advancing society. Network Rail plays a major role in providing such accessible public environments with station access and journeys per annum topping 1 billion (DFT, 2004). Add to this the market potential for 35% of the population who at present do not access the rail network and you have two very strong drivers to investigate rail network accessibility. With a specific interest in inclusive design a study has been based on one area of accessibility – passenger information systems and wayfinding considerations in the journey chain.

When making a journey a passenger needs to take a series of steps that enable them to reach their destination. From pre-journey (learning and planning) through to arrival, the passenger relies on a barrier free environment to enable successful passage. Whilst any such barriers may affect ease of use for the young able bodied passenger, to those with any of a variety of impairments they may prevent them from simply getting from A to B. In investigating such barriers and accessibility this research sets out to answer the following questions. What issues about wayfinding information design are generated when observing boundary case users of varied abilities accessing the train network? Also, can Inclusive Design principles improve wayfinding and information accessibility for users of all ability levels in the context of the modern rail industry?

When considering the role of Network Rail in accessibility issues, the DDA (1995) states that the service infrastructure provider must be seen to think through the passenger journey and identify the barriers to people with different impairments. In so doing they must think creatively about removing these barriers. Access must no longer be a 'bolt on' and disability must no longer be assumed to be just wheelchair users. In consideration of the above, the SRA, in their 'Railways for all' document (2005) state that the main focus should be on four specific key points for accessibility:

1. Customer Information Systems
2. Public Address
3. Step-free access, and
4. Training.

Background

Wayfinding

When finding their way in an environment and getting from A to B, a user relies on the ability to spatially orientate themselves and wayfind (Passini, 1984). To avoid exasperation and enable successful carriage, environments must be designed to allow easy flow within a system. Unclear and ambiguous design can place barriers to users, which may exasperate the able bodied and block those with impairments from carrying out their journey. Any such barriers may arise not only from physical restrictions but also from informational or cognitive inaccessibility.

In developing an 'action plan' in simple environments, one can execute this via a 'perception-action' framework consisting of behaviour, e.g. 'to go to', and a characteristic place e.g. the escalator. However in complex environments such as an underground station, wayfinding information design must be planned on the basis of sequential or step-to-step processing. In such environments if design is ambiguous and landmarks are poorly presented then circulation can be unnecessarily blocked.

Inclusive design and the older adult

The population over the age of 65 is set to grow from 400 million to 1.3 billion by 2050 (Summers, 2001). There is support for a 'Third Age' (Laslett, 1989, Coleman 2001) where the older population is more active, holds valuable market power and holds a continued productive role in society. However many 'Third Agers' are poor, and struggling to get by financially and/or physically in an environment that can enforce disability out of impairment (Ricability, 2005). With only 3% of older users actually using the rail network regularly (Mollenkopt, 1997) a large percentage of the population are not accessing rail services. It is important to question whether this has anything to do with their ability to access the environment and, in the case of this study, how information systems and wayfinding guidance affects experience management and possible reasons for lack of usage?

When considering the older user and impairment, decline is an insidious process subject to variables such as health, nutrition, individual differences and work and exercise though life. Over time, ageing leads to visual decline (with increased susceptibility to glare, weaker accommodation and convergence), hearing loss (by the age of 50 there is often significant loss to leading to impairment) and cognitive deficits (with episodic memory loss and poor ability to cope with divided attention (just two of many possible examples).

Despite the heterogeneous nature of ageing it is observed in us all at some stage, and the case for disability has often been made in relation to design via the restrictive, rather than de-restrictive environment (Fisk, 1993) that surrounds the individual. However, good design can make the less severe of these impairments invisible, and it is here that inclusive design practice can enable rather than exclude the older user (Gyi et al., 2000). If inclusive design principles are implemented in the design process from conception, end products (and systems) can enable rather than disable the older (or young impaired) user. Products designed in this way that account for the older user and those with most difficulties will themselves satisfy the younger user (Woodhuysen, 1993 Coleman 2001, Ricability, 2005). The television remote control or the easygrip™ handles prove excellent examples of such principles designed for those with difficulty and benefiting all.

Alongside older users, a comparison group was also involved in this study, with regards to information accessibility. These are younger visually impaired users, who rely primarily not on the visual but on other perceptual systems to wayfind, and therefore encounter very different problems.

There are approximately two million blind or visually impaired people in the UK with the loss of sight proving the most common form of disability. In considering that twenty five percent of passengers across Europe struggle to travel by train due to visual deficit (Tiresias – 2005, COST 335) the access to this comparison group of younger users, offered the researcher another perspective on systems of wayfinding and information usage that required consideration.

Human factors, psychology and design

Don Norman's consideration of principles important to successful design guide this research, with a framework for discussion of findings around design and information provision (e.g. via visibility, constraints, feedback, affordances and mapping). The product or system user can be seen to take the information from visibility, affordances, feedback, constraints and mappings to build conceptual models which, via mental imagery, support the users' understanding of the product or system state. If products or environments are poorly designed it is common for users to blame themselves and their perceived 'stupidity' or 'clumsiness' rather than design failures. Norman however focuses on systems design weaknesses represented by the gulfs of execution and evaluation. This approach is represented in the data gathered in this study.

The complex naturalistic environments that train terminals provide do not support the naïve understanding of processing via 'input–process–output' models. This representation by the likes of Fischcoff (1988) would suggest that individuals specify costs and benefits, choosing the 'ideal' solution based for example upon a mathematical choice approach. Rather, in an understanding more akin to Hollnagel's (2003) Joint Cognitive Systems, or Klein's (1993) Naturalistic Decision Making, the dynamic 'journey chain' with its associated environments, artefacts and individuals is more representative. Here the passenger 'satisfices' (Simon 1956) based not only on limited information from within their head but from the complex built environment of physical cues, constraints and affordances that 'flow' or 'jolt' an individual through the system. Klein's understanding of ill structured and uncertain environments in which a user has ill-defined or competing goals formed within time stresses, fits the way that people are now seen to 'act a little' 'think a little' 'act a little'. This can be seen in naturalistic environments such as the one investigated here, which fits well with the sequential processing described earlier.

Methods of study

In using information and wayfinding within complex environments the journey chain begins at point A (the desire to take a journey) and ends at point B (successful completion of the journey). In between these points the traveller requires information for: 1. Learning, 2. Planning, 3. Starting, 4. Entering, 5. Ticketing, 6. Waiting, 7.Boarding, 8. Journeying, 9 Arriving and 10 Continuing (Myerson 2001). All ten stages require the provision, assimilation and use of information. Whilst there is much research that could be carried out on all of these stages, including pre-booking, planning and assessment of information (via web and phone based information services such as National Rail Enquiries), this study focuses on stages 4-9 where the passenger accesses station infrastructure-based information. The research set out to gain insight into behaviours that may assist in 'bridging the gap' with regards to the designer's (Network Rail) and the users' (older and visually impaired) representations of the system.

Purposive information from sampling within set boundaries and a constructed framework was used, based on knowledge from prior research. This theory-driven approach sampled participants using a snowball/chain method (Miles and Huberman, 1994) from contacts at the RNIB and

older adults groups in Islington, London. The use of observation and interviewing techniques and generative processes for relatively few individuals (due to limited access to older and visually impaired users) led to a focus on depth and quality rather than breadth and quantity.

To pilot the journey, a shadowing/walkthrough process was taken using expert users (from Network Rail's Ergonomics team) who carried out walkthroughs using empathic tools. Verbal protocol knowledge elicitation gathered expert users' decision processes regarding their information and wayfinding experiences. Using Hitchcock et al.'s (2001) 'third age suit' for inspiration, 'empathic glasses' and earplugs simulated sensory impairments via a 'sensory ageing technique'. Glasses were designed to simulate cataracts based upon RNIB website images (2005) with reduced acuity, increased glare sensitivity and visual field restrictions. Earplugs crudely mimicked hearing loss. This technique allowed stakeholders to view the world outside of the 5-95th percentile perspective and enabled an empathic understanding from living the situation for themselves.

Following the pilot study and discussions with subject matter experts at Network Rail a journey was set. Users were observed taking the set journey and this process was recorded using audio and video equipment and the subsequent data collected, analysed, interpreted and represented. Six participants completed the same route in London and the southeast region with a single route being chosen to analyse how information and wayfinding was utilised in one major station, one interchange station and an interchange with London Underground. The in-depth analysis of one route, whilst not directly being transferable to all stations, provided a case-study approach regarding a major station analysis and comparative approaches to systems use. It also facilitated the use of group workshops with older users to reflect on the experience.

In conjunction with the observational data, an analysis (in an annotated table format) of the Guidelines for Design and Architecture from the Major Stations (2004) was carried out in light of the data gained in the participant journeys.

Following the journey shadowing process, individuals attended user workshops where general issues around transport accessibility and the journey were discussed alongside some creative work to gain users' 'feelings' and understanding of the process. These results, along with observational analysis at three stations - Euston, Paddington and Watford Junction, fed into the discussion and analysis of the project.

The data were gathered by multiple converging techniques as follows:

1. Decision tree (Euston overland compared to Euston London Underground Limited – L.U.L).
2. Annotated Station Observations – Three station observations were undertaken at Euston and Watford Junction Station (the two main stations in this study), and Paddington Station which is seen as a 'best practice station' (Network Rail, 2004).
3. Informal interviews – The researcher gained access to many stakeholders (inside and outside Network Rail) involved in the provision of travel information, design and day-to-day workings of the infrastructure. These meetings took place with prior notice and the researcher prepared informal semi-structured questioning around topics, covered either in person or by telephone.
4. Link Analysis – A Spatial Link Diagram format was chosen to represent spatial movement within the routes taken by users in the process of wayfinding around Euston station.
5. User workshop – Older users attended a workshop in which data provided complemented the contextual information gathered from the journey shadowing (Fulton-Suri and McLelland, 2005).

Results

The six participants who took part in the study had a variety of impairments, which challenged their ability to wayfind and access information. The two younger visually impaired users were both fully mobile but with varied limited vision (with one being a guide dog user and one being a cane user). The four older participants ranged from age 63 to 84 with varied visual, mobility, hearing and balance related deficits. Each user encountered unique issues on each journey. The environments, in both the underground and overground at Euston station, were analysed and represented in a decision tree format to aid the understanding of decision choices within each environment. Euston rail station proved a far more complex environment due to a wider set of options at each decision point when wayfinding, whereas the major tunnelling and architectural constraints guide the user within the underground station (Figure 3.1).

The data gathered from the verbal protocol recordings and journey walkthroughs provided 'hotspots' for each user. After transcription and coding, themes represented via comparative timelines were formed that illustrated specific areas for concern and improvement.

Verbal protocol data also provided rich personalised accounts of problems encountered on the journey: frustration encountered by Elizabeth (visually impaired) who pointed out that 'I shouldn't have to be asked if I could see properly' (by an LUL staff member, when being led with her guide dog); Ahmed and Hamu (older participants) choosing the train to Watford 'seeing it is the third stop', when it is actually the 16th, due to misunderstood scrolling effect on the Customer information screen-CIS and blaming their own stupidity. This representation provided an understanding of the individual effects that design influences and a complex time pressured environment can have on the journey process. Commercial units blocking clear access to information led some of the older

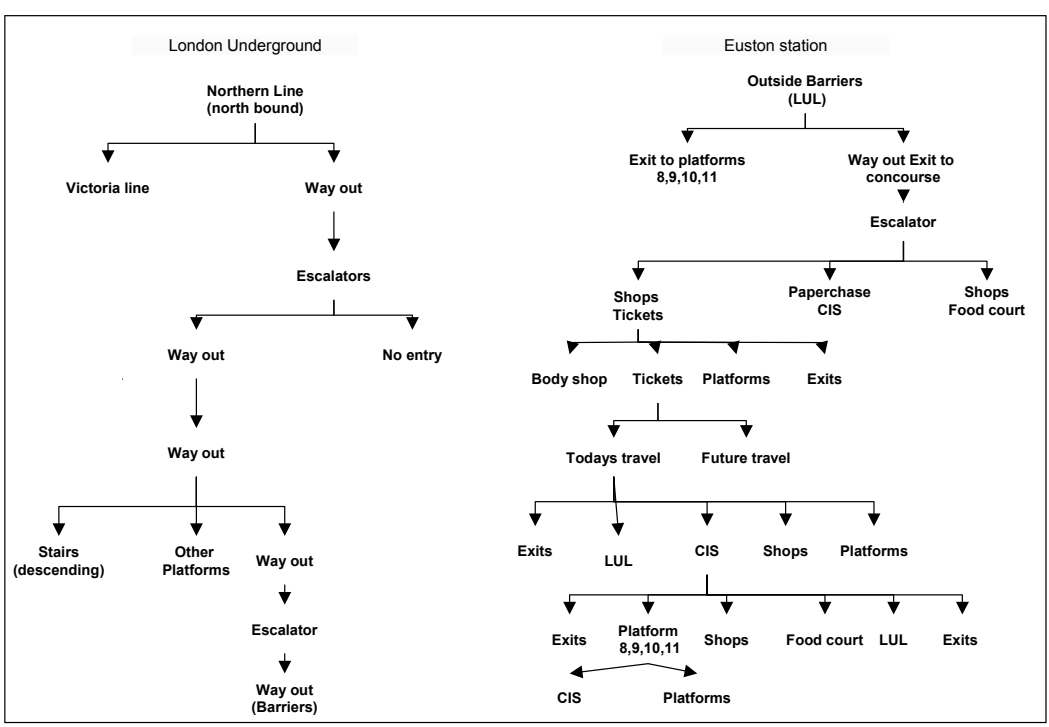


Figure 3.1 Decision tree for wayfinding

participants to wander around the environment, whilst the visually impaired wandered due to poor tactile and auditory cues. These issues represent only a tip of the iceberg.

Perceived problem points or 'hotspots' encountered in the journey process are represented below from most to least problematic:

1. Front of train information
2. CIS information
3. Line of sight
3 Signs provide too much information
3. VDU monitors
3. Route maps
3. Paper timetables.
8. Directional arrows on signage
9. Station maps
10. Too much general complexity
11. Ticket Barriers
12. Station architecture
13. P.A, Staff
14. Information – where to find it?
15. Other passengers

Discussion

All of the participants formed plans in order to navigate from A to B relying on sequential information provided at each decision point en route, due to their naivety in the environment. Of primary importance to the older users in the study (who could access visual information) was the complicating factor of poor line of sight in both Euston Rail and L.U.L. systems. Whilst being directly related to unchangeable architectural constraints within L.U.L., within Euston mainline station it was mainly due to commercial unit placement and related commercial pressures. At Watford Junction the homogeneity of the physical architecture led to individuals developing poor conceptual models of their spatial whereabouts, and with few landmarks to anchor their location wayfinding relied totally on signage and assistance (Figure 3.2).

a b c

Figure 3.2 Main Station sites: Watford Junction (a),Paddington (b) and Euston (c)

Line of sight problems led to wandering behaviour and trial and error responses when wayfinding. Wayfinding techniques of the participants were at times unsuccessful due to a number of barriers, which in themselves may be due to a number of design issues.

The information points/kiosks in the major stations provided good information content from the staff who manned them, however the lack of consistency of the physical presentation and poor lighting meant that the ageing adult had difficulty in finding the correct location for the information that they required (Figure 3.3).

Furthermore, information from station paper timetables was deemed too complex and the text too small. To older users, pictograms were often not visible (due to size) when considering the other competing information in which they were often engulfed.

Another key issue was the visibility of ticket barriers in L.U.L. and Euston station. The time constraints enforced on passengers to rapidly flow through barriers led to rapid decision making requirements up to and beyond the capacity of our participants. Ticket barrier interface design proved confusing to older users and those who were unfamiliar with the system (e.g. foreign passengers). One participant Rita, managed to insert her freedom pass (oyster card) in the machine and render it out of use for over 5 minutes. On a more positive note, exit signage visibility was generally very clear due to its consistency (yellow on black).

Architecturally, ramps effectively linked the concourse, and the platforms and escalators and stairs positively afforded exit and entrance to interchange areas via architectural cues. Participants were guided by these and also relied on the flow of the passenger population, but at times to their detriment. These cues could only be accessed by the visually impaired if guided by staff, as tactile cues were limited in the mainline stations. Poor architectural management and signage at Watford junction on the other hand directed users down what appeared to be a back alley leading to confusion and the feeling of being 'lost' in the environment.

System feedback was a major stumbling block for all users especially at decision points. Information on the front/side of the trains was often inconsistent or non-existent, whilst information conflicts – such as between platform number information and car stop numbers – at times led to unclear understanding of the system state (Figure 3.4). The lack of tactile information was apparent in L.U.L. and over ground stations and thus the visually impaired were often not kept 'in the loop'.

Older participants cited London underground's consistent use of colour as re-assuring in relation to mapping, whilst in Euston station culturally embedded mapping techniques led to all older users using wayfinding signage (static) from the top down, and not the bottom up as designed by Network Rail (2002).

Figure 3.3 Where to get information? – Varied information points across the stations

Figure 3.4 Car stop signage can be confused with platform numbering

The execution of users' action plans were complicated by line of sight issues and the multiplicity of forms of information presentation – the CIS, paper timetables, maps, informal/formal instructions and advice from other people. Signage positioning at decision points proved problematic at Watford Junction, where unsupported points left the user lost at times. Physical restrictions to executing plans could be seen at ticket machines and ticket barriers. This all leads to a gulf of execution. When evaluating, passengers found specific difficulties with front of train information, map orientation, CIS and on-board information. Whilst assistance was at times excellent, at other times it was confusing and even incorrect (especially when informal information was accessed). This leads to a gulf of evaluation. In many cases, both the older and the visually impaired users blamed themselves for design failures in the system (Norman, 1999). When failing to access the information effectively on the CIS or VDUs, the older users blamed their own 'stupidity' or 'clumsiness', they blamed their 'eye sight' and 'inexperience'. When complexity and lack of line of sight called upon them to wander in trial and error, they called for themselves to be more 'patient'.

The 'best practice' station (Network Rail, 2004), Paddington, revealed a number of positive benefits with regards to information design, but also some poorer elements of accessibility and wayfinding (Table 3.1).

This suggests that each station has strengths and weaknesses, and individually needs assessment.

Potential improvements

Norman (1999) suggests seven principles to transform a difficult environment into a more simple one to use, as follows:

Table 3.1 Best practice appraisal

Paddington positives	Paddington negatives
1. Superior line of sight than at Euston. 2. Less competition in wayfinding signage and advertising. 3. Commercial units, generally, are situated away from the platforms. 4. Good clear platform signage is present – lit and large and consistent. 5. Staff visibility is clear and consistent (yellow vests). 6. Level access in the main concourse through to the taxi rank. 7. Separated retail area. 8. V.D.U.s on each platform – consistent and clear. 9. Excellent mobility assistance unit.	1. Generationally platforms 11-14 have a very poor appearance and appear detached from the main station. 2. Confusing entrance when approaching from the bridge – Hammersmith and City Underground entrance. 3. Long walk distances to access platforms 11-14. 4. Two C.I.S. screens whilst being beneficial for line of sight and access, have a secondary issue in that they lead to lateral pinch points within Paddington station regarding flow.

1. Improving knowledge in the world.
   a. Make information points consistent.
   b. Increase front of train information provision.
   c. Make CIS presentation in the major stations consistent.
   d. Make VDU screen information consistent and improve visibility.
   e. Improve PA clarity.
   f. Make tactile guidance more available and more consistent.
   g. Improve line of sight for passengers in the concourse.
2. Simplify the structure of tasks.
   a. Zone differentiate – The use of block colour code zones to improve consistency – (for example at station entrance and the ticket office).
   b. Provide a clear line of sight to support users ability to form cognitive maps of the environment.
   c. Improve the clarity of information.
   d. Minimise time pressures in information provision e.g. tickertape and scrolling.
   e. Improve barrier design in terms of visibility and support users mental models of usage e.g. ensure constraints do not allow users to input incorrect tickets/cards.
   f. Provide static wayfinding information in smaller chunks with salient information provided from the top down.
   g. Make timetables simpler and increase text size.
   h. Provide station maps based on simplicity and always orientate them in the direction of the station as approached.
3. Bridge the gulf of execution and evaluation (see earlier explanation).
4. Provide effective mapping. Use colour more effectively as in L.U.L.
5. Exploit the power of constraints – use line of sight and architectural constraints positively to assist wayfinding so that the user can be supported in assisting flow through the station.

When complicated by competing commercial units however negative constraint effects complicate the environment and limit a user's ability to find their way easily.

6. Design for error. Be aware that poorly mapped environments can lead to passengers wandering into areas of the station that are unsafe and out of bounds. Clarification of such zones is imperative for safety.
7. When all else fails standardise. This should be considered across zone clarity, CIS, information points, staff information, uniforms, maintenance, and line of sight.

These potential improvements will assist all users but especially those who rely on their abilities within the constraints of reduced vision, mobility, auditory abilities and cognitive processing capacities.

In simplifying the process of station wayfinding the ease of line of sight usage seems paramount. This could be further improved by a clarity of zoning based on block colour representation that illustrates to users (including those with visual impairment who often use natural light or block colours to navigate) to see where specific zones are.

Landmarks provide meeting points for passengers, with the current international meeting point symbol appearing unsuccessful as passengers were observed using commercial landmarks, e.g. WH Smith, which due to their position were seen to impede station circulation flow. Effectively positioned landmarks could remove this problem.

On platform and on-board train information is required in a consistent manner for all users. Station and route maps could be simplified, enlarged and lit. Also the removal of time pressures, exacerbated by tickertape and scrolling information from the CIS boards, should be considered.

Conclusions

Major train terminals present the passenger with an environment that prevents simple perception-action processing, due to the complex, time pressured and dynamic environment.

Older passengers may be dealing with certain physiological changes and to ensure that they are not excluded from these environments, designers must consider their needs in ensuring accessibility. The results of this study show that not only are older individual's expectations less, due to their perceived difficulties, but also the design of these environments is far from ideal for their varying capabilities. Whilst the DDA (1995) suggest that infrastructure providers must think creatively in removing barriers to such users, this work has illustrated areas where there is still much more that can be done.

This research incorporated direct feedback to the Network Rail Design and Architectural team and to the older adult population of Islington via talks to pensioner groups in the community. Whilst the Design and Architectural guidelines (2004, 2002) and SRA's guidelines (2005) are theoretically sound, what is apparent is that 'in the wild' all is not quite as clear-cut when considering information accessibility. The nature of an increasingly complex environment makes wayfinding and information provision a tough issue to manage and one in which generational improvements will require fighting for.

References

Coleman, R (2001). *Living Longer – The new context for design*, published by The Design Council pp. 1-55.

Disability Discrimination Act (1995), at http://www.drc-gb.org/thelaw/thedda.asp.

Department for Transport, (2004) National Rail Conference, www.dft.gov.uk/stellent/groups.

Fischcoff, (1988). Quoted in Wickens, C.D, Gordon, S.E, Liu, Y, (1998) *An Introduction to Human Factors Engineering*, Longman, New York, p185.

Fisk, J (1993). Design for the elderly: A biological perspective, Applied Ergonomics, 24 (1), pp. 47-50.

Fulton-Suri, J. and McLelland I., (2005). *Involving people in design*, quoted in Wilson, Evaluation of Human work, pp281-333.pp. 8-13.

Good Grips ™, in Keates, S. and Clarkson, J. (2004). *Countering design exclusion: An introduction to inclusive design*. Springer, UK.

Gyi, D. Porter, M. and Case, K. (2000). Design practice and 'designing for all', Proceeding of the IEA, 2000/HFES 2000 congress.

Hitchcock, D.R., Lockyer S., Cook, S. and Quigley, C. (2001). *Third age usability and safety – an ergonomics contribution to design*, International Journal of Human-Computer Studies, 55, pp. 635-643.

Hollnagel, E. (2003). *Handbook of Cognitive Task Design*, Lawrence Erlbaum Associates, London.

Klein, G.A., Orsanu, J. Calderwood, R. and Zambor, C.E. (1993). *Decision making in action: Models and methods*, Ablex publishing corporation, Norwood, New Jersey.

Laslett, P., (1989). *A fresh map of life: the emergence of the third age*, Weidenfield and Nicholson, p. 71.

Miles, M.B. and Huberman, A.M., (1994). *An expanded sourcebook: Qualitative Data Analysis,* 2nd ed, Sage, London.

Mollenkopf, H., (1997). *Cost 335 – Passengers' accessibility of heavy rail systems*, proceedings of the seminar, Brussels 30-31 October. EC directorate – General for transport. pp. 23-32.

Myerson, J. (2001). IDEO ;Masters of Innovation, Laurence-King. London.

Network Rail, (2004). *Railway Estates managed stations design guidelines*, Network Rail, Railway estates.

Norman, D.A. (1999). *The Design of Everyday Things*, 2nd Edition, Basic Books, New York, pp 1-219. a.

Passini, R. (1980). *Wayfinding: A Conceptual Framework*, Urban Ecology, 5, pp. 17-31.

Rail Track (2002). *Major Stations Wayfinding: Summary design guidelines*, Issue 7, 15.8.02, Design Research Unit.

Ricability (2005). *Meeting the needs of older and disabled customers*, http:/www.ricability.org.uk/reports/report-design/guidelinesforproductdesign.

RNIB (2005). Images, http://www.tiresias.org/controls/visual_impairments.html.

RSSB (2005). *Research into signage and wayfinding at stations*, unpublished

Simon, (1957) quoted in Wickens, C.D., Gordon, S.E, Liu, Y, (1998). *An Introduction to Human Factors Engineering*, Longman, New York, p. 185.

Strategic Rail Authority (SRA), March 2005. *Railways for All :A consultation document – delivering more accessible railways.*

Summers, A. (2001). From '*Living Longer – The context for new design'*, Published by The Design Council.

Tiresias (2005) documents at http://www.tiresias.org/reports/inter.htm.

Woodhuysen, J., (1993). *A call for Transgenerational design*, Applied Ergonomics, 24, (1) pp. 44-46.

This page intentionally left blank

Chapter 4

Improving the Content and Placement of Anti-Trespass Signs

Nicola Forsdike, Claire Turner, Fiona Bellerby, Sue Nelson and Paul McGuire

Introduction

The need to consider the design of anti-trespass signs was re-highlighted by a fatal accident to a trespassing and fun-seeking teenager who was killed when he was hit by a train. Although there was adequate fencing that the point by which he gained access, his mother proposed that had there been a warning sign at the point of entry this might have acted as an additional deterrent.

The placing of 'do not trespass' signs has not traditionally extended to locations such as fence lines adjacent to bridge abutments, rather they have been placed at platform ends, access points and level crossings. Moreover, a recent project undertaken by Halcrow and Human Engineering for the RSSB, which looked at trespass and access from the platform end, raised questions as to the adequacy of the messages conveyed by the current signage. This doubt is strengthened by other studies which have shown that messages must be closely tailored to their target audience if the railway is to influence behavioural change in the public.

Trespass on Britain's railway was responsible for 34 deaths in 2004 (Source: RSSB Annual Safety Performance Report 2004). Of these deaths, 2 were children (both 12 years of age). While the past 5 years have seen a downward trend in trespass, incidence remains high. Over 11,000 incidents were recorded in the Safety Management Information System (SMIS) last year alone, and the actual level of trespass on the railway is thought to be much higher than this. A research study conducted in 2001 concluded that there are an estimated 10 million offences of criminal trespass on Britain's railways committed annually by children and 1.3 million children who are current and repeat offenders.

The prevalence of youth trespass ensures it remains high on the rail safety agenda and the industry takes a multi-faceted approach to dealing with it, including educational campaigns and cross-industry partnerships. The cornerstone of the approach has been the 4Es strategy that has led to rail trespass being managed in line with four key strategies, that is:

- Enabling – Developing the framework for giving people the tools to do the job
- Education – Awareness of the issues among young people and other relevant audiences
- Engineering – Finding engineering solutions and designing out the problem
- Enforcement – Deterring, detecting and apprehending offenders.

Objectives

The objectives of the research were to establish:

- The effectiveness of the existing policy regarding the design and placement of do not trespass signage, and to identify possible improvements.
- The extent to which current do not trespass signs convey the consequences of trespass effectively
- Whether the current placement arrangements of do not trespass signs is effective
- The development of alternative, more explicit, do not trespass signs
- Testing of these signs to determine which design best conveys to target audiences the potential consequences of trespass
- Potential benefits that might be obtained by placement of the best alternative do not trespass signs at locations where traditional do not trespass signs are located
- Benefits of extending the provision of the alternative signs at other locations, where those intent on trespass access the railway
- Development of a cost effective, criteria based, approach to the deployment of alternative design do not trespass sign(s).

Method

The study approach comprised three main phases; an inception phase, encompassing the review of existing knowledge and leading to the development of messages and images for testing; qualitative research amongst the target audience, taking the form of focus groups, and; analysis, recommendations and the drafting of a protocol for practitioners.

Inception

Review of existing knowledge

Trespass is prevalent across a range of industries and a range of countries. The study sought initially to benefit from both experience internal to the rail industry and also from seeking knowledge from other industries and countries. In particular, the following sources were explored:

- Existing research studies, including work on the profile of trespassers (including age and motivation)
- Existing UK signage, both rail and non-rail (for example construction industry signage)
- Existing signage from other countries, both rail and non-rail
- UK anti-trespass initiatives, including work done by the Community Safety Support Unit (CSSU) and Network Rail
- Overseas anti-trespass initiatives, including Direction 2006 in Canada and Operation Lifesaver in the USA
- RSSB's Safety Management Incident System.

The review of research uncovered a number of characteristics common to young trespassers, which are summarised in Table 4.1.

Table 4.1 Profile of a young trespasser

Category	Characteristic
Age and gender	The majority of offences are committed by males aged between 8 and 14 (www.trackoff.org) Females typically are less likely to commit a trespass offence although there is research to suggest this trend may be changing. In the last 5 years the percentage of female child fatalities has risen from 9% to 24% (RSSB, 2003)
Reasons for trespassing	Incidence of trespass increases during spring, summer and autumn, during school holidays, especially Easter and summer. A survey of children (Risk Solutions, 2004) revealed the primary reasons for trespassing are: 'Taking a shortcut' – 27% 'Nothing else to do' – 28% 'My friends wanted me to' – 19% 'Lots of people play / hang out – 11% Research carried out by the Rail Passenger Council (2001) suggests that: It is unlikely that they are many, if any, young people whose criminal activities are focused on the railways; i.e. a vandal is just as likely to damage private cars as trains. Amongst young people, specific patterns of criminal behaviour might only last for a few weeks. In addition, a lot of trespassers are known to engage in acts of vandalism. The two crimes are inextricably linked in terms of the railway.
Location	A MORI (1998) poll identified that those youths attending schools closest to the tracks are the most likely to trespass and vandalise railway property, and the critical distance at which the number of offences begins to fall significantly is about two miles.

Images were obtained of existing UK anti-trespass signage, firstly from the rail environment. The signs are shown in context, in Figure 4.1.

Findings from this first phase confirmed that the UK is unusual in placing responsibility for preventing access on the railway, rather than on the individual. However, several overseas administrations use signage, for example at platform ends, to warn people of the dangers of trespass on the line. Examples are shown in Figure 4.2.

UK anti-trespass initiatives

Our consultation both on this and previous studies, has identified a number of initiatives being undertaken into deterring trespass. One such is a current project by the CSSU. In response to an approach from teachers, the CSSU is currently working with schools on a project that looks into the clarity and effectiveness of rail warning signs including for the dangers of voltage and electricity. Various symbols and designs have been tested amongst the study community on-line. Respondents have been asked to evaluate various designs and suggest where the signs should be placed. To date, school children have suggested fences, gantries on overhead cables, bridges and bridges that pass overhead cables.

Figure 4.1 Anti-trespass signs

Network Rail treats the issue of trespass with gravity; their summer campaign "No Messin" has formed part of the input to our primary research (see below). Other complementary work is being undertaken by Train Operating Companies (TOCs), including campaigns by MerseyRail that take them into a range of community groups from youth groups to mother and toddler groups.

Development of signs for testing

Based on the findings of our review, a range of images/text was drawn up and agreed with RSSB for further research. The range was drawn from firstly, existing signage without alteration and secondly, revised concepts based on best practice from other industries or countries. In addition, images were incorporated from work currently being led by the CSSU with schools, under which school children are themselves designing messages and images that could be used on signage. All

the designs were mocked up into flash cards to provide graphic images to be used as the basis for discussion in the focus groups. Signs were developed for use nationally and included signs that address specifically the electrocution risk in third rail areas. A variety of different concepts were selected to test a range of stimuli, including:

- Text alone
- Shape of signs
- Size of signs
- Different colour combinations
- Imagery
- Mix of text and imagery

Examples of the material developed for testing in the focus groups are presented in Figures 4.2 to 4.8.

Figure 4.2 Anti-trespass signage from outside the UK

Figure 4.3 Existing UK anti-trespass signs

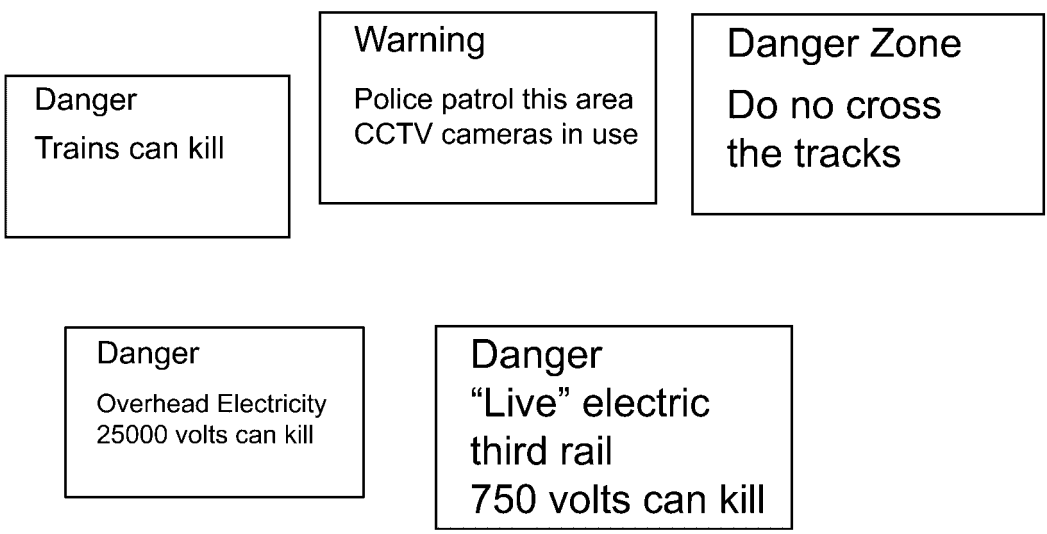


Figure 4.4 New text-only signs

Figure 4.5 New colours and shapes

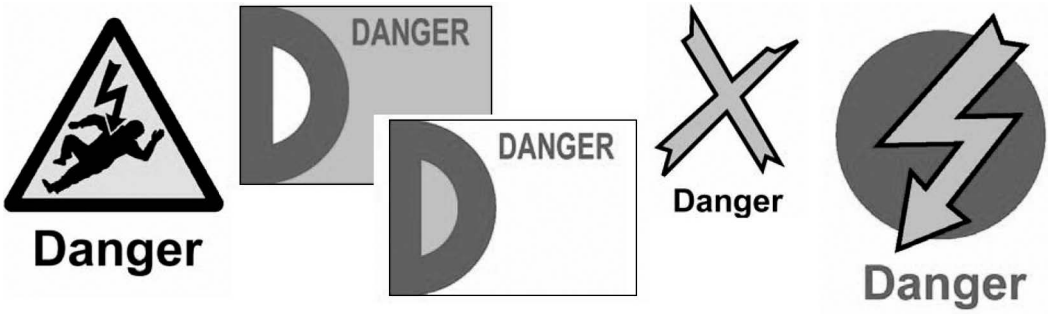


Figure 4.6 New graphics

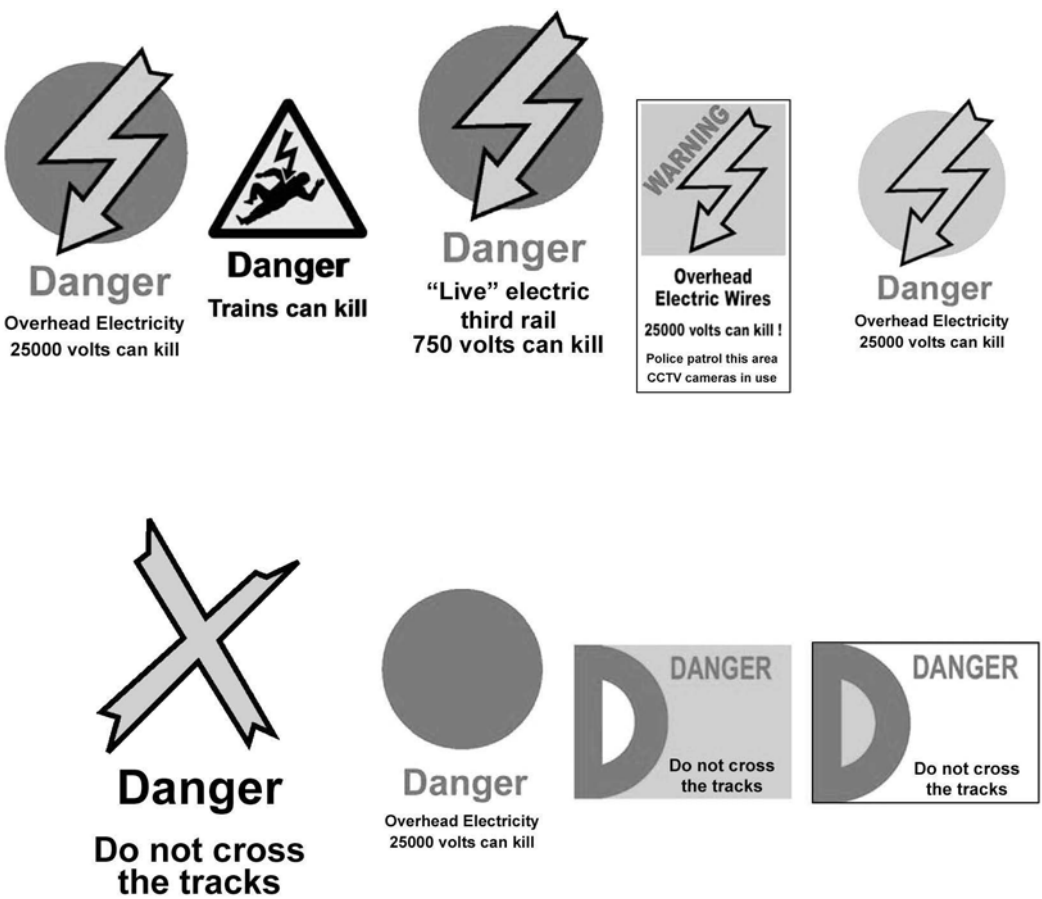


Figure 4.7 New graphics and text

Figure 4.8 Existing UK posters and youth campaigns

Focus groups

The crux of this study was testing the reaction of the target audience (that is, young people) to the stimulus material. Youth research is a specialist area, requiring extremely experienced moderators for both undertaking the research and analysing the results. Young respondents are sensitive to being patronised and a responsive approach, especially with this type of target audience, was required. For this reason, a specialist partner, Quaestor Market Research, was engaged.

Based on the findings of the previous phase, a recruitment profile was built up for the focus groups. Accordingly, recruitment focused on youths who:

a) Are known to congregate in large groups (due to the large influence of peer pressure).
b) Come from areas where there are few other attractions for children to engage in e.g. community centres, clubs, activity camps during holidays etc.
c) Attend schools within 2 miles of a railway line.
d) Could be considered generally "good" kids who trespass because they are bored with nothing else to do and are likely to hold the view that "If I'm caught my mum and dad will be angry and I will be in trouble with the police". (Signs are most likely to have a bigger impact with this group than with existing frequent trespass offenders for whom the threat of getting into trouble with their parents, and even the police, is not sufficient disincentive).

Friendship pairs were recruited to encourage discussion and aid participation and reliability. Groups were limited to no more than eight, to encourage active discussion, debate and, importantly, honest open responses to the topic area. Given the age range of the target market coupled with the 'sensitive' nature of the research (i.e. mindsets and attitudes towards trespassing) the original plan was to undertake 90-minute focus groups, each focusing on a small age range. Participants in the

focus groups were split into the following groups, allowing more homogenous age breaks in the youth sample to be focused upon, giving moderation, peer cohesion and analysis benefits:

- 7–9 years old
- 10–12 years old
- 13–15 years old
- 16–18 years old

The younger age groups were identified as most at risk for being tempted onto the railways, whereas the older groups were identified as most likely to commit trespass.

These groups were augmented with a group of parents, who either had children who currently trespass or who were worried that their children may trespass on the railways in the future. As hoped, this added a valuable dimension and introduced an angle on the issue of trespass and the effect of sign design that has acted as a useful catalyst for the analysis.

To encourage open participation, all groups were single sex except for the parent group, which was mixed gender. Based on the findings of the review stage, it was decided that the focus groups would concentrate on males, as being most at risk from trespass; however, the review also raised concerns about the activities of older females, with the result that an additional focus group was added, of girls aged 16-18.

Recruitment for the groups took place in locations where trespass onto the railway was identified as likely to be a risk. An initial analysis of RSSB's safety management information system (SMIS) suggested that over 11,000 incidents of trespass occur per annum and that geographically, these are very widespread. Further insight was therefore sought, and Network Rail's Community Liaison teams consulted to identify the most promising locations for research.

Three contrasting trespass hotspots were identified and deliberately selected to allow for exploration of third rail and non-third rail issues, and possible differing regional attitudes to child/youth "freedom". Table 4.2 below illustrates the location and composition of research groups.

Table 4.2 Focus groups

Location	Age	Gender
Crawley, Sussex	7–9	Male
	13–15	Female
Cardiff East	10–12	Male
	16–18	Female
Leeds	16–18	Male
	Parents	Mixed

The content of the focus groups was meticulously planned and covered the topics shown in Table 4.3.

Table 4.3 Focus group structure

Discussion Areas	Time
Intro/warm up	10 minutes
Purpose of the session	5 minutes
Trespass behaviour	15 minutes
Awareness of signage	10 minutes
Evaluation of new and current signage	40 minutes
Summary advice	10 minutes

Detailed discussion guides were drawn up prior to the focus groups to ensure that the following topics were explored at length:

- Current patterns and attitudes to trespass on the railway
- Attitudes towards different types of signage e.g. signs conveying a strong personal warning, rather than a "don't do" message that can encourage misbehaviour
- Reactions to specific verbal messages, both current and potential
- Reactions to specific visual messages
- The extent to which particular combinations of visual and messages have the potential to alter behaviour
- Whether sign size plays a role in perceived importance of the message conveyed.

To generate interest a number of projective techniques were employed within the groups such as a 'sign quiz' (using current and proposed designs), to assess reactions to and understanding of trespassing. The different types of signs were also rated to allow a quantitative basis for comparison of the different options.

Results

Initial findings suggest that the output is likely to be thought provoking and likely to lead to considerable debate within the industry – and ultimately, we hope, to a change in practice. In particular, initial results confirm that current signage, whilst meeting statutory requirements, is ineffective in actually reducing trespass. There was general consensus across all groups that the signage did not convey the extent of the risk and threat posed by trespass on the railway. Not surprisingly there was variation within the different age groups with regard to the type of stimulus that might prevent trespass. However, there may be a case for much starker signage, conveying a very simple message through both words and image.

There appeared to be some awareness of the risk of trespass amongst target groups; however, there was a dis-connect. Although apparently well aware that trains could kill you if they hit you, none of the respondents knew anyone to whom this had happened. As such, none of them believed that it would happen to them. Some spontaneously mentioned school talks on other safety issues, but none could recall a visit that covered the risk of trespassing on the railway.

Worryingly, the parents also tended to underestimate the risk, with some admitting to taking shortcuts over railway tracks with their children (an activity evidenced in the SMIS records), thus

re-enforcing the attitudes and behaviours. One outcome of this is that, unconscious themselves of the risk, they have never taught their children the dangers of railway trespass.

Different groups were found to respond to different threats. Of all the factors deterring older teenagers, that most likely to affect their behaviour was the risk of being picked up by the police and getting a criminal record. Signage alerting them to police patrols and other activity is thus potentially effective; however, once again, unless this is backed up by experience then the threat will lose its effectiveness. Signage alone is not sufficient.

Finally, early analysis suggests that many incidents of trespass and access arise as a result of boredom. Younger children lacked even the basics of somewhere to play football, choosing, for example to play in the station car park, a large open space, which was well lit on dark evenings (again, SMIS lists many incidences of children and teenagers playing football on station platforms). Older teenagers are desperate for a space of their own ("somewhere where we can just be ourselves") where they can hang out and talk, smoke and drink alcohol away from their parents. This suggests that the issue is not just a railway issue, but involves the wider community.

Analysis and recommendations

Once the results of the primary research are fully analysed and written up, draft recommendations will be developed for discussion with the RSSB and stakeholders.

In particular, recommendations are likely to consider:

- potential changes to text and imagery
- holistic context – need for education and enforcement
- wider approach, potentially looking at displacement activity
- any further testing required.

Finally, a protocol will be written up for use by practitioners. It is envisaged that this will include a cost effective, criteria based, approach to the deployment of alternative design do not trespass sign(s) and will guide the user to select the most appropriate locations for deployment of such signage. It will allow the prioritisation of investment based on evidence gathered both from the focus groups and the literature review carried out in the inception phase of this project.

References

http://www.trackoff.org/research.html.
http://www.railwaycrime.org/facts_hotspots.asp.
MORI (1998). *Trespass & Vandalism*. A project for Railtrack plc.
Rail Safety Standards Board (RSSB) Annual Safety Performance Report 2003
Rail Passengers Committee Scotland (2001). *Trespass and Vandalism: Overview of HotSpots*. October 2001.
Risk Solutions (2004). *Route Crime – Industry Perceptions*. A Report for Railway Safety, June 2004.

This page intentionally left blank

Chapter 5

Travel Information vs Commercial Signboards: The Battle between Travel and Shopping

Answina D.P. Zwaagstra

Introduction

Rail Human Factors focuses largely on working conditions. But apart from personnel, there is another important 'rail human factor': the passengers. During our consultancy work, mainly for the Dutch Railways, many projects and studies are conducted that focus on passenger needs. For example , there have been studies on travel information, pictographs, ticket vending machines, social safety in stations etc. As can be expected, there are topics where passenger needs are conflicting with the needs of personnel or the needs of commercial benefits.

Here we describe a project that involves the issues of travel information and of way finding versus commercial information in railway stations.

Three parties were involved in this project:

1. ProRail, the organization that owns the station buildings and is responsible for the interior and maintenance of the buildings.
2. NS Reizigers, the train operating company that is responsible for an optimal flow in station buildings in order to realize efficient rail traffic.
3. The NS Stations, the organization that is responsible for all commercial activities in railway stations and has to make the station buildings profitable.

These three parties all have their own, sometimes conflicting, interests. ProRail's main concern is that the character of the building as being a railway station, is guaranteed at all times. In the eyes of NS Reizigers, a station is a place for transfer and the primary purpose of a station building is to meet this function. NS Stations considers a railway station as a place where many people are present requiring travel related facilities (trains, ticket office, luggage depot etc.) but also other services like shops, restaurants etc. Staying in a station building should be as pleasant as possible.

Scope

Because of the conflicting interests of the three parties, there is a need for an objective method to define the right balance between the of types of information. There is limited space for information presentation and also placement of information in a specific location influences the effectiveness (conspicuity, clearness) of other information.

The methodology has to be applicable to all station buildings, of all shapes and sizes (Figures 5.1 and 5.2) and should outline how choices have to be made concerning the application of revenue-earning advertising in relation to travel information. Furthermore, the three parties agreed upon an objective method that should give an independent guideline and therefore needs to be based on cognitive ergonomic principles.

Figure 5.1 A big station building the historic Amsterdam Central Station

Figure 5.2 The smaller Hengelo Station

The purpose is to develop a method that will be described in a handbook. The handbook will only be a success if all parties agree on the method described.

This handbook will be a tool for facility managers, who are responsible for the layout of a station building. Facility managers can use the method in new station buildings as well as for redesign of the current layout. For each part of the station building the handbook gives guidance how to apply information presentation methods and explains the basis of the psychological principles beneath them. The handbook provides insight into which means of information are suitable in the different parts of the station and why.

Approach

In a brainstorm session with representatives of the parties involved, basic principles were defined. The most important starting point is the mutual understanding on the relationship between travel information and commercial information: both are equal. Without travellers, there is no sale, without profitable activities there is no station.

Other principles where the three parties agreed were:

- Guarantee an optimal flow of travellers;
- Create a positive visual experience of the station area;
- Guarantee physical and social safety;
- Maintain architectural harmony in the station building.

In order to develop a pragmatic handbook that describes the method for correct placement of information, the following steps were taken:

- Assessment of the current situation on several stations;
- Formation of a work group where all involved parties participated; Participation of the three parties was essential to sustain commitment to the outcome of the project.
- Development of human factors based principles relevant for information presentation;
- Research on human behaviour in passenger flow and in information processing;
- Integration of human factors principles with the interests of the other workgroup members;
- Abstracting station building features in a visual presentation;
- Application of human factors based principles on the means of information.
- Drawing up the handbook;
- Testing the handbook in several stations.

As the figures 5.3 to 5.6 illustrate, travel information and commercial information are often sited close together and can overshadow each other or make each other less visible. The outcome of the scoping research showed that apart from Dutch psychologist H.J.G. Zwaga, there is not much research done on this subject (Boersema and Zwaga, 1985; Zwaga and Easterby, 1984; Zwaga et al., 1998).

The method that Intergo developed is based on human factors principles. It is important to provide passengers or visitors the station building only with the information they need in a way that is effective, efficient and satisfactory. The information should therefore be easy to find, easy to read and should not be disturbed by information that is not relevant at that moment. In short, the approach is based on knowledge of:

- Visual perception, eye movement;
- Conspicuity;
- Decision making;
- Human behaviour in traffic;
- Information needs;
- Effects of the environment on human behaviour;
- Effects of experience on visual perception and way finding.

Figure 5.3 Travel information and commercial information are in the same line of sight

Figure 5.4 Travel information is square to the passenger flow, advertising is parallel to the passenger flow

Figure 5.5 Confusion: The bookstore advertises right above ticket vending machines

Figure 5.6 Travel information and advertising have the same orientation with the advertisement dwarfing travel information

With the information from studying the current situation in station buildings and a thorough understanding of human behaviour in way finding, there was sufficient knowledge to resolve the issues of travel information versus commercial information, and develop a handbook. Several definition rounds in the workgroup followed to make a concrete proposal for the handbook:

- Domains: the parts of the station building where information can be placed;
- Principles for the way in which information must be placed;
- Defining relevant characteristics of information.

Results

Domains

In order to use an unambiguous method, station buildings are divided into three areas (station hall, platforms and tunnels and exterior) that are separated in several so-called domains. A domain is an area or surface which is suitable for the application of travel- and/or commercial information. The division of a station building in domains is determined by the relevance of these entities to information presentation. So every part that hypothetically can be used as a place for information is defined as a domain through members of the work group and an advertising company (see Figure 5.7 for examples).

Domains are often physical locations like a wall or a pillar, but can also be an area or zone in a station building. One domain has a priority over another domain (for example, the domain 'pillar' may not obstruct the domain 'information zone/airspace') depending on the relative importance of a particular domain in the way finding system in a station. A domain is defined as:

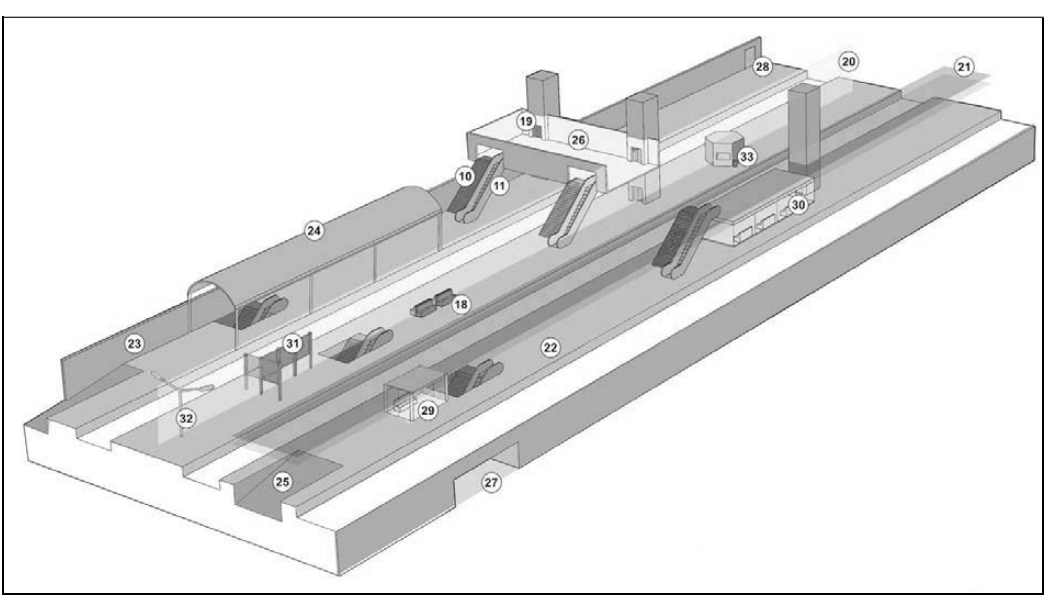


Figure 5.7 Illustration of domains in the area 'Platform and tunnels'

- A constructive or spacious entity;
- A static element, it doesn't move or change;
- A accurate specified unit, in terms of colour, size etc.;
- A unit that is perceptible or apparent.

Different domains are indicated; e.g. platform walls (23), platform roof (24), subway (27), platform shelter (29), waiting room (30) escalator (10,11), lift (19).

Basic principles for placement

When positioning information in a domain, four basic principles have to be considered:

1. Passenger flow; it is important that travellers and visitors to the station building can easily find their way through the station and there are no blockages
2. Safety; Not just social safety (in terms of crime or crowding) should be taken into account but also physical safety, like preventing the risk of falling.
3. Visual perception: nice to look at, pleasant, orderly, tranquil.
4. Architecturally consonant: the structure and outline of a domain has to remain clear. Certain forms of information are obviously only applicable at suitable places, so placing information on windows or works of art is not allowed for example.

The principles are not easy to work with directly so they are translated into information characteristics. Most of the time these characteristics are physical entities based on psychological principles.

Characteristics of information

Placement

- Orientation: squared or parallel to a wall or passenger flow. Information applied squared to the wall is best viewed from the passenger flow. People can see and read information while they walk. Travel information must be place squared to the flow, and commercial information parallel to the flow.
- Position in relation to the passenger flow. Travel information has to be readable from the main passenger flow. Information can be placed in the flow or along the side. Travel information should be in the passenger flow (if it hangs from the ceiling).
- Unobstructural view.

Different displays of information should not block each other's visibility:

- Transparency
  Transparent or translucent parts of the station building should not be covered by information displays. Transparency of a building has specific architectural, social safety or functional purposes, like the ability to view arriving trains.

Discrimination of separate means of information

- Distances between commercial information and travel information.
  The distance between different forms of commercial information have to be smaller than between commercial and travel information. As well as the content, a clear discrimination can be made between commercial and travel information through separations. The distance between all information is at least 750mm, for a sufficient separation between means of information.

Content

- Information density.
  Too much information invites people to stop walking and read and can therefore hinder passenger flow. The amount of information should be readable during walking at a normal pace.
- Emotional content.
  A strong emotional content in text or image can distract the attention away from neighbouring travel information too much or causes congestion of the passenger flow. Decisions on this must be made by the Dutch Railways.
- Recognition, disturbance.
  Colour combinations and the use of arrows indicating retail, that can be confused with directional way finding signage, are not allowed.
- Reference to retail.
  Reference to retail is allowed as long as it doesn't interfere with reference to travel information. Obviously, reference to retail must have distinct layout and colouring from travel information.

Format

- Form
  A variety of forms complicates recognition. This aspect is largely covered because means of information have mostly standard sizes. Use standard sizes for information presentation.

- Size, volume or surface
  3 rules
  a) Contours of a station building should not be covered but have to remain visible and recognizable.
  b) Conspicuity is mainly determined by size. Revenue–earning advertising should never oversize travel information if they are in each other's vicinity (fig 6).
  c) Limiting the number of different sizes to a maximum of two or three sizes is desirable.

Colour and colour contrast

- Colour has a strong influence on recognition of means of information and therefore the ease with which the desired information can be found. The different means of information (travel and commercial information) are indicated by using a standard poster frame for each type of information (in the Netherlands blue for travel information, black for revenue-earning advertising).
- Colour combination give a strong association with certain brands, whether these are retail or transport operators. (e.g. red and yellow is McDonalds, yellow and blue is the Dutch Railways). These colour combinations should be reserved for specific applications. Apart from the psychological aspect of confusion there is the issue of trademark protection.

– Colours influence each other. Avoid placing information on a background with the same colour.

Luminance

– Information displays should not be lit more than three times clearer than their surroundings. When the difference in luminance is too great, the information attracts too much attention and makes perception of non-illuminated displays harder.
– For the benefit of good readability, the contrast between letters or signs and the background has to have a minimum of 3 : 1 (maximum 30 : 1)

Movement

– Movement attracts attention. Moving means of information (like rotation displays) are not allowed in the vicinity of passenger flow or travel information. Obviously, reloading dynamic information displays are allowed.

For every domain, characteristics of the means of information that are relevant are specified and indication is given as to which means of information could be suitable.

As an example to illustrate this, the box contains description of the domain 'platform wall'.

Platform walls
1. Description of the domain

This domain incorporates all walls that in some way are adjacent to platforms. This can be a wall along:

· a side-boarding platform
· escalators,
· buildings on a platform.

Windows or glazed walls are not included in this domain. Platform walls can be in line with or squared to the platform.

2. Part of the surface that is suitable for information presentation:

It is allowed to apply means of information on this domain as long as the view on coming or going trains is not obstructed and it does not hinder passenger flow on the platform.

The distance between travel information and commercial information has to be bigger than the in-between distance. Travel information must have a clearly different manifestation from commercial information; and vise versa.

3. Domains relative to each other:

Travel information is only allowed if there is no other travel information within 900mm distance, on the same or a neighbouring domain.

4. Proportions of travel information and commercial information:

The balance between travel information and commercial signing is 50/50.

5. Relevant aspects of travel information

Brightness: Illuminated and with sufficient contrast with background;
Placing: not in direct vicinity of platform entrance because of blocking;
Placing: Direction indicating panels must be oriented squared to the wall unless transfer information is already mounted on the ceiling of the platform;
Placing: station name boards: parallel to the track.

Examples

- Train departure indicator
- Railway map
- Direction indicating panels
- Station name board

6. Relevant aspects of commercial information

Brightness not more than three times as clear as its surrounding.
Placing orientation against the wall. Or, if more than 300mm above the ground, squared to the wall is possible.
Content: signage to retail is possible as long as the lay out and colour do not resemble travel information.

Examples

- Signboards
- Light boxes
- Plasma screens
- Neon
- Stickers

Conclusion

With the development of the guidelines, we provided a tool for facility managers that will help them in determining the correct way of placing revenue-earning advertising and travel information in stations and how these means won't have to be at odds with each other. In this way, not only are the needs of the train travellers taken into account, also the need of displaying commercial information in order to attract customers is guaranteed, thereby meeting commercial benefits.

This handbook for the lay out of station buildings has been enthusiastically received. An agreement between the three parties will be signed in which they will commit themselves to the guidelines in the handbook. They are pleased that there is clarity on this issue, finally. In a pilot setting Intergo has reviewed a few stations according the handbook and has pointed out several improvements. Obviously it is not the intention of Intergo or the involved parties to provide a rigid set of rules. The handbook has to give guidance and provide arguments to resolve issues concerning the 'battle between travel and shopping'!

References

Boersema, T and Zwaga, H. (1985). The Influence of Advertisements on the Conspicuity of Routing Information; Utrecht: Psychological Labotory, University of Utrecht.

Zwaga, H. (Ed) (November 1998) Visual Information for Everyday Use: Design and Research Perspectives; Publisher: Taylor & Francis; ISBN: 0748406700.

Zwaga, H., and Easterby, R. S. (Eds.) (1984). Developing effective symbols for public information; Information design: The design and evaluation of signs and printed material (pp. 277-297). New York: J. Wiley & Sons.

This page intentionally left blank

PART III
Driver Performance and Workload

This page intentionally left blank

Chapter 6

Assessing the Impact of Increased Numbers of CCTV Images on Driver Only Operation of a Train

J. Wood, M. Freer, E. Grimes, N. Brook-Carter

Introduction and background

In driver only operation (passenger) (DOO[P]) of passenger trains, the driver's direct visual check that all train doors are clear, before moving out of the station, is currently carried out using platform mounted CCTV monitors, mirrors, or a physical look back along the length of the train. One TOC (train operating company) has introduced an in-cab CCTV system which involves fitting cameras to the side of the train and relaying the images into the drivers cab, see Figure 6.1.

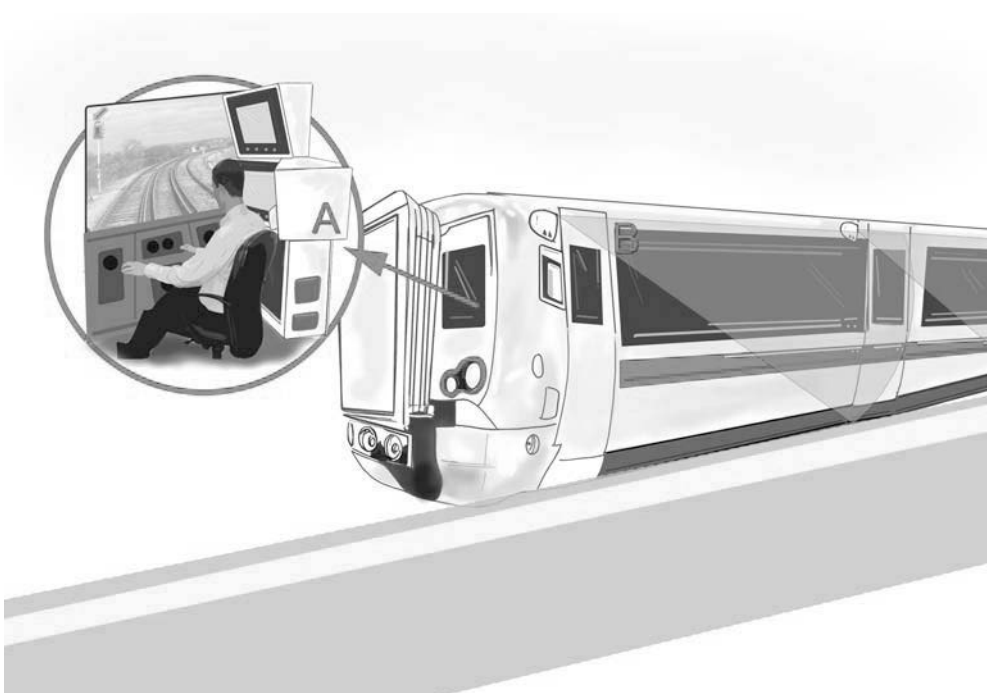


Figure 6.1 Arrangement for DOO(P) CCTV on Rolling Stock A) arrangement in cab B) viewing cone for externally mounted camera

Current railway standards (GE/RT8060) allow for up to 4 images to be displayed on a single monitor. Since each camera covers one coach, this restricts the maximum train length which can be operated with such a system to 4 carriages with a single monitor, and 8 carriages with 2 monitors in the cab. Space for locating additional equipment in the cab is restricted – the driver does still need to see out and operate the controls! At the most two flat screen monitors could be mounted in an ergonomically acceptable position in this cab. The wish by the TOC to run train sets of up to 12 carriages, using DOO(P), was clearly not practical without an easing of the existing standard restricting the number of images on a single display to four.

The objective of the CCD study was to examine whether increasing the number of images on a single screen would have any impact on driver performance in detecting items trapped in the train doors.

Preparatory work and experimental design

Early on in the programme it was agreed that determining the impact of increasing the number of images on a single monitor would be examined under laboratory conditions using a range of simulated targets. Working with the rolling stock engineers, the process of door closures and the existing safety measures, were reviewed. It became clear, from these discussions, that the scanning of the doors via a CCTV monitor was not the primary measure for ensuring passenger safety – interlocking of door closure with the application of power prevented the train from being moved off if obstructions of greater than 2.5cm were encountered. This still left the possibility of passengers being trapped by doors which had closed, say around the straps of a rucksack, where the primary safety systems would present a misleading signal in the cab. This would show that satisfactory door closure had been achieved and would allow the driver to apply power. For this reason a visual check of the closed doors, before taking power is always carried out.

Some further working assumptions were documented. It was agreed that only the 'safety' related aspects of DOO(P) CCTV would be considered. Discussions with the train operators had shown that the DOO(P) system could also be used for 'service' related train running – classically the driver looking for passengers running down the staircase leading to the platform. With this information the driver could make service related decisions about whether to let people onto the train or not. These considerations were excluded from the study. It was also agreed that the experiment would be based around the current protocols for DOO(P) operation. These took account of the fact that images were only switched on when the driver had stopped at the station and opened the train doors, and switched off almost immediately after the train had started to move off.

The railway standard suggests use of a flat 'grey board', measuring 1m x 0.5m as a target for checking the image quality of CCTV systems. It was agreed that such a board would be included as one of a series of 'targets' to be included in the trials. Following discussions with the project team the other targets were agreed – these are illustrated in Fig 2. These included a trapped person waving their arms, a buggy and a trapped person who was unaware they had been trapped, to simulate a passenger trapped by the straps of their rucksack – rucksack inside the train, passenger outside on the platform.

An initial study, carried out for the rolling stock manufacturer, specified how images could be presented to permit up to 12 images over 2 screens.[1] This included suggestions on how to deal with the blank spaces, where no image was to be expected, for train sets made up of less than 12 carriages. The experimental study replicated the actual presentation of images in the cab, Fig 6.3.

1 Evaluation of DOO CCTV Monitors Bombardier GoVia Class 375, HEL/BT/02686/RT1, 19 Feb 2002, Issue: 01.

Figure 6.2 Range of 'Targets' used during recording of test material: From top and clockwise 'trapped waving', trapped not moving', 'child', 'grey board', 'buggy'

Figure 6.3 Layout of DOO(P) images – left to right – 4, 6, 8 and 12 carriages

Two options were considered for recording the source material to be used for the experimental work. Recordings made with a 12-car train at the depot would allow the team under very controlled conditions to record their material. This would, however, require that 'passengers' would need to be simulated using actors and the variability arising from different stations and lighting conditions would be lost. This was weighed against the other option of using a live train with all the safety implications that this would entail. The decision was made to use a live train and to design a procedural protocol which would address potential safety issues that the project team had identified. The use of a live train, on the Victoria to Brighton route, allowed for variations in stations, lighting, and passenger numbers to be included in the source material. By selecting running times, starting in the afternoon, it was possible to cover both daytime and night time conditions.

A 12-car train was used for the trial. The experimental team was dispersed along the train and the following procedure was used for recording the targets/'incidents'. Once the train had stopped at the station, and passengers had finished boarding and alighting the train normally, the guard and driver coordinated the closing of the doors and satisfied themselves that the train was safe to go, following normal procedures. At this point a member of the experimental team would exit though a door controlled by a member of staff – with or without the appropriate 'targets' – and simulate an incident alongside the train. The time spent was about 10–20 seconds – long enough to get a satisfactory recording. They then returned via guard's door and when safely aboard, the train moved off, after the guard had again checked the doors. Apart from a slight delay in departing it is unlikely that the majority of the passengers were even aware of what was happening, although there had been some amusing comments about the paper mache 5-year-old child target!

The video material was edited into randomised sequences. These included a mix of target types, stations and pictures with no targets. The option of randomising the train length for each set of images presented to the participant [Figure 6.3], was considered but discarded as unrealistic, as drivers would normally operate a specific number of cars over a whole journey. Therefore, drivers were presented with randomised sets of conditions for one particular train length at a time.

Finally, before moving off a driver checks that the platform end signal is green. When considering the experimental design the team were aware that this needed to be taken into account, since it was an integral part of the visual task of the departure sequence. It was felt important that in the experimental set up an additional task, to that of checking the DOO(P) images, needed to be factored in. It was decided that it did not need to be a traditional signal but could be replaced by a randomly computer generated number which had to be read out, to replicate the viewing and processing of the information.

Experimental set-up and conduct of trials

The experimental arrangement is presented in Figure 6.4. Two flat panel screens (TFT), to the right of the operator, presented the DOO(P) images, and to the left a further screen presented a randomised number as an alternative to the platform mounted signal.

Each monitor was faced with a mock-up of the real in-cab monitor front piece, including control icon graphics, in order to provide a realistic looking experimental set-up, and to ensure that the active monitor areas were accurately sized and spaced.

By using a small pad in front of the screens the subjects could initiate a sequence and subsequently record their decision – clear to move off or target detected.

Twenty one drivers participated with a mean level of DOO(P) experience of 10 months. There was also variation in whether drivers had used the in-cab CCTV system operationally, as an operational trial period had been granted, but not all participants had used the images for DOO(P). Each subject was exposed to four sequence sets, representing 4, 6, 8 and 12 cars, including the

mix of target variables described earlier. The subjects were instructed to scan the images and decide whether to take power – since the doors were clear – or to halt the process as a target had been detected. Subjects were not paced and no scanning techniques were suggested. Prior to the recorded series of runs subjects went through a familiarisation run.

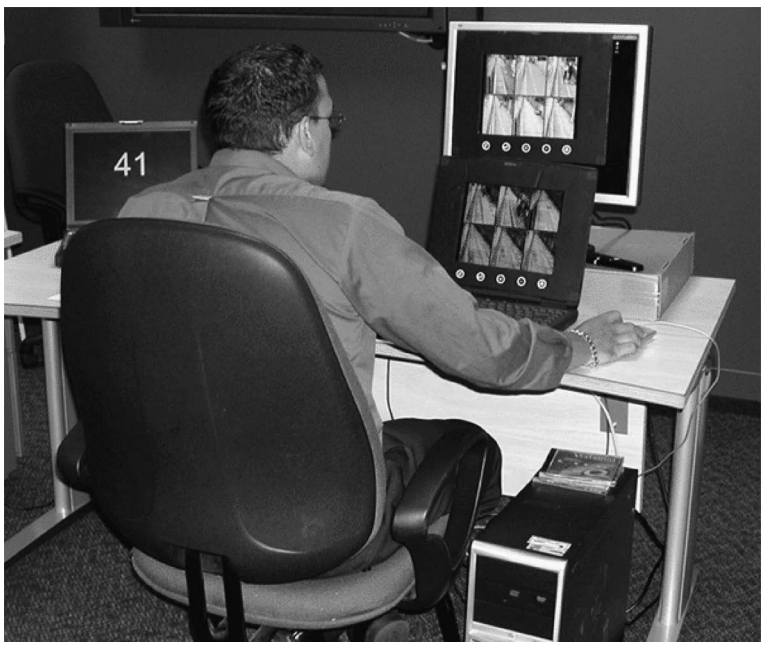


Figure 6.4 Experimental set-up for DOO(P) trials

The experimental protocol is presented in Figure 6.5 and started with a still image to provide context.

Following the completion of the trials a debriefing session was held, during which questions were asked about any particular scanning techniques the drivers had adopted, and whether they had encountered any particular problems in undertaking the tests.

Results

Detailed statistical analyses were conducted on the data collected. Of primary interest to the experimental team was whether the introduction of additional images on a DOO(P) monitor would have any impact on driver performance.

No differences were found in the numbers of targets missed when different numbers of images were presented. The grey test card and the trapped but unaware targets seemed to have the most 'misses' associated with them but there were no statistical differences in the failures whether it was 4, 6, 8 or 12 images. It was considered that both of these targets were unrealistic, as a person trapped is unlikely to remain 'unaware' for a period of time. When the grey test card and 'trapped but unaware' results were extracted from the data-set, and the analysis re-run, the differences were not statistically different though the numbers showed a non-significant 'dip' with 8 car presentations, see Table 6.1.

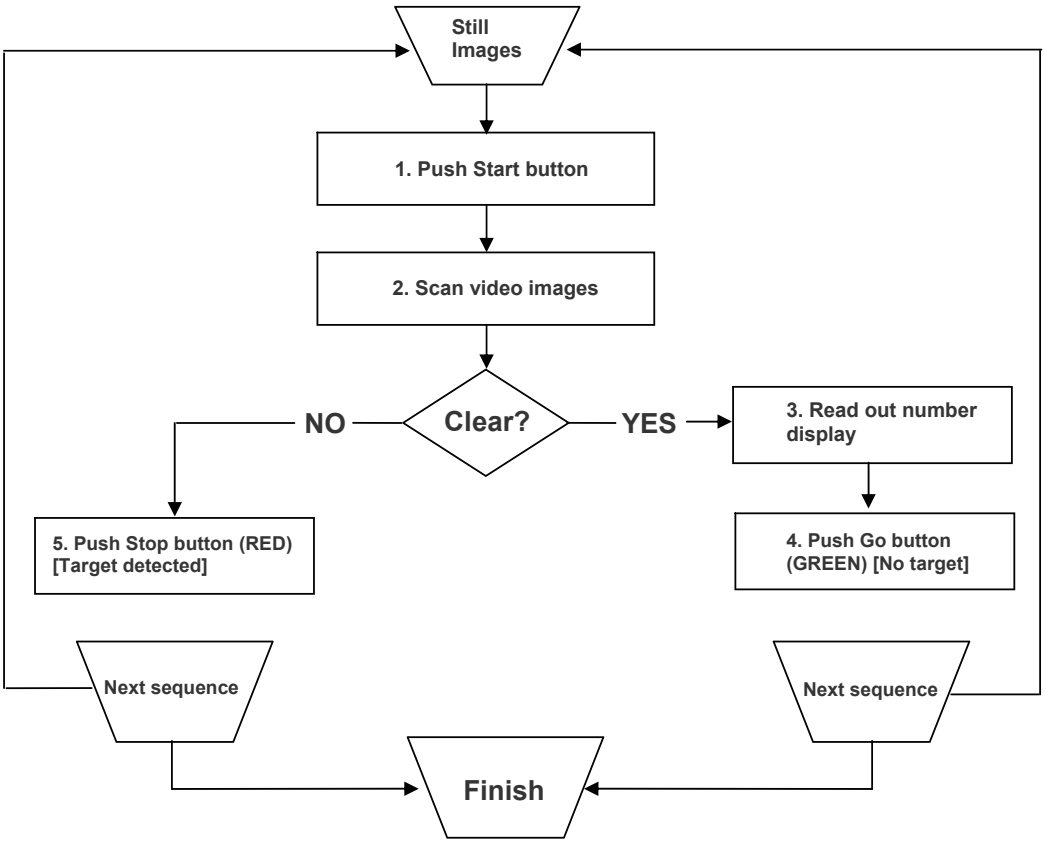


Figure 6.5 Experimental protocol for DOO(P) trials

Table 6.1 Target detection reliability scores excluding 'grey board' and 'unaware' targets for different numbers of images

	Target detection reliability
4 car	99%
6 car	98%
8 car	93%
12 car	98%
Overall	97%

When the data were analysed further it became apparent that the difference in target detection reliability for 8 image presentation might be related to the number of missed 'frozen' images, where the target was the detection of image 'freeze'. This appeared to be linked to the layout of images when 8 are presented. The flashing icon, to indicate that the system is working correctly, is

presented in the top left of the first slot on each monitor. As 8 images are presented over the bottom 3 slots of the top monitor, and 5 images on the bottom monitor (Figure 6.3), the driver would have to look at an additional slot to check the images on the top monitor are operating correctly.

'Response time' differences were statistically significant, see Table 6.2, though as far as overall railway operations are concerned the small increases would have little impact on service provision, however dwell times at stations will be reviewed, for longer trains, in the light of this work.

The mean age of the drivers who participated was 40.5 with a range of 28–59 – no differences were found when age and numbers of targets missed were examined. Similarly no correlation was found between driving experience (1–30 years) and targets missed and response times. Whilst most drivers had some DOO(P) experience there were 6 who had none – no significant correlation was found between DOO(P) experience and targets missed and response times.

Table 6.2 Response times for different car sets and numbers of images

		95% Confidence Interval	
	Mean	Lower Bound	Upper Bound
Four car	6.46	2.43	10.40
Six car	7.04	2.21	11.90
Eight car	7.99	2.73	13.25
Twelve car	9.92	4.49	15.45

Although the impact of differences in vision were not part of the experiment, a record was kept of the nature of any correction to vision required by drivers including short sight and the use of bifocals. No significant differences were found between drivers with corrected sight, and targets missed, though there was some evidence that those with bifocals and longsighted operators might be slightly slower than their colleagues not requiring any correction to their vision.

The mean error rate for those who claimed to be adopting a strategy for their scanning was found to be slightly better than those who claimed to be taking a more random approach. Statistically this was not significant but may represent an area worth exploring further.

Conclusions and discussion of results

The study concluded that viewing up to 12 images does not affect the reliability of target detection, with the field of view and image quality provided. The detection of targets on arrays with 8 images seemed more susceptible to error. More images require more time to check, which may result in some review of dwell times. Longsighted drivers and those wearing bifocals take longer than their colleagues but are no less reliable.

The results show that the number of images on an individual monitor can be increased, from the current limit of 4 to 6, without a significant degradation in performance.

A design study had identified the arrangement of image presentation, for each possible number of cars, prior to the fitment or operational use of the equipment. This included the arrangement of 8 image presentation for 8 car formation. The arrangement of images for 8 car presentation appears to be more complex than that for 4, 6, or 12 car formation, therefore the search pattern involved may be more complex. The increased error rate for 8 image presentation associated with detection of frozen images, suggests that there is a need to review the image layout.

Whilst answering the specific questions posed by RSSB the results provided indications of where other improvements might be achieved. The grey target used in the existing standard for DOO(P) should be assessed for its effectiveness for representing real targets. The target is useful as a means of checking the image quality of a CCTV system, as it is relatively difficult to detect, however it is not realistic as an example of a target one would expect a driver to detect as part of the DOO(P) task. The potential benefits to be achieved from structured scanning protocols should be examined further.

Acknowledgements

CCD and RSSB would like to thank the Southern Train's team for their assistance in recording the video images during the live run and for also making available technical assistance during the setting up of the trials.

Chapter 7

Understanding Driver Route Knowledge

R. Luther, H. Livingstone, T. Gipson and E. Grimes

Introduction

As a result of a series of incidents on the UK railway system, including the Ladbroke Grove accident, RSSB commissioned a study to investigate train driver route knowledge. The purpose of the study was to provide detailed recommendations for route knowledge training based on a sound model of the psychological processes involved in the use of route knowledge.

Currently, route knowledge is defined by the rail industry as 'the types of information required to be remembered, in order to operate over a route'(GO/RC 3551). It is generally acknowledged, however, that information 'remembered' must be used in real-time as a driver negotiates a route. This view indicates that route knowledge is more than the information held within a driver's memory; it is the skilful use of this information to interpret the current and future state of the environment. Therefore, any model of the psychological processes involved in route knowledge should incorporate psychological constructs such as situation awareness and attention as well as considering issues related to memory.

While extensive research on topics such as operator situation awareness, attention, and decision making has been undertaken in industries such as aviation and air traffic control, these constructs have not yet been widely studied within the rail industry (Li & Hamilton, 2003). Due to the scarcity of rail specific information, any consideration of these issues must begin by utilising knowledge gained within other industries. However, the applicability of this information to the rail industry cannot be assumed, and its relevance must be verified. An example of this approach is provided by Endsley (1999) who modelled the situation awareness requirements of infantry soldiers using psychological models developed in non-infantry environments as a starting point.

We outline two methodologies applied during the Route Knowledge study which utilised knowledge from non-rail environments. *Methodology 1* utilised fundamental literature, available train driver performance data and information from industry experts to develop a model of train driver route knowledge based on the concept of situation awareness. *Methodology 2* utilised information from non-rail industries regarding the impact of advanced technology on operator performance to support an analysis of the potential impacts of new technology on train drivers.

Methodology 1 – Development of a model of train driver route knowledge

In order to meet the study aim of defining and investigating train driver route knowledge requirements the research team developed a model of the psychological processes associated with route knowledge. The model was required to identify the relevant psychological processes and describe the relationship between them in a manner comprehensible to the non-psychologist. A two stage approach was used to develop the model:

Stage 1 – Development of an outline model of train driver route knowledge and investigation of the relationship between model components;

Stage 2 – Assessment of the validity of the model to the rail context.
Stage 3 – Development of an Outline Model of Train Driver Route Knowledge

In order to develop a model of train driver route knowledge it was necessary to identify the key psychological processes associated with route knowledge and to develop a description of the relationships between these processes. The following steps were undertaken to develop the model:

- Interviews with rail industry experts to gather information about route knowledge in order to identify relevant psychological processes for further consideration;
- A review of fundamental psychological literature to identify models of human information processing incorporating psychological processes identified as relevant during interviews with industry experts;
- Selection of appropriate models from fundamental literature for use as input into the Route Knowledge Model;
- Development and population of the Route Knowledge Model.

Interviews with rail industry experts indicated that while remembering the features of a route was a core component of route knowledge, other factors such as the ability to use this knowledge to interpret the current and future state of the environment were also relevant. This indicated that psychological processes such as situation awareness, attention, and memory were all relevant to route knowledge. Therefore, a search was conducted for models that described these processes. A range of models was identified, however, a model of situation awareness developed by Endsley (1995) was chosen as the most appropriate to provide a basis for the route knowledge model as it included all of the psychological processes of interest (see Figure 7.1).

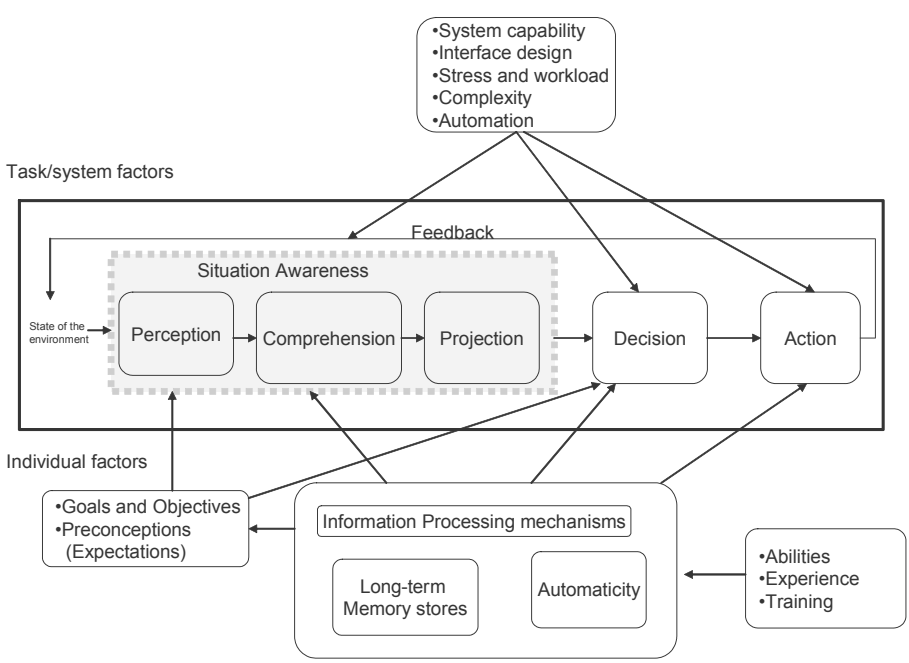


Figure 7.1 Model of Situation Awareness in Dynamic Decision Making (Endsley, 1995)

While Endsley's model contained the psychological processes required, it was complex and needed to be simplified to meet the requirement of being comprehensible to the non-psychologist. As a result a simplified Route Knowledge Model was developed using Endsley's model as a basis (see Figure 7.2). The core psychological and information processing components of the model developed by Endsley (situation awareness elements, decision, and action) were retained as this offered an easily understood structure. In order to simplify the factors influencing situation awareness contained in Endsley's model these were divided into:

- Individual Factors – factors that are intrinsic to the individual;
- External Factors – factors that are extrinsic to the individual.

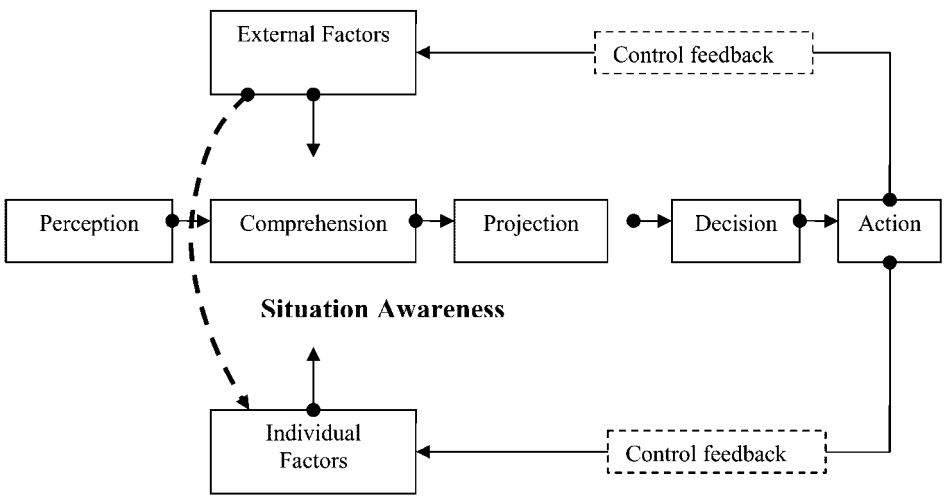


Figure 7.2 Train Driver Route Knowledge Model

The outline model was used to guide an extensive review of research literature to identify specific factors to populate model components, for example, stress was added as an individual factor. Figure 7.3 shows all of the factors identified under each model component[1]. The literature review also provided a large body of information about the relationship between model components (e.g. between stress and situation awareness).

In order to simplify the storage of information about the relationships between model factors and components, the Route Knowledge Model was used to structure a relational database which was populated with natural language statements derived from the review of the research literature (see Figure 7.4). The relationships were restricted using the permissible links defined in the Route Knowledge Model. More than one statement defining a link could be entered into the database and statements could reflect either a positive or negative influence on situation awareness. Having both positive and negative statements from the fundamental literature enabled the complexity of the relationships within the model to be highlighted. For example, literature indicated that stress had the potential to have both a positive and negative impact on situation awareness.

1 Note that Figure 7.3 also contains model components identified during stage three of the model development where rail industry data were incorporated into the model (e.g. control speed, move off).

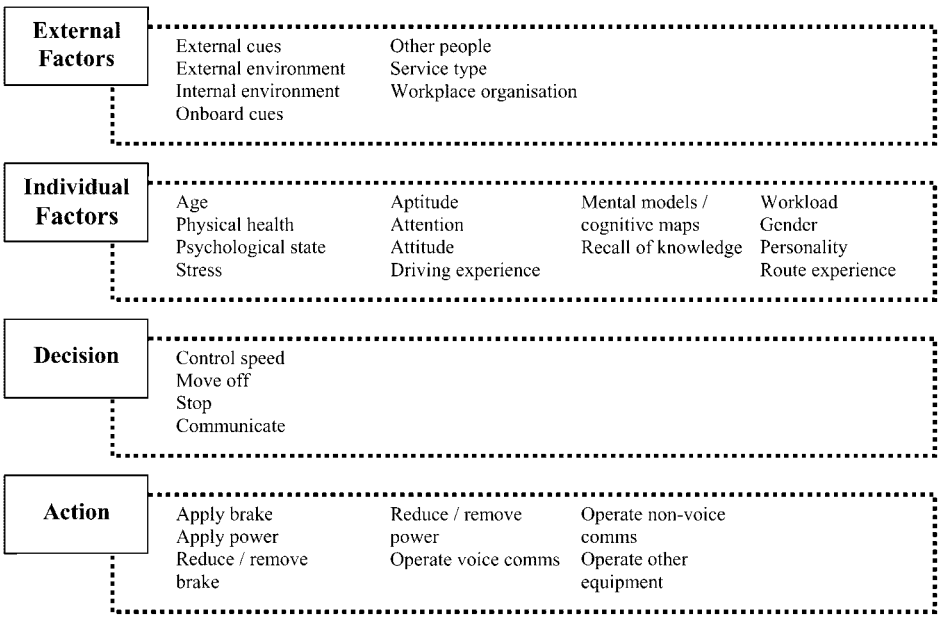


Figure 7.3 Factors contained within Train Driver Route Knowledge Model components

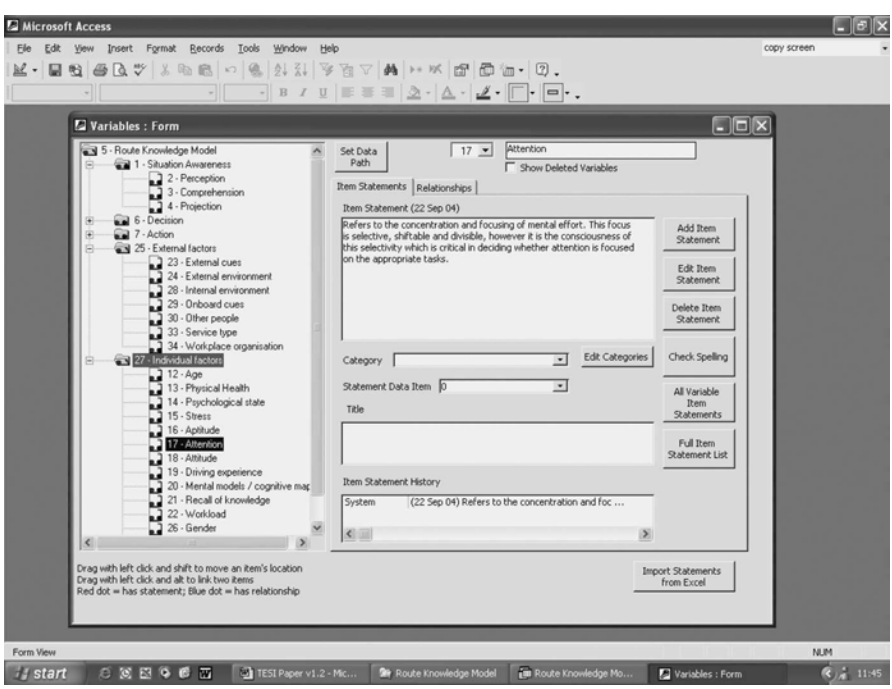


Figure 7.4 Screen shot of relational database

Stage 2 – Assessment of the validity of the model to the rail context

Stage 1 of the development of the Route Knowledge Model was conducted using information gathered from research conducted in a range of industries. Therefore, it was necessary to assess the validity of the Route Knowledge Model to the rail context. In addition, it was necessary to identify factors relevant to the rail industry to populate some model components (e.g. the factors contained in the action components of the model, see Figure 7.3). The Route Knowledge Model was validated and further populated using information gathered during an analysis of route knowledge errors. The error analysis was based upon two sorts of data:

- Rail incident reports;
- Interviews with rail subject matter experts, regarding safety related errors and their possible causes.

Statements derived from rail sources were entered into the relational database and the source of the statement was recorded. As a result, it was possible to assess whether model factors, model components and the relationships between them, as described by the review of literature from other industries, were supported by rail industry data.

Application of the Route Knowledge Model

The data contained in the Route Knowledge Model relational database, derived from fundamental literature and rail based sources were used to support the following:

- *Development of route knowledge training objectives:* statements contained within the database were reviewed and those issues that could be supported by training were identified. A content analysis was conducted on statements where training related issues were identified and key themes developed. These themes were used to frame the training objectives;
- *Prioritisation of training objectives*: the validation of model components using incident data provided a means by which a risk based prioritisation of training objectives could be conducted. Training objectives with error analysis data linked to them were given a high training priority, as the presence of error data indicated that the objective was associated with either an actual or potential incident.
- *Development of knowledge, skills and behaviours (KSB)*: KSB associated with training objectives were developed based upon the information contained within the database.

Methodology 2 – Analysis of the impact of future railway operations on train driver psychological processes

In order to meet the study aim of considering how new rail technology might impact on route knowledge training and assessment requirements, the research team developed a methodology to assess the risks that new technology posed to train driver situation awareness. The methodology contained three main steps:

Stage 1 – Analysis of planned changes to rail operations;
Stage 2 – Analysis of the impact of advanced technology on operator situation awareness in other industries;

Stage 3 – Analysis of the impact of identified technologies on driver route knowledge and identification of mitigation strategies.

Stage 1 – Analysis of planned changes to rail operations

The aim of Stage 1 was to gather information about advanced rail technology and select technologies for inclusion in the study. A search for new rail technology was undertaken using three sources, an internet search, a review of existing studies, and consultation with rail industry experts. Once a technology had been identified, a short description was developed to assist in assessing it for possible inclusion in the study. To be included, each technology had to meet two criteria:

1. The technology had the potential to impact upon the route knowledge related decisions and actions as outlined in the Route Knowledge Model.
2. The technology has not currently been broadly implemented within the UK rail network but had been considered by either railway bodies or technology providers.

Technologies that met these criteria were presented to RSSB and those selected were put forward for detailed consideration in later stages.

Stage 2 –Analysis of the impact of advanced technology on operator situation awareness in other industries

Due to the lack of information within the rail industry about the effect of advanced technology on the factors contained in the Route Knowledge Model (e.g. situation awareness) a review of information gained in other industries was conducted.

Twenty-five studies from a range of industries including aviation, road transport, air traffic control and petrochemical were identified for review. Each study was reviewed to extract statements about how new technology might affect operator performance. Because the aim of this activity was to gather information about the effect of technology on operator route knowledge, the Route Knowledge Model was used to structure the review. This ensured that only risks (directly or indirectly) related to the model constructs were documented. The relationships predicated by the model structure (see Figure 7.2) provided a set of 'rules' that statements extracted from the studies had to comply with. On this basis the studies were reviewed for the following types of statements:

- The impact of advanced technology on situation awareness;
- The impact of advanced technology on any of the models individual factors;
- The impact of advanced technology on any of the model's external factors.

In total 181 statements were identified. These statements were stored in the route knowledge model relational database providing information about how advanced technology could affect each relationship within the model. In addition, a content analysis was conducted on the 181 statements to provide consolidated, descriptive statements of the 'key risks' that advanced technology posed to the psychological processes described by the model. Examples of key risks included:

- Automation/high technology can reduce operator situation awareness due to poor system design (poor feedback, complex modes, and mode changes without notification);
- If technology is unreliable (e.g. only detects and displays 5/10 events) operators will have to devote attention to gathering missed information from other sources (e.g. through outside observation) rather than attending to other tasks;

- Technology that places the operator in a monitoring role can induce reduced vigilance, both to the system being monitored and to other cues available outside the system.

Stage 3 – Analysis of impact of identified technologies on driver route knowledge and identification of mitigation strategies

The 'key risks' identified during Stage 2 were used to assess each of the selected rail technologies identified during Stage 1 in terms of their possible effect on driver route knowledge. This process ensured that the relevance of each risk to each rail technology was considered (rather than assuming all the risks found from other industries would be relevant). This assessment involved the following activities:

1. Discussion with rail industry experts about the key risks related to each identified technology;
2. Review of rail-based studies on the impact of new technology on drivers;
3. Analysis of the key risks associated with each technology, based on information gathered;
4. Identification of mitigation strategies.

Discussion guides for interviews with rail industry experts were developed by creating probing questions associated with each key risk identified during Stage 2, ensuring that each risk was fully discussed. Using the discussion guide, interviews were conducted with rail industry experts who had specific knowledge of each of the technologies under consideration. Where possible, this information was supplemented by information gathered from existing studies of each technology.

An analysis of information gathered from industry experts and rail-based studies was undertaken to establish whether any of the risks identified during the 'other industry' review were relevant to the application of the technology in the rail industry. Information relating to a key risk was recorded in a risk review template. Evidence supporting or contradicting the risk was recorded. During this analysis, data from industry experts and rail-based studies was also reviewed to ascertain whether any risks discussed were not covered by the list of key risks developed from reviewing studies from other industries. Any extra risks identified were added to the list of key risks. Therefore, this analysis was also used to evaluate the applicability of the list of key risks for the rail industry.

Those risks identified as relevant to each technology were analysed to determine how they might be avoided or mitigated. Mitigation strategies were identified using a systems-based approach to ensure that a comprehensive range of mitigation and avoidance strategies were considered for each risk (rather than simply looking at training interventions).

Conclusions

Two methodologies for utilising knowledge from non-rail environments were developed during a study of train driver route knowledge undertaken for RSSB. Both approaches provided structured processes for gathering relevant knowledge from other industries and assessing its validity to the rail industry.

Methodology one involved the development of a Route Knowledge Model which described the psychological processes involved in the use of route knowledge. The development of the model ensured that full use was made of a large body of existing knowledge available from psychological literature. The use of a relational database to store information about the relationships between model components ensured a large body of qualitative data (in the form of statements) could be stored and searched with relative ease. The key benefits of the relational database were that it enabled

complex relationships between model factors to be mapped with relative simplicity and provided a clear audit trail of which relationships were validated with rail industry data. The development of the model ensured that the route knowledge training objectives and associated KSB developed during the study were well grounded within existing applied psychological literature but were also validated by information gathered from within the rail industry.

The Route Knowledge Model was developed specifically to describe the psychological processes associated with route knowledge. The approach could however, be potentially applied to other complex, cognitive tasks to support the identification of training objectives. The data collected from fundamental literature to populate the relational database within this study is essentially generic, and therefore applicable to all safety critical industries, the model being adapted by:

- validating the internal and external factors;
- developing industry specific decisions and actions;
- collecting industry specific error data to populate the relationships between the model components.

Methodology two involved the assessment of the impact of new technology on route knowledge. The method that was developed supported the identification of a range of route knowledge related risks posed by new technology. The key benefit of this methodology was that by using the Route Knowledge Model to structure a search of key risks posed by advanced technology in other industries, a broad range of possible risks to route knowledge were identified. The identification of these possible risks ensured that a detailed and structured questionnaire for use when discussing rail technologies with industry experts was developed. This approach ensured that a broad range of issues were discussed and considered during interviews and that discussion was focused around factors that represented risks to route knowledge.

The approach used in methodology two was akin to an Early Human Factors Analysis (EHFA). EHFA is an established process used within the defence industry to identify and assess human related risks early in the development cycle. Human related issues are captured and logged in a risk register during the early phases of a project. Using the risk register, mitigation strategies can be identified and integrated into the business case. The approach is particularly powerful when projects are first starting up as it structures thinking about issues and risks. The method used in this study enabled the identification of a specific set of human factors risks, those related to route knowledge as defined by the Route Knowledge Model, to be identified and logged for a range of technologies.

References

Endsley, M.R. (1995). A taxonomy of Situation Awareness Errors', in R. Fuller, N. Johnston and N.McDonald (eds), *Human Factors in Aviation Operations*, Ashgate, Aldershot, pp. 287-292.

Endsley, M.R., Holder, L.D., Leibrecht, B.C., Garland, D.J., Wampler, R.L., and Matthews, M.D. (1999). Modelling and Measuring Situation Awareness in the Infantry Operational Environment. *Report Prepared For: U.S Army Research Institute for the Behavioural and Social Sciences.*

Li, G., & Hamilton, I. (2003). Railway Signal Sighting: Scientific Evidence for Sighting Principles. Report Prepared for Railway Safety (ref: 03-T130-HUEL-01-TRT).

Railway Safety Approved Code of Practice for Train Driving GO/RC3551.

Chapter 8

Robot Trains: Results of Focus Groups with Remote Control Operators in the United States and Canada

Stephen Reinach

Introduction

In a recent effort to reduce operating costs and increase safety and efficiency, U.S. Class I freight railroads (line haul railroads with gross operating revenues above US$277.7 million; Association of American Railroads, 2005) have begun to implement RCL (remote control locomotive) operations in and around railroad switching yards. An RCO (remote control operator) wears a portable operator control unit (OCU; see Figure 8.1), usually by means of a vest (see Figure 8.2), and controls the movement of a locomotive while on the ground or riding the locomotive or rail car. The OCU sends commands via radio frequency to an onboard control computer (OCC) mounted on the locomotive. The OCC, in turn, instructs the locomotive to actuate locomotive functions such as braking, acceleration, and blowing the horn.

Typically, two RCOs work together as a crew, although other crew configurations exist. This varies from a traditional yard crew configuration, where typically 1-2 switchmen on the ground provide radio, hand or lantern instructions to a third crewmember, a locomotive engineer onboard the locomotive who controls the locomotive's movement. Operating environments include yards, industrial spurs and sidings, and some main tracks.

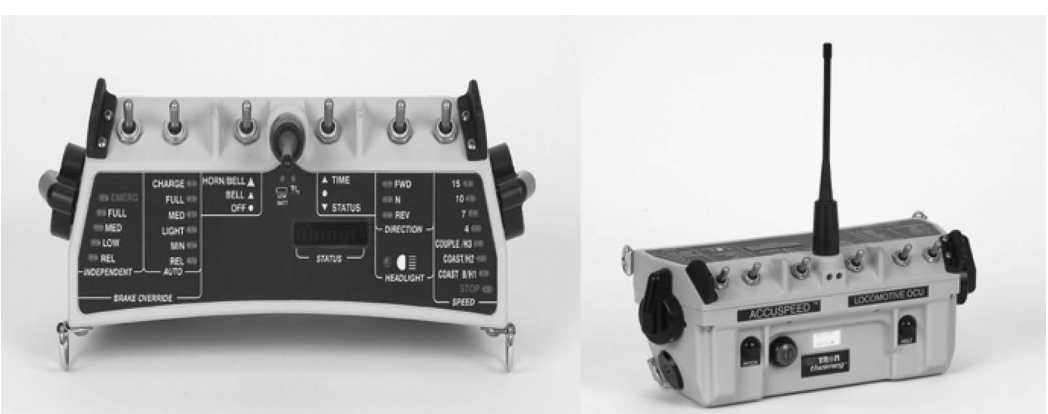


Figure 8.1 OCU top (left) and front (right) views (Courtesy of Cattron-Theimeg, Inc. 2004. Reprinted with permission)

Although remote control technology has been available for decades, the safety implications of using this technology in U.S. railroad yards and reducing crew size in yard switching operations remain unknown. Proponents of RCL operations suggest that controlling the locomotive from the ground affords the RCO the best vantage point and that the technology reduces or eliminates miscommunication errors between the locomotive engineer onboard the locomotive and a switchman on the ground. Opponents of the technology raise a number of safety concerns, such as inadequately trained and experienced RCOs.

FRA (Federal Railroad Administration) has begun to collect RCL operations-related train accident and incident data. However, due to the recent implementation of RCL operations on a large scale in the United States (beginning in early 2002) and the more recent FRA requirement for railroads to specifically report the involvement of this technology and RCOs in train accidents/incidents (effective May 01, 2003; FRA, 2003), this data collection process will require several years before ample data are available to analyze.

To understand some of the potential safety implications associated with RCL operations in U.S. freight railroad switching yards, FRA's Office of Research and Development Human Factors Program and Office of Safety launched several research studies to examine RCL operations. These studies included a risk assessment comparing RCL operations to conventional yard switching operations, human-centered accident/incident investigations, and focus groups with RCOs. This chapter discusses the results of focus groups with RCOs in the United States and Canada. Focus groups provided a snapshot taken in the very early stages of RCL implementation in the U.S. railroad industry.

The objectives of conducting the focus groups were to gather information on RCO experiences with RCL operations, discern RCL operations safety-related issues, identify RCL operations lessons learned and best practices, and solicit suggestions for how to improve RCL operations from those who were most familiar with the equipment and operations.

Figure 8.2 RCO coupling rail cars

Methods

Focus groups (see Krueger, 1994; Morgan, 1997) are a qualitative data collection research method in which small groups of individuals meet and discuss a number of pre-identified issues of interest. The focus group facilitator encourages participants to answer from their own experiences. Focus groups are advantageous in that there is no one correct answer, and no attempt is made to quantify the results. The benefits of focus groups are in the richness, or quality, of information gathered, and the broad range and depth of information and insights, sometimes unanticipated, that can be obtained from participant interaction. Focus groups tap participants' experiences, opinions, and attitudes toward a topic, and they are well-suited to examine RCO experiences and identify industry best practices.

To obtain a broad picture of RCL operations, it was important to look at a wide array of RCL operational experiences. Several criteria were established to help tap into a range of RCO experiences across the United States and Canada. These criteria included:

- Identification of focus group locations where RCL operations had been implemented by at least two railroads (and thus, different operational experiences) to enable RCOs from multiple carriers to attend. Since RCL operations were so new, it was not always possible to satisfy this criterion.
- Identification of at least one focus group location east of the Mississippi River and one west of the Mississippi River to account for cultural and operational differences among the different railroads east and west of the Mississippi River.
- Conduct of focus groups with switchmen and locomotive engineers. Both switchmen and locomotive engineers can qualify to become RCOs; however, whereas locomotive engineers have received significant training and experience controlling locomotives, switchmen typically have no training or experience controlling locomotives.
- Conduct of at least one set of focus groups in Canada since Canadian Class I railroads have been using RCL technology for over a decade in some yards. It was anticipated that U.S. railroads would benefit from Canadian RCO experiences, especially lessons learned.

These criteria were used to establish the focus groups and enabled researchers to examine a diverse cross section of RCO experiences in the United States and Canada. Researchers worked with the two operating craft unions that represent RCOs in the United States and Canada to identify focus group locations and select focus group participants.

Several criteria were established to assist local union officials in recruiting focus group participants. Focus group participants had to (1) be certified RCOs, (2) have at least 3 months of on-the-job experience as an RCO, (3) be willing to share their experiences and opinions about RCL operations, and (4) not hold an elected or appointed union position on a committee of adjustment (e.g., local chair) or legislative board (e.g., local legislative representative).

Three focus groups were conducted at each of the four different locations. Focus group times varied at each location to give RCOs working different shifts an opportunity to participate, given possible operational differences across shifts (e.g., night-time versus daytime operation). Each focus group lasted 1 ½ hours. Participants were asked to complete a background questionnaire before starting the focus group. Focus groups were guided by a pre-established set of questions. At the completion of each session, participants were compensated for their time. A total of nine focus groups was conducted in three locations across the United States, and an additional three focus groups were conducted in one Canadian city. Each focus group used the same set of questions, which concentrated around five major topics: implementation of RCL operations, RCO training,

current RCL operations and safety, the influence of prior switchman/engineer experience on RCL operations, and other-than-yard RCL operations. Questions asked RCOs to relate their experiences to these five topics. No attempt was made to validate any statements made by RCOs.

Participant profile

A total of 78 RCOs participated in 12 focus groups conducted between March–May, 2003. Participating RCOs came from seven different railroads, six Class I railroads and one regional railroad. Of the 78 RCOs, four were women. Participating RCOs ranged in age from 23-58 and had an average age of 40. Table 1 presents railroad operating experience data for participating RCOs. Although the average amount of engineer experience was 36 months, a majority of the participants (50 of the 78 RCOs, or 64 percent) had no locomotive engineer experience at all. The average amount of RCO experience, 34 months, was also inflated by the inclusion of Canadian RCOs, all of whom had significantly more RCO experience (average 79 months) than their U.S. counterparts (average 14 months) due to the more recent introduction and implementation of RCL operations in the United States.

Table 8.1 Participating RCO railroad experience

	Average (mo)	Range (mo)	Median (mo)
Railroad experience	174	14-425	143
Yard experience	144	14-425	107.5
Engineer experience	36	0-300	0
RCO experience	34	4-156	13

Results

Results are organized into two major subheadings: key issues and suggestions to improve RCL operations.

Key issues

The focus groups identified a number of key issues among the five topics. The following sections present some of the more significant key issues.

RCO training and preparedness

Most RCOs felt that training was inadequate to prepare them for the job. In general, RCO training lasted 2 weeks, with approximately 1 week of classroom training followed by 1 week of on-the-job training (OJT), although some locations offered more OJT. One of the specific criticisms focused on a lack of instructor experience with RCL operations. RCOs noted that often classroom instructors had only 2 weeks of RCL operations training themselves and did not have any actual RCO experience. Noted one RCO, “It’s like the blind leading the blind.”

Another criticism centered on the limited scope of the training. Many student RCOs were trained on only one job or one part of the yard but were expected to work any job in any part of the yard after they completed their training. A number of RCOs felt that each location in a yard had unique operating requirements. A third concern focused on peer training. RCOs pointed out that some peer trainers, qualified RCOs, did not receive any train-the-trainer training; sometimes they did not let the student RCO operate the OCU during a shift because they did not want to be responsible for any incidents in which the student was operating the OCU; or they had just completed their own RCL training and therefore had little, if any, RCL operations experience to share.

RCOs were asked how prepared they felt at the conclusion of their training. A few RCOs felt adequately prepared; however, the majority felt they were "not at all" prepared or "hardly prepared at all," and they were not comfortable at the end of their training. RCOs were asked how much hands-on experience was necessary before they felt adequately prepared to operate an RCL. Responses ranged from less than 1 week to 4 months per job, with most indicating they felt adequately prepared after 1 month on each job or location in the yard. Notes one RCO, it "…took about a month to understand what these cars are gonna do." Many of the RCOs pointed out that, due to operational and territory differences among different yard jobs (e.g., bowl, hump, industry), the time it took to get comfortable was location-specific. For example, it may take 1 month to become comfortable working a bowl job and another month to become comfortable at the hump. RCOs also suggested that prior experience as a switchman or locomotive engineer helped them become more comfortable (a majority of RCOs were switchmen before qualifying as RCOs, while a few RCOs were locomotive engineers before qualifying as RCOs). One RCO explained, "How fast you get comfortable with [RCL operations] depends on how much [switchman] experience you have."

Those RCOs who were locomotive engineers noted that their engineer experience helped them to understand how weight affected a cut of cars' ability to brake, which helped them to know when to apply the brake(s). These RCOs also noted that their engineer experience helped them to pay more attention to the grade, enabled them to diagnose locomotive problems, helped them to know when it was necessary to apply sand, and provided them with an understanding of the air brakes. One engineer noted that experience also helps "you plan ahead… and anticipate." RCOs with prior switchman experience felt that their experience on the ground helped them to understand switching and track configurations. RCOs noted that, during their training, experience as an engineer or switchmen enabled them to focus on the mechanics of operating the OCU. They felt that, without this experience, RCO students would be overtaxed by having to learn track configurations and how to switch at the same time they are learning how to operate the OCU.

Equipment reliability

RCOs described several types of reliability problems associated with the RCL equipment, including communication failures between the OCU and OCC, frequent error messages, delays in RCL response, and RCL overspeed (where the locomotive exceeds the selected speed). RCOs noted two notable communication-related problems. First, RCOs expressed frustration with communication delays between an input to the OCU and the time it took for the locomotive to respond. The delay can affect start-up of the locomotive or stopping. One example that was given involved initiating a locomotive movement, where an RCO commanded his locomotive to move, but a delay occurred. After a period of time, the RCO assumed the input had not been received by the OCC, so he began to make a new input, just as the locomotive began to move. This caused a fault, forcing the locomotive to stop. Perhaps more significantly, a cut of cars may travel several car lengths between the time an OCU command to stop is made and the time the locomotive begins to brake. Separately, RCOs

explained that sometimes an OCC would receive an input by the OCU, and other times the same OCC would not respond to the very same command. In one case, an RCO gave an example where he said it took him seven attempts to get the locomotive to apply the emergency brakes. Another RCO noted that once she fell while wearing an OCU and the OCU did not send the locomotive into emergency, which the OCU is designed to do as part of the tilt time-out function. Other RCOs noted that a locomotive would just stop because of a communication failure. RCOs noted that the use of repeaters around the yard has eliminated many of these communication failures, but some communication failures still occurred.

Inadvertent activation of the OCU

RCOs reported frequent problems with inadvertent activation of the OCU. For example, RCOs noted that wearing thick, winter gloves can cause an RCO to inadvertently bump an adjacent switch while making an input to the OCU. Another problem was bumping into rail equipment (e.g., a locomotive grab iron or cab door) while wearing the OCU. RCOs felt that the shape of the box facilitates inadvertent activation since some of the OCU's switches (e.g., the speed selector or independent brake paddles on either side of the OCU) are exposed or not well protected. Furthermore, an RCO may not receive any feedback about a change to the locomotive status (that a change has occurred) when an inadvertent activation occurs. For example, if an RCO sets the speed selector to 4 mph, then looks up to begin monitoring the cut of cars, and the speed selector then gets bumped up to 10 mph, the RCO may not have any idea about the change in speed selection until the cut of cars continues to build speed beyond 4 mph or the RCO looks down at the OCU. Further, if an RCO is riding the locomotive or a rail car, and an inadvertent activation occurs that causes the locomotive's brakes to be applied, there may not be any warning to the RCO that the locomotive or rail car is about to decelerate.

Operator situation awareness

RCOs report that operating the locomotive on the ground and away from the locomotive has reduced some of the critical feedback cues (visual and kinesthetic) available to cab-based operators (i.e., an engineer) and consequently reduces their situation awareness. Three distinct types of reduced situation awareness were identified. These are discussed below:

- *Loss of locomotive orientation awareness*. Some RCOs explained that it was sometimes difficult to remember which end of the locomotive was facing which direction. Since locomotives can operate in both directions, every locomotive has an "F" end. The "F" indicates the forward direction for the locomotive and corresponds to the forward position of the reverser. Setting the OCU reverser to forward will command the locomotive to move in the direction that the "F" end is oriented, while reverse will cause the locomotive to move in the opposite direction. An RCO must know which end of the locomotive is the "F" end and what direction the "F" end is oriented to move the locomotive in the proper direction. If an RCO forgets which way the locomotive is oriented and is unable to determine which end the "F" is located on the locomotive, it is possible for the RCO to set the locomotive in the wrong direction before making a move. Notes one RCO, "It's easy to get it [locomotive] in the wrong direction."
- *Lack of immediate locomotive response feedback*. If an RCO is working away from the locomotive and the locomotive is out of sight, such as around a curve, or occluded due to a bridge or rail cars in an adjacent track, an RCO may not know for a period of time whether or not the locomotive has responded to an OCU command. In other words, locomotive response feedback is delayed. If an RCO moves a cut of cars in the opposite direction as

desired, he or she may not become aware of this problem for a period of time.

- *Loss of movement awareness*. RCOs noted that it was possible for an RCO to not be aware of a cut of cars that has broken in two or cars that have derailed, since RCOs lack kinesthetic feedback and may lack visual feedback as well. Thus, it is possible for the RCO to not feel the resistance (i.e., dragging) of derailed cars, or may not feel or see that his or her cut of cars has broken in two. An RCO may hear a locomotive laboring harder to overcome the dragging cars, but, unless the RCO can see the derailed cars, he or she may believe that the locomotive is compensating for poor traction or struggling to travel up a grade, or the RCO may simply assume the increased laboring is because of the OCC's algorithm for throttle and brake inputs.

Additional tasks and responsibilities

RCOs felt that the extra tasks beyond the traditional switchmen or engineer responsibilities: were a burden, could lead to excessive workload, and were a concern to some RCOs. Noted one RCO, "When I get a switch list in my hand, I forget I'm running an engine." Another RCO explained that there are now so many things on an RCO's mind that he or she can become distracted and then it is easy to overlook something. RCOs also noted that, although they have added tasks and responsibilities, as well as the added weight of the OCU, they are not given additional rest breaks to recover from the added responsibilities and weight; this was a frustration to some.

Train handling

According to RCOs, the OCC does a poor job of train handling compared to a conventional locomotive engineer. Specifically, RCOs describe the OCC's train handling as the alternating application of the throttle and brake. In other words, after an RCO selects a given speed, the OCC revs up the throttle, then firmly applies the brakes, revs up the throttle, then firmly applies the brakes, and so on, creating a "herky jerky" response from the locomotive as it attempts to match (or hunt for) the selected speed. According to RCOs, this can cause a rough ride, potentially creating enough slack action (jerk) to throw an RCO to the ground if he or she is not expecting the change in acceleration. This also makes coupling cars very difficult when the RCO has stopped short of the coupling location because the locomotive may rev up the engine to initiate movement, and, by the time the brakes are applied to stop the locomotive, it has already made a hard coupling, slamming into the car to be coupled. RCOs noted that it was not possible to feather either the throttle or the brakes using the OCU, as locomotive engineers can do under conventional switching operations, to make very gradual movements. Spotting cars at an industry is especially difficult, several RCOs noted, especially when cars must be spotted within 2-3 feet of a location. If cars are close to the end of a track, the RCO also risks running into or through the end block if the locomotive surges before the brakes are applied and the cut of cars can be stopped. RCOs at one location noted that an accelerator exciter has been installed on some locomotives to enable the locomotive to accelerate more quickly in response to an OCU command. However, according to RCOs, the exciter can make controlling short moves more difficult.

Extreme and inclement weather

RCOs explained that extreme weather, both hot and cold, affected locomotive performance. For example, OCU batteries tended to have a shorter charge period in cold weather. RCOs also described times when the extreme cold caused an OCC to freeze or the extreme heat caused an OCC to overheat. Some of the heat-related problems have been fixed by placing fans inside the locomotive cab to keep the OCC equipment cooler. RCOs also noted that it was difficult to operate

the OCU wearing gloves due to inadvertent activation of adjacent buttons and not receiving critical tactile feedback to help locate a particular button or to inform the RCO that the button had been pressed. RCOs also noted that it was hard to find the vigilance reset button when wearing gloves since the button was almost flush with one side of the box. RCOs pointed out that an earlier generation of OCUs had a large mushroom-shaped vigilance reset button located on the top of the OCU that could easily be depressed in all weather conditions and provided clear feedback that the button had been pressed.

Suggestions to improve RCL operations

In addition to identifying key issues present in RCL operations, RCOs recommended a number of suggestions to improve RCL operations. The following sections present some of these suggestions.

Improve RCO training

Some RCO suggestions to improve RCO training included using more experienced and knowledgeable instructors; reducing the amount of classroom training spent on the mechanics of OCU operation and increasing the amount of time spent in OJT; covering the entire range of locations, operations, and configurations of cuts of cars that RCOs will encounter, during OJT; requiring RCOs to have a minimum amount of ground experience as a switchman before becoming an RCO; and offering refresher training to those who have not operated an RCL for a certain period of time.

Improve remote control equipment reliability and response

The most frequently cited suggested improvements included prevention of inadvertent activation of OCU controls (e.g., by placing guards on the OCU or designing the switches to move horizontally rather than vertically), more responsive equipment, and additional control over, and feedback from, the locomotive (e.g., add a display onto the OCU to indicate when the locomotive has responded to the command). Another notable suggestion was to make the OCUs and radios hands-free and voice activated. Thus, for example, an RCO could say, "come ahead 10 [car lengths]," as if talking to an engineer, and the OCU would send a command to the OCC to initiate the move.

Improve remote control practices and procedures

Suggestions included requiring RCOs to protect the point at all times and providing familiarity training to those who work with and around RCL operations, such as yardmasters, trainmasters, and conventional yard crews. The goal of the training would be to increase awareness and understanding of RCL operations so that these individuals can work safely with and around RCL crews.

Improve railroad facilities to support RCL operations

Suggested improvements included increased maintenance of switches and switch leads, widening the separation between tracks, smaller ballast to walk on around the tracks, and more yard lighting.

Make adjustments for other-than-yard operations

RCOs identified three core areas where RCL operations should be improved before a railroad considers taking RCL operations out beyond a railroad yard or surrounding area. The three areas of improvement included: more extensive training (e.g., to cover train handling, air brakes,

locomotive systems and troubleshooting, communications protocols, and territory familiarization); more reliable and responsive equipment (e.g., the locomotive's brakes must respond immediately to an OCU input); and more information on, and control over, the locomotive and cut of cars (e.g., information on air pressure status and access to dynamic brakes). RCOs wanted as much knowledge and control over the locomotive and cut of cars as locomotive engineers have when operating a locomotive conventionally.

Discussion

This chapter highlights findings from focus groups conducted with RCOs in 2003. A number of these findings reflect the very early stages of RCL implementation in the U.S. railroad industry. As such, undoubtedly some of the issues that are discussed have subsequently been addressed by the U.S. railroad industry. Furthermore, those who participated in the focus groups were not statistically sampled to be representative of all RCOs in the United States or Canada. Thus, while these RCOs provided significant insights into RCL operations issues, the results may not be representative of all RCL operations or all RCO experiences.

The introduction of RCL technology represents a significant change in how railroads operate and could be as revolutionary as the introduction of the diesel locomotive or the elimination of the caboose. Focus groups with RCOs provided invaluable insights into some of the early issues the railroad industry is facing and identified some possible strategies for helping to ensure safe RCL operations. Based on the focus group results, some of the more notable areas that might warrant further consideration include RCO training and preparation, reliability of the equipment, and RCO situation awareness. Improvements in these areas are likely to pay immediate dividends.

Acknowledgements

Thanks to Dr. Thomas Raslear, FRA Office of Research and Development, and Mr. John Conklin, FRA Office of Safety, for their sponsorship, technical support and guidance during the conduct of this study. Thanks to Ms. Sarah Acton, formerly of Foster-Miller, for assisting in the conduct of the focus groups, and to Dr. Frederick Gamst for providing subject matter expertise. The study would not have been possible without the significant support from, and participation of, numerous individuals from the United Transportation Union, including Mr. Rick Marceau and Mr. James Stem, and the Brotherhood of Locomotive Engineers and Trainmen, including Mr. Robert Harvey. Lastly, thanks to Ms. Judith Gertler and Ms. Susan McDonough for reviewing a draft manuscript and providing important feedback. This work was performed under FRA Contract No. DTFR53-01-D-00029. The views and opinions expressed here do not necessarily represent those of the sponsoring agency or participants.

References

Association of American Railroads (2005). *Class I railroad statistics*. Retrieved September 6, 2005 from http://www.aar.org/PubCommon/Documents/AboutTheIndustry/Statistics.pdf.

Federal Railroad Administration. (2003). *FRA guide for preparing accident/incident reports*. Report No. DOT/FRA/RRS-22. Washington, DC: FRA Office of Safety.

Krueger, R. (1994). *Focus groups: A practical guide for applied research. Second edition*. Thousand Oaks, CA: Sage Publications.

Morgan, D. (1997). *Focus groups as qualitative research, second edition*. Qualitative Research Methods Series 16. Thousand Oaks, CA: Sage Publications.

This page intentionally left blank

PART IV
Driving and Cognition

This page intentionally left blank

Chapter 9

Cognitive Workload of Train Drivers

Ilse Gillis

Introduction

For many years, machines were designed without taking account of human capacities with regard to processing information, the speed of manual performance, etc. The use of machinery, moreover, requires taking the right decisions at the right time on the basis of the information provided. Both factors result in a complex environment which may lead unconsciously to taking the wrong decisions. These “mistakes” may lead to incidents and, in the worse case, accidents. It is assumed that human errors are the main cause of 60% to 90% of serious accidents and incidents in nuclear facilities, process control and air traffic (Reason, 1997; Rouse & Rouse, 1983). 60% to 70% of incidents and accidents within the Belgian railway system are due to human error. The consequences vary from ignoring a signal without this causing any material damage to a crash involving two trains.

The rise in the use of the railways and subsequent increase in time pressure place ever increasing demands on the human information processing system. The mental processing of all information that the train driver (TD) gets to see and hear, produces a load that may result in an overload. Taking note of the signals along the tracks, checking the speed on the speedometer, stopping at the right stations within a strict timetable are all aspects that have to be taken into account by the TD. He must continually divide his attention between his primary task of driving the train and secondary task of taking note of information coming from his surroundings. Apart from the above external factors, the TD's mental load is also influenced by internal factors or, in other words, factors that are specific to the TD. Tiredness, being distracted, use of stimulants and medication are all examples of internal factors which may lead to human errors.

Internal factors such as being distracted or losing one's concentration have in recent years been expressed in the word “vigilance”. There is no straightforward definition for this term. Hupbach (Hupbach, 2004) defined vigilance as *“...the ability to attend to a set of stimuli over a certain period of time in order to detect the appearance of a target stimulus (signal)”*. The first major study into this subject was from Yerkes and Dodson (1908). In their ‘arousal’ theory, they show that there is a link between the level of performance and the level of vigilance. The relation between performance and the level of ‘arousal’ is presented as an inverted U where optimal performance occurs by optimal vigilance. The researchers Smit, Eling and Coenen (Smit, Eling & Coenen, 2004) demonstrated that a higher mental load led to decreased vigilance. In this way, they show that the "arousal" theory is not all encompassing and further investigation into influencing factors of vigilance will be necessary.

When carrying out their tasks, train drivers must be continually vigilant. Every signal or sign given must be noted, anticipated and registered. Vigilance alone is therefore not enough to be able to faultlessly complete the task. The information provided is not always immediately enforceable. Being vigilant and taking note of the information is not sufficient at that moment. A large amount of the information received must be registered for a specific period to be

incorporated in the task later on. This special “registration” is done by the prospective memory (Brandimonte, Einstein & Mc Daniel, 1996) – in other words, a memory that reminds us to carry out a “postponed” task.

Prospective memory is a fairly recent term used in cognitive ergonomics and psychology and consequently the number of investigations carried out is limited. However, there have been studies done into the influence on prospective memory of load, delay and verbal repetition in a dynamic environment (Stone, Dismukes & Remington, 2001). The study was carried out on air traffic controllers. An increase in the task load resulted in an increase in errors in prospective memory tasks (PMT). Delays and limiting verbal repetition of prospective memory tasks had no influence on performance of the PMT. In 2001, an investigation was started in the University of Trieste in Italy into the differences between a PMT and a vigilance task (Brandimonte, Ferrante, Feresin & Delbello, 2001). The investigation demonstrated that there was a great difference between a vigilance task and a PMT. In the first case, a stimulus was continually searched for while; when carrying out a PMT, stimulus was only sought at specific moments. In the case of a time-based PMT (Eysenck & Keane, 1995), the stimulus was sought at a specific moment in time. In the case of an event-based PMT (Einstein & Mc Daniel, 1990), a specific event/incident would be the reason to look for a stimulus. The investigation carried out by Otani (Otani, 1997) can be regarded as an application of the above investigation. Otani showed that carrying out a PMT actually had no effect on the performance of an ongoing task. In other words, only at specific time or event moments was there any mental "investment" in PMT. Examining the effects of workload on the carrying out of the PMT shows that a higher workload of the ongoing task (main task) results in a decreased performance. The retention or activation of the intention occurs by means of cues (stimuli). Vortac, Edwards, Fuller and Manning (1995) said in their study that an external cue that is present in the prospective memory during the retention interval does not give any improved results compared to when there is no external cue present.

In this study, the mental load of the TD has been determined by applied cognitive task analysis (CTA). Klein and Mitello (1998) defined CTA as a description of the necessary cognitive capacities to successfully complete a task. Carrying out a CTA can be done in various ways depending on the objective (Wei & Salvendy, 2004). In this study, a protocol analysis (Sanderson et al., 1989) has been applied to summarise the cognitive tasks. The study concentrated on prospective memory and vigilance tasks in a station environment. The reason for this choice is due to the results of an investigation (Gillis, 2005) into the causes and location of incidents and accidents on the Belgian railways.

Various phases were used (Figure 9.1).

- *Phase 1*: Two TDs were observed while carrying out practical exercises on Simbase, the part-task simulator. Their actions were recorded by a digital video camera.
- *Phase 2*: One of the experienced TDs from phase 1 commented on the video images. He summed up the actions, the reason for the actions, the thoughts, focus of attention, the visual and auditive signals, etc. Everything was recorded by a digital video camera and a dictafone.
- *Phase 3:* The data from phases 1 and 2 were examined and analysed. Based on a timeline, it was indicated at what time a physical, visual and auditive task was carried out. The influence of these tasks on the prospective memory and on vigilance was indicated on the same graph.
- *Phase 4:* A third experienced TD drove the same route as the TDs in phase 1 in the presence of an instructor.This time the route was not driven in a simulator but in reality. The reactions of the TD were recorded by two digital cameras – one camera for the actions alone and

one for the route. During the recording, notes were taken by an observer. Any extra aspects specific to the actual situation could thus be registered.

- *Phase 5:* The recorded images were commented on by the instructor who took part in phase 4. He summed up the actions, the reasons for the actions, the thoughts, focus of attention, the visual and auditive signals, etc. This phase was recorded by a digital video camera and a dictaphone.
- *Phase 6:* The data from phases 4 and 5 were examined and analysed. On the basis of a timeline, it was indicated at what time a physical, visual and auditive task was carried out. The influence of these tasks on prospective memory and vigilance were indicated on the same graph.

The investigation resulted in a graphic depiction of the mental load of the task of the TD.

The timelines indicated visually where the "critical overload points" were to be found. Due to the many visual spatial tasks that have to be done at the same time, this could lead to negative interference. As a result, one or more tasks will not be done or will be done less well.

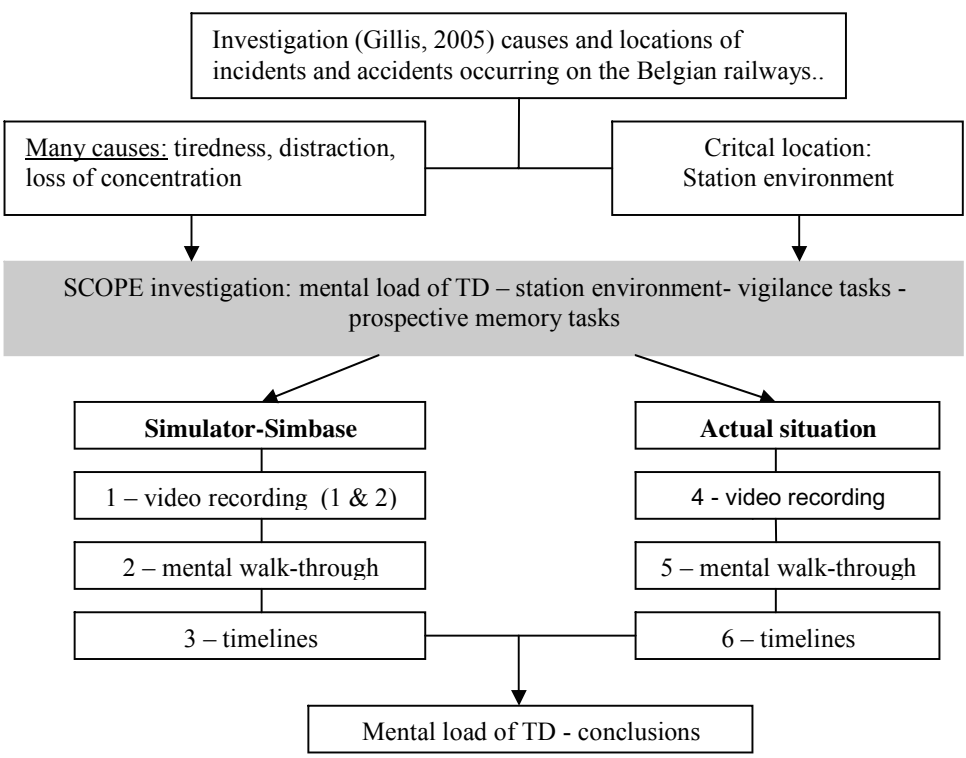


Figure 9.1 Overview of the various phases of the investigation

Materials and method

Participants

Three experienced TDs working on passenger and freight trains and one instructor took part in the study. Inexperienced TDs would have had too strong an influence on the results. Their lack of practical experience would have led to extra mental load due to their looking for information. The selected TDs were chosen at random. The instructor was chosen on the basis of experience, material and line knowledge.

Material

Simulator – Simbase: Simbase is a part-task simulator that, since 2004, has been used for the practical part of the training of a train driver. A computer program provides the simulation of possible driving situations. The simulations were developed by experienced TDs. Apart from the basic program that provides the visualisation of a specific route, there is an extra vigilance program. This latter program tests the vigilance of the TD by asking questions arising from previous situations. During the questioning, the simulation is stopped. The route chosen for this investigation consists of five station environments. From the analysis of the incidents/accidents, which form the basis of this investigation, it appears that the station environment is "hazardous". The simulator is fitted with the TBL 1 (Transmission Balise locomotive system) guidance system.

Computer Simbase: In phase 1, every action carried out by the TDs is registered and stored in the computer using specific software. This registration was not linked to time units but to distance units. The stored data were transformed into an Excel table in which the actions were indicated per distance unit.

Actual situation – control unit: A control unit with a TBL1 guidance system was chosen. The route taken using this control unit was similar to that in Simbase. Extra parameters were available in this system compared to Simbase. The following could thus be included: calls via mobile phone, interaction with the train guard, calls by radio, passengers getting off and on the train, various weather conditions, etc.

Digital video cameras: Recordings in phases 1, 2 and 5 were made using a digital camera. Situation 4 was recorded with two digital cameras. Recording during phase 1 was necessary as the data were linked to distance.

Dictaphone: Verbal explanations given in phases 2 and 5 were recorded using a dictaphone.

Procedure

In the training centre, route 2.2 integration 1 was chosen from existing simulations. This is a route in which five station environments have been processed. The first and last stations were not included in the investigation. Both were too deviant from a normal station environment with arrivals and departures.

Those TDs who took part in phase 1 of the investigation were experienced TDs who participated in the Simbase research team. Upon arrival in the training centre, the TDs were first informed about the objective of the investigation. They were given the assignment to carry out the chosen

simulation exercise. During the assignment, their performance was on the one hand recorded with a digital video camera and on the other recorded and stored in the Simbase computer. To avoid any extra mental pressure, they were assured that the resulting video images and computer data would not be used for evaluation purposes.

The data registered in the computer were transformed into a blank Excel table where every action was linked to a distance point expressed in km. The table was extended by extra data from the route such as station name sign, type of signal and signal image, etc. Using this recording and the registered images, an initial overview was made of the visual, auditive and physical tasks. In the same graph, the influence of these tasks on the TD's vigilance and prospective memory was shown.

In phase 2, the recorded video images were commented on by one of the TDs from phase 1. These data were recorded using a digital video camera and a dictaphone. An indication was given what the tasks were, why they were carried out, the focus of attention, the TD's thoughts, etc. This was done for station environments 2, 3 and 4 in the simulation program. The recorded images and discussion were noted again together with images from phase 1. The first overview of the tasks were extended and adjusted. The mental load was adjusted on the basis of the changes in the tasks. The result was twice three overviews with a graphic depiction of the tasks and the linked mental load.

For the "actual situation" phase 3, a route was used that was similar to the simulation program. In order to make comparison of the data possible, identical material was sought. In other words, a train was chosen with a TBL1 guidance system. The simulator's control unit is a replica of the control unit of a type 96 passenger train. Train type 96 was also chosen for the investigation. The chosen route included various stations similar to the simulated exercise. On arriving at his control unit, the TD received an explanation of the aim of the investigation. He was told that the recorded video images would not be used for evaluation purposes. In this way, any additional mental pressure was kept to a minimum. No data were recorded via the computer as there was no computer present in the control unit. Two digital cameras were used in this phase. One camera recorded what was happening outside and was set up next to the TD. A second camera recorded all the TD's actions and was set up in the same way as in Simbase. An instructor was always present in the control unit during the actual recording. This was necessary to be able to ascertain any divergent behaviour and to rectify it. The same instructor provided an explanation about the image in the next phase. Notes were taken during the journey by an observer to register extra influencing parameters. After the journey, the recorded images were examined and commented on. The instructor indicated which actions were undertaken and why, the focus of attention, the TD's thoughts, etc. In this phase, a digital camera and a dictaphone were once again used. In this case, only the station environments were scrutinised.

The recorded images and collected data from phases 4 and 5 were examined and analysed. All visual, auditive and physical actions were set out in graphs including the mental load required to perform these tasks, especially observations, perceptions and memory with regard to knowledge and prospective memory. The "critical points" were determined. In a final step, both situations, exercises on the simulator and the actual situation were compared to each other. Extra tasks that needed to be carried out in the actual situation were indicated.

Results and discussion

The investigation consisted of various phases. For the three situations analysed, two in a simulation environment and one in an actual environment, the TD's mental load was determined. The number of vigilance tasks and prospective memory tasks that the TD has to process in a station environment

were indicated on a timeline. Possible interference between the "mental" tasks was examined by determining the nature of the information.

The data collected from phases 1, 2, 4 and 5 were processed and schematically indicated in three graphs – timelines, total mental load and interfering tasks.

Timelines

The timelines consist of an upper part, above the grey bar with an indication of time and the lowest situated under the grey bar. An extract is shown in Figure 9.2 of the timeline resulting from the third investigation and the actual situation in stations 5 and 6. Every investigation carried out resulted in a corresponding timeline. The results obtained are provided in the extensive study (Gillis, 2005). The uppermost part gives an overview of the information sources that the TD has to process as well as the physical actions that he carries out as a response to these sources of information. The information is provided visually or auditively. This distribution is shown in the graph. The lowest part, mental load, provides an overview of the influence of the information and the tasks on the mental information processing system of human beings. This was subdivided into control loop, observation, perception, memory – knowledge and prospective memory.

A TD must be continuously vigilant and have recourse to his line and material knowledge. For this reason, these aspects were not treated as a separate section in assessing mental load. They are part of observation, perception and memory – knowledge. This method of working was required to avoid showing load as a succession of coloured bars.

Result

From the graphic depiction obtained from the various studies, it is clear that a large number of physical actions have to be carried out by the TD as a response to the information obtained. Some of the physical actions performed give a visual or auditive "reaction". Pressing the warning light after getting a double yellow signal, for instance, results in a lit up warning light as a "reaction" which becomes visual information for the TD. Every movement of the traction handle causes a change in the speed of the train and as a result the TD must constantly check the speedometer to estimate the consequences of his action. At certain times, various information is given in addition to standard information such as speed, prevailing stress, small or large movements, etc. The TDs in the first instance had to note all the information that was offered, make sense of this information based on their own knowledge and then store it in order to utilise it some time later. In the latter case, prospective memory is resorted to. In other words, every source of information must, continually or just once, be processed mentally. In the lower part of the timelines, it should be noted that in the case of a greater flow of information, increased mental load is produced. Some tasks or information sources resulted in a longer mental load. An example of this is the double yellow signal. This information requires future action, a prospective memory task and consequently the load will start when the signal is seen and stop when the action has been completed. It should be noted that prospective tasks linked to additional signalling such as speed restrictions are indicated as "short-lived" prospective tasks in this investigation. This means that they are indicated as prospective tasks but their duration has not been taken into consideration. This was due to lack of time and a limited knowledge of the details of the regulations.

Total mental load.

Certain suppositions were made to determine total mental load. Firstly, it was assumed that all tasks use the same resources which by definition cannot be correct. Automatic tasks, for instance, use less resources than controlled tasks, difficult tasks require more resources than easy ones, etc. Interfering tasks may lead to reduced performance in one of the tasks or in both tasks. Secondly, it was assumed that time sharing between the two tasks would not cause any "attentional switch-cost" (Wickens, & Hollands, 1999). This assumption is, based on the literature, also a false one. However, as until now it has been impossible to follow every mental action, this assumption is essential. The ultimate aim of this analysis was to determine the critical points of the TD's tasks in a station environment. The timelines were used as the basis to determine total mental load. A period of fifteen seconds was used to show total mental load schematically. As the time in a station environment on average lasts four minutes, every station environment is divided into about sixteen time periods. Figure 9.2 gives an extract of total mental load at stations 5 and 6 in the actual situation.

Result

It is clear from the various graphic depictions that the load is very high just before entering and just after leaving the station environment. The results indicate that the mental load increases if a

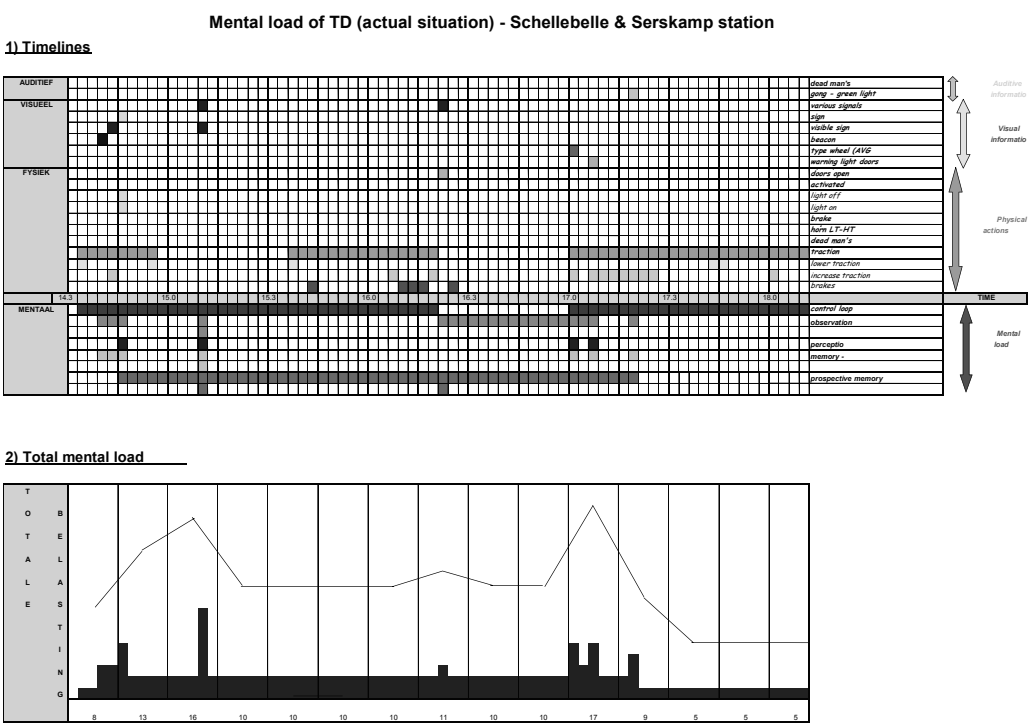


Figure 9.2 Timeline and total mental load – extract from the application to an actual situation

double yellow signal is given before arriving at the station. After leaving the station, the mental load is higher than when a green signal is given before leaving the station. This may provide an explanation for ignoring signals after the TD stops at the station when a double yellow signal is given as an entry signal.

Task interference

Carrying out various tasks at the same time may lead to task interference. In this case, a specific task or various tasks are carried out sub-optimally as the allocation of resources makes it impossible to carry them out optimally. According to Wickens' multiple-resources-model (1984), there are various dimensions which determine whether or not tasks interfere with one another. In this investigation, those tasks which make use of visual spatial resources are compared to tasks which make use of visual verbal resources. The same applies to auditive spatial and auditive verbal resources. As opposed to the previous analysis of total mental load, the upper part of the timeline is used for interfering tasks. The auditive and visual information and any resulting physical tasks are divided into a visual spatial or verbal task or an auditive spatial or verbal task. The result of the analysis of the various investigations is given in the complete study (Gillis, 2005). Figure 9.3 gives an extract from the third investigation, the actual situation in stations 5 and 6.

Result

The investigation shows that the TD's tasks are mainly visual spatial tasks. Up to four visual spatial tasks were noted within a three-second interval. This exceeds human capacities as only one place can be actively perceived. As a result, the TD must continuously switch from one task to another.

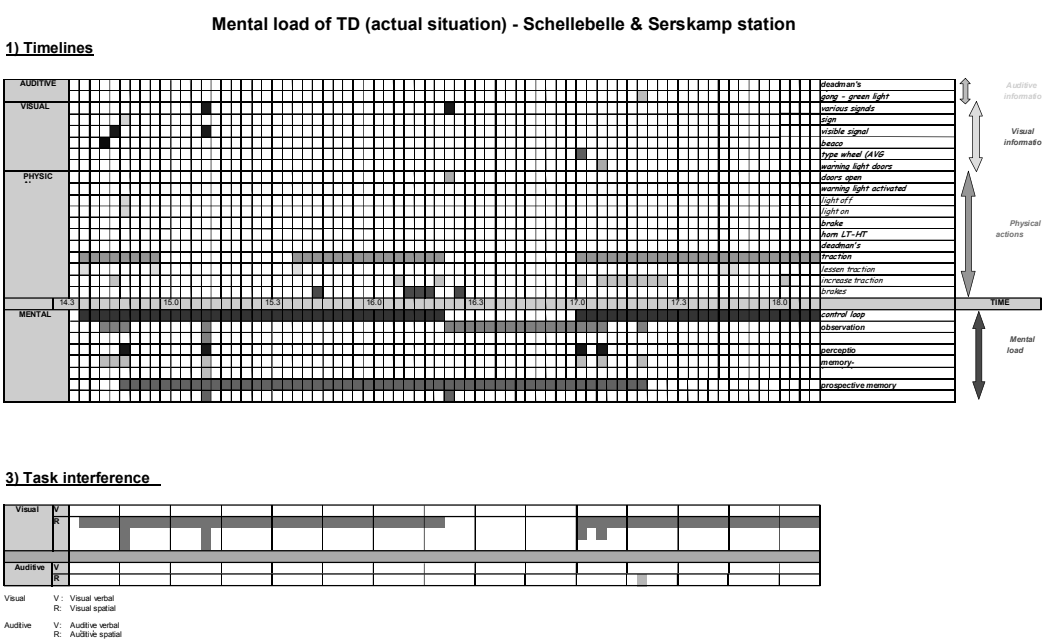


Figure 9.3 Timeline and task interference – Extract from application to an actual situation

This investigation shows that task switching does not require any extra resources. However, there is an additional important parameter when switching tasks. Moray and Sheridan (1986, 1972) ascertained in their study that task switching does not happen so quickly. There is a tendency to spend time on lower priority tasks while sometimes switching to high priority tasks occurs too late. If we apply this to the results of this investigation, it is possible that certain visual tasks remain undetected. In the worst case, this may led to incidents and accidents. In order to correctly interpret interference of these tasks on overall performance, further investigation should be done using an eye movement device. This follows the activities of the eye and determines on which task the TD has focused his attention and for how long. The auditive spatial tasks mainly occur when the confirmation gong of the green signal is activated which became clear in the actual situation. Another auditive spatial task occurs when the gong of the vigilance device, the dead man's pedal, is activated. Compared to other guidance systems, the frequency of this gong in TBL 1 is not every 60 seconds. The reason for this is that with the TBL1 guidance systems, every action done by the TD is regarded as a sign of vigilance. For this reason, in the various studies, the gong is only heard when the train is standing in the station or after long periods without traction or other actions. The graphic depiction of interfering tasks shows that there are periods when the number of tasks with the same resources are increased.

Comparison of simulator to actual situation

An important difference was noted in the station itself. In this location, the TD was confronted with a few extra tasks in the actual situation which did not occur in the simulators. During the journey, anomalies may occur such as electricity cuts, being given additional orders, calls from the train guard or dispatching department, workmen on the line, exceeding a closed position after following out an order, helping a colleague in trouble, etc. Each of these tasks will result in an additional mental load. In some cases, this will lead to a mental peak load level or even overload.

Conclusion

Three investigations were carried out to determine the mental load of train drivers in a station environment. The aim of these investigations was to ascertain if there was a mental load explanation for the high percentage of incidents and accidents in this particular environment.

In the first two investigations, the TD's mental load was determined when carrying out an exercise in the Simbase simulator. Both experienced TDs used in the investigation were given the same route that passed three different stations. The third investigation was an actual situation where an experienced TD travelled on a specific route using an identical guidance system to the one used in the simulator. This consisted of 11 station environments. The results of the investigation showed that no mental overload was noted either when arriving or leaving the station. A rather striking aspect was that for each of the tasks, a brief mental load was registered but only for the control loop and PMT. The control loop may be regarded as a vigilance task as the TD must continuously take care not to exceed the speed limit and must be aware of any speed limit restrictions. The time needed to carry out the various investigations was limited and it was therefore not possible to ascertain the theory postulated by Yerkes and Dodson (1908). Each investigation was carried in the morning or at the beginning of a shift. The vigilance of each TD was therefore optimal and this also applied to their performance. An increased load was noted on the resulting timelines when arriving at and leaving the station. When a station is approached, normally the TD sounds his horn to warn passengers that a train is approaching. The timelines showed on various occasions that the horn

was not sounded when the train approached the station. It can therefore be postulated that increased mental load leads to a lessening of vigilance. (Smit, Eling & Coenen, 2004). Nevertheless, this is an incorrect assertion. Sounding the horn must be regarded as a prospective memory task and not as a vigilance task. Not carrying out this task is a failure of the event-based prospective memory. In this investigation, the results of the study done by Vortac, Edwards, Fuller and Manning (1995) confirmed the results obtained for the warning light. The warning light which reminds the TD of a limitation when approaching a station fails as an external cue. Because the light remains on during the retention period, it melts into the environment with the result that when leaving the station the TD will ignore the warning light. To fix the TD's attention on the warning light, it must be activated a few seconds before the train departs. If this done, the TD will fix his attention on the warning light and give it the appropriate interpretation.

This investigation shows that there is no direct mental overload for the TD in a station environment. However, it was noted that there are many visual spatial tasks some of which have to be carried out simultaneously. Depending on human information processing with all its limitations, this may lead to losing information or wrong interpretation. The TD must be capable of managing his tasks and be able to switch from less important to more important tasks on time. Priorities have to be set. Investigation into determining the TD's visual follow-up using an eye movement device would indicate whether or not the appropriate duration and attention will be given to a specific task.

Prospective memory is a fairly recent investigative tool which should certainly be further examined in the context of this application especially examining the performance of PMT depending on the age of the TD. Various studies can be found in the literature which examine the relationship between both aspects. It is expected that the increased number of incidents and accidents with experienced and therefore older TDs can be explained. However, this will have to be demonstrated through additional investigation.

References

Brandimonte, M. A., Ferrante, D., Feresin, C., and Delbello, R. (2001). Dissociating prospective memory from vigilance processes. *Psicologica, 22*, 97-113.

Brandimonte, M.A., Einstein, G.O., and Mc Daniel, M.A. (1996). Prospective Memory: Theory and applications. Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Einstein, G.O., and Mc Daniel, M.A. (1990). Normal ageing and prospective memory. *Memory, cognition 16*, 717-726.

Eysenck, M.W., and Keane, M.T. (1995). Cognitive psychology. London: Lawrence Erlbaum.

Gillis, I. (Ed.). (2005). [Cognitive workload of train drivers]. Unpublished thesis by Ehsal Brussels.

Hupbach, A. (Ed.). (2004). *Attention. The allocation of a limited resource*.
(http:// www.arizona.edu).

Klein, G., and Mitello, L.G. (1998). Cognitive task analysis. In workshop of *human factors and ergonomics society,* 42nd Annual meeting, 12, Chicago, Illinois.

Moray, N. (1986). Monitoring behviour and supervisory control. In K.R. Boff, L. Kaufman, and J.P. Thomas (Eds.), *Handbook of perception and performance* (vol. 2, pg. 40-1-30-51).New York: Wiley.

Otani, H. , Landau, Josh D., Libkuman, Terry M., Kazen, Joseph K., St. Louis, Paul J., and Throne George W. (1997). Prospective memory and divided attention. *Memory*, *5 (3)*, 343-369.

Sanderson, P.M., James, J.M., and Seidler, K. (1989). SHAPA: *An interactive software environment for protocol analysis*. EPRL-89-08. University of illinois at Urbana-Champaign. Urbana. IL.

Sheridan, T. (1972). On how often the supervisor should sample. *IEEE Transactions on systems*, Science, and Cybernetics, SSC-6, 140-145.

Smit, Annika S., Eling, Paul A.T.M., and Coenen, Anton M.L. (2004). Mental effort causes vigilance decrease due to resource depletion. *Acta psychological*, *115*, 35-42.

Stone, M., Dismukes, K., and Remington, R. (2001). Prospective memory in dynamic environments: Effect of load, delay and phonological rehearsal. *Memory*, *9(3)*, 165-176.

Vortac, O.U., Edwards, M.B., and Manning, C.A. (1995). Function of external cues in prospective memory. *Memory*, *3*, 201-219.

Wei, J., and Salvendy, G. (2004). Cognitive task analysis methods for job and task design: review and reappraisal. *Behaviour and information technology*, *23(4)*, pp. 273-299.

Yerkes, R.M., and Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurological Psychology, 18*, 459-482.

This page intentionally left blank

Chapter 10

Train Drivers and Fatal Accidents on the Rails: Psychological Aspects and Safety

Valdimar Briem, Sonia de Lima and Camilla Siotis

Introduction

Swedish railways are considered fairly safe with regard to collision accidents, and many measures are taken to prevent such events. Nevertheless, the annual frequency of accidents with people killed or seriously injured is strikingly high, and more than half of these accidents happen under circumstances that suggest suicide or attempted suicide. If rail workers' accidents are included in the figures, the incidence of serious accidents is around 100 per year – an average of about two deaths or disabilities per week on the railway lines, according to the Swedish National Railway Administration's latest statistics report (Banverket, 2004) – see Table 10.1.

Table 10.1 Number of accidents on the Swedish railways, including suicide or attempted suicide, from 1997 to 2003. Passenger accidents, including passage to or from trains at stations, train collision, and electric accidents are not included. Adapted from Banverket (2004)

	1997	1998	1999	2000	2001	2002	2003
Killed	96	71	73	69	83	80	76
Injured	27	13	22	39	34	14	19
Sum	123	84	95	108	117	94	95

The consequences of these accidents are extensive for everyone involved, including the train drivers, who risk experiencing at least one such fatal accident during their working life. Despite this, the psychological and safety consequences for the drivers has not been thoroughly documented. In Sweden, sick leave among underground train drivers who had experienced fatal accidents was documented with a one-year follow-up by Theorell et al. (1994), and a Swedish-Norwegian study (Karlehagen et al., 1993; Malt et al., 1993) has documented the effects of serious rail accidents on train drivers' psychological health in the Swedish Railway's Northern Region. According to this study, over 50% of the drivers displayed acute stress symptoms, which diminished over time (one-year follow-up). British researchers (Tranah & Farmer, 1994) found that approximately 17% of underground train drivers in London who had been involved in fatal collisions showed stress symptoms. One month after the accident, 30-40% of these drivers still suffered from stress, but a year later these symptoms had diminished significantly. The Swedish-Norwegian and British results coincide to some extent, but considerable uncertainty exists regarding both the long- and short-term effects of these traumatic experiences.

Witnessing a fatal accident causes severe stress in most people, effects that can appear immediately (Acute Stress Disorder, ASD: up to four weeks) or later (Posttraumatic Stress Disorder, PTSD). According to international standards (DSM-IV) (American Psychiatric Association, 1999) PTSD is defined by four criteria:

Criterion A: The individual must have experienced a catastrophic or traumatic incident of such severity – generally regarded as exceeding the limits of normal human endurance – that a traumatic crisis can occur as a result (e.g. witnessing a fatal accident).

Criterion B: The individual subsequently has recurrent memories of the event, which may be more or less realistic (*Intrusion*).

Criterion C: The individual consistently avoids stimuli associated with the trauma and/or has numbed or reduced general reaction capacity (*Avoidance*).

Criterion D: The individual shows recurrent symptoms of emotional activation that were not present before the trauma (*Hyperactivation*).

It should be noted that PTSD is not a psychiatric illness, but a syndrome that may occur in otherwise healthy individuals as a result of an external event. In this context, the relationship between stress and coping strategies also provides good insight into individual differences in the ability to cope with stress. Myrtek et al. (1994) have discussed the coping strategies used by train drivers involved in rail accidents.

Other salient psychological consequences of traumatic events include cognitive disturbances, insomnia, lack of concentration and various practical errors. All this can impair both individuals' work performance and ability to function optimally in their private lives. For instance, Australian researchers (Edkins & Pollock, 1997) found that train drivers had difficulty concentrating on their work routines, which often involve long shifts and monotonous tasks, and that the drivers' psychosocial status was influenced by various factors, including dissatisfaction with their work situation. Other investigators have discussed negative safety and psychological consequences related to train drivers' working conditions (Cocks, Fletcher, & Dawson, 1989; Parkes & Lansdown, 2000).

The primary aim of this study is to gain a clearer insight into the occurrence and nature of problems related to serious and fatal accidents on the railway tracks and their effects on train drivers' psychological and physical health. Another aim is to examine the consequences of these factors for the drivers' ability to satisfactorily perform their work without errors and mistakes. The results are expected to give an indication of how train drivers' health, well-being, and cognitive abilities are affected in the long term, and whether this can impair their work performance, thus reducing the safety of rail transport.

Method

Participants

Eighteen train drivers at the Swedish State Railways (SJ) Southern Region were interviewed. Two of these, who had very much shorter job experience than the others, were not included in subsequent statistical analyses.

***Material*:**

a. Interviews:

   i. *Background*: Name, Age, Number of years as train driver, Family status, Number of children, Number of serious accidents on the railway, Number of fatal accidents.
   ii. Subjective *vulnerability*: (i) Number of minor accidents during preceding year, (ii) Accident proneness: Less, equally many, or more accidents than other people (-1 / 0 / +1), (iii) Probability (%) of having a major accident next year.
   iii. Semi-structured, *in-depth interview*: Included detailed description of serious accidents, description of physical and psychological state before, during, and after the accident, as well as private and work relationships.

b. Tests:

   i. *Coping Resources Inventory (CRI*) (Psykologiförlaget, 1991). Contains five subscales which reflect different aspects of the individual's ability to recover from stress:

      - *Cognitive* (COG): Positive view of oneself and others, optimistic attitude in general.
      - *Social* (SOC): Ability to be a member of a social network.
      - *Emotional* (EMO): Acceptance of and ability to express emotions.
      - *Spiritual/phiosophical* (S/P): Regulation of actions by stable religious, family or cultural values or personal philosophy.
      - *Physical* (PHY): Regular participation in health-bringing activities.

   i. *Sense of Coherence (SOC*) (Antonovsky, 1991; Frenz, Carey, & Jorgensen, 1993): Contains five subscales concerning an individual's ability to understand and accept his life situation:

      - *Comprehension: Clear ideas about own philosophy of life and emotions.*
      - *Life interest: Experience of meaningfulness and continuity in intercourse with surroundings.*
      - *Self-efficacy: Ability to accept present life-situation and continue with own existence.*
      - *Interpersonal trust: Acceptance of others' actions that do not agree with own expectations.*
      - *Predictability: Belief that events in life are expected and predictable.*

   i. *Impact of Event Scale – Revised (IES-R*) (Horowitz, Wilner, & Alvarez, 1979; Weiss & Marmar, 1997): Contains three subscales which, apart from the traumatic experience itself (*Criterion A*), define PTSD according to DSM-IV:

      - *Criterion B: Intrusion*
      - *Criterion C: Avoidance*
      - *Criterion D: Hyperactivation*

Procedure

Data were collected at the volunteer train drivers' work place at SJ AB in Malmö. The drivers were interviewed singly, before or after their work shifts.

Results

Quantitative annalyses

Descriptive data are shown in Table 10.2. Of the 16 male drivers, thirteen were married, three single. Fifteen had children, one was childless. Eight had experienced fatal accidents, seven other serious accidents, and one no serious accident.

Table 10.2 Descriptive data for the train drivers (*n* = 16)

Type	Variable	mean	s.d.	range	missing
	Age	48,1	4,5	40 - 56	0
Background	Years in job	23,7	3,7	17 - 29	0
	Number of children	2,6	1,7	0 - 6	0
	Serious accidents	1.9	1.2	0 - 4	1
Accident	Fatal accidents	0.9	1.0	0 - 3	1
	Time off (days)	4.7	5.0	0 - 14	3
	Minor accidents	2	3.3	0 - 10	0
Vulnerability	Accident proneness	-0.4	0.6	-1 - +1	0
	Probability (%)	21.9	26.2	0 - 95	0

Test properties: For *CRI,* the statistical properties on all five scales were similar to the standardisation data, except that here *s.d.*'s were only half as big. For *SOC,* points on all five scales were considerably higher than in the cited studies, and the *s.d.*'s were considerably smaller. Valid answers on the *IES-R* test were obtained from 13 participants, including all those who had had a fatal accident, as well as five who had had a serious accident. The scores were somewhat lower than what one would expect for a serious PTSD.

Correlations ($n = 16$, $p < 0.7$):

i. *Background and Vulnerability*: (i) Between accident Probability (%) and Age, number of Minor accidents, and Accident proneness (Figure 10.1).
ii. Tests (*CRI*/*SOC*): For *CRI*, (i) between COG and all the other coping variables, and (ii) between SOC and EMO. For *SOC*, (i) between Comprehension and three other coherence variables: Life interest, Self-efficacy, and Interpersonal trust, and (ii) between Life interest and Self efficacy.
iii. Vulnerability and *CRI*/*SOC*: (i) Between number of Minor accidents and PHY and Life interest, (ii) between Accident proneness and EMO and Predictability, (iii) between accident Probability (%) and Life interest, and (iv) between PHY and Life interest (Figure 10.1).

ANOVAs with Fatal accidents (No/Yes, $n = 7/8$, $p < .05$) as independent variable:

i. Accident variables: No significant relationships.
ii. Tests (*CRI/SOC;* subscales as repeated measures): For *CRI*, no significant differences. For *SOC*, no significant difference for whole test ($p = 0.9$), but significance on two subscales, Comprehension and Self-efficacy.

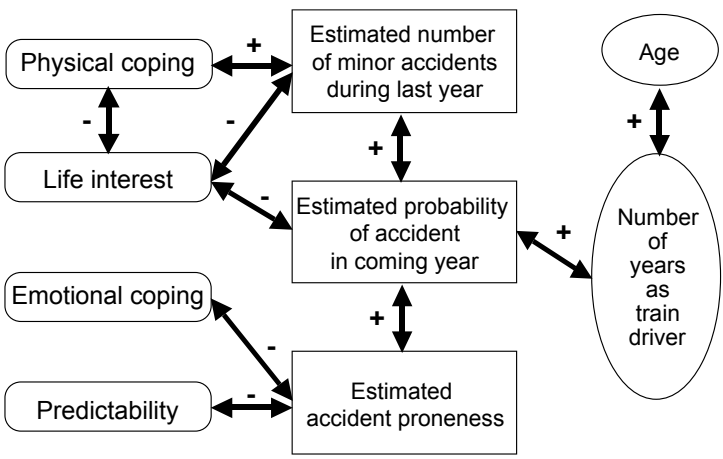


Figure 10.1 Associations between variables and test scores

ANOVAs with Family status (Single/Married, $n = 3/12$, $p < .05$) as independent variable in relation to estimated vulnerability and test outcomes (*CRI/SOC*):

i. Frequency of Minor accidents was 3.5 times higher for Singles than Marrieds. (Significant with Fatal accidents as a covariate.)
ii. Accident Probability (%) was 3 times higher for Singles than Marrieds. (Effect stronger with Fatal accidents as a covariate, $p < 0.01$.)
iii. Accident proneness showed no significant relationships.
iv. *CRI*: No significant relationships. *SOC:* Marrieds scored higher than Singles on all scales (total score: Singles = 146.7 and Marrieds = 165.2). MANOVA indicates a significant difference for the whole test and three subscales, Comprehension, Life interest, and Interpersonal trust.

Stepwise regression analysis of *IES-R* (three subscales, $n = 13$, $p < .05$): It was possible to predict PTSD indications on the *IES-R* subscales from background and test variables. *Intrusion*: 98% of total variation explained by 4 variables, Severe accidents, Probability (%), Predictability, and Comprehension; *Avoidance*: 79% of total variation explained by 2 variables, COG and Interpersonal Trust; *Hyperactivation*: 82% of total variation explained by 3 variables, Accident proneness, Probability (%), and Comprehension.

In-depth interviews

The accidents

The train drivers gave detailed accounts of each serious accident that they had experienced while at train controls, including thoughts, memories, fears, and visual and auditory impressions of the event. The drivers specified the circumstances of the accident, the length of time since the event had taken place, time of year, time of day, the location of the accident, driving conditions, stations near-by, and train origin and destination. They also gave objective descriptions and subjective impressions of the person that had suffered the accident.

The immediate aftermath

In most fatal accidents, the communal rescue service was responsible for collecting the deceased and cleaning up at the scene. Sometimes the fire-brigade or police were called in, and in cases where there was a survivor, he or she was taken to hospital in an ambulance. In accidents with a passenger train, other staff on board normally took charge of the situation as soon as possible. Otherwise the driver himself would alert staff at the nearest station and contact the rescue service, then sit in his cabin and wait to be relieved.

Before the 1980's, no established routines existed for assisting drivers after serious rail accidents, and usually the driver had to resume his work and drive on as soon as possible afterwards. At the beginning of the 1980's, the drivers themselves formed a peer association to help those who had experienced serious accidents. The association and the union came to an agreement with the rail company management, then wholly government owned, about measures to be taken to help drivers in cases of serious accidents. From the mid-1980's on, a train driver involved in a rail accident has without exception been relieved of his driving duties by a colleague who drove the train to the next station.

Step by step, fixed routines have been established. As soon as possible after an accident, the driver is joined by a trained "peer helper", who (in SJ) is one of a network of drivers on call night and day. The peer helper listens to the driver's account of the incident, and offers explanations of the driver's reactions. The physical reactions may include crying, shivering, vomiting, and involuntary urination or defecation, while typical psychological reactions related to the trauma may include feelings of unreality and helplessness. mental fatigue, grief, anger, and self-doubt: "Could I have done otherwise?" Afterwards, the helper accompanies the driver to his home, where the driver tells his family what had happened.

All the drivers praised highly the help given by their peers. They declared that the peer association was singularly effective and absolutely necessary for helping after a serious or fatal accident.

Nowadays, train drivers who have had such accidents are also sent to the company's medical service, where they are examined by a doctor who together with the management decides when the driver will return to work which, according to the drivers, may be from no time at all to 14 days, in most cases 4 – 5 days. The driver receives full pay during the time he is off work.

Reactions to serious and fatal accidents

Eight of the 16 drivers had experienced at least one fatal accident, and all except one of the others had experienced at least one serious accident where the victim had survived. When asked about what they would do if they were to experience another serious accident on the rails, they would

typically answer: "I cannot plan how I shall react. There is no rule-book. It happens so quickly." "It is impossible to veer off if there is someone who wants to kill himself. You cannot help." "There are always minor incidents – it happens once a week or so. One comes along at full throttle – it only takes a few seconds."

The immediate experience of the accident was never easy. Some of the drivers believed that it was easier to endure if you shut your eyes when it was clear that the collision was unavoidable. Others would not use this strategy, as they considered it necessary to check the instrument panel at that point. Looking away was also a strategy mentioned by some, and some considered blowing the whistle when they discovered the person on the rails to be a way of trying to prevent the inevitable.

When asked how it affects them when a colleague has had a fatal accident, their reactions vary. Most become ill at ease, wondering when it is going to be their turn: "Last December there were many fatal accidents, and then I noticed that I was always on edge." "One is more observant, tense. I had nightmares. I fear it will be my turn next." "A feeling of unease – it is there all the time. It has been happening so often lately".

Generally, the drivers are painfully conscious of the ever-present risk of a fatal accident, and that the risk is big. From the time they started their training, they have known about this eventuality, and they are continually reminded of it by the frequent accidents of others. They frequently examine the accident statistics and work out the likelihood of having such accidents themselves. They pity other drivers who experience fatal accidents, and maybe remember how they themselves were affected. Many drivers told of a paradoxical feeling of "relief" at hearing about others' accidents, that it was over for this time, but also of increased uncertainty and apprehension.

Accident sequelae

The train drivers learn of fatal accidents in the area through a general mobile phone alert: "When one hears that traffic has been stopped along a certain stretch, one knows that something serious has happened." It is not announced officially who the driver was, but for the most part, such information is passed along the "grapevine". Most of the drivers prefer it to be that way.

The drivers place great trust in the peer helpers, who are from their ranks, who have themselves unintentionally run over somebody on the rails, and know what it is all about. They know that any train driver who has such an accident will be well taken care of.

Generally, the drivers did not think that the traumatic experience affected their work, but confessed that they often felt tense. Most of them were of the opinion that certain times of the year were much more difficult than others, the worst periods being examination time at high school and college, the midsummer festivities, and the days before Christmas.

Memory of the accident

Most of the train drivers realised that the fatal accident was permanently etched into their memory, and that memories can flash back without warning, in particular on receiving news of yet another fatal accident. What most of the drivers remember clearly is the moments just before the accident happened, who it was that relieved them, and the peer helper who came to fetch them. Otherwise the drivers often had a feeling of a memory gap regarding the accident.

Some of the drivers said they frequently thought about accident, while others said that they seldom or never thought about it, sometimes adding that they had thought about it "mostly the following day", or remarked something like "for a month, when I was driving I was very tense and vigilant".

Some drivers commented that "one never gets over it" and "this is something you have to live with". They also commented that some memories of the accident change and blur, while others keep popping up. Some said that they had had difficulties with their feelings immediately after the accident. One driver commented that "the more accidents you experience, the less sensitive you get", while another driver stated that "you get more vulnerable" for each accident you experience.

Comments such as "it sits in the subconscious" or "one is especially tense on certain stretches" were common. Most drivers maintained that sometimes "the pictures pop up" as clear as if it had happened yesterday. Some details "are glued on the retina", and one driver commented that when he passes another train that has a particular number, he gets a nauseous feeling, just as if it had been the particular train he had driven over the person. It is also practically impossible for the drivers to avoid driving along the same stretch some time again. They experience this in different ways, depending on the time elapsed since the accident and on how traumatic it was, but detailed memories of the event and their own reactions seem inevitable. Because of this, there is a firm rule that, as soon as he feels he is up to it, the driver goes along as a passenger with another driver who is driving on the stretch, in an attempt to induce habituation of the stress response.

Many of the drivers commented that by far the worst thing that could happen was for them to drive over a child. There was a big difference between driving over a "suicide" or if it was a "pure accident". In the former instance they see the train as an instrument used by somebody for taking his life, in which case the driver generally comes to the conclusion that he did all he could to prevent the acccident. Such an explanation does not prevent self-recrimination, guilt, doubts, and anxiety from emerging from time to time. Often a suicide leaves a note for his or her family, and it brings a feeling of relief for the driver when the police confirm this.

Discussion

The train drivers were unanimous in saying that they were satisfied with their job, a statement corroborated by their long working lives. The high *SOC* scores would also indicate that most of the drivers indeed feel they are "in their right niche in life". The homogenity of the group is also reflected in the relatively low variation in the drivers' test scores and clear relationships between test results and other variables.

The long working life also means that most of the 18 drivers had experienced a serious accident on the rails, some more than one. Both the test and interview results show that a fatal accident usually leaves the drivers with deep psychological scars, and that they afterwards use various strategies for dealing with recurring memories of these events.

The drivers are in many ways a tightly knit group with intimate knowledge about one another. On account of the frequent rail accidents, the drivers also constitute a singularly vulnerable group, and they are conscious of their own and other drivers' vulnerability and can be compared with other jobs such as the police or military. You know and accept the odds from the start, but still hope that bad things will not happen. When they repeatedly do, it is uncertain whether drivers become more or less affected for each new occurrence.

"Some time it will be my turn!"

Two sets of attitudes emerge when associations between test variables and vulnerability are examined (Figure 10.1). Some of the drivers seem to consider themselves to be quite accident prone, but are less able to cope emotionally, accept or show their feelings, or predict what is going to happen in their lives and how they will react. There are also those who report many recent, minor accidents, are interested in leading a healthy life and keep physically fit, but otherwise find

little meaning and coherence in their lives. Drivers showing either of these sets of attitudes deem it likely that they will soon have a serious accident.

This taps into a major current in the life of the train drivers, the clear possibility that they will in fact experience at least one major accident on the rails during their working life. This affects the drivers in different ways, some more adversely than others. We also find that the more years the drivers have been in their jobs, the more likely they are to expect a serious accident to happen soon. The drivers who had not experienced a fatal accident commented that it was only a matter of time before it happened. It seems that for them the probability increases as time passes. The opposite belief also showed up, that if one had already experienced a fatal accident, there was less subjective probability of it happening again. Such reactions seem to be universally human. One sees parallel behaviour in soldiers under artillery attack during wars, who throw themselves down into fresh bomb craters in the belief that it is less likely that a bomb will land there again.

Beliefs of this kind among train drivers may be related to or maybe a consequence of great and small accidents that they actually experience in their job. It is also interesting to note the drivers' interest in accident statistics and their own and others' accidents. Feeling unable to influence the situation, the drivers may see themselves as predestined to cause another person's death at some time during their career. They feel both dissatisfied and helpless about their future, but are eager to obtain a better insight.

Drivers who had experienced a fatal accident had lower scores than the others on both *CRI* and *SOC*, and thought of themselves as more accident prone. This indicates a considerable, negative, psychological effect of the accidents. *CRI* and *SOC* subscales were also negatively correlated with PTSD indicators (*IES*): Low COG and low Interpersonal Trust implied *Intrusion*, low Predictability and Comprehension implied *Avoidance*, and low Comprehension implied *Hyperactivation*. Thus, drivers with poorly developed coping strategies and sense of coherence in their lives also show PTSD indication, and this was characteristic of the drivers who had experienced fatal accidents. There were also strong associations between the *IES* subscales and some of the other variables, indicating some serious underlying problems.

Drivers who feel especially vulnerable to accidents feel less engaged in their personal life, they are sometimes health fixated, but at the same time blunted emotionally, and they sometimes feel uncertain about their relationship to other people and out of social context. Their personal well-being and quality of life is often comparatively low, and is in some cases characterised by anxiety or depression. Generally, they appear to be suffering from a long-term PTSD.

The psychosocial situation and job relations

The psychosocial situation both privately and at work profoundly influences the psychological well-being of the drivers. This is also likely to be reflected in the quality of their work, as has been indicated previously (Edkins & Pollock, 1997). There were general indications of lowered work capacity among the train drivers on account of the irregular hours, which often resulted in irregular diurnal rhythm and lack of sleep. Improved work situation in this respect would almost certainly lessen this kind of stress. This is likely to be particularly important during the period when a driver is recuperating after the accident and trying to cope with emergent PTSD. An improved work schedule would also have beneficial effects on the driver's social life and home environment. Drivers who have a family that gives support in difficult situations show a greater sense of coherence and meaning in their lives – the lower quality of life that sometimes characterises those who live alone is particularly visible for those drivers who have experienced a fatal accident.

However, there was some indication that the drivers do not fully utilise the possibilities they have for rest after a serious accident. There may exist some group pressure to return to work at the earliest opportunity, and the position of the management is not clear. While some of the drivers

maintained that an early return to work took their mind off the acccident and helped them resolve the trauma, some of those who had been off work for less than a week after the accident were of the opinion that a longer period of rest would have been better.

Conclusions

The main results concern the psychological after-effects of having had a serious or fatal accident on the rails. Associations obtained between test results and other variables indicate long-term traumatic effects of these events, with a reduced psychological well-being as a consequence. While there are strong indications that the drivers develop a chronic PTSD after fatal accidents, the after-effects of other, non-fatal, serious accidents are not as pronounced. The drivers employ various coping stategies and defence mechanisms to ameliorate the negative psychological effects. It is also clear that support from family, peer group, and management is of great importance in the healing process. However, it is not possible to accurately assess the resultant safety effects on the basis of the present data.

References

American Psychiatric Association. (1999). *Mini-D IV. Diagnostiska kriterier enligt DSM-IV. (Andra svenska utgåvan)*. Danderyd: Pilgrim Press.

Antonovsky, A. (1991). *Hälsans mysterium* (M. Elfstadius and f. a. M. C. o. L.-G. Lundh, Trans.). Stockholm: Natur och kultur.

Banverket. (2004). *Statistik över olyckor på statens spåranläggningar år 2003.* Borlänge: Banverkerket.

Cocks, R. A., Fletcher, A., and Dawson, D. (1989). Trauma in the tube: The problem of railway suicide and its consequences. Field-based validations of a work-related fatigue model based on hours of work. *Stress Medicine, 5*(2), 93-97.

Edkins, G. D., and Pollock, C. M. (1997). The influence of sustained attention on railway accidents. *Accid Anal Prev, 29*(4), 533-539.

Frenz, A. W., Carey, M. P., and Jorgensen, R. S. (1993). Psychometric Evaluation of Antonovsky's Sense of Coherence Scale. *Psychometric Assessment, 5*(2), 145-153.

Horowitz, M., Wilner, N., and Alvarez, W. (1979). Impact of Event Scale: a measure of subjective stress. *Psychosom Med, 41*(3), 209-218.

Karlehagen, S., Malt, U. F., Hoff, H., Tibell, E., Herrstromer, U., Hildingson, K., et al. (1993). The effect of major railway accidents on the psychological health of train drivers--II. A longitudinal study of the one-year outcome after the accident. *J Psychosom Res, 37*(8), 807-817.

Malt, U. F., Karlehagen, S., Hoff, H., Herrstromer, U., Hildingson, K., Tibell, E., et al. (1993). The effect of major railway accidents on the psychological health of train drivers – I. Acute psychological responses to accident. *J Psychosom Res, 37*(8), 793-805.

Myrtek, M., Itte, H., Zimmermann, W., and Bruegner, G. (1994). Psychische Bewaeltigung von Unfaellen bei Lokomotivfuehrern: Die Relevanz von Copingfragebogen zur Erfassung von funktionalen und dysfunktionalen Copingprozessen. *Zeitschrift fuer Klinische Psychologie Forschung und Praxis, 23*(4), 293-304.

Parkes, A., and Lansdown, T. (2000). Driver error is... ...a system error. *RAIL BULLETIN, 12*(2), 10-12.

Psykologiförlaget. (1991). *Coping Resources Inventory - CRI*. Stockholm: Psykologiförlaget, AB.

Theorell, T., Leymann, H., Jodko, M., Konarski, K., and Norbeck, H. E. (1994). 'Person under train' incidents from the subway driver's point of view--a prospective 1-year follow-up study: the design, and medical and psychiatric data. *Soc Sci Med, 38*(3), 471-475.

Tranah, T., and Farmer, R. D. (1994). Psychological reactions of drivers to railway suicide. *Soc Sci Med, 38*(3), 459-469.

Weiss, D., and Marmar, C. (1997). The Impact of Events Scale - Revised. In J. Wilson and T. Keane (eds.), *Assessing psychological trauma and PTSD*. New York: Guildford.

This page intentionally left blank

Chapter 11

The Cognitive Tasks of the Driver: The Approach and Passage Through Diverging Junctions

Amanda C. Elliott, Sarah D. Garner and Elaine Grimes

Introduction

In the UK, there are a number of different permitted means to signal junctions. This introduces some inconsistencies to the driver on the approach, and when driving through a turnout, in terms of both information presentation and meaning. The main potential problems which arise are Signals Passed at Danger (SPADs) overspeeding and that train drivers cannot remedy mis-routeings. The result of mis-routeing a train include the train and route being incompatible or the driver not being "signed for", i.e. has not been trained and assessed over the route. Mis-routeings always result in lost time but could also lead to the need to set the train back against signals and, in a worse-case, a fouling of stock and infrastructure.

The UK uses a route-signalling system such that drivers are signalled for the route they are to take. Most other countries in Europe use a speed-signalling system, which tells the driver what speed to drive at over the next section. Thus the UK has some idiosyncrasies to manage with respect to junction signalling. One important issue is that, although signallers have a duty to set the correct route, the driver must prevent a mis-routeing by stopping the train before the junction wherever possible (Rule Book, 2003). In any case, the driver must always drive to the signal aspects shown; must drive through the turnout at the correct speed for the route set; and must stop the train if it has been mis-routed. Performing all of these activities correctly relies upon the driver *knowing* which route they should be taking, the speed over the junction and, in complex areas, the different means of achieving it. Therefore, research work carried out for this project sought to understand how and when the driver determines what route is actually set, so that they can drive the train accordingly.

Here we describe the method used to understand the information the driver requires and how they use this information on the approach to a junction, so that junction signalling options can be considered from a human-centred perspective. The consideration of junction signalling options, to improve the efficiency of junction signalling without detriment to safety, is currently work in progress, and not presented here.

Method

The project used a task analysis (TA) as the method for understanding the cognitive tasks of the driver on the approach and passage through junctions. The aim of the task analysis was to identify how drivers approach junction signals by describing the tasks that they perform in order to achieve their goal of driving a train safely and efficiently. The task analysis was limited to a descriptive

overview of the tasks associated with the approach to and driving through junctions and does not purport to be a full analysis of all driving tasks. For our needs, the scope of the task analysis was limited to the following driving tasks:

- Signal and aspect recognition and understanding;
- Route recognition and understanding;
- Speed regulation whilst approaching junctions.

Task analyses from previous work conducted for RSSB (HEL, 2003; QinetiQ, 2003) were made available. However, the re-use of these task analyses was limited, as it became apparent early in the work that neither of the task analyses available dealt with the cognitive driving tasks in sufficient depth for this project to be able to use them as they were. As such, effort was required to develop a project specific task analysis. Therefore, initial task concepts were developed by the project team and a time-line analysis of the order in which drivers carried out their tasks was produced.

To provide further depth, a series of meetings were held with various Train Operating Companies (TOCs) in the early stages of the project in order to understand their issues and concerns relating to junction signalling from a user's point of view. It was important that the project gained a true understanding of what the driver really does, and the information that they need, as there are many theories as to how a driver deals with junction signalling.

Discussions and brainstorming sessions were also held with the project team and operations experts to facilitate the development of the task analysis. An analysis of Safety Management Information System (SMIS) data relating to junctions was also included as an input to our comprehension at this point.

In order to validate the initial analysis and models, we undertook a set of interviews with 23 drivers, at 4 different depots (3 different TOCs). These interviews were analysed and were used to produce a framework, which updated and validated the tasks and the interconnections at both the higher (situational) level and cognitive level. In addition, a wider review was carried out on the draft task analysis, in particular by RSSB. This review provided further pointers for consistency and accuracy, resulting in a better-presented, coherent analysis.

Understanding how all of the elements in the driver's environment interact and affect what is likely to happen next was key to presenting a full cognitive model of the driving task. Such a situation awareness model (SA) was developed as a high-level model of the tasks that the driver has to perform. Each of the individual tasks within the high-level model was then further re-described in activity-flow diagrams. This activity-flow approach uses simple flowchart principles employed in logic or programme flowcharts. This choice of TA method was made due to the highly iterative nature of the tasks reviewed. It was considered that a Hierarchical Task Analysis approach was not flexible enough to be able to break down the cognitive task in an easily demonstrable, logical way.

Situational Awareness Model

The output from this project resulted in an overview of how the driver's tasks and elements in their environment interact. This situational view provides an indication to the driver of what is likely to happen next, when combined with their route knowledge and experience. The model shown in Figure 11.1 explains our findings, using the different SA levels, as described by Endsley (1999). The high-level task relationships can be understood by considering the approach to the junction signal and the pieces of information the driver will be absorbing and using. In the first SA level of perceiving, the driver will be taking in and identifying relevant pieces of information, which could

be a variety of external stimuli such as the feel of track, noting the presence of a particular bridge or speed board, or the information on the signal itself. Emphasis in this model is on SA level 3, prediction, with the SA level 2 (understanding) taking up less of the driver's cognitive activities. We found that all response levels (decision and performance actions) were shaped by the prediction tasks.

From our thorough review of drivers and their tasks, we developed a set of task analyses. The task analyses were presented as activity-flow diagrams, the two most pertinent diagrams are shown in Figures 11.2 and 11.3.

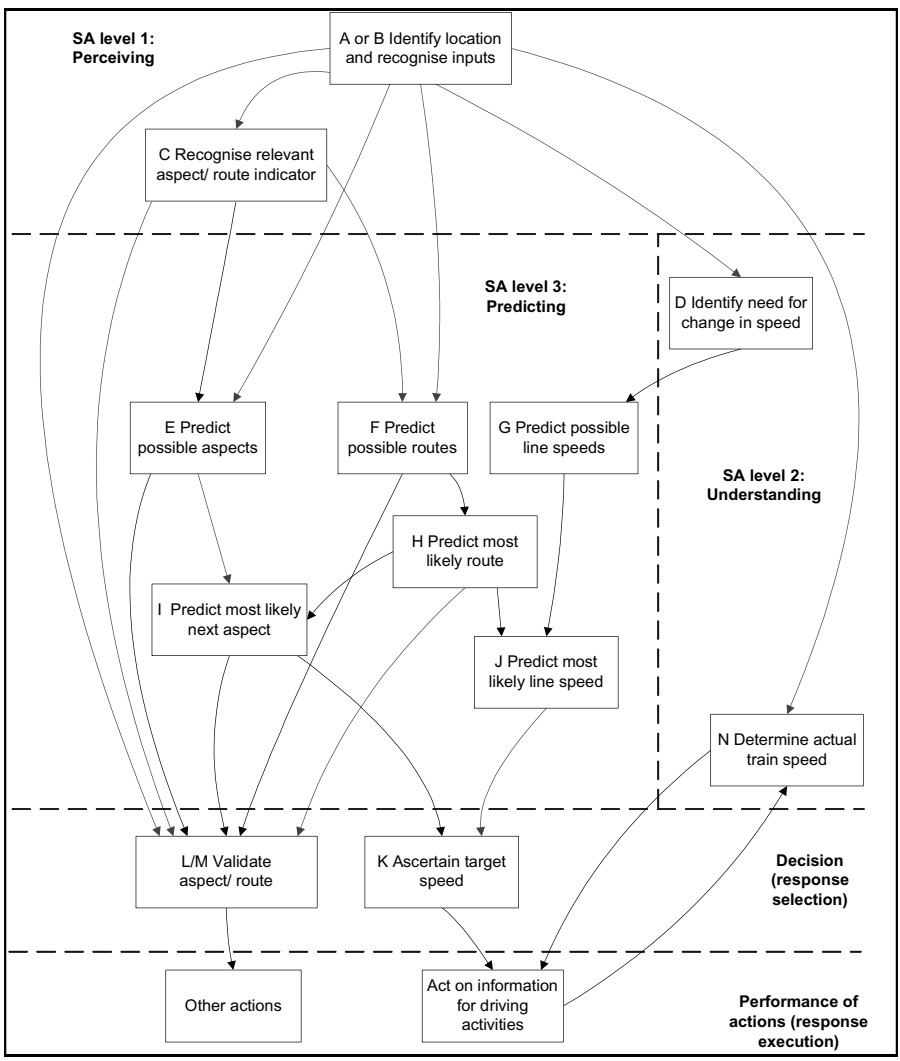


Figure 11.1 The high-level tasks of approaching and driving through junctions, and their task relationships to each other using the demarcations within Endsley's model of situational awareness (SA)

Task analysis – Identification tasks

Figure 11.2 depicts the relationships between the identification tasks A, B, C and D, which are described in detail below.

Identify relevant location (Task A) and recognise correct inputs (Task B)

There is iterative relationship between identifying a relevant location (Task A) and recognising specific inputs (Task B); in order to know where one is, the driver must find a known or relevant feature, and in order to determine if the input is relevant in the first place, one must knows one's location. Task A encapsulates how the driver receives external inputs and checks them against their knowledge of the route. This is a continuous process throughout the driver's journey and should result in the driver knowing where they are. Drivers stated in interviews that if this continuous process becomes disrupted, it was uncomfortable not knowing where they were, showing a strong focus on the need for route knowledge. This explains that drivers use cues to identify location and associate these with other actions such as braking. The interviews recorded that it is the norm for items that were not part of the fabric of the railway to be used to identify location, using either general landscapes or specific recognition cues to locate oneself.

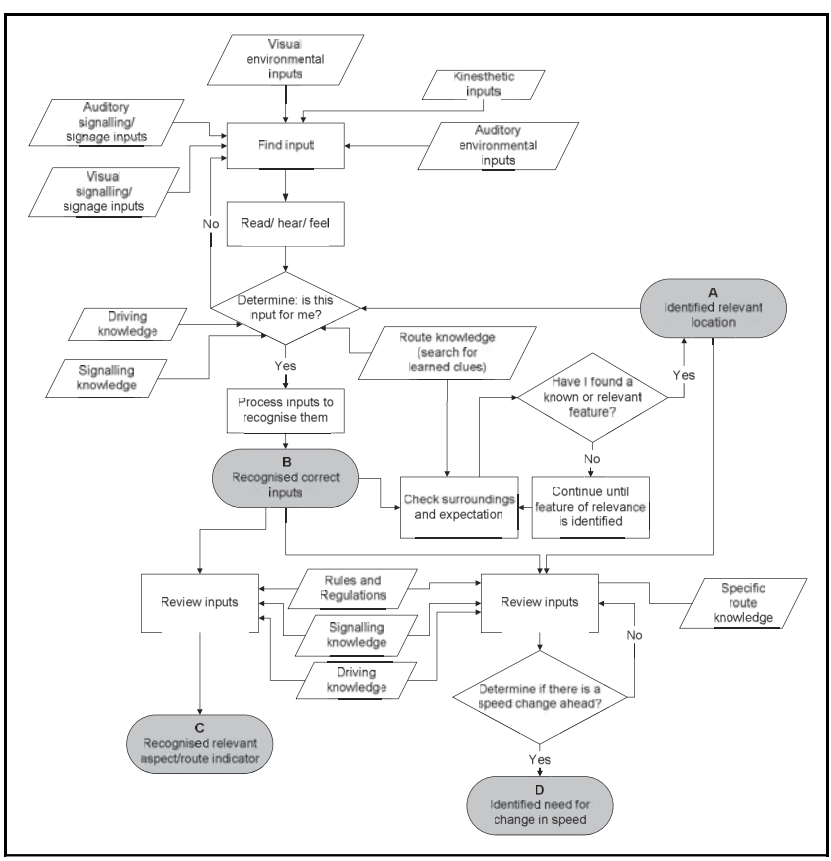


Figure 11.2 A detailed activity-flow analysis for the identification tasks of the train driver

Task B is about the driver absorbing the correct visual, auditory, kinaesthetic inputs and making sense of them in context, before using them for the tasks that follow. These are the foundation for other tasks in this analysis and an error here will propagate throughout the remainder of the driver tasks and activities. This task was identified during the interviews; it was not included in our original analysis. It became clear that a generic level activity of becoming aware of an input (find input), filtering it (is this input for me?) and then taking it in (read/listen/feel) was used for most forms of information. Drivers are presented with many sources of information; therefore they selectively attend to, and process, information that is of relevance to them. In the high-level situational awareness diagram (Figure 11.1) Tasks A and B are shown together. However, it is important to note that Tasks A and B are distinctly different to one another, as the driver can recognise inputs under Task B, which are not related to the location. Also, the driver may identify their location (through Task A) without processing other inputs (through Task B).

Recognise relevant aspect/ route indicator (Task C)

Once located, the driver will be able to apply their knowledge to what they see, to recognise what the input actually means to them. In general, Task B can be considered sufficient for many of the driving inputs. However, we found that the driver will commonly go through a deeper processing phase to understand what a signal aspect or route indicator means. The reason for this deeper processing is because Task C provides input to prediction and validation tasks and different aspects and route indications can have more complex meanings. For example, if the driver sees a yellow aspect, the instinctive meaning is "prepare to slow"; however, if it is yellow and flashing the meaning is changed to "drive to clear the junction and slow if necessary for the junction speed".

Identify need for change in speed (Task D)

On any rail network, there is the regular need to revise the speed of a train. To this end, a driver may use a speed board or warning indicator to provide the input to this task. However, generally, they will also take a cue from the external environment, such as permanent structures at points close to where they expect to change speed. During the interviews, the drivers reported that they know where the speed changes are, from their route knowledge. Therefore, they did not drive to the speed boards, in the same way they do not drive from signal to signal. They recall from memory the speed change ahead and adjust their speed prior to reaching the speed board, only using it to confirm their prediction. For deceleration, speed boards are presented too late to brake to the speed board, so recall of the correct place to commence braking is essential. For acceleration, the point at which the driver can accelerate to the higher line speed changes for the length of train.

Note that although these direct inputs could lead to the driver actuating a change in speed, the real change in speed or velocity is also affected by a number of prediction tasks, which are described next and are shown in Figure 11.3.

Task analysis – Prediction tasks

Drivers reported that they drive proactively, rather then reactively, and thus predict the possible signal aspects ahead. The following narrative details the two levels of prediction tasks; the prediction of possible and the most likely outcomes, as shown in Figure 11.3.

Predict possible aspects (Task E); routes (Task F); and line speeds (Task G)

Having recognised the signal aspect and carried out the deeper processing in Task C, the driver will use the relevance of the location and his/ her knowledge to predict all of the legitimate next aspects, for the next signal and often for a number of signals ahead during Task E. This task relies upon the driver accurately remembering the signal aspect or route indicator (Task C) and having identified their route correctly (Task A).

In line with proactive driving practices, which the drivers reported, all possible routes are also predicted as Task F, so that at a later stage they do not reject a legitimately viable alternative. The driver uses their location and last seen aspect along with route and rule information to predict the most likely route. For example, if the driver sees a double yellow they might predict that they are being routed off the main line (if the train was not being routed off the mainline, a green might be expected).

Task G relates to the anticipation of speeds that the driver may be required to drive at in the near future. The driver is aware that they need to make speed changes regularly, from driving and route knowledge. This task is necessary because UK signalling is route not speed based. For example, when the driver knows they are approaching a crossover, they will know the two possible speeds for each potential route. They need to predict these speeds as the speeds will not be

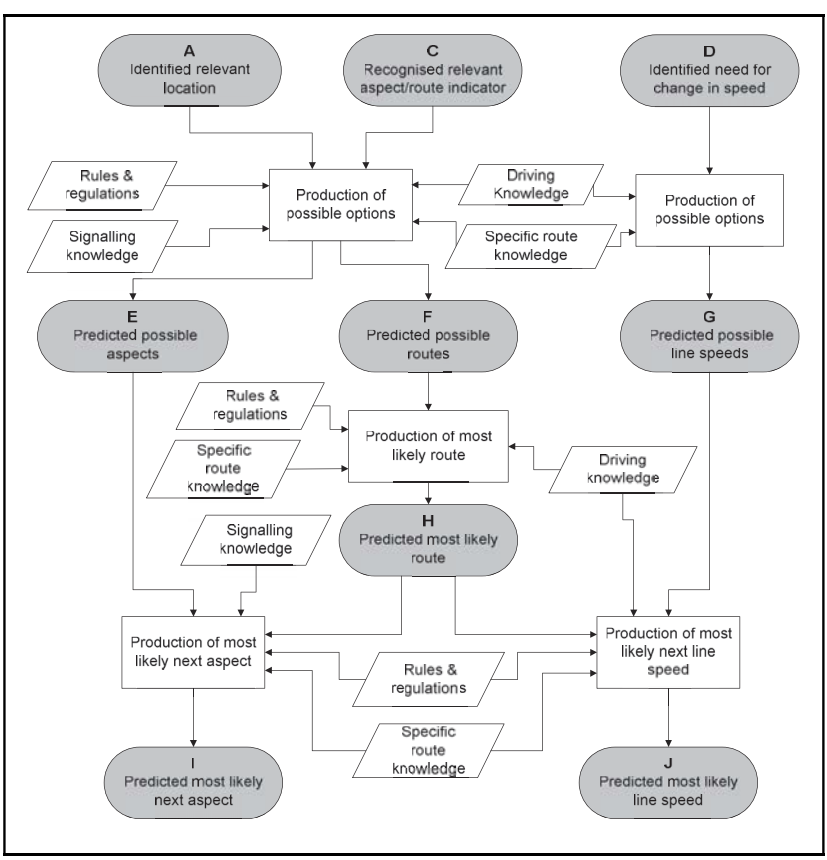


Figure 11.3 A detailed activity-flow analysis for the prediction tasks of the train driver

explicitly presented well ahead of the junction itself. In reality, this task is very subconscious and is a transient link to Task J, where the route information makes the difference to which speed the driver will predict is most likely. Drivers stated that if they were unsure of the speed, they would always prepare themselves to drive with caution.

Predict most likely route (Task H), next aspect (Task I) and line speed (TaskJ)

As with the other prediction tasks, these tasks rely upon short term memory inputs. From the possible outcomes, the driver will pick one as the most likely. This may be carried out seamlessly with the tasks of predicting the possible routes and aspects (so that the driver may not consciously separate the two cognitive activities), or at some time later, when more information or time is available. This will form the basis of an expectation that the driver will have, which will influence their further actions. This applies equally to places where the train may be stationary (e.g. platform) or whilst on the move.

As these tasks were analysed, it became clear that routes and signal aspects were strongly related. Thus, the analysis explains that the driver is likely to use the predicted most likely route to help determine the most likely next aspect. The tasks are, in many respects, iterative; the actual signal (Task C) provides an input to predicted possible route (Task F), which leads to the prediction of the most likely route (Task H), which helps the driver then determine what the most likely next aspect will be (Task I). Note that the validation tasks (Tasks L & M) loop back to Task C, and the whole cycle starts again. The relationship between Task I and Task H was difficult to determine categorically. This was because drivers themselves found it difficult to decide whether aspects help to determine which route is set (Task I feeds Task H), or vice versa (Task H feeds Task I). As such, the relationships presented for these tasks provide one acceptable and workable description.

To undertake Task H, the driver normally takes their predicted possible routes and compares them with their usual pattern and any modifiers. Modifiers may be real-time modifiers (e.g. "the freight [train] hasn't passed me yet and I'm likely to get routed in front of it") or temporary longer-term modifiers (e.g. I know that there are engineering works on that line at the moment). This prediction may be continually updated and reassessed as the driver continues down their route, and additional inputs are fed into the process.

A further prediction task for driving on the approach to and through junctions is the prediction of the most likely line speed. Having predicted the possible line speeds in Task G, the driver will pick a most likely line speed, according to the most likely route (predicted in Task H). The driver will update their prediction of the most likely line speed modified, as further confirmatory or contradictory evidence arises.

Discussion

A driver's primary aim is to drive safely on the approach to junctions, as at other times. During the interviews drivers stated that there are two vital issues for them to drive optimally for junction signals: knowledge of whether the next signal aspect will be red and knowledge of what route is set. In turn, these two issues determine what speed the train should be driven at (e.g. for the turnout) and whether to stop. Stopping the train because of the red signal or because of a routeing issue can be described as a function of the amount of time the driver has to react to an unplanned or unpredicted aspect or indication.

This project puts forward the concept that drivers do not react to each individual signal but plan ahead; the majority of signals that are seen after the prediction then confirm what the drivers have already predicted and planned for. This planning or prediction will happen to a greater or

lesser extent depending on factors such as route driven, stopping pattern, training, current weather conditions, traction type, operating condition (normal/abnormal) and individual driving styles.

The high-level situational model shows a breakdown of the cognitive levels of driver's tasks. Drivers are able to predict things quickly using learnt information – they are able to predict well because they are experienced in both driving and the route and have knowledge of signalling principles in practice. They are able to make predictions based upon their skilled behaviour and expectancy (knowledge of what has gone before).

The model shows that information is used to progressively confirm the predictions that the driver has made. It also indicates that the final validation is not made until the junction signal. Any of the predictions can only be validated when presented with a suitable external stimulus; these are represented on the SA model as Tasks L and M. The driver will recognise appropriate inputs in order to confirm or modify previous predictions. In the majority of cases, the cycle of recognition, prediction, further prediction and validation is carried out by the driver on a continuous basis without any conflicts arising. If the *predicted* aspect or route indicator is not exactly the same as that actually seen – the driver will immediately review what they have recognised as the input against other legitimate possible predictions (from Tasks E or F, as appropriate). They will then act according to the actual input, *not* the predicted one, if it appears valid to them. If it is not a valid input, the driver will take a course of action to stop and validate the aspect or the route indicator by contacting the signaller or control, as necessary.

From the interviews, the drivers stated that they will use whatever signal aspect or route indication information is available to improve the reliability and consistency in their prediction of the signal aspect or route indicator. For example, the use of advance warnings of route, such as Preliminary Route Indicators, flashing yellows and splitting distances are used to provide the driver with a more accurate prediction. The driver will usually become more confident of their longer-term predictions as aspects are passed. However, they cannot validate the predictions as correct until the junction signal becomes visible.

The driver may not be able to form an appropriate prediction in terms of timing or content if the processes described in the task analysis are disrupted. Disruptions may come from changes to the normal or expected course of events. This may explain why some SPADs occur during disrupted conditions or perturbations.

The tasks that require recognised inputs, such as Tasks B and C are dependent upon the information being presented in a timely fashion, with reliable and consistent meaning. Possible errors here mean that the driver may not be able to form an appropriate prediction or may not be able to react to an unexpected indication that does not match the prediction. These factors associated with information presentation may well be attributed to the high number of occasions where the driver does not stop before the junction signal in the reported cases of mis-routeings. It can be understood, therefore, that errors may occur where the driver has correctly recognised the input but then mis-recalled the previous aspect, or incorrectly predicted the next aspect; leading to the driver to be presented with an unexpected aspect.

This model is a simplified representation of the information processing and prediction activities that a driver undertakes to maintain their situation awareness and drive the route safely and efficiently. There may be several of these loops continuing in parallel, in different stages or for different lengths of time. Importantly, the concept of identification and prediction is clearly shown and should be used to influence any recommendations for future junction signalling.

Observations

The application of the cognitive task analysis to the task of driving through junctions has been particularly useful in understanding how drivers use speed and signal information presented to

them. This knowledge has been useful in considering junction signalling options and how they these options can support the driver in forming and maintaining appropriate situation awareness.

In summary, the following are the main observations from this work:

- Drivers have two vital issues for them driving optimally: knowledge of whether it will be red and knowledge of what route is set;
- Drivers continually predict the next possible outcome and action;
- Drivers may use a large number of inputs for identification tasks;
- Drivers are heavily reliant upon taught and acquired route and driving knowledge;
- Drivers plan ahead and include contingency (understanding possible legitimate options) in their planning;
- Drivers can be completing several different prediction and validation activities at the same time;
- Exactly the same information (e.g. a single signal aspect) can be processed by the driver for use in prediction as well as validation and response;
- The final outcome is heavily reliant upon the outputs from the earlier tasks being correct;
- ·Incorrect or mis-recalled predictions can lead to the driver seeing an "unexpected" aspect or indication, or in the worst-case missing this information;
- The situation awareness of the driver can be affected by errors at all levels from identification and recognition, to processing and decision-making for all prediction and validation tasks.

Acknowledgements

This work has been carried out as part of RSSB Project T274 "A Human-centred approach to Junction Signalling and Route/Train Incompatibility Protection". Lloyd's Register Rail Limited and RSSB would like to thank all of the Train Operating Companies and train drivers who took part in this study.

References

Endsley, M.R. (1999). *Situation awareness in aviation systems.* In Garland, D.J., Wise, J.A., and Hopkin, V.D. (Eds.), Handbook of Aviation Human Factors, Mahwah, NJ: Erlbaum.

HEL, (2003), Stage 1.2 Hierarchical Task Analysis of the UK Train Driving Task. Produced for Railway Safety. Human Engineering Limited (Ref. HEL/RS/02768a/RT3) 22nd January 2003, Issue: 01.

QinetiQ, (2003), *Interim User Information Requirements Report*. Produced for RSSB under Contract No Rserv 296. (Ref. QINETIQ/KI/CHS/CR032109/1.0)

Rule Book GE/RT8000. Module S2 Section 5. June 2003, Issue 1.

This page intentionally left blank

Chapter 12

A Train Driving Simulator Experiment to Investigate Driver Fault Diagnosis

W.H. Gibson, M.W. Halliday, L. Sutton, J. Shelton, and M. Bond

Introduction

Train driving cab simulators have been recognised as an important component of train driver training (Keon, 1991; Kappler, 1993) and this was emphasised in the UK at the Ladbroke Grove Inquiry (Cullen, 2001). Subsequently, many UK train operators have adopted simulator training as part of their competency management systems and the Rail Safety and Standards Board has developed guidance on this topic (RSSB, 2005). Within the wider context of railway operations simulators have been used for research and human factors design assessments in the context of signaller train control or train dispatch (Lenior, 1993; Nichols and Crawford, 2002; and Malsch et al., 2004). Young (2003) identified the importance of train driving cab simulators to human factors research related to train drivers. However, while published examples do exist of the use of train cab simulators for research (Roach et al., 2001; Hamilton et al., 2004) they are relatively rare, particularly in the context of the assessment of design changes to train cab interfaces. This study describes an experiment which captured train driver performance data when driving a train driving cab simulator (Halliday and Gibson, 2005). Driver performance was compared for three different interface designs for the diagnosis of non-driver initiated brake demands, particularly those related to the Train Protection and Warning System (TPWS).

TPWS is one of the key train safety systems used in the UK. TPWS automatically applies the train brakes in certain situations when the train speed is too high or the train has passed a signal at danger. It has been identified that train drivers can misdiagnose the cues presented in the cab after a TPWS brake application has occurred and in some cases ignore these cues. These failures can have consequences, in particular, if the driver does not identify that a signal has been passed at danger and the train continues to a point of conflict. To reduce the frequency of occurrence of these failures, improvements to the design of the interface have been proposed in order to improve driver diagnosis performance regarding TPWS initiated brake demands.

It should be noted that the controls and displays for the TPWS system are also related to the Automatic Warning System (AWS). The AWS is another key safety system which automatically applies a train's brakes, in this instance, if the driver does not respond in a timely manner to a warning horn presented for events such as restrictive signal aspects or speed restrictions. The new interface designs therefore have included changes to the driver interfaces for both AWS and TPWS.

The use of a full train cab simulator was identified as the best method to compare the proposed designs because:

- TPWS and AWS activations occur rarely during normal driving, whereas they can be presented at a higher frequency in a simulator.

- Train drivers can use both the internal cab interface and external cues (e.g. lineside signals) to diagnose the cause of a brake demand. Therefore, the use of a simple desktop simulator without external visual cues would have been inadequate and would also have suffered from a lack of face validity for industry stakeholders.

The current and new TPWS driver interfaces

The original interface design is termed variant 1 in this chapter and the two new designs variant 2 and variant 3. The weaknesses in the current design, which were considered likely contributors to driver misdiagnosis, etc, were rectified with the two new designs and these are described below:

1. Variant 1 does not provide an audible alert to the driver that a TPWS activation has occurred. Both of the new variants include the same unique audible alert from an intermittent horn to alert the driver to the fact the brake demand is occurring due to a TPWS activation.
2. No lasting unique visual indicator is provided in variant 1 to identify TPWS activations and differentiate the activation from AWS. In addition, there is no differentiation of whether the TPWS activation is related to a TPWS Overspeed Sensor (OSS) or TPWS Train Stop (TS). The TS generally indicates that a signal has been passed at danger and is therefore more safety critical than OSS which applies the brakes if the train speed exceeds defined criteria. This differentiation is achieved with light state indications for the two new variants which differentiate TS and OSS and are presented using the same red brake light indicator used for AWS. This is the critical difference between the two new design variants. The light states for the existing and two new variants are summarised below:
3. For the current design, the brakes automatically release and the brake light goes off once 60 seconds has elapsed from the start of the TPWS or AWS brake demand following acknowledgement. This means that in some cases the brakes and indicators can revert to off before the driver has correctly diagnosed that TPWS has activated. For both new variants, this has been changed such that any light remains illuminated until the brakes are positively released by the driver.
4. For the current design there was no positive acknowledgement required by the driver for AWS following an AWS brake demand to release the brakes. For TPWS, acknowledgement was required to release the brakes, but this was by pressing the AWS button which is routinely used for AWS operation and therefore does not have a strong association with TPWS brake demands. To rectify these issues, new methods for driver actions to positively release the brakes were developed for both TPWS and AWS. These were the same for variants 2 and 3.

Table 12.1 Red brake light illumination state for the design variants following acknowledgement

	Variant 1 (Current)	Variant 2	Variant 3
TPWS TS	Steady	Fast flash	Flash
TPWS OSS	Steady	Slow flash	Steady
AWS	Steady	Steady	None

It should be noted that the designs for the new cab interfaces were not developed from human factors first principles but were constrained by existing engineering and cab interfaces. This was because the design basis specified allowed only relatively minor changes to the in-cab equipment and underlying engineering systems. This was due to cost-benefit arguments required by industry and also was appropriate in the context of longer term cab design developments such as the European Rail Traffic Management System.

Method

Key variables tested in the experiment

This section describes the two main independent variables tested in the experiment. The following section describes the output measures which were used to assess performance on these factors (dependent variables). The main aim of the experiment was to compare driver performance using the three different interface variants described above. Therefore, the first key independent variable tested in the experiment was the different variants. This was treated as a between-subject variable (i.e. each participant was tested on only one variant), to minimise the impact of experience effects which could arise if treated as a within-subject variable (where each participant would be tested on two or more variants in sequence).

The second key independent variable was the type of brake demand experienced by the driver. This was treated as a within-subject variable, with each participant experiencing all the different brake demand types. In addition to confusion between AWS and TPWS activations, AWS and TPWS activations can be misdiagnosed as other train safety system activations and therefore, examples of these were included within the experiments. Due to practical constraints within the simulator these other faults were limited to MCB circuit breaker trips only. The four types of brake demand experienced by drivers were therefore: AWS, TPWS Train Stop (TPWS TS), TPWS Overspeed Sensor (TPWS OSS) and MCB circuit breaker trips. 24 brake demands, six brake demands of each type, were experienced by drivers during the experiment. The number of brake demands experienced was limited largely by the time available to run the simulations. It should be noted that due to initial errors with the design of the simulation route, data for some brake demands were not collected. The sequence of brake demands was counterbalanced to avoid order effects using the Latin Squares approach.

In addition to these two independent variables considered in this chapter, it should be noted that in order to test driver performance with the interfaces fully it was also necessary to manipulate the visual cues external to the cab which the drivers experienced. The three types of external cue used were:

- Supporting – where the brake demand experienced was supported by external cues. For example, a TPWS TS occurred and this was supported by the driver experiencing passing a red signal.
- Misleading – where the external cues were contrary to the brake demand experienced. For example, a TPWS OSS occurred but the driver experienced passing a red signal.
- None – No external cues were presented to the driver and they had to rely on the in-cab interface.

Dependent variables

The dependent variables reported in this chapter are:

- Driver brake demand diagnosis accuracy. Drivers verbally reported information after experiencing brake demands as to what had caused the brake demand and what cues they had used to make that diagnosis.
- Driver workload ratings. Driver subjective ratings of workload were collected using the 9 point Integrated Workload Scale (Pickup et al., 2004).

Experimental protocol

Participants were randomly assigned to one of the three interface variants. Participants received training specific to their assigned interface variant. This was conducted by a training specialist from Southern Trains. The drivers were instructed to drive the route normally, but made aware that they would experience a number of brake demands during the route which would have to be diagnosed. They were also informed that the information presented within the cab would always be correct. This was required to ensure that drivers did not diagnose brake demands on the basis of cues external to the cab rather than based on the cab interface.

Familiarisation with the train cab simulator was then undertaken. The four simulators at Southern Trains at Selhurst were used, which are identical configurable simulators of a Class 170. The external cab information is presented as video animated graphics on the cab window and scenes were drawn from a library of routes for trains travelling between Victoria Station and Brighton via Quarry Line tunnel.

Subjects drove along a pre-designed route which was identical for each participant. After each of the 24 brake demands experienced along the route, the driver was instructed to make a call on the cab radio to an experimental administrator. The administrator asked the driver a predefined set of questions regarding the brake demand and collected the driver IWS rating. The IWS rating was based on asking the driver the following question: "Look at the prompt card provided and select the descriptor which most represents your opinion of your workload when diagnosing the brake demand". The length of the trial for each driver depended on his/her driving style and length of discussion following each brake demand. The experimental trial, including a ten minute brake in the middle, generally took approximately 2 hours.

In addition to the verbal debriefs after each brake demand, video recordings of the drivers' actions, verbal recordings of communication with the experimenter and detailed time-stamped information on driver interaction with train cab equipment were collected. Demographic data and various subjective usability evaluations were collected from the drivers both before and after the experiment.

Subjects

All the subjects were qualified UK drivers. They were selected on an opportunity basis, depending on the availability of staff from train operating companies. 27 drivers took part in the study, 25 male and 2 female. The drivers' average age was 43 with a minimum of 23 and maximum of 59. Average experience as a driver was 12 years with a minimum of 6 months and a maximum of 36 years. The drivers were drawn from four passenger train operating companies covering both suburban and intercity type services.

Results

Brake demand diagnosis likelihood

A balanced ANOVA was undertaken and significant differences were identified for differences between interface types ($F = 36.26$, $p < 0.00$) and brake demand types ($F = 3.91$, $p = 0.01$). A significant interaction was also identified between interface type and brake demand type ($F = 10.12$, $p < 0.00$). A significant interaction means that the different interfaces had different performance profiles across the different brake demand types.

The average number of correct diagnoses was improved from 65% for the current system to 91% – 92% for variants 2 and 3 respectively. Tukey's HSD post-hoc test identified a significant difference between variant 1 when compared to both variants 2 and 3, but no significant difference between variants 2 and 3. Both variants 2 and 3 achieved an improvement (approximately 25%) compared to variant 1 (the current system) in the correct diagnosis of brake demands.

For the differences between the different types of brake demand it is the significant interaction which is of most importance. This is presented graphically in Figure 12.1.

The key points to note from these data are:

- The current variant one interface is particularly weak for correct identification of TPWS TS and TPWS OSS brake activations.
- Both variants 2 and 3 provide much improved reliability for identification of TPWS TS and OSS brake applications, which was a key objective for the study. This is particularly the case for TPWS TS, which is the most critical indication type.
- When comparing variants 2 and 3, they have contrasting weaknesses. Variant 2 leads to lower levels of reliability for AWS brake demand diagnosis compared with variant 3, while variant 3 leads to reduced reliability of diagnosis of the MCB faults which were not related to AWS or TPWS, compared with variant 2 (and also variant 1).

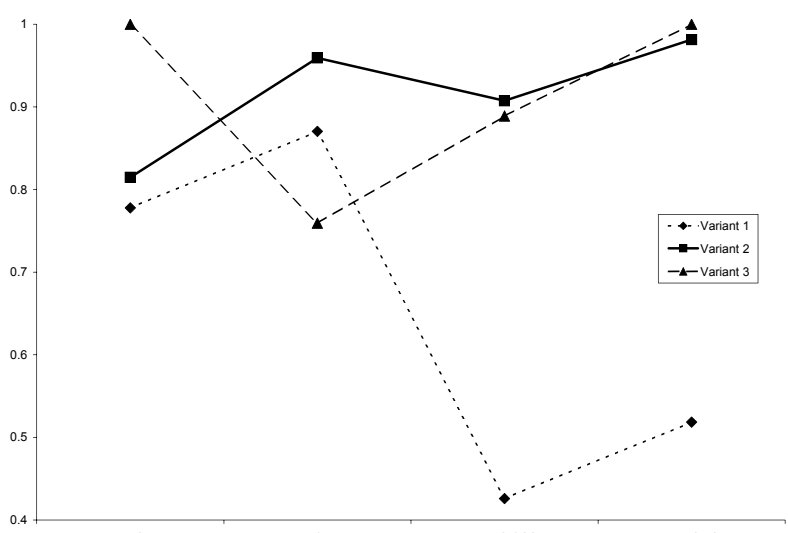


Figure 12.1 Interaction plot for interface type for the different brake demands

In order to explore this issue further, confusion matrices were developed which describe the types of misdiagnosis which occurred.

Table 12.2 Confusion Matrix for variant 2 incorrect diagnoses

	Diagnosis						
Brake demand Experienced	AWS	OSS	TS	TPWS	Not sure/other	All Demands Experienced	Total
OSS			5			54	5
TS		1				47	1
AWS		1	5	1	3	54	10
MCB	2					46	2
Total	2	2	10	1	3	201	18

Table 12.3 Confusion Matrix for variant 2 incorrect diagnoses

	Diagnosis						
Brake demand Experienced	AWS	OSS	TS	TPWS	Not sure/other	All Demands Experienced	Total
OSS			6			54	6
TS						54	
AWS						54	
MCB	13					54	13
Total	13		6			216	19

In broad terms, it is anticipated that the misdiagnoses associated with variants 2 and 3 relate to the following design weaknesses:

- Variant 2. The use of a single light in three states to differentiate between AWS, TPWS OSS and TPWS TS. The three states are steady, slow flashing and fast flashing. While human factors guidance was followed in relation to flash rates, it is clearly not optimal to have four meanings (if light not illuminated is counted as a fourth) to a single visual indicator. This provides an initial explanation as to why 16 of the 18 observed errors occur between OSS, TS and AWS.
- Variant 3. The lack of a unique indicator light for AWS indications. The dominant misdiagnosis type is incorrectly identifying MCB brake demands (as representative of 'other') as AWS brake demands. It is believed the lack of a unique illumination related to AWS led to a tendency for drivers to treat any brake demand without illumination as AWS, thus classifying the MCB errors as AWS brake demands. The other six errors may still relate to the use of single light to mean two things by being steady or flashing.

Further analysis of both sets of misdiagnosis data have been undertaken by review of the video data and voice recordings of responses to post-brake demand debrief questions. These data have highlighted that for both variants 2 and 3 participants have generally correctly perceived the inputs but then made incorrect decisions based on these inputs. These therefore appear to be rule based errors which would be expected to relate to the complexity of the diagnostic process or the training and retention of the diagnostic rules by participants. While equivalent training durations and types were provided for all three variants, these were time-limited. Further research has been proposed which looks at learnability and retention of information related to the two new variants. It should be noted that other learning effects were not identified in the data. This finding does not diminish the fact that despite this potential training issue the different interfaces led to different types of misdiagnosis, as seen in the confusion matrices.

Workload ratings

Analysis of the workload data revealed statistically significant differences between the variants (F=3.8, p=0.027). The data are illustrated graphically in Figure 12.2, with low values representing low workload on the Y axis and high values representing high workload. Further post-hoc analysis using Tukey's pairwise comparisons identified a statistically significant difference between variant 1 when compared with variants 2 and 3, but no differences between variants 2 and 3. These data highlight issues with variant 1 which lead to higher workload. It is interesting to note that the scores provided by participants were generally low across all three variants. This is expected to be because the diagnosis required was not viewed as complex by many of the participants.

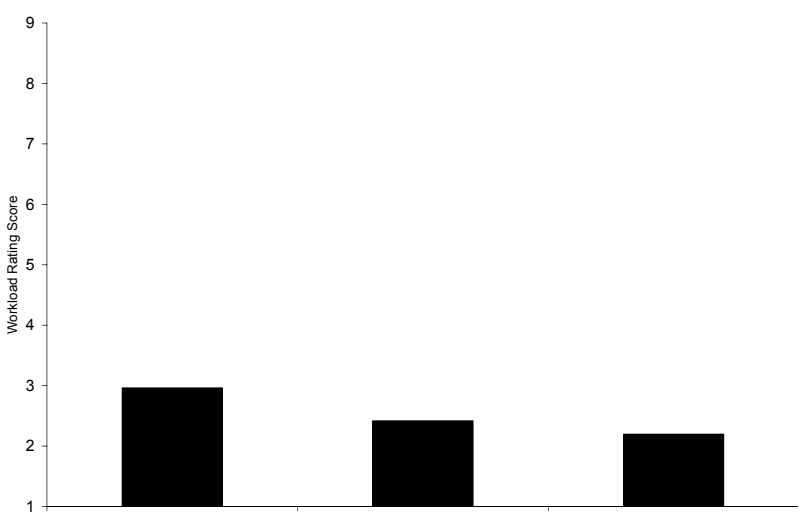


Figure 12.2 Average workload rating by variant

Conclusions

This research has demonstrated the important contribution which simulator experiments can make to the development of safer train cab designs. Both driver diagnosis performance and workload scores highlight the benefits which can be gained from the two new proposed design variants. However, the study was not able to differentiate between the two new variants in pure performance terms as significant differences were not identified between the dependent variables for the two new variants. However, by analysing the errors which occurred, it has been possible to describe the specific types of error associated with the two new interfaces and their likely causes, for example, the use of a single light in a number of states to provide different meanings.

While providing improved performance, errors did still occur for variants 2 and 3 and this is expected to be in part related to limitations in the new designs. These limitations arise out of the requirement to provide a cost-effective solution within current engineering cab design constraints. Clearly, a further factor is that human diagnosis in these types of situation can never be perfect. Therefore a decision has to be made as to whether these predicted improvements are adequate. The data generated in the simulator are an important source of data which can inform risk-based decision making as to whether the new designs provide sufficient improvement.

This study data also provided important data on human reliability in a problem-solving task. These data will be added to the CORE-DATA database of human error probabilities, along with other railway data and data collected in other domains (Gibson et al., 2006a, 2006b). These types of data are central to developing our understanding of the contribution of human error to risk and for underpinning risk assessment processes.

The experimental study described in this chapter forms part of an ongoing research initiative which aims to provide a practical solution to current deficiencies in the cab interface. To this end further work has been ongoing. An additional simulator trial has been undertaken. This examined the two new interfaces in the context of more complex multiple brake demand scenarios. The analysis has supported the results of this first simulator trial in terms of driver brake demand diagnosis, but has still not enabled differentiation between the two variants. In addition a detailed human error analysis has been undertaken to consider the two new design variants (Gibson, 2005). These data are currently under review by industry and will be used as the basis for possible changes to cab design across UK train operating companies.

Acknowledgements

This work was funded by the Rail Safety and Standards Board and the input from the RSSB Human Factors team is gratefully acknowledged. The report authors would like to thank the drivers from London Lines, Midland Mainline, South West Trains and Southern who took part in the study. Also, the study would not have been possible without the support of Southern, through the use of their simulators.

References

Cullen, W.D. (2000) *The Ladbroke Grove Rail Inquiry Volume 1*. H.M. Stationery Office, London.

Gibson, W.H. and Halliday, M.W. (2005) *Human Error Analysis To Compare Two Proposed Driver Interfaces for TPWS* RSSB report submitted to the rail industry for review.

Gibson, W.H., Hickling, B. and Kirwan, B. (2006a) *Collection of Communication Error Probability Data for the ASAS Trial Eurocontrol Report* (in press).

Gibson, W.H. Megaw, E.D., Young, M and Lowe, E. (2006b) 'A Taxonomy of Human Communication Errors and Application to Railway Track Maintenance' in *Cognition, Technology and Work* (in press).

Halliday, M.W. and Gibson, W.H. (2005) *TPWS Development and Evaluation: Human Factors Trial Report,* Thales report submitted to the rail industry for review.

Hamilton, W.I., Lowe, C., and Blanchard, H. (2004) 'Cognitive Task Analysis and Performance Modelling for Early Human Systems Integration' in Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society, New Orleans, Louisiana, September 20-24, 2004. Human Factors and Ergonomics Society, Santa Monica, California, pp. 2406-2410.

Kappler, W.D. (1993) 'Views on the Role of Simulation in Driver Training' in Proceedings of the XII European Annual Conference on Human Decision Making and Manual Control. Kassel, Germany, June 22-24, 1993.

Keon, L. (1991) 'Roles of Simulators in the Railway Industry: Human Factors - Trainees-Trainers' in Proceedings of the Second Seminar in Transportation Ergonomics. October 24-25, 1990, Edited by R.M. Heron. Transport Canada, Montreal, Publication No. TP 8855, pp. 161-165.

Malsch, N., Sheridan, T. and Multer, J. (2004) *Impact of Data Link Technology on Railroad Dispatching Operations.* DOT/FRA/ORD-04/11. Washington, D.C: U.S. Department of Transportation, Federal Railroad Administration.

Nichols, S. and Crawford, S. (2002) 'Simulation of Industrial Decision Making: Lessons Learnt from Field Studies in Process Control and Railway Signalling Environments' in de Waard, D., Brookhuis, K.A. .Moraal, J. and Toffetti, A. (Eds) *Human Factors in Transportation, Communication, Health, and the Workplace*. Shaker Publishing, Maastricht, The Netherlands, pp 455 - 470.

Pickup el al (2004) *Train driver IWS Tool Guidance Note T147*, RSSB: London.

Rail Safety and Standards Board (2005) *Good Practice Guide on Simulation as a Tool for Training and Assessment* RS/501 Issue 1. Rail Safety and Standards Board, London. (http://www.rgsonline.co.uk/docushare/dsweb/Get/ Rail-7458/501.pdf)

Roach, G.D. et al. (2001) 'Comparing the Effects of Fatigue and Alcohol Consumption on Locomotive Engineers' Performance in a Rail Simulator' in *Journal of Human Ergology* 30 (1-2), pp. 125-130.

Young, M.S. (2003) 'Development of a Railway Safety Research Simulator' in McCabe, P.T. (ed) *Contemporary Ergonomics* 2003. Taylor & Francis, London, pp. 378 - 384.

This page intentionally left blank

PART V
Train Cab and Interfaces: Simulation and Design

This page intentionally left blank

Chapter 13

An Ergonomics Methodology for Retrofitting Equipment in Train Cabs

Katie Buchanan, Gary Davis and Richard Roels

Introduction

Retrofitting a new piece of equipment in the diversity of UK train cabs is a significant engineering challenge. When the equipment is a new radio system that must be operable by the full range of train drivers, from their various driving postures, the challenge is one of ergonomics as much as engineering.

There are some recent examples that have highlighted the importance of considering ergonomics during cab fitment activities. For example, the positioning of Train Protection Warning Systems (TPWS) has been suggested to be a contributory factor in "reset and go" errors. Some TPWS devices are located outside drivers' acceptable vision, with 50% of drivers reporting that the positioning of the TPWS human machine interface (HMI) could be improved (QinetiQ, 2004). Similarly, drivers have reported difficulties in reading the display of the Cab Secure Radio (CSR) in the Class 450 cab because the radio is not angled towards the driver.

The GSM-R (Global System Mobile Communications-Railway) voice radio system is to be introduced by duty holders for communications between the driver and the signaller. To facilitate the fitment of a new GSM-R radio system on the UK rail network, the Rail Safety and Standard Board (RSSB) commissioned Davis Associates Ltd (DA) to provide human factors guidance for the cab fitment process. We present the main findings of this work, which can be divided into 3 main sections:

1. Human factors principles – the definition of human factors principles for GSM-R HMI location within the cab.
2. Applying the human factors principles – the identification of ergonomics analysis methods through which these principles could be applied.
3. The survey process – the development of a preferred survey process through which optimum fitment locations could be identified.

Human factors principles

A set of human factors principles were defined for the location of the GSM-R HMI, which comprises a control display unit (CDU), handset and speaker. These principles followed human factors best practice and were tested and refined through a series of cab fitment surveys. These principles covered matters relating to:

- the use of the radio in relation to all other driver tasks;
- drivers' acceptable vision and reach zones (for the full range of driver anthropometry and working postures);

- access and egress to the driver's seated position;
- access and sightlines to existing cab displays, controls and other essential items, such as paper-based information;
- exterior sightline requirements;
- light emissions from the CDU (with respect to reflected glare);
- orientation of the CDU in respect to the driver's line of sight; and
- access for maintenance purposes.

Central to these was the requirement for fitment locations to be within acceptable reach and vision for a range of driver anthropometry (5th percentile female to 95th percentile male) when adopting realistic driving postures (e.g. some smaller drivers have to sit on the front edge of the seat in order to reach the primary controls).

Applying the human factors principles

There is little point developing principles unless they can be applied and assessed in an accurate and consistent manner. As a result, once the principles were defined, the research focused on developing and evaluating practical analysis methods that would allow them to be used to assess alternative fitment locations in any type of train cab.

The driver packages in cars, trucks, aircraft and trains should in theory be designed around the driver and a key datum should be the seat reference point (SRP). The SRP lies the on the centre line of the driver's seat at the junction of two planes: one representing the backrest, and the other representing the seat pan/base. If the driving package has been well designed, surveying the range of seat adjustment alone would be enough to determine the range of driver eye and shoulder and therefore the optimum reach and vision zones for any new equipment.

Unfortunately, it is very evident from our surveys on this and previous projects that the wide diversity of train cabs in service in the UK are not all well designed. In many cases, the driver is forced to sit forward of the SRP in order to reach poorly positioned primary driving controls. The implication for the GSM-R fitment project was that the precise range of driver eye and shoulder points could not be predicted from the SRP alone. The true working postures had to be determined before the human factors principles for GSM-R HMI fitment could be applied. During a pilot study, a number of different analysis methods were tested on a series of different cabs, in order to identify those that would be suitable for use during the nationwide fitment of GSM-R HMI.

Analysis methods

At the core of the human factors principles was the requirement for the HMI fitment locations to accommodate for the range in driver body sizes, from the 5th percentile female (approx height of 4ft 1") to the 95th percentile male (approx height of 6ft 2"). Some form of standardised measure for these body sizes was needed and the research identified three possible approaches (see Figure 13.1), all of which involve the use of manikins in some form:

i) The use of an anthropometric rig to be installed in the driving seat;
ii) The use of user participants of defined anthropometric proportions; and
iii) The use of computer-aided drawings and software human manikins.

These manikins should be constructed, or in the case of user participants recruited, based on UK anthropometric data. In particular the following key dimensions had to be considered: popliteal

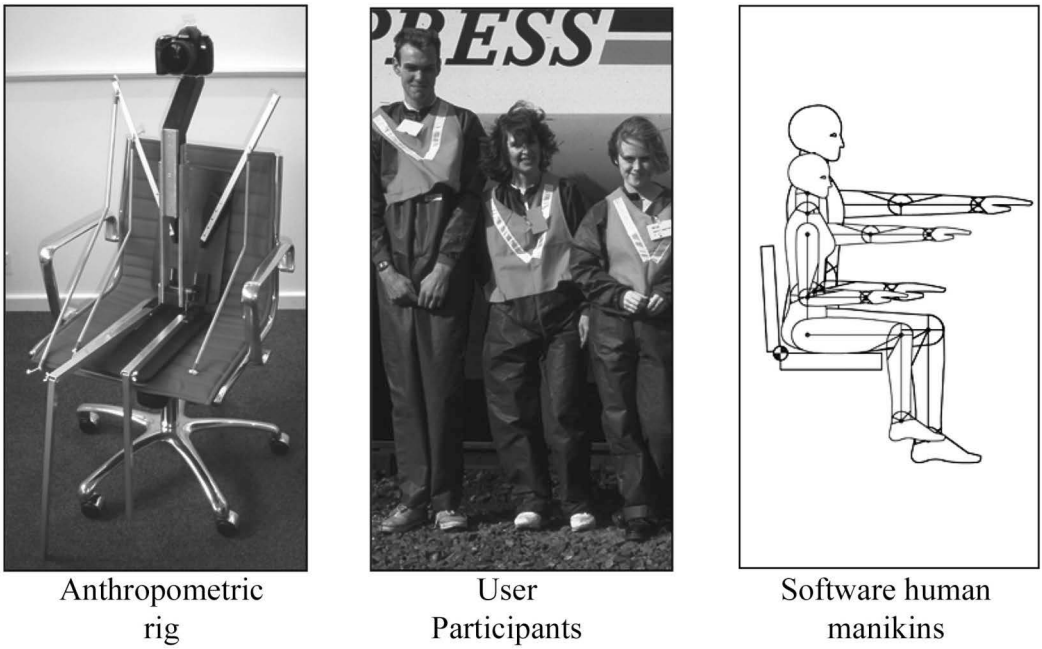


Figure 13.1 The three possible survey methods

height, buttock-popliteal length, sitting eye height , shoulder pivot locations (height, width and fore-aft position), arm length (shoulder-to-finger-tip with articulated elbow joint).

The manikins are therefore required to test the driver reach and vision principles from drivers' realistic working postures. In order to achieve this, the manikins should be placed in a posture which allows them to reach the:

- Driver Safety Device (DSD) foot pedal with their foot (where applicable)
- Primary driving controls with their hands.

This may require adjusting the driver's seat height accordingly. However the research highlighted that in many cases this will result in the 5th percentile female manikin being perched on the front edge of the seat, without back support. In cabs that allow drivers to operate the train whilst standing, assessment of vision lines should also be taken from the perspective of the 95th percentile standing male.

Anthropometric rig

The development of an anthropometric rig enabled the on-site assessment of driver reach and vision principles (see Figure 13.2). The anthropometric rig incorporated the key anthropometric dimensions for the 5th percentile female and the 95th percentile male, including shoulder pivot points, arms, hips and legs. Additionally, the rig was equipped with a mount of adjustable height that was used for the vision zone projector and wide angled photography.

To assess reach to surfaces or objects within the cab, the rig was placed on the driver's seat. The seat height was adjusted so that the 5th percentile female's "heel" touched the DSD pedal and/or

The use of the rig to assess reach

The use of rig to project the vision zone

Figure 13.2 An anthropometric rig in use

footrest and her hands could reach the primary driving controls. The 5th percentile female's left or right arm was then extended as far as possible towards the potential fitment location, to see if it was within reach. The process was repeated for the 95th percentile male with re-adjustment of seat height.

The rig was also used to assess acceptable vision angles in two different ways, by using direct projection of the vision zone and by using wide-angle photography. The first approach involved projecting light though masked domes into the darkened cab, this immediately highlighted which areas of the cab were within the acceptable vision zone, as stated in the human factors principles. These domes were mounted at the sitting eye heights of the 5th percentile female and the 95th percentile male.

The rig was also used as a platform from which to capture wide-angle photographic images from the drivers' eye points, for a second method of assessing the driver's acceptable vision zone. The final stage of this method took place after the surveys, when a calibrated grid with the vision zone was superimposed onto each image and to see which possible fitment locations fell within the acceptable vision angles.

Table 13.1 illustrates that whilst this process is relatively quick and provides on-the-spot guidance it has a major disadvantage in that it cannot accurately represent true driving postures in poorly designed cabs.

An alternative technique to the anthropometric rig is to recruit participants of specific body measurements to act as human manikins. Figure 13.3 shows a photograph of a group of participants that were used for a similar sizing trial on a previous project. Recruited participants can adopt realistic working postures in the cab from which they can then be asked to reach to potential fitment locations, to see if they are within their reach. In order to assess vision angles accurately a sighting device would be required so as to allow participants to assess the acceptable vision lines to the fitment locations.

Figure 13.3 A group of users

Table 13.1 The advantages and disadvantages of using an anthropometric rig

The use of the anthropometric rig	Advantages	Disadvantages
1) Identify potential fitment locations. 2) Place the rig on the seat and adjusts the rig to allow the primary controls to be reached (DSD and traction controls). 3) Use the arms of the rig to test reach principles. 4) Use the rig to assess vision by either: a) Projecting the vision zone from the range of driver eye points to test vision lines to the locations. b) Taking photographs from the range of driver's eye points. Review the photographs by super-imposing the vision zone on to the photograph.	- A relatively quick process once installed in the cab. - Provides on-the-spot guidance	- Relies on the appropriate placement of the rig on the seat and cannot reliably replicate the true working postures of drivers. - Requires training to ensure consistent deployment. - Difficult to document the results for checking. - Design and build costs for a more robust anthropometric rig would be considerable. - Transportation of the rig. - Availability of the rig(s).

Whilst this process is relatively quick and would once again provide immediate guidance, logistical problems may arise in recruiting people of specific body dimensions and ensuring their availability once recruited (Table 13.2). In addition, a vision line sighting device would be required.

Table 13.2 The advantages and disadvantages of using user participants

The use of user participants	Advantages	Disadvantages
1) Identify potential fitment locations. **2)** Two people of specific body dimensions are asked to sit in working postures in the cab. **3)** The participants are then used to assess reach to each of the locations. **4)** The participants are then used to assess vision to each of the locations using a sighting device.	- A relatively quick process once on site. - On-the-spot guidance. - Accurate assessment of reach and vision from drivers realistic working postures.	- Requires at least two (preferably more) trained participants of the required anthropometric proportion. - Requires sighting device to test vision lines. - Site access logistics and health and safety risks - Availability of the same two participants. - Cost of participants' time.

Software human manikins

The final technique involves the use of computer-generated software manikins. This would obviously require a CAD drawing/model of the cab in which the software manikins can be inserted. Once these manikins have been placed in realistic working postures the reach and vision principles can be accurately and consistently assessed (see Figure 13.4).

This assessment can either be undertaken in a 2D or 3D CAD environment. Whilst 3D assessment would require the use of a more expensive/sophisticated software package it does have a number of additional benefits over 2D assessment, as it:

- requires less technical drawing abilities; and
- allows for a more reliable simulation of postures.

Unlike the other types of assessment techniques the use of graphical manikins does not provide on-site guidance. However there are a number of advantages particularly in terms of consistency and accuracy in applying the HF principles across different cab classes (Table 13.3).

The preferred human factors analysis method

Based on the above analysis, it was concluded that the most effective and reliable method was the use of the CAD drawing/model and software human manikins.

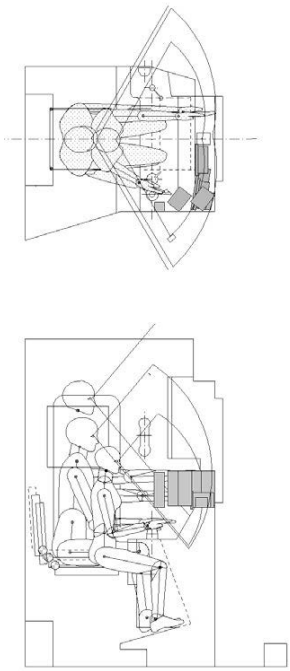


Figure 13.4 The use of computer-aided drawings and software human manikins

Table 13.3 The advantages and disadvantages of using software manikins

The use of software Manikins	Advantages	Disadvantages
1) Potential fitment locations are identified on a CAD drawing/model of the cab. 2) A 5th percentile female and 95th percentile male software manikins are placed in working postures in the cab in the CAD drawing. 3) Assessment of driver reach and vision principles are made using the manikins, with other principles also assessed.	- Accurate and consistent assessment of all feasible options. - Allows assessment based on drivers' realistic working postures. - Provides traceable and defensible results.	- Requires either access to existing CAD drawings or measurements of the cab to be taken. - Requires suitable CAD system with human manikins. - Requires trained CAD operator with abilities to correctly manipulate human manikins.

1. It provides an analysis method through which the human factors principles can be accurately and consistently applied.
2. In contrast to the use of an anthropometric rig and user participants it would not require on-site ergonomics expertise.
3. There is less room for error and individual interpretation.
4. CAD models can be altered immediately which is particularly beneficial when taking into account slight variations within cab classes.
5. The CAD drawings provides a permanent record of the survey.
6. Provides more accessible evidence for quality checking.
7. It provides traceable results and documentary evidence for assurance review by approval bodies.

The survey process

After the preferred analysis method was identified, the next challenge was to develop a robust and practical survey process that could be applied to train cabs all over the UK. This survey process was developed and refined by conducting pilot surveys of three further classes of cab.

The three part survey process had to be applied on a class-by-class basis:

1. An engineering survey to identify and rate, in terms of ease of installation and maintenance, all the technically feasible fitment locations for the CDU, handset and loudspeaker.
2. A subsequent ergonomics assessment to identify which of the fitment locations comply with the human factors principles.
3. Rationalisation of the conclusions drawn from the engineering and ergonomics assessments, in order to identify the fitment locations which best meet both sets of requirements.

The survey process has 8 key steps:

Step 1: *Source any existing CAD data*
To identify and get access to any existing CAD models of the cab.

Step 2: *Measure and record cab dimensions*
To measure and record critical facets of the driver's workspace if CAD drawing not available. Detailed measurements of the cab to be taken in the cab with reference to defined datums. The measurements must be sufficient to create a 2D/3D CAD model of the cab.

Step 3: *Consult a Train Operating Company (TOC) representative(s)*
To identify any operational issues with the potential fitment locations. A TOC representative(s) should be interviewed regarding the feasibility of the possible fitment locations. Particular areas of interest are:

- whether any equipment in the cab is/or will become obsolete;
- whether there are any plans for the installation of any new equipment;
- guidance on the acceptability of moving particular items of equipment;
- any operational issues with potential fitment locations.

Step 4: *Identify HMI fitment locations and rate engineering difficulty.*
To identify all CDU, handset and loudspeaker fitment locations and rate the difficulty of installation and maintenance access. All possible fitment areas should be rated against an engineering fitment scale, photographed and key dimensions taken (Figure 13.5). Particular areas that should be considered during location identification and rating were:

Figure 13.5 An example of photographs taken of possible fitment locations for a mock-up of GSM-R CDU and handset

- mounting arrangement e.g. space constraints;
- interference with existing equipment, either physical or electrical;
- cable runs from CDU to radio tray or to the handset cradle; and
- ease of access for maintenance purposes.

Step 5: *Produce a 2D/3D CAD model*
To produce a 2D drawing (or 3D CAD model) of the cab, using the cab measurements taken in step 2 or use the existing cab drawing.

Step 6: *Insert potential GSM-R equipment locations into the 2D/3D CAD model*
To locate all the possible fitment locations within the CAD model.

Step 7: *Use the CAD model to assess the fitment locations against human factors principles*
To identify the optimum fitment location for CDU, handset and loudspeaker, through testing the fitment locations identified in the engineering survey against the human factors principles. Software human manikins should be placed in realistic working postures in the cabs, to accurately assess reach and vision principles.

Step 8: *Identify optimum location*
To identify an optimum location for the CDU, handset and loudspeaker within the cab that satisfies both ergonomics and engineering requirements. The survey and analysis team involved in the assessment should convene to discuss the CDU, handset and loudspeaker fitment options identified. Each location should be discussed in respect to:

- the results of the ergonomics assessment; and
- the location's engineering fitment score.
-

The survey and analysis team should then agree upon the optimum fitment location for the CDU, handset and loudspeaker. The outcome of this meeting should be documented to provide a traceable record.

Conclusion

This project has defined the human factors principles that should be applied when fitting the GSM-R HMI into existing train cabs. The principles take account of reach and vision zones from the range of drivers' body sizes, whilst adopting realistic working postures. The project has also

identified and evaluated three ergonomics analysis methods for the application of the human factors principles and has highlighted the preferred method for the GSM-R project. The preferred method is a CAD-based assessment using software human manikins. This has the great advantages of accuracy, consistency and traceability. The alternative methods - the use of an ergonomics rig in each train cab, and the use of participants of predetermined anthropometric proportions - cannot match the positive attributes of the preferred method.

The project has also suggested an 8-step survey process for the identification of the optimum fitment solution - a solution that is the optimum compromise between the often conflicting engineering and ergonomics requirements. For contractual reasons, the process could only be a suggestion - the chosen contractors would be at liberty to propose alternative methods as long as they could demonstrate that the human factors principles were being correctly and consistently applied.

Reference

QinetiQ, 2004, Driver Error Data Collection Project: Final Report to RSSB.

Chapter 14

The Waterfall Train Accident: Implications and Lessons Learnt

Andrew S. McIntosh and Graham Edkins

Introduction

The immediate causes of the Waterfall train accident and contributing factors are presented in another chapter from the authors in this book. Here we focus on the lessons learnt and implications of the accident. The Special Commission of Inquiry and the previous inquiry into the Glenbrook accident, also in New South Wales, made numerous recommendations regarding all levels of rail operations (McInerney, 2001; 2005). The accident revealed that the combination of medical standards, training, design, train and driver safety systems, and safety management systems were inadequate to control the risks inherent in rail operation.

In general, one important lesson learnt was that ergonomics and human factors can provide valuable input into the design and operation of rail systems. However, in this case, and other examples that were reviewed during the Inquiry, input was restricted largely to occupational health and was often of insufficient depth to provide critical input. In addition, rail standards specifications that referred to anthropometric data, eg. the footpedal device shall not be able to be accidentally circumvented by the dead weight of the operating limb or limbs within the male 97.5 percentile band, failed to provide support for designers that gave examples of relevant population data and their application. This failure was evident in other areas, as described here.

Train and driver safety systems

The accident revealed that the Tangara deadman system did not reduce to an acceptable level the risk of a train remaining under power if the driver became incapacitated. Furthermore, it found that warnings regarding this flaw had not been addressed in the Tangara's operational history.

A review of design documents, reports and SRA correspondence revealed that the flaws had been identified as potential problems early in the Tangara's operational life. The design process and selection of pedal activation and deactivation forces was not documented, neither were the forces clearly specified. Early in the Tangara's operational history there was a major redesign and refit of the driver's console. A 1991 Worksafe project provided valuable ergonomic input that lead to the modification of the driver's console, controls and displays (Stevenson et al, 2000). The function of the deadman's foot pedal as a central component of the Tangara's safety systems appears to have been largely ignored in these projects. The focus of the pedal redesign was improving comfort and reducing the risk of lower limb musculoskeletal disorders, even though it was explicitly stated that safety had the priority over other considerations. Therefore, the primary importance of safety was recognised on the one hand, but design efforts appeared to focus on ergonomic considerations in the area of occupational health. Human factors data and design documents are required to assist in the translation of intent to practice.

Victorian survey of driver safety systems

The findings of the inquiry also triggered other rail operators and safety regulators to review the operation of their systems. A survey of Victorian suburban train drivers was conducted by the authors to support a review of driver and vigilance operating systems conducted by the Victoria Department of Infrastructure. Some aims of the survey were to examine whether driver population and circumvention issues were an isolated NSW artefact or more widespread. The survey consisted of a structured questionnaire completed by the drivers and measurement of stature, body mass and pedal forces using a simulated seat-pedal arrangement. Participation was voluntary and anonymous. One hundred and twelve drivers participated and basic population and anthropometric data are presented in Table 14.1.

Table 14.1 The foot pedal force was measured with participants wearing their normal shoes

	N	Mean ± SD
Experience as Suburban Driver (yrs)	110	14 ± 8
Age (yrs)	111	45 ± 8
Height (cm)	111	175 ± 8
Weight (kg)	111	87 ± 13
Vertical Foot-Pedal Force (N)	107	79 ± 19

The trains used by the drivers had both a foot and hand pilot valve (deadman system). The mean and median percent of time in which the foot pilot valve was used was 53% and 60% compared to 47% and 40% for the hand pilot valve. The foot pilot valve control and the plunger type hand pilot valve control were considered the best of the four controls, with the twist style being the worst.

Drivers were asked whether they experienced discomfort while driving and to describe the nature of it, eg. location of symptoms and duration. Seventy-five percent of the drivers did experience discomfort or pain while driving and using the pilot valve, with the upper limb most affected (56%) Seventy percent of drivers found that changing posture relieved the pain or discomfort, however there was a small residual (14%) who complained of pain/discomfort lasting greater than 24 hours after the end of a shift. This highlighted the need to consider comfort and occupational health issues in system design, although without sacrificing the intended utility of the safety device, and demonstrates how uncoordinated design activities commence.

Surveys of driving postures and feet positions on the pedal showed a large range of positions for each individual and across the population. Fourteen percent of the drivers reported wedging the heel of their shoe into the pedal mechanism, thus jamming it in an activated position. While most drivers reported that the deadman system had triggered braking at some point while driving, normally due to bumpy track, they also reported that it rarely deactivated if they were drowsy or inattentive. With regards to a question asking drivers to recall what would happen if they rested one or two feet on the pedal with no muscular effort, 52% reported that the pedal would remain activated with one foot and 26% reported it would remain activated with two feet.

Figure 14.1 is a comparison of operational forces for Victoria suburban trains and the measured foot-pedal forces from the driver population sampled. The figure does not consider specific foot

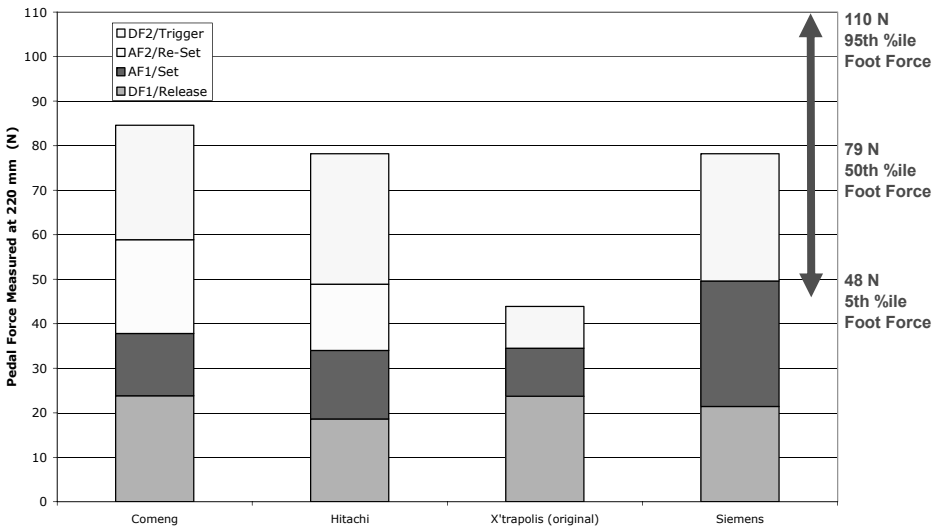


Figure 14.1 Comparison of measured range of foot forces (with shoes) to activation and deactivation forces on Melbourne trains. In comparison the deactivation forces on the Tangara were 50 N (df1) and 199 N (df2).

placements or the lever action of the pedals. It confirms that without further adjustment for foot position on the pedal or the use of two feet, about 50% of the drivers could, through the weight of the leg, simply hold the pedal in an activated position. The results are consistent with the driver surveys.

The Victorian survey confirmed that the deadman system problems identified in the Waterfall accident were present in the Victorian driver population and train fleet. The possibility of drivers inadvertently circumventing the deadman function by wedging their foot/feet in the pedal, the failure of the pedal to deactivate when the driver was relaxed, and the biomechanical analyses pointed to similar latent conditions. In comparison to the Victorian trains, the deactivation force range on the Tangara was much greater. As a result, the deadman device installed on suburban trains in Victoria is being progressively enhanced with a task linked vigilance system.

Risk management

One of the most disturbing pieces of evidence in the Inquiry was that the risk of an incapacitated driver passing a signal at stop was assessed to be "Barely Tolerable" for Outer Suburban Tangaras (OST) as it had been for Double Deck Intercity Cars (DDIC). While task linked vigilance systems had been progressively installed on the DDICs from November 1999 to treat those risks, OSTs were not modified. Furthermore, 'ergonomic' assessments of the drivers' consoles and controls focussed on occupational health criteria, not on safety, and there appeared to have been no formal human factors evaluation of the prototype task linked vigilance system. Human factors considerations are important in risk management.

On-time running

When the broader streams of evidence and information from the Inquiry, from official sources and in the press were combined, it became apparent that the culture of 'on time' running was of great significance at the time of the Waterfall accident. One of the recommendations from the Glenbrook inquiry was that data loggers be fitted to all trains. The G7 unit had a data logger, but it had not been activated at the time of the accident. Therefore, relevant data that would have accelerated the investigation were not available. One of the obstacles to activating the data loggers was the perception that drivers may have exceeded speed boards to meet the timetable. Therefore there was resistance to their introduction. Further, the culture of speeding may have influenced the G7's guard not to 'pull the tail' (emergency brake), as it may have not appeared to be abnormal that the train was travelling too fast, until too late.

The program to install and activate data loggers appears to correspond with the deterioration in peak period on-time running to about 80% in December 2003 reaching only 48% in February 2004 (ITSRR, 2004). On-time running was defined as being < 4 minutes late. This problem was worse in off-peak services and known well by Sydney rail commuters. The Hon. Michael Costa was reported in Hansard to have said that "we have said consistently that data loggers have reduced the running speeds of trains because they operate almost as fixed speed cameras" (Costa, 2004). Implicit is an acknowledgement that running in excess of the speed boards was commonplace and, probably, necessary to meet the timetable. An analysis conducted by the Independent Transport Safety and Reliability Regulator (ITSRR), the Rail Safety Regulator, suggested that there may have been a 2% slow down in transit time due to behavioural change and that other factors may have been equally or more important, eg. dwell time, crew management and incident management (ITSRR, 2004).

Medical standards

Given the health risk factors outlined in the previous chapter for the driver, the medical examination regime for train crew was examined to determine whether it was sufficiently rigorous enough to identify such candidates and mange them accordingly.

The medical procedures used by the rail organisation were compared to various industry standards including those of other Australian rail operators, civil aviation and current practices for drivers of commercial vehicles. This review identified that the medical procedures in place at the time of the Waterfall accident, were less than satisfactory based on industry benchmarks.

The medical guidelines were very broad, providing only limited detail in determining medical fitness for work. For example, there was no recommended course of action for staff found to be borderline risk in regards to heart disease, diabetes, sleep apnoea or are drug and alcohol dependent. Clinical interventions were largely left to the discretion of the medical examiner. There was no formal process for the selection and authorisation of medical practitioners to undertake occupational medical assessments, and because there was no induction process the system relied on "word of mouth" for handovers of rail staff. There was also no central or in-house expertise within the rail organisation for medical practitioners to consult with regarding special cases. In particular, the medical guidelines provided little direction with regard to the identification of psychological health issues or detailed requirements about follow-up procedures for staff involved in critical stress incidents and who might experience symptoms of post traumatic stress disorder (PTSD).

With regard to obesity, the standards referred only to applicants for employment who must have a Body Mass Index (BMI) less than 30. There were no guidelines for existing employees and no detail about what interventions were required for current employees who have a BMI greater than 30. This

investigation also identified a pattern of internal correspondence between management, in which significant deficiencies with the standards were acknowledged but not acted upon over a seven-year period.

It was also observed that there had been a 10 kg increase in the mean body mass of the SRA drivers in the decade preceding the Waterfall accident. The combination of increased driver age, potential increased risk of incapacitation related to increased body mass, and the function of the deadman system are latent conditions that cross the medical, engineering and human factors domains.

In summary, the medical standards used by the rail organisation were characterised by:

- a lack of guidelines on the qualifications, selection and monitoring of authorised medical practitioners,
- limited guidelines in relation to determining medical fitness for work for a number of medical conditions including cardiovascular disease, obesity, sleep disorders, psychological health and drug and alcohol dependency,
- a failure to address contemporary health issues (eg. HIV) and to account for more recent advances in occupational medicine and clinical practice.
- less frequent periodical assessments required under operator standards compared to the vast majority of other standards.

Fatigue management

The rail organisations approach to fatigue management was found to be deficient. They did not apply FAID™ to the periodic or daily "local" level rosters to determine if work rosters were fatigue inducing. In practice master level rosters change significantly at the local level due to shift swaps between train crews, or depot managers attempting to fill the gaps from personnel absenteeism fluctuations. In other words, often the master level rosters are a poor representation of the actual hours worked of a given train guard or driver.

Furthermore, rostering practices within the organisation allowed drivers to be rostered to return to work on an early morning shift following a period of significant leave (as had the driver of the Waterfall train), and thus did not take into account issues associated with circadian disruption or other fatigue elements at the micro level. No guidelines, training or education material has been provided to train crew to help them manage a return to work under these circumstances.

Training

While the organisation had conducted some joint safety training for guards and drivers, there was no evidence that this training encouraged these two parties to operate effectively as a team. Other industries such as aviation, had implemented crew resource management (CRM) training to encourage better coordination and communication (Helmreich and Foushee, 1993).

Training within the organisation had typically been directed at drivers and guards, operating as individuals in isolation, rather than as a team. Had the organisation conducted a training needs analysis, or utilised a risk based approach to curriculum development, hazards such as authority gradients between drivers and guards, the need for critical decision making skills, the importance of standard phraseology for communication, and staff reticence to report safety issues, may have been identified and addressed.

The role of guards in relation to safety was ill defined and overly focused on customer service requirements. Procedures and specific training in guards being able to recognise a speeding train, initiating the appropriate action and their role in emergency situations were deficient.

Safety culture

Although the organisation had a documented safety management system, it was the application of this system in practice that was not clearly evident. A proactive and systems wide approach to safety hazard identification and a prevailing attitude amongst many senior managers that the Waterfall accident was an "unfortunate accident", were indicators that there was not a dominate safety "consciousness" throughout the organisation. In contrast, the main imperative driving the organisational culture appeared to be on-time running. The imbalance between safety and on-time running became a factor that influenced day to day decision making and ultimately led to many of the causal factors identified in the Waterfall investigation.

Conclusions

A holistic human factors-risk management-engineering approach is required to assess issues related to safety and reliability in rail. The accident and subsequent inquiry showed that without consideration for all of these approaches, latent conditions will not be identified, or will be created.

There are many human factors lessons arising from this investigation for transport operators, policy makers, investigators and regulators alike. In particular:

- Designers and engineers require relevant human factors data, as do those involved in system procurement.
- Systems and components cannot be modified to satisfy occupational health objectives without consideration for safety.
- The management of fatigue involves more than monitoring employee work rosters but should include comprehensive education, guidelines, policies, quality assurance processes and targeted training at supervisory level staff to ensure that workers are fit for duty.
- Medical examination standards for safety workers need to be periodically reviewed to ensure that they are predictive and preventive of potentially incapacitating medical conditions and updated in line with contemporary advancements in medical science.
- That initial and recurrent training programs practically reinforce the concept of teamwork to ensure that individuals do not operate independently but are supportive and cooperative in the conduct of their duties. In addition, specific procedures and a supporting periodic training regimen for teams in regard to critical decision-making skills such as that adopted within the aviation and healthcare industries.
- The importance of ongoing independent reviews of an organisation's safety culture to identify values, norms and attitudes that may have a potentially adverse influence on safety.

Acknowledgements

The authors were engaged as consultants to the Special Commission of Inquiry into the Waterfall Rail Accident. The chapter reflects the author's opinions not those of the Inquiry. The authors acknowledge the exemplary work undertaken by many involved in the Inquiry and the specific

support provided by Justice McInerney, his legal team, and Mr. Bob Lauby. The Victorian study was funded by the Department of Transport.

References

Costa, The Hon. Michael (2004). NSW Legislative Council Hansard: Millennium train driver training, Sep 15 2004.

Helmreich, R.L. and Foushee, H.C. (1993). *Why crew resource management? Empirical and theoretical basis of human factors training in aviation*. In E. Weiner, B. Kanki, and R. Helmreich (eds), Cockpit Resource Management (pp 3-45). San Diego, CA: Academic Press.

ITSRR (2004). *Review of on-time running of Cityrail services*, Published by Independent Transport Safety and Reliability Regulator (ITSRR), Service Reliability Division, Issue Date: June 2004, Reference Number: SR 01/04.

McInerney, P.A. (2001). Special Commission of Inquiry into the Glenbrook Rail Accident Final Report, April 2001. The Honourable Peter Aloysius McInerney.

McInerney, P.A., 2005. Special Commission of Inquiry into the Waterfall Rail Accident Final Report Volumes 1 and 2, January 2005. The Honourable Peter Aloysius McInerney QC

Stevenson, M.G., Coleman, N., Long, A.F., and Williamson, A.M. Assessment, re-design and evaluation of changes to the driver's cab in a suburban electric train, Applied Ergonomics, 2000; 31: 499-506.

This page intentionally left blank

Chapter 15

Determining User Requirements for a Human Factors Research Train Driver Simulator

Thomas K. Yates, Sarah C. Sharples, Ged Morrisroe and Theresa Clarke

Introduction

Simulation is a widely used technology that takes many forms. It can range from abstract computer modelling of, for example, fluids structures which give purely mathematical answers or computer based visual representations of prototypes to virtual simulations of events which can be either interactive or passively observed.

Simulation is a tool which is widely used in most domains and with particular success in engineering for uses such as prototyping, testing and evaluating (Chan, 2003; Deviprasad and Kesavadas, 2003). The reason why simulation is so popular is because there are huge cost and time benefits to be gained. It can be much quicker to build a computerised representation of a prototype to evaluate and remove any potential problems; this will be considerably cheaper than actually building the prototype. There are also some safety benefits to be obtained since using a virtual environment to recreate safety critical situations is considerably safer, because it possible to practise tasks in an environment where there are no repercussions for failure. Simulation also makes it possible to recreate some situations that would be too expensive or impractical to recreate in the real world such as emergency scenario training.

This research, which is funded by the Engineering and Physical Sciences Research Council (EPSRC) and Network Rail, concentrates on train driver simulators. Network Rail is responsible for maintaining and renewing the rail infrastructure, but ultimately it will be the train drivers who will be affected by these changes. Therefore, Network Rail requires some tool to determine the impact of the infrastructure changes on the driver, essentially to assess the driveability and safety of routes post-infrastructure change. There are currently systems in place to ensure that no changes are made which will affect the driver adversely and during the repositioning of signals. For example, there is a Signal Sighting Committee which assesses the impact of the intended changes and signal sighting simulation software is currently used by Network Rail to assist with this process.

The successful use of this software suggests that a simulator could be useful in situations where Network Rail intend to introduce a change in their infrastructure and would like to assess the human factors issues related to these changes prior to any work being undertaken. These changes could be altering parts of the rail infrastructure such as signalling, track layout, signage location, adjusting signage design or assessing the impact of changes in signalling technologies on drivers.

The uses for the simulator could range even further beyond factors that are related directly to the infrastructure to factors that affect the rail industry as a whole. This would increase the usefulness of the simulator and it could be used for various areas of research, such as assessing the impact of cab design on drivers, cab layout, effects of shift pattern on drivers etc.

Simulators for human factors research

Within the rail industry, there are a number of areas of human factors being researched and there is a need for some of this research to be undertaken in a simulator. There are various reasons for this such as determining the impact of new technologies which have no safety case. The rail industry is a safety critical one and there are special control arrangements for access to the train driver and cab, also it is disruptive to the train operating companies and Network Rail to give researchers access to their drivers, signallers and controllers. In addition, the very nature of some of the studies is to interrupt the participants mid-task and as such it would be inappropriate to allow this type of research to be undertaken in-service.

There is no rail cab simulator in the United Kingdom which is specifically for research purposes; it is well known to human factors researchers working in the rail industry that having such a simulator would be very useful, in some projects essential, when undertaking human factors research. As part of a Rail Standards and Safety Board initiative there was some progress made towards developing and commissioning such a research specific simulator (Young, 2003). The initiative is currently not being pursued, so this research hopes to complement and extend this early work. There are various train driver training simulators around the United Kingdom, however. Young (2003) stated that these simulators are not suitable for research for various differing reasons: the lack of time available in the simulators; the lack of scenario flexibility; and the limited data analysis possible. As highlighted by Young (2003) the requirements for a research simulator are different to those for a training simulator.

There are still some unresolved issues regarding simulation, particularly with regard to how realistic the simulators have to be in order to produce valid results from experiments. So before any designing or building of simulators commences, there are many factors to consider as there are no guidelines for the fidelity requirements for train driver simulators for human factors research.

Problems with rail simulators

Developing and building a train simulator for research purposes is a significant task because there are many aspects which must be appreciated and understood before a design is decided upon. The researchers working on this project undertook a brainstorming exercise and some of the unanswered questions that emerged are:

- what type of train cab should we develop?
- should it be a cab specific or general design?
- should it be a Electric Multiple Units (EMU) or a Class 43?
- should it be a replica or a real cab?
- what should be visually represented?
- should it be an exact replication of the train driver's world or should it be a partial representation?
- what type of studies is it going to be used for?
- what research equipment such as data collection equipment needs to be included?
- what type of data need to be collected and in what format?
- what feedback should be included in the design: sound, motion etc?
- what weather conditions are required?

Network Rail have expressed a desire for a simulator to be portable, this will increase the scope of the simulator and ensure it can be available to more research projects. A simulator would be most

appropriate if it was specifically designed and built with a particular research topic or study in mind as the simulator would most likely fill all the requirements of that study, but taking this approach to simulator design would be problematic. The main reason why this would be the case is that it would be too expensive, except possibly in large long-term projects such as major resignalling works. In addition, when the project was completed the expensive simulator would be surplus to requirements, because the simulator would have been designed to answer the particular problem that needed answering in that study. This may have been, for example, the driveability of a new signal layout; it is therefore unlikely that the simulator would be useful for other projects without requiring expensive re-designing or re-programming.

It would therefore seem appropriate to build a simulator which would be suitable for all types of studies for which it could be used or, perhaps more realistically, designed in such a way that it would be flexible and therefore could be easily changed depending upon the type of study that was being undertaken. This approach in itself creates other issues. In order to determine what types of studies could be undertaken on the simulator we looked at completed human factors and simulation studies (eg Sanders, 1991; Brook-Carter et al., 2003; Luke et al., 2005; Newman et al., 2005; Smith et al., 2005).

The problem that arises from using past simulator requirements is that any design decisions would be based on the requirements of these past studies, which will not actually use the simulator. Therefore, we would have already introduced a dated element to the design. Consequently, it was decided to get the wider views and opinions of the human factors community to resolve some of these issues and help develop a User Requirements Specification.

User requirements specification

The end product from this work is a User Requirements Specification of a generic train driver simulator for human factors research. This specification was obtained by asking the opinions, using a predetermined and formal procedure, of a cross section of the human factors researchers currently working in the rail industry. In addition to this, it was necessary to review the studies that have been undertaken on a train driver simulator and also to ascertain if there are any completed studies that would have been suitable for a simulator if one had been available. If so, what would have been the ideal specification of this simulator in order for it to have best achieved the aims and objectives of that research.

This user-centred approach to the information gathering and consultation process will create a shared vision for the simulator design, ensuring that the eventual simulator produced from the specification has had end-to-end input from the stakeholders. This will result in a simulator that is useful, effective and beneficial to human factors researchers.

Due to the varied nature of work currently being undertaken in the United Kingdom on rail human factors, it is unlikely that any simulator developed would meet the needs of all the human factors researchers who would potentially want to use it. This is due to various issues such as, for example, the data logging requirements: in a human factors study it may be important to accurately record the drivers' interactions and response times to events, whereas with a training exercise these interactions may not be as important. However, there will be some common requirements for the majority of the projects and it is these that we are endeavouring to determine. This will hopefully develop a specification which will satisfy the majority of the needs in the most cost effective way. This is the main reason why this user centred approach was chosen, because if all the users' needs are taken into account and if possible catered for, then it is likely that the simulator will be more useful as an asset.

The end user for this simulator will be the human factors researchers, also called in this work the Subject Matter Experts, who will be undertaking the studies; these people are anyone who is running a study with the aim of obtaining, observing, measuring and analysing the reaction of the train driver to some controllable variables. These Subject Matter Experts could conceivably be from any of the agencies or consultancies who undertake rail related projects and be from any background ranging from the technical through applied to fundamental research.

Method

The data elicitation process that will provide the information for the User Requirements Specification involved: designing the interview to ensure the questions were suitable and appropriate, drawing up a confidentiality agreement, determining who we would like to talk to and then contacting these participants to see if they consented to being interviewed, undertaking the interview and then analysing the data and drawing up a specification.

Participants

Correspondence in the form of e-mail was sent to Subject Matter Experts who were known to the research group as a whole; these Subject Matter Experts were selected because it was thought that they were working in the field of human factors and that they also had an interest in the rail industry. This interest was to be primarily from a research point of view, meaning they must currently be undertaking a rail related project or be commissioned to undertake any rail related projects in the near future.

Confidentiality

During the interview, but prior to the questions stage, the interviewee was presented with a confidentiality agreement, which stated that all the data collected and their responses would remain strictly confidential. The interviewees' answers would not be identifiable by name or any other personal details and results data will be presented so as not to identify any of the interviewees. If they agreed with the measures implemented to ensure their confidentiality, then they were asked their permission for the interview to be tape recorded.

Sequence for the questions

Following the procedure as recommended by Robson (2002), the sequence of the interview was as set out below. The prompts and probes were pre-determined for consistency; were the same during each interview and the use of these were accurately recorded.

- Introduction: Interviewer introduced himself, explained purpose of the interview, assured confidentiality and asked for permission to audio record the interview and to make notes.
- Questions: They were asked a series of questions on demographics, projects and simulators.
- Any other comments: The interviewee was given an opportunity to ask any questions or make any other comments.
- Closure: The interviewees were thanked for their time.

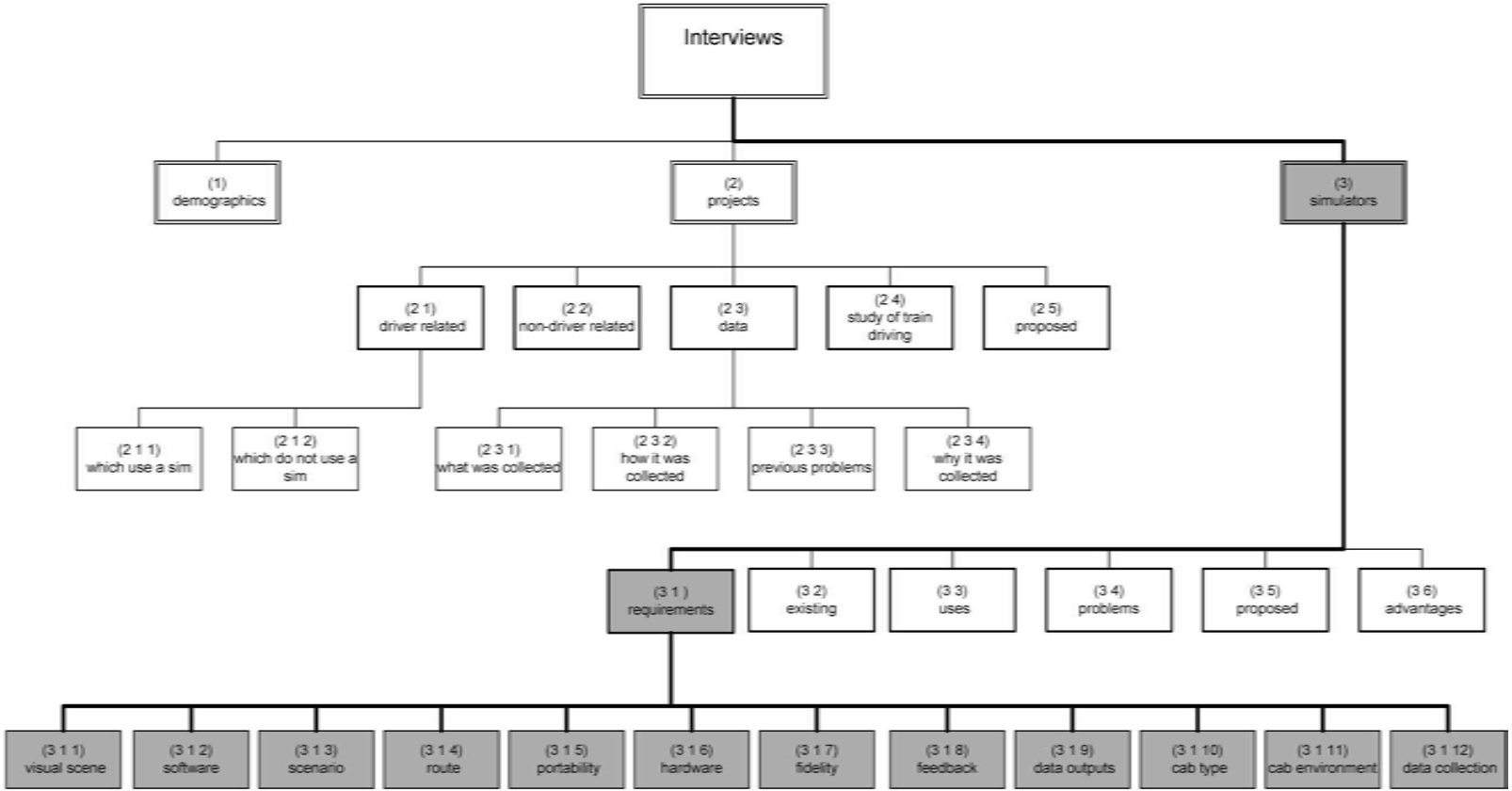


Figure 15.1 Three level node structure of raw interview data. Grey path shows data analysed in this chapter

Interviews

The interview was designed as semi-structured, the wording slightly changed and particular explanations given or omitted depending upon of what is appropriate for the interviewee. The three main topics of questions (see Figure 15.1) were:

- Demographics, details about the interviewees' experience in the field of human factors, their job position and their length of service.
- Projects, types of projects worked on by the interviewees and the aims and objectives of those studies. They were then asked about: the data that were collected from those studies, the methods used to collect these data, did they feel they would have benefited if any of these methods had been automated and finally were they prevented from collecting any data due to lack of technology, time and resources.

The interviewees were asked about the types of projects they had worked on, to try and obtain a cross section of the type of work that has been completed. It was hoped that this would give an understanding of the variety of work currently being undertaken in this field.

The interviewees were asked about the data collection methodology they used in their studies. This would give an understanding of the types of raw data collected in a study rather than the fine-tuned, tidied up and analysed data that are reported in the chapters. This information gives us an idea of the types of data and the collection techniques that would be used, as this would affect the design of the simulator.

- Simulator, the interviewees were then asked about simulators, first if any of the projects they have been involved with were undertaken in part or in full on a simulator. They were asked to provide more details about how they had used the simulator and details about the age, cab control type, simulated track details and any other details that they could recall. They were then asked: to provide details about the type of study they had undertaken on this simulator; their opinion of the simulator that they used and if it had met their expectations; and how it could have been improved in order to help them with their research.

The interviewees were asked what aspects of the visual scene in the simulator were important to their study. It was hoped that this question would reveal the aspects of the visual scene that are actually used by a driver and also what aspects of the visual scene were important to which study.

The interviewees were then asked if they had been provided with unlimited resources what would be their ideal specification for a simulator. This hypothetical question was to try and determine what features would it be useful to have in a simulator. This was found to not be a particularly useful question on its own, because it was yielding answers like 'everything' which were not particularly constructive comments. So it was complemented by the question: if there were limited resources, what would be your ideal specification of a simulator. This follow up question was asked to determine which of the items, stated in answer to the previous question, were really important to them.

Finally the interviewees were asked about any data collection techniques which would be useful if built into the simulator. This was to determine what type of data collection tools and equipment it would be beneficial to include in the simulator design.

Results and analysis

The raw interview data were transferred into NVivo and then coded into sections by undertaking various passes on the data. This analysis process was based on the Theme Based Content Analysis technique (Neale and Nichols, 2001). The data were split into three top-level headings: demographics, projects and simulators. Below we report only on the simulator requirements data. The summary of the requirements in Table 15.1 and Table 15.2, and in the text below shows the number of times a requirement was stated by the interviewees in parentheses, which gives an indication of the relative importance of each simulator requirement.

Visual scene – Some (3) of the interviewees were of the opinion that everything that the driver could see out of the window was important and should be included in the virtual environment. It was noted that train drivers, to assist with their gauging of speed and also to determine their location, use non-rail related objects such as local landmarks (3). They also use the commonly seen rail infrastructure such as: the rail (4), AWS magnets (1), signs (3), signals (4), mile markers (1), overhead lines (1), gantries (1) and stations (2) so these were important to include in the model.

Feedback – Most (10) of the interviewees thought that the sound of the train and of the external environment were critical for a good simulation. Some (8) thought that haptics or any form of physical movement of the cab would be desirable, but were aware of the costs involved so did not think it was essential in order to undertake most studies. One interviewee did note that for certain studies, for example adhesion studies, having some train cab motion would be necessary to ensure the results of the study were transferable.

Data collection – Many (4) of the interviewees stated that they would want a video recording and play back facility as part of the data collection equipment. They would require a recording of the driver and what the driver can see out of the cab, preferably with eye point of gaze data (4) and other data overlaid on this recording (2). The data the simulator needs to collect, all of which need to be time stamped (4), are continuous variables such as: braking profiles (2), speed profiles (2), signal aspect (1), line speed limit (1) and distance travelled (1). It was stated (2) it would be useful if the simulator's data collection facility has the flexibility to be configured to collect different results for different experiments.

Portability – Some interviewees (3) thought that having the simulator as a portable machine would be useful in order to take the simulator to depots, this would ensure that there were no shortages of drivers to participate in studies. However, some (3) thought the opposite and thought that having a simulator that is portable would compromise on the quality and it would be better to have it fixed in one central location.

Data outputs – The interviewees thought that the format of the outputted data was important. Most interviewees stated that it was critical to ensure that any data collected were compatible with standard software packages, such as Microsoft Office (4) and standard analysis packages such as SPSS (1).

Cab type – Some interviewees (2) thought that we should develop a Class 43 whereas some (2) thought that we should make the cab a replica of an EMU. Most people (5) said that having a modular type design would be the most useful, as it would allow them to construct a cab of their own choice depending upon the requirements of their study.

Table 15.1 Summary of simulator requirements (numbers in parentheses indicate frequency of comments)

Visual scene	Software	Scenario	Route	Portability	Hardware	Fidelity
signals (4) track (4) everything (3) signs (3) local landmarks (3) station (2) gantries (1) seasonal changes (1) AWS magnets (1) overhead line (1) mile markers (1) stanchions (1)	pause function (2) software robustness (1) simplicity (1)	weather (3) flexibility (3) blockages on track (2) degraded scenarios (2) low visibility (2) time changes (1) loss of train brakes (1) emergency scenarios (1) normal scenarios (1) communications (1) timetable (1) level crossings (1) passenger actions (1)	flexibility (4) geospecific routes (3) geotypical routes (3)	portable (3) fixed London (2) fixed central England (1)	full immersion (1) wide visual field (1) wrap around screen (1) video recorders (1) communications (1)	high fidelity required for training (1) low fidelity for route learning (1) low fidelity for rules based driver training (1)

Table 15.2 Summary of simulator requirements (cont'd)

Feedback	Data outputs	Cab type	Cab environment	Data collection	Data collection (cont'd)
sound important (10) haptic feedback preferable but not essential (8) acceleration characteristics of train important (3) haptics important (2) haptics compromise possible (1) haptics necessary for adhesion studies (1) braking characteristics of train important (1)	compatible with excel (2) compatibility with standard software packages (1) compatible with word (1) compatible with SPSS (1) speed and braking profiles (1) video recordings of route (1) video recordings of driver (1)	modular or flexible cab design (5) Class 43 (2) EMU (2) flexible controls layout (2) generic cab layout (2) accurate replica controls (1) adjustable wind screen (1)	full cab interior (2) correct smell (1) correct temperature (1) flexibility (1) windscreen adjustable (1)	reaction times (5) video recording (4) driver's view recorded (4) eye point of gaze data (4) key events time stamped (4) recording of any driver interaction with system (3) flexibility to include secondary tasks (3) flexibility (2) sampling rate control (2) speed profile (2) braking profile (2) play back facility with other data incorporated (2)	print out data (2) downloadable data (1) automatic data collection (1) play back facility with eye point of gaze data (1) brake pressure (1) continuous variables recorded (1) distance travelled (1) position in space (1) driver recorded (1) SPAD errors (1) recorded variables adjustable (1) line speed limit recorded (1) overspeeding errors (1) signal aspect recording (1) record of errors (1)

Routes – Some interviewees (3) thought that the routes available in the simulator should be actual representations of real routes (geospecific) and others (3) thought that having a route which was typical of certain areas, such as countryside or cityscape, would be sufficient for their work (geotypical). Most (4) of the interviewees thought that flexibility to design their own routes easily was important in order to meet the demands of an experiment.

Scenarios – The scenarios that the interviewees wanted the simulator to run were examples of emergency, degraded and normal running conditions. These included things such as: adverse weather conditions (3), blockages on track (2), low visibility (2), passenger actions such as people jumping onto the track (1) and also passenger actions on the station (1). Having the ability to easily programme new events into the scenarios was also considered important (3).

Hardware – The hardware that some of the interviewees wanted to be included in the simulator were a full wrap around screen to ensure good immersion of the driver (1), communications devices such as the radio (1) and also video cameras for some of the data collection techniques (1).

Software – The interviewees (2) wanted the simulator software to be robust and easy to use. They (2) also mentioned that having a pause function would be useful so they could interrupt a scenario when necessary.

Cab environment – The interviewees (2) thought that it was important to have a full cab interior and also include anything into the cab environment that made the cab more realistic to the driver. These were factors such as the appropriate temperature (1) and smell (1) for the scenario that was running.

Fidelity – Some of the interviewees (2) stated that high fidelity was necessary for training and also that low fidelity would be adequate for some, if not most, human factors studies.

Conclusions

This research took a user-centred approach to determining the ideal specification of a train driver simulator for human factors research. We sought the views and opinions of a cross section of the human factors community working in the rail industry and now have a comprehensive list of simulator requirements.

The next stage to this work is to cross-reference these requirements with the interviewees' demographics and also with the projects that the subjects had undertaken and are planning to undertake. This could highlight any patterns in the data and help us to fine tune the simulator requirements.

These data will be used to build a battery of simulations of varying fidelity using the *requirements specification* as the basis of any design decisions. These simulations will be used for experiments to help identify the appropriate level of simulator fidelity for a particular study or piece of human factors research. This will assist with the development of a quantifiable measurement technique of fidelity and allow us to propose a validation process to evaluate a train driver simulator's fitness for purpose.

Acknowledgements

The authors would like to thank all the interviewees for their assistance with this project and for being so generous with their time during the interview process.

References

Brook-Carter, N., Parkes, A. M., Mills, A. (2003). A methodology to investigate train driver visual strategies. Proceedings of the First European Conference on Rail Human Factors, York, 13-15 October 2003.

Brook-Carter, N., Parkes, A. M., Mills, A. (2004). Piloting a Method to Investigate the Thought Processes behind Train Driver Visual Strategies. *Contemporary Ergonomics*. pp 271-275.

Chan, D. S. K. (2003). Simulation modelling in virtual manufacturing analysis for integrated product and process design. *Assembly Automation*. 23. 1. pp 69–74

Deviprasad, T., Kesavadas, T. (2003). Virtual prototyping of assembly components using process modelling. *Journal of Manufacturing Systems*, 22, 1, pp 16.

Luke, T., Brook-Carter, N., Smith, R., Parkes, A. M., Mills, A., Grimes, E. (2005). The effect of signal infrastructure on train driver visual behaviour on signal approaches. *Contemporary Ergonomics*. pp 578-582.

Neale, H., Nichols, S. (2001) Theme based content analysis: a flexible method for virtual environment evaluation. *Int. J. Human Computer Studies*. 55. pp 167-189.

Newman, M., Hamilton, W. I., Maxwell, I. (2005). Validation of minimum reading time requirements for railway signal reading. *Contemporary Ergonomics*. pp 598-602.

Robson, C. (2002). *Real World Research*. Second Edition. Blackwell Publishers Inc. Massachusetts. pp 277.

Sanders, A. F. (1991). Simulation as a tool in the measurement of human performance. Ergonomics. 34. 8. pp 995-1025.

Smith, R. K., Luke, T., Brook-Carter, N., Parkes, A. M., Mills, A., Grimes, E. (2005). Investigating train drivers' perceptions of their visual environment. *Contemporary Ergonomics*. pp 553-557.

Young, M. S. (2003). Development of a railway safety research simulator. *Contemporary Ergonomics*. pp 379-384.

This page intentionally left blank

Chapter 16

A Facility for Testing ERTMS/ETCS Conformity and Human Factors

Klaus P. Jaschke, Katrin Hartwig, Michael Meyer zu Hörste, Karsten Lemmer

Introduction

The national railway systems across Europe are due to be advanced to a harmonised and interoperable system. This complex and long-lasting process will be fulfilled in a lot of single steps, and mainly:

1. The technical interoperability will ensure that a train coming from one country going to another country can use the infrastructure with one single system.
2. The operational interoperability is based on the technical interoperability and will be created by a harmonised set of operational rules. The new or modified rules must be analysed with regards to changed behaviour of staff.
3. The safety of the operation must be analysed in the next step. This includes extended analysis of the human influences.

Currently, the first implementations of the European Train Control System (ETCS) of the European Rail Traffic Management System (ERTMS) are being launched. In Spain, for instance, on the line from Madrid to Barcelona, on-board units of different suppliers must be interoperable with trackside subsystems from another supplier. To avoid additional effort in time and cost, the interoperability of these components will be tested in a lab before the testing and operation on the track. Different tests have to be performed: tests of technical safety, reliability and availability on the one hand, tests of operational safety and conformity with the specifications on the other hand.

Such a simulation environment which can cover both cases is the core of a railway laboratory, called RailSiTe®.

Concept and architecture of the RailSiTe®

Concept

The RailSiTe® is a multipurpose platform for scientific as well as for industrial experiments. The RailSiTe® provides the complete chain from the signaller to the train driver (Meyer zu Hörste et al, 2003). The global approach is shown in Figure 16.1.

Due to the modular and flexible architecture the RailSiTe® can be used for several different purposes. Some of them are:

- The test of conformity with the requirements.
- The evaluation of operational safety of a complete railway operation control system.

- Performance evaluation of a modified or new train control system.
- Analysis of an automation concept for trains.
- Hardware-in-the-loop simulation including real parts of a system.
- Experiments related to human factors of automation approaches.
- Experiments related to stress and workload of the operators.
- Experiments related to safety-critical errors of the operators.

The RailSiTe® is designed to fulfil the requirements related to reproducibility as well as to flexibility. This is realised by a modular architecture.

Error models and simulation of erroneous system behaviour

The basic laboratory testing approach is [Illgen *et al.* 2000]:

- Functional behaviour is tested against the functional requirements without introducing disturbances. This kind of experiments can be used to test the conformity with the functional or technical system requirements.
- Behaviour is tested under disturbed, environmental influences against the appropriate safety and availability requirements. This kind of experiment can be used to test operational safety or availability. The stability and performance of operation concepts can be evaluated as well.

The relevant sources and typical modes of errors or disturbances have been studied carefully. Some important sources have already been identified. All modules or subsystems which perform an information transmission in a continuous or discrete way will be equipped with technological error models. The basic classification contains five types of errors:

- Delay of a message
- Change of the sequence of messages
- Loss of a message
- Change of message content
- Sending of a wrong message

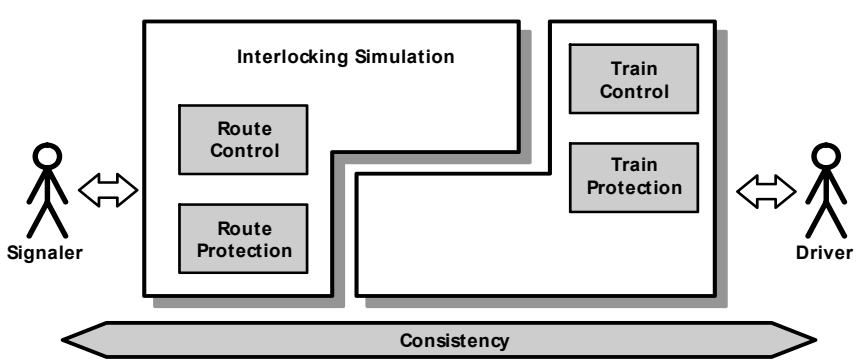


Figure 16.1 RailSiTe® concept

This behaviour can be represented by fixed values calculated as a stochastic mean value or a stochastic value defined by a given distribution function. They can be independent of the environment or defined by location or time. These different behaviours will be represented by specific "Transmission Error Modules" which can be chosen related to the purpose of the experiment. Their implementation will start with the definition of generic modules, which fulfil the basic error modes. Two basic generic error-modules are radio communication errors and transponder errors. For the use of testing they will be parameterised, e. g., to represent the behaviour of GSM-R or the Euro-Balise. In the following steps of implementation, these modules will be extended and refined.

Test scenarios

Experiments and tests can have different goals: performance of a train operation system or the conformity of a train control system with the specification. Even if the goals of the experiments or tests are different, the basis for all of them is basically the railway operation. The railway operation is represented by operational sequences, which can be concatenated to simulation train runs. For these train runs a description of the track is necessary. For the performance evaluation of a real line this description can be taken from the real line. For the evaluation of a system the use of track descriptions related to real lines is not efficient, because a lot of special operations are rare on real lines and there is a small number of variations of operations on one line. The use of a generic track for time-efficient evaluation of a train control system saves time and effort. The generic track is based on the same operational scenarios as the simulation of the operation on a real line.

The scenarios which are to be tested can be operational scenarios, which test a specific operation situation, and interoperability scenarios used to check the conformity of one implementation as well as the correct interaction of two or more implementations. The basic test sequences can be derived from a formal specification [Meyer zu Hörste *et al.* 2000], they can be written completely manually or they can be created in a mixed automatic and manual process.

Formalization of tests

The reproducibility and the automation of tests lead to the need for a concept to formalise and to automate the test sequences. For instance the test cases of ERTMS/ETCS can be shown as a graph with approximately 1200 nodes. The manual generation of test sequences is not efficient and can lead to errors. The proof of completeness needs a lot of effort as well. To avoid this problem a specific algorithm based on graph theory has been used to improve the generation of test sequences.

Modularity

The RailSiTe® environment is defined in a very modular way to ensure both the research aspect and the interoperability validation purpose. The infrastructure and the physical train are replaced by simulation modules. They generate the inputs needed to stimulate the train control system. All other inputs like route map, timetable and all kinds of drivers or operator inputs are sent by interface generators. All of them are acting in real time to ensure real time behaviour of the simulation environment.

The environment has not been created to fulfil certification purposes, but it can be used to analyse technical interoperability. An overview of the basic architecture of the first steps of implementation is given in Figure 16.2.

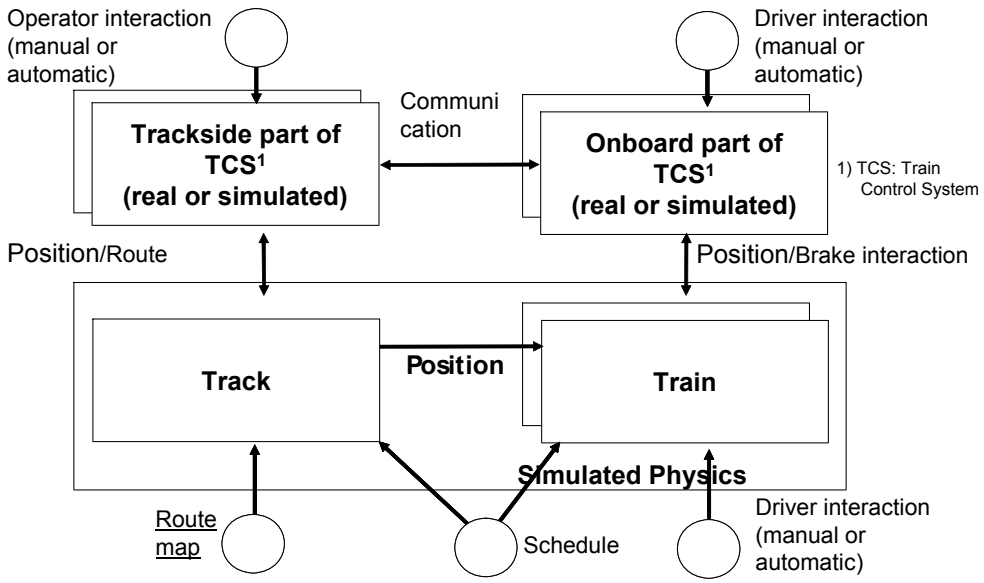


Figure 16.2 Basic simulation concept of the RailSiTe®

Steps of implementation

Due to the complexity of the entire system, the start of operation is split into several steps. The implementation is based on the state of the art of train control systems.

Step 1: The on-board subsystem and the related simulation environment.
Step 2: The trackside subsystem with the related simulation environment, the interface of the interlocking and the connection to the on-board subsystem.
Step 3: Implementation and adaptation of a driver desk [EuDD 2003]. To fulfil the driving simulation, a driver's view visualisation will be added too.

Further steps:

- Connection to a level crossing simulation facility.
- Improvement of the interlocking interface.
- Implementation of the ERTMS/ETCS level 3 moving block functionality.
- Connection to a 3D-movement platform.

Hardware-in-the-loop simulation

One of the applications is the hardware-in-the-loop simulation to check the conformity of a new or improved component of a train control system. Due to the modular set-up of the RailSiTe® the simulation of the on-board or trackside subsystem can be replaced by a real system. A specific computer performs the interface adaptation for the real system.

Several configurations are relevant to check the interoperability, e.g., of ERTMS/ETCS on-board units (OBU) and Radio Block Centres (RBC):

- The evaluation of the interaction between one on-board part and one trackside part of a train control system. This configuration will be used to check the functionality and the conformity with the requirements. If the parts are implemented by different suppliers this configuration will be use to check the interoperability.
- Another configuration containing two on board parts and one trackside will be used to check the interaction of two trains over the common trackside.
- A third configuration with one on board unit and two trackside systems is implemented to evaluate the national or international hand-over of a train.

The RailSiTe® is prepared for two on board units and two trackside systems to have the possibility to implement all these configurations.

Analysis of human factors' influence on the operational safety

Aim

The railway has got – like all modes of public transport – high safety requirements. They lead on the one hand to high technical safety and a complex safety case of new systems and on the other hand to the need for well-trained staff. The migration to a new technical system like ETCS as well as the modification of the operational rules can lead to hazardous situations for two reasons: First the operator has to use a new, unfamiliar user interface; second, and potentially more important, there could be a new safety concept behind the new system and the accompanying operational rules, so that the operator has to change his behaviour generally.

Consequently the objectives for the design of new user interfaces, taking into account knowledge about human factors, are twofold. One aim is the minimisation of risks and hazards linked to them. The other ambition is to create a convenient work place for the staff, to provide a healthy work environment. To reach these objectives the optimal workload for all involved needs to be found, and the work place has to be designed according to the needs of the activities to be carried out and the duration the personal has to be in the workplace. It is also not possible for the signaller or driver to stop and have a look at the manual before taking the next action, which would cause unacceptable delays or even hazards.

Information about performance of the operators

The influence of these factors can be investigated by examination of psychological and physiological data, gathered in tests in the laboratory.

Information of interest for the design of the work places for driver and operator is:

- response time;
- psychological and physiological demand (overload or underload);
- visibility of signs and signals;
- recognisability of given information (duration from perception of sign to recognition of information);
- unambiguousness and comprehensibility of information
- clarity and understandability of rules.

Approach

A driver's desk and an interlocking operator's work place have been integrated in the lab for the analysis of human factors, to record the performance data for the evaluation of operational scenarios and related instructions. Information about suitable designs for the work places, signs and signals can be obtained. The possibility to investigate the behaviour of drivers and operators in a laboratory allows examining the behaviour of different persons in exactly the same reproducible scenarios and situations as often as necessary. It is also possible to change the scenario in small steps and examine the influence of every single change or the combination of certain factors. Furthermore the normal railway operation is not affected by the tests, i.e. no delays or hazards are caused by the tests. In addition tests can be run to evaluate the performance of the personnel in normal and extreme operational and situational scenarios, while neither the personnel nor other persons nor material are in danger. The experience gathered by the institute's staff in their work with driver assistance systems in the automotive field, psychological examinations and evaluation of data collected in field tests, will play an important role.

By means of cameras and other sensors the reactions of the test participants are recordable. The recorded data can be evaluated from different points of view and provide real-time information about reactions. Surveys during a test distract the person under evaluation from his task and questionnaires, filled in after the test, can provide only information retrieved from the memory of those interviewed.

Objectives of the investigations can be e.g.:

- The design of work places and driver desks,
- The design of signs and optical and acoustical signals,
- The avoidance of disadvantageous backgrounds for signs and optical signals,
- The optimisation of mental workload of operator and driver, to avoid mental overload and underload,
- The optimisation of operational rules, e. g. to minimise ambiguous rules and misunderstandings,
- The validation of human factors of automation approaches,
- Experiments related to operational stress and workload of the operators,
- Experiments for safety-critical errors of the operators,
- Knowledge about the influence of medicinal drugs or alcohol.

To gather information about the performance of the personnel in normal as well as unusual situations examinations can be carried out as long-time stress tests. Also extreme situations with poor vision, limited adhesion or operational situations such as many delays or blocked tracks, can be simulated by the appropriate behaviour of the simulator and the simulated environment. In this way situations can be found that cause high stress for the personnel and can particularly lead to errors. With the knowledge gained from these examinations technical devices to support the staff in these situations can be evaluated and optimised.

Sensors, measured data and methods

For the examination of factors that influence human behaviour and performance, a variety of sensors can be installed in the laboratory, according to the requirements of the particular method of investigation. There are two different purposes for the use of sensors: first the acquisition of information about what the operator has recognised and done, and second the acquisition of the physical and psychological status of the operator. For the first purposes are classical logging

facilities used to log the inputs and outputs of the system, but in addition an eye tracking system can be used to acquire which events, system outputs and optical signals could be recognised by the signaller or driver. For the second purpose we have data about the activity of muscles, the heart, the resistance of the skin, the electrical activity of the brain, and the eyes. From these data information about the performance of the test person can be deduced. Other methods and tools coming from psychology can be used to provide subjective ratings of the situation.

Summary

One of the first independent operational labs for the testing of interoperability and conformity will be the RailSiTe® of the German Aerospace Center (DLR) in Braunschweig. In the RailSiTe® independent tests, e.g., of ERTMS/ETCS implementations and operational strategies according to the European Requirements and Standards, can be performed. The train movement and the infrastructure are realised by means of simulation as well as all the inputs from the driver and the operator. As a basis for error representation and inclusion, specific transmission error generators have been developed. For test automation specific concepts are used to ensure realistic and reproducible testing. For the evaluation of the influence of human factors on the safety of the operation a set of methods and tools coming from the psychology are prepared to be used in combination with the RailSiTe®.

References

EuDD (2003). *European Drivers Desk*. www.eudd.org.

Fairclough, S. H. (1993). *Psychophysiological measures of workload and stress*. In A.M. Parkes & S. Franzèn (Eds.), Driving Future Vehicles. (pp. 377-390). London: Taylor & Francis.

Illgen, I., M. Meyer zu Hörste, B. Ptok, E. Schnieder, H. Schrom, and I. Weber (2000): *Distributed multi-train simulation with integration of real component simulators*. Proc. of Computers in Railways (COMPRAIL VII) (B. Mellitt, R.J. Hill, J.Allan, G. Sciutto, and C.A: Brebbia, Eds.) WIT Press, Southampton.

Meyer zu Hörste, M., K. P. Jaschke and K. Lemmer (2003): *A Test facility for ERTMS/ETCS conformity*. Proc. Formal Methods for Railway Operations and Control Systems. (G. Tarnai and E. Schnieder), L'Harmattan, Budapest.

Meyer zu Hörste, M., E. Schnieder and H.-M. Schulz (2000). *Test-case Generation from Petri-net Specifications of a Train Control System.* Proc. Computers in Railways (COMPRAIL VII) (B. Mellitt, R.J. Hill, J.Allan, G. Sciutto, and C.A: Brebbia, Eds.) WIT Press, Southampton.

Schieber, F.; Harms, M.L.; Berkhout, J., Spangler, D.S. (2005). *Precision and accuracy of video-based measurements of driver's gaze location* (pdf, 29.09.2005). Internet: http://www.usd.edu/~schieber/pdf/gaze.pdf (29.09.2005, 17:13).

Trick, L.M.; Enns, J.T.; Mills, J.; Vavrik, J. *Paying attention behind the wheel: a framework for studying the role of attention in driving*; Theoretical Issues in Ergonomics Science, Volume 5, Number 5, September-October 2004, pp. 385-424(40); Taylor and Francis Ltd; 2004

UNISIG (2003a). *ERTMS/ETCS – Class 1: UNISIG Functional Requirements for an on board Reference Test Facility*. Ref.: Subset-094-0, Issue: Version 1.0.0, Date: 14.04.2003

UNISIG, (2003b). *ERTMS/ETCS – Class 1: Test Sequences 1 – 65*. Ref.: Subset-076-6-3-001 – Subset 076-6-0-65, Issue: 1.0.0

This page intentionally left blank

Chapter 17

Cardboard to Computers: An Evolution of Design Visualisation

Karen Wright

Introduction

Traditional two-dimensional (2D) layouts and elevations are usually insufficient for anybody other than those with expertise in architectural design to gain an adequate appreciation of the real spatial environment that they represent. Given that the majority of end-users do not possess architectural expertise this presents somewhat of a stumbling block when trying to get them to make objective evaluations on the suitability of the design of new workplaces.

Transferring these 2D drawings into a three-dimensional (3D) model, either in a physically static mock-up or a computer-based graphical mock-up, can immediately make the design 'come to life' for both ergonomist and end-user alike. Below we describe how traditional mock-up methods have evolved, where the strength in the different techniques lie and how these can be used in a cost-effective manor on projects to get constructive evaluation and criticism from end-users.

Mock-ups and end user participation

The terms 'mock-up' and 'simulation' are used for a range of techniques that involve the creation of a three-dimensional representation of a yet-to-be-built product or system. The purpose of a mock-up is to enhance end-user participation early in the design cycle. The ergonomist can observe the end-user as they verbally walk-through typical operational tasks and identify any ergonomic issues with the proposed design at a stage where changes can be made with little cost impact.

Cardboard boxes

Many ergonomic textbooks feature 1960s sepia photographs of an ergonomist and end-users surrounded by cardboard boxes. However comic this image was to myself, and doubtless other ergonomic students, the fact that 21st century textbook revisions feature similar colour photographs bears testimony to usefulness of this mock-up method.

Life size models of proposed equipment designs can be created from cardboard, foam or wood depending on the budget, practicality and nature of the mock-up testing. Cardboard is a fairly popular medium to use due to the low cost and level of skill required to make the mock-up elements. Physical mock-ups can vary greatly in the level of fidelity. For example, at one end of the scale a number of square boxes could be stacked up to where a new signalling panel would be positioned at little cost. At the other end of the scale the cardboard could be modelled to include the foot well space, the angle of tilt on the panel front and include a colour facia indicating all of the signalling functions, which will take much longer to create.

Whilst these methods are still widely used in many industries where the development of prototypes are essential prior to manufacture, recent developments in the cost and power of computers has generated a new method of visualising designs.

Here comes the science bit – 3D CAD models

The first stage in creating a 3D CAD model requires 2D floor plans, elevations, and hand-drawn sketches of the proposed design. Some of these design data may have already been generated by project designers or engineers and will often exist in standard CAD applications (e.g. .dwg Auto-CAD files). These files can then be easily imported directly into a 3D modelling and rendering software program (e.g. 3D Studio MAX).

At this stage, the model will only contain 2D objects. These models are then modified to add the third dimension. Various other objects – for example VDU screens, keyboards and telephones – can also be added to the model at this stage. One of the more important objects that need to be added are the human models. These can be shown in a range of different sizes and different postures to help visualise the space and use of the proposed design.

To achieve a photo-realistic 3D scene, the real-world properties of objects are simulated. The surface material of an object can be created from with the 3D program library or from digital photographs and then applied to the 3D object. For example, a digital photograph of a wooden surface can be imported then mapped onto the surface of a 3D desk. This type of detail adds realism to the final model. The 3D program allows simulation of the reflectivity, smoothness, and transparency of various objects to further enhance realism.

Another important aspect of modelling a 3D environment is light simulation. The lights used in the 3D program can cast accurate shadows and mimic light decay as in the real world. The 3D programs normally provide several different types of lights including a spotlight, direct light and ambient light.

Figure 17.1 Example light

When the surface materials and lights have been added, the model can be animated with a virtual camera that moves though the 3D model along a predetermined path. The camera has many different lens sizes and zoom functions which can reproduce an end-users viewpoint. Once the desired motion of the camera is set up, each frame is generated and rendered, after which the animation can be shown to end-users in real time at 30 frames per second.

An additional feature on some 3D software is the ability for the user to interact with the 3D environment. Users can move though the 3D scene by manipulating the mouse. This technique is more photo-realistic than a typical walk-through video and allows the user to investigate the 3D environment at their own pace. The camera can be moved and controlled in a number of ways to reflect typical views as a person walks through any environment. For example, the standard forwards and backwards motion (which brings objects into more or less detail), a rotation motion (akin to turning around on the spot and viewing 360° around yourself) and a tilting motion (looking up to the ceiling and down to the floor). For individuals which are less experienced with these controls a number of pre-set camera positions can be easily selected but also allow the end-users to freely interact and immerse themselves in the model.

Human models in 3D CAD

The 3D CAD modelling described in this chapter does not propose that fully adjustable human models are used to test the designs that are being mocked up. There are currently a range of more sophisticated 3D software tools, such as SAMMIE and JACK, that can represent the full range of human anthropometrics and human bio-mechanics capability and limitations. These tools have had great success in the manufacturing industries. These 3D software have powerful programs that can utilise a chosen anthropometric and biomechanical database enabling human models in the 3D environment to be a range of sizes, shapes and postures. The programs are able to indicate how the human would move, highlighting collision points and reach volumes for optimal equipment positioning. However, the 3D CAD modelling described in this chapter is cheaper and quicker to use.

In the author's experience of working on UK railway projects, the level of involvement of the ergonomist on typical projects will not support nor arguably require this level of input/modelling. In the majority of railway control room projects the physical form of the equipment is not novel. There is normally a renewal/replacement or expansion in the existing equipment. Therefore, the type of 3D modelling that would need to be applied in a manufacturing project such as the design of a new train cab require detailed representation of human bio-mechanics would be overly expensive and time consuming in a project looking at the implications of adding a new VDU screen to a workstation. It is not to say that there will be not be ergonomic issues that need to be addressed, but simply that the magnitude of the issues is different in design terms.

Certain human attributes however, can be quickly and easily represented by the 3D modelling described in this chapter. This is where the interaction between a 3D CAD modeller and the ergonomist is valuable. The ergonomist can supply data on the anthropometrics of the human form that they wish to see represented in the 3D environment. Using a standard human model the 3D modeller can quickly modify the gross anthropometrics such as stature, body depth etc. Human models of 5th and 95th percentile stature of the target population are included in most 3D environments to give a sense of scale.

The reach envelopes of the 3D humans can also be simulated as 'objects' in the same manner as any other object in the 3D environment. Similarly a standard cone of sight can be added. Usually the reach and zone of sight objects are given a semi-transparent surface to indicate the representative nature of these 'objects'. The camera lens to indicate the field of view can also be manipulated to indicate the user's typical field of view in the animation and virtual walk-through.

Advantages of the evolution in design visualisation

Cheaper, powerful computers and versatile software have enabled the 3D CAD modelling described in this chapter to be much more accessible. 3D CAD models can be quickly developed by individuals with only little CAD knowledge and experience. The real benefit in this type of mock-up however is the ability to recreate the environmental factors – proving the user with a clear 'look and feel' of the final workplace. Whilst the physical mock-ups have a high level of realism in spatial terms the 3D mock-ups have a high level of realism in visual environment terms. This realism can be extremely effective in engaging even the most sceptical end-users in the assessment of a proposed design.

In terms of costs, developing a 3D model can cheaper and quicker that developing a physical mock-up. When numerous design options need to be assessed the cost savings for the 3D approach come into their own. Different versions of the base 3D CAD model can be quickly created to illustrate the various design options. This enables end-users to be involved in the process of option selection at an early stage in the design lifecycle and enables many design improvements to be identified before a final validation exercise is conducted on a more costly physical mock-up.

A case study of cardboard and computer modelling

Signal box merger

As part of the Network Rail renewals programme in the North East Region, the possibility of co-locating the signalling functions from the West Burton signal box to the Thrumpton signal box on one combined entry and exit (NX) panel was investigated. The West Burton signal box controls a small area of the mainline as well as the branchline moving trains to and from the West Burton power station. Thrumpton signal box controls freight and passengers services over a fairly simple mainline layout with one station stop. The project wanted to know if a combined signalling area of control could be satisfactorily represented on a panel, if the panel would physically fit in the Thrumpton signal box and if it could be adequately controlled and monitored by one signaller.

In line with the requirements of the Network Rail standard RT/E/P/24020: 'Incorporating Ergonomics Within Engineering Design', baseline studies were conducted on the two signal boxes. Data were collected on the workload levels of the signallers, equipment operability issues and the working environment. Detailed measurements were taken of the signal boxes and the existing panels. The initial baseline assessment determined that it would be possible for a single signaller to operate the combined areas of control at the current traffic levels. Furthermore, it was identified that the footprint of the new combined panel would not greatly increase the existing equipment array if certain equipment items were built into the panel itself.

The project supplied information on the manufacturer's standard panel dimensions and a drawing of the face plate with the combined signalling area of control was developed. The face plate was reviewed by the ergonomist and several modifications were made including moving alarm buttons to within the reach of the 5th percentile, re-positioned some of the point switches to ensure a direct correlation with their corresponding point indicators and embedding the TRUST and Train Describer VDU's into the panel in a central location.

Full scale mock-up

A full scale cardboard mock-up of the new signalling panel was requested and created. The cardboard panel was freestanding and colour printouts of the signalling faceplate and the embedded CCTV

and TRUST screens were attached to the tilted front. In addition, other freestanding work surfaces were created that enabled the boundaries of the floor space with the signal box to be formed.

A user trial was held with four signallers, all with experience of the two signal box control areas. The signallers 'stepped-into' the full scale physical mock-up of the signal box and were asked to perform a verbal walk-through of normal, abnormal and emergency signalling scenarios. These scenarios were designed to test whether the Signallers would have any problems with the increased area of control and the corresponding increase in panel length. While the signalling representatives were verbally walking through the steps in the scenarios, the ergonomists were observing their actions and postures and questioning if there were any problems or issues relating to the face plate, panel or signal box layout.

A number of useful findings resulted from the scenarios. The signallers were quick to point out that the position of the control pedestal for the local level crossing would make it impossible for them to operate the crossing whilst viewing the far end of the new panel. Thus a design recommendation to have the controls and a CCTV image on the panel was proposed. The signallers welcomed the design to embed the TRUST, train describer and CCTV screens into the panel. Previously this equipment was in a filing cabinet where the door blocked the view of the panel, or on top of the panel at a viewing height that caused discomfort with the keyboard lead trailing down the panel controls!

Using a full scale mock-up of the panel faceplate was invaluable. Even the ergonomists who had been working with the concept of a panel at just over 5½ metres in length only really appreciated the size when they printed out the faceplate in full. The realism of the full scale panel for the Signallers was demonstrated on a couple of occasions when they leant on the cardboard ledge! Thankfully a sturdy cardboard modelling design meant that the model remained intact.

One key lesson learnt with the full scale mock-up was understanding which were the most essential features of the environment to re-create for the signallers to provide sufficient feedback. The freestanding panel was not deemed essential and the same information could possibly have been achieved by attaching the faceplate to a wall and adding a panel ledge.

Figure 17.2 Full scale mock-up

3D CAD mock up

In addition to creating a full scale mock-up, 3D CAD models of the proposed designs were created. The dimensions of the room and panel were provided to the 3D CAD developer. The faceplate drawing was converted to a .jpg image and incorporated into the virtual 3D panel. This meant that the signallers could see all the same information about the signalling controls and indicators in the 'virtual' signal box as they did in the physical mock-up.

The still images, animations and virtual signal box were presented on a laptop at the signallers' user trial. The signallers were shown the pre-set animation a few times, then they were 'shown around' the virtual signal box. The ergonomist controlled the movement around the virtual signal box at the request of the signallers and the panel design was discussed further.

Figure 17.3 3D CAD mock-up

Lessons learnt

The cardboard mock up helped the signallers get a good feel for the length of the panel that they would be operating in the future. However, the 3D CAD model enabled the signallers to clearly envisage how their signal box would look once the new panel was installed. The level of realism provided in the 3D CAD model enabled them to grasp how their new workplace would look and feel. Viewing the new panel in-situ in the 3D CAD model made the design come to life. In addition, many members of the project team who were not familiar with the signal box felt that they gained a much clearer understanding of the project implications after viewing the model. Therefore, the 3D CAD model went on to provide a secondary benefit after the one day workshop to improve the communication of the design within the project team itself.

The 3D CAD model contained the same information as the physical mock-up therefore the discussion on the face plate continued while the signallers were viewing the panel in the 3D environment. For example, from these discussions the signallers requested that the system alarm buttons be located together, frequently used CCTV controls be placed lower down and the block bells be located in a central location on the panel.

Conclusions

This chapter is not suggesting that one method of design visualisation should replace another as the strengths of each can be seen to be complementary. In many respects employing both methods can serve to increase the level of confidence in the final design solution.

The evolution in the production of mock-ups has enabled the traditional cardboard mock-up to be supplemented with 3D realistic images of proposed designs. While the full size cardboard mock-up enables the end-user to generate a good spatial appreciation of a design, the 3D CAD mock-up enables them to gain a good visual appreciation of the design to fully visualise their future working environment. The quick and inexpensive to develop 3D images provides the ergonomist with another tool to ensure end-users actively participate in the design cycle at early stages which leads to more workable environments.

Acknowledgements

The author would like to thank Eric Waters from Balfour Beatty for permission to publish the work on the Thrumpton Renewal Project, the Network Rail signalling representatives, the Mott MacDonald cardboard modelling team, faceplate design team and the 3D visualisation team.

This page intentionally left blank

PART VI
Routes, Signage, Signals and Drivability

This page intentionally left blank

Chapter 18

Early Route Drivability Assessment in Support of Railway Investment

W. Ian Hamilton, Emma Lowe and Charlotte Hill

Railway investment process

Network Rail is responsible for the operation, maintenance, renewal and enhancement of the national rail network in order to meet its statutory and regulatory requirements and the aspirations of its stakeholders. To fulfil this goal, Network Rail must make intelligent investment and engineering decisions that deliver schemes that can be operated safely, reliably and efficiently.

The Guide to Railway Investment Projects (GRIP) (Network Rail, 2005) describes how Network Rail manages and controls projects, throughout their lifecycle to ensure success and compatibility with existing rail operations. All qualifying renewals, enhancements and capital expenditure projects must comply with the GRIP process. The GRIP process is structured with an overarching policy manual that is applicable to all projects. Similarly there is an underpinning project management manual that is applied to all projects. The process is also supported by:

- An Enhancements and Developments Manual,
- A Renewals Development Manual,
- The Delivery Manual,
- The Third Party Manual, and
- The Investment Regulations, with guidance on obtaining funding authority.

All projects must be able to demonstrate their compliance with these processes. This is achieved through a seven stage project lifecycle. Each stage is required to deliver an agreed set of products. The lifecycle is illustrated in Figure 18.1.

Each of the stages represents a discrete business or technical milestone and GRIP defines the products that should be produced at that stage. For each product GRIP defines its purpose, input dependencies, format and content, and there is a responsibilities matrix that provides specific guidance on the preparation of products and cross-references to applicable standards and templates. Each stage is also subject to a gate review to confirm that all management, business and technical milestones are achieved.

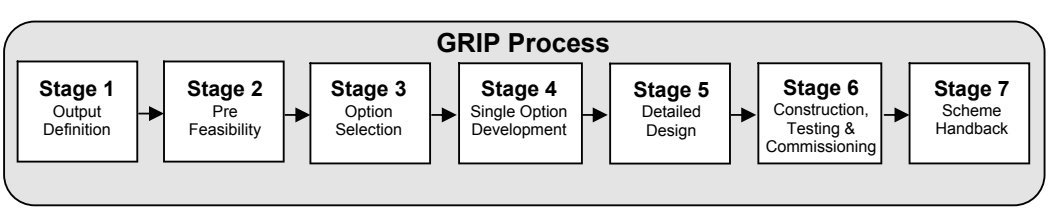


Figure 18.1 GRIP process project lifecycle

Human factors implications for investment

New railway infrastructure projects must be able to demonstrate that anticipated operational revenues will justify the capital investment for their development. In the case of projects such as Thameslink 2000 and Cross Rail, operational revenues will depend on running frequent, high occupancy services, stopping at numerous stations, with short dwell times. These requirements introduce route design and signalling scheme complications that place greater workload demand on drivers and signallers. Consequently, human factors can be a key determinant in the ultimate success of the scheme.

Track upgrades and new signalling scheme projects are under pressure to deliver shorter, more reliable journey times, more frequent services, higher track capacity, greater levels of safety and reliability, and increased value for money. This can result in scheme designs that feature higher operating speeds combined with more frequent speed changes. Stock specific signage is also a feature of the proliferation of lineside signs, all of which demands vigilance and discrimination by the driver, placing a greater reliance on route knowledge. There can also be increased complexity from more variable routing. This means more lineside signage and signalling, plus a host of other junctions and associated indicators. New cab displays and communications systems also contribute to the information processing demands on drivers. And reduced station dwell times and journey turn-around times can add time pressure to normal duties.

Understanding these demands and the drivers' capacity limitations for vigilance, attention, information processing and memory are key inputs into the design activity. There is considerable evidence that high workload can reduce performance effectiveness and increase the propensity for error, prevent effective recovery from mistakes, and impose occupational stress that can affect health.

Network Rail acknowledges the importance of human factors for investment projects and has produced a specific standard (RT/E/P/24020) and guidance (RT/E/G/00027) on incorporating ergonomics (human factors) within engineering design projects (Network Rail, 2003a and 2003b). These deal with the requirements to consider the human interface, and the phases of human

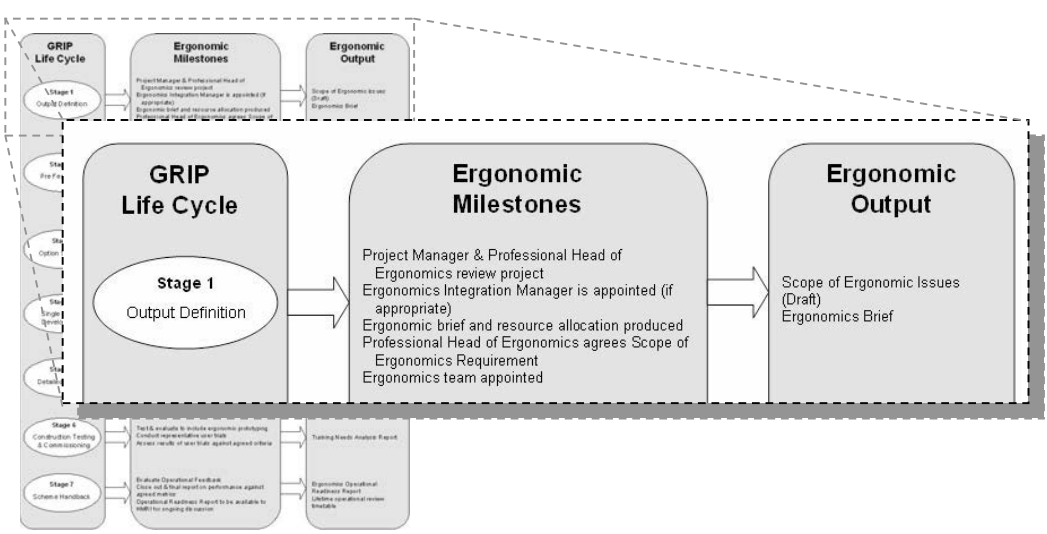


Figure 18.2 Engineering lifecycle with ergonomics milestones and output

factors work are fully aligned with the GRIP process – see Figure 18.2. In particular, the guidance document provides details of the expected ergonomics technical milestones and outputs that are relevant to the stages of the GRIP lifecycle.

One such output is the workload assessment report. Workload assessments are undertaken to establish the feasibility of the demands imposed on the operator by the proposed design of the user interface. The route drivability technique (RDT) provides a means of assessing workload imposed on drivers by design of the infrastructure and proposed operating conditions.

Route drivability = workload

Drivability is influenced by the design of a signalling scheme, because the driver's task is determined by the infrastructure features and characteristics of a route. For example, a driver will read a speed sign, and respond by accelerating after the train length has passed the sign. A driver may then pass a radio change marker board, read it, and respond by checking that the radio has changed to the correct channel. The type of infrastructure features and characteristics of a route will govern the tasks that a driver performs. The location of these features and the linespeed of the route determine when these tasks occur and the time available to perform them.

Previous studies showed how increasing the number of infrastructure features on a route increases the number of tasks to be performed by the driver, and increasing the linespeed reduces the amount of time available to perform these tasks. This, in turn, increases driver workload, which may reduce drivability, all other things being equal (Li et al., 2003; Hamilton et al., 2004a). Thus, decisions taken about the design of a route will directly affect the workload of the driver, and analysis of workload as a dependent variable of route design can be diagnostic of the drivability of the route when it is built.

At present drivability can only be assessed late in the design lifecycle, and tends to be evaluated as part of other processes such as risk assessment and signal sighting where there is a reliance on structured expert judgement. Consequently, the arguments for design modification on the basis of drivability may be difficult and impracticable at mature stages of delivery, and so other engineering and budgetary factors may take priority. The RDT provides a means of making this evaluation early and on the basis of objective, quantified criteria.

Behavioural building blocks

The scientific basis to the RDT, its development and main features of a general model of human information processing capabilities applied to railway engineering are described by Hamilton and Clarke (2003). The model utilises cognitive task analysis and modelling techniques to describe driver performance in relation to infrastructure design features and operational conditions. The model can predict performance time, workload and error consequences of different operational conditions.

Features on the route prompt driver activities. A feature will prompt a group of tasks – the "behavioural building block" – a meaningful group of goal-directed tasks. The building blocks are "anchored" to features identified and selected for modelling. The construction of corresponding behavioural "building blocks" that describe how the driver responds to each route feature, and how they are "anchored" or related to the corresponding features, was undertaken as a scoping study (Hamilton et al., 2004b). The scoping study presented a hierarchical task analysis that described the core driving activities identified in the relevant Railway Group Standard (GO/RT3251). The task analysis covered revenue service operation, including the movement of passengers and freight.

As there are a very large number of different features present on the UK railway, and the driving task covers many different circumstances, it was decided to capture the most important aspects of the task to an appropriate level of detail. Features were prioritised according to the following criteria:

- How much the feature, and its associated task, affects the drivability of the route, as determined by previous work and consultation with Network Rail.
- How much is known from theory or observation about how it affects the driver.
- Whether it can be modelled, either quantitatively or qualitatively.

An example of an anchored building block is presented in Figure 18.3. The anchor is a critical event that prompts the behaviour, like the sighting of a signal or the sounding of an alarm. The anchor may occur at the start or end, but most often occurs partway through the task. Anticipation and expectation play an important role in expert driver performance, and these parts of the task, called the "early" component, will be performed in advance of the prompt. The parts of the task that occur after the anchor event are called the "late" component. The task is represented on a time axis, the length of the task indicating the performance time.

GOMS framework

The drivability methodology is based on the Goals-Operators-Methods-Selection Rules model of skilled human performance (Card, Moran and Newell, 1983). It assumes that drivers perform the driving task in a goal-directed manner, and use rules to select the most appropriate behaviour to complete the task. The hierarchical task analysis of train driving explicitly describes the goals the driver aims to meet, and the methods they have available to them. Simple rules can be constructed to model driver behaviour in relation to task events (for example speed changes, maintaining progress, response to particular features). These have been shown in the validation study to be closely related to real driver behaviour, especially when that behaviour is clearly bounded by task

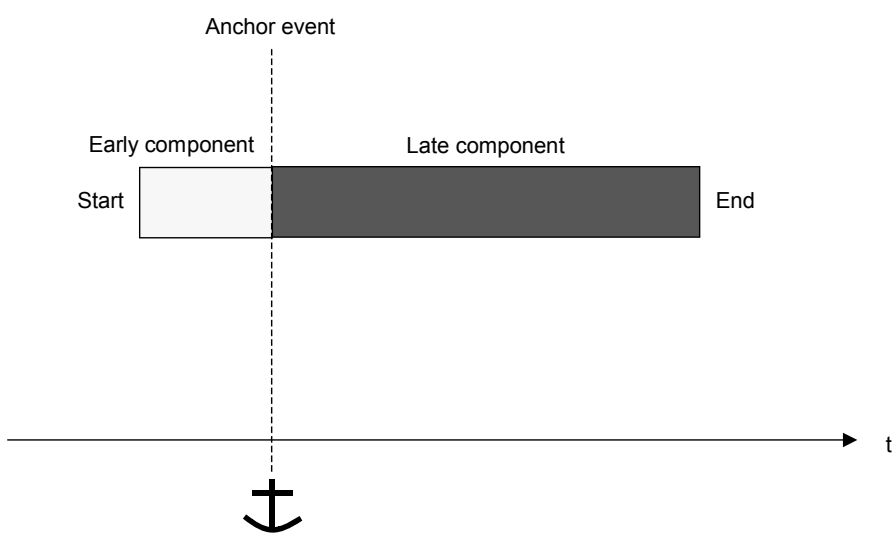


Figure 18.3 A task-group anchored to an event

goals and environmental constraints (for example, braking for a new speed restriction or stopping at a station to meet the service pattern).

Perceive-Recognise-Decide-Act cycle

The perceive-recognise-decide-act cycle (PRDA) provides a general strategy for driver responses to task events. The PRDA cycle divides behaviour into three stages of action:

- A perceptual stage – when the feature is detected and recognised.
- A processing stage – when the response to the feature is selected.
- A response stage – when the response is made. The response may be a physical act, but may also be a cognitive decision that can influence future behaviour.

The model assumes that drivers have a plan of how they intend to drive the route, based on experience and knowledge of how they usually drive the route, and constantly updated to allow modifications in strategy as required. The driving plan allows the driver to predict future events and prepare for them in advance, indeed certain activities require this for efficient performance. The driver continually checks that information and driving events are consistent with the driving plan. New information or unexpected events are incorporated into the plan through the cognitive re-planning task. To reflect this, three additional stages are also included:

- An anticipation stage – when the driver prepares for the activity.
- A search stage – the driver actively searches for the event in order to release the behaviour.
- A checking stage – following the event the driver checks that his action is appropriate and that any information collected is consistent with (or added to) his driving plan.

Therefore the general strategy used in the behavioural building blocks can be summarised by the following tasks:

- Anticipate feature
- Search for feature
- Perceive feature
- Recognise feature
- Respond to feature
- Check/update plan.

Where evidence existed (from studies or driver interview) that further behaviours would be required for the driver to perform a particular activity these were added, but the guiding principle was that the behavioural description would be the simplest one that could account for realistic performance. For the purpose of modelling, behavioural building blocks are broken down into action descriptions. These are classified into either visual, auditory, cognitive or psychomotor behaviour. This so called VACP task analysis is a standard method in human factors which is associated with techniques for performance time estimation and workload analysis, (e.g. demand ratings). Further details are provided in Reference 6.

The following infrastructure features were identified as priority features for inclusion in the methodology:

- Signals, banner repeaters and route indicators
- AWS magnets

- General information boards, radio change and whistle boards
- Permissible speed signs and advanced warning boards
- Gradients and neutral sections
- Stations, platforms and stopping cards.

Table 18.1 shows an example of the 'AWS Horn Response' behavioural building block (Hamilton et al., 2004c).

Table 18.1 VACP Description of the "AWS Horn Response" Behavioural Building Block

Task	Action description	VACP	Demand	Performance time range (s)
Anticipate AWS horn	Anticipate AWS horn about to sound	Psychomotor – Discrete Actuation	1	0.076 to 0.420
		Cognitive – Encoding/ Decoding, Recall	2	0.167 to 0.527
AWS Horn Anchor	-	-	-	-
Hear AWS horn	Hears and recognises the horn	Auditory – Detect/ Register Sound	1	0.075 to 0.370
		Cognitive – Sign/ Signal Recognition	1	0.100 to 0.540
Check AWS alert	Compare AWS alert with expectations	Cognitive – Evaluation/ Judgement (Single Aspect)	2	0.167 to 0.527
Cancel AWS horn	Return hand to desk	Psychomotor – Discrete Actuation	1	0.076 to 0.420

Summary results of RDT validation study

The measures of drivability are the workload variables of time pressure, task demand level and cognitive conflict between concurrent tasks. The RDT's performance time and workload prediction features have been validated for use in drivability assessment. A cognitive task analysis was used to model driver behaviour in response to route features encountered when driving through the core section of the London Thameslink Railway. Performance time and workload predictions were then derived using the various performance time and workload prediction algorithms in the cognitive model.

The modelled scenario was then driven in both a train cab and route simulator (of the same route) with 17 driver subjects. Observational data were obtained for driver task performance, a secondary task measure of workload (i.e. a self-paced mental arithmetic test), and for subjective ratings of workload (using NASA TLX and Instantaneous Self Assessment). These data were then used to compare the drivers' behaviour with the predicted driving strategy. Task performance times were compared with predicted execution times, and predicted workload values were compared with secondary task and subjective data.

The behavioural rules in the cognitive task analysis had predicted, with a high degree of accuracy, how all 17 drivers responded to events. For example, it was possible to predict with precision, the speed profile driven by drivers over the entire route. This is illustrated in Figure 18.4, showing the close match between the predicted driving speed profile for the route (the bold line) and that driven by the drivers in the simulator (the lighter lines). The other line shows the speed profile for a real train on the same route. The modelled driving behaviour, simulated driving, and real world driving speeds all coincide.

The kind of behaviour that proved most difficult to predict accurately was closed-loop control behaviour, such as maintaining a constant train speed. Although the goal of maintaining speed was clearly revealed by the cognitive task analysis, there was a high degree of individual difference in the traction power adjustment strategy employed by drivers. However, the strength of the match to the overall speed profile driven on the route as a whole (Figure 18.4) suggests that this difficulty does not undermine the general utility of the method.

The behavioural performance time algorithms provided reasonable estimates of actual performance for most behaviours. There was a very close match between predicted typical times (Median) for most overt driving behaviour and actual observed behaviour – see Table 18.2. The ranges too were quite accurately predicted, although it is clear that the algorithms tend to over-estimate the slowest (Slow) performance times. This consistent error can become a significant over-estimate when timings are added together for extended sequences of action. For example, station departure preparation actions were predicted as taking typically 7.9s but were observed to have taken only 6s. Further work provides some evidence that predictive accuracy for slowest times can be improved by using a set of algorithms corrected using z-scores (Hamilton and Newman, 2003). The predicted workload values for time pressure, demand and conflict were compared with the subjective and secondary task measures from the simulator study. The RDT produced many more workload peaks than were observed in the simulator data.

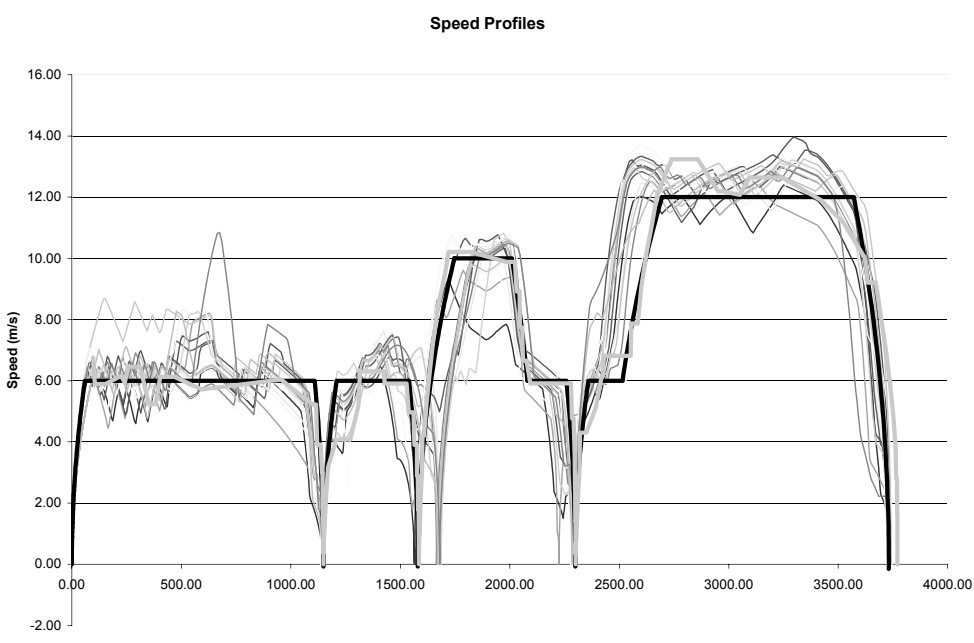


Figure 18.4 Speed profiles

Table 18.2 Comparing predicted (fast, slow and median) and observed (min, max and mean) timings

Task	Min	Fast	Mean	Median	Max	Slow
Move power controller one notch (hand on)	0.17	0.23	0.55	0.52	1.08	1.06
Move power controller one notch (with reach)	0.83	0.57	1.10	1.09	1.63	1.92
Adjust power controller two notches (hand on)	0.42	0.39	0.94	0.90	2.25	1.87
Adjust power controller two notches (reach)	1.33	0.53	1.47	1.10	1.83	2.11
Move brake one notch (hand on)	0.13	0.24	0.46	0.53	0.92	1.07
Move brake one notch (with reach)	0.96	0.58	1.42	1.11	2.29	1.94
Move brake two notches (hand on)	0.75	0.39	1.48	0.91	3.00	1.88
Respond to AWS horn (with anticipation)	0.25	0.20	0.46	0.50	0.71	1.05
Respond to AWS horn (no anticipation)	0.71	0.61	1.17	1.17	1.88	2.04
Set DRA	1.21	0.65	1.50	1.31	1.79	2.52
Take out DRA	0.96	0.65	1.77	1.31	2.58	2.52
Move master switch one position	0.96	0.71	1.36	1.26	2.29	2.05
Sound horn (hand on)	0.92	0.70	1.47	1.48	1.96	2.84
Sound horn (with reach)	1.17	1.11	2.09	2.16	3.38	3.83
Pearson correlations	**0.79**		**0.90**		**0.77**	
Spearman correlations	**0.69**		**0.81**		**0.60**	

However, for every peak in the subjective and secondary task data there was a corresponding peak in the predictions (Figure 18.5). Both the predicted measures and the observed measures show peaks at arrival into stations and during speed control under high vigilance demand periods. It is suggested that the additional peaks in predicted data reveal the greater sensitivity of the predictive methods to short duration fluctuations in workload. This is probably due to the higher sampling rate of the predictive method compared to the intermittent and post-hoc methods of data collection relied upon for the subjective reporting and secondary task methods. Overall the results of the validation study have provided sufficient evidence to justify the continued development of the RDT.

RDT in support of GRIP

The RDT now exists as a software supported method for use in predicting the drivability of a route (Figure 18.6). The RDT is implemented as a Microsoft Access database application that allows human factors specialists and signal scheme designers to select the features of a planned route and quickly assess the feasibility of the driving task. The basic RDT procedure involves:

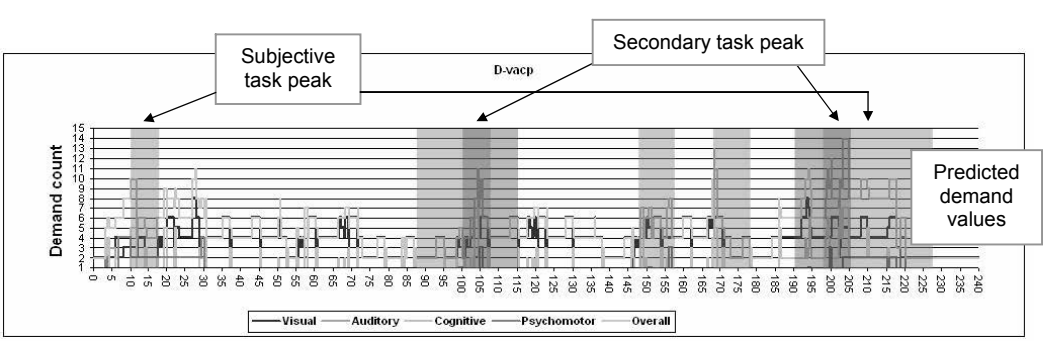


Figure 18.5 Workload profiles for demand values

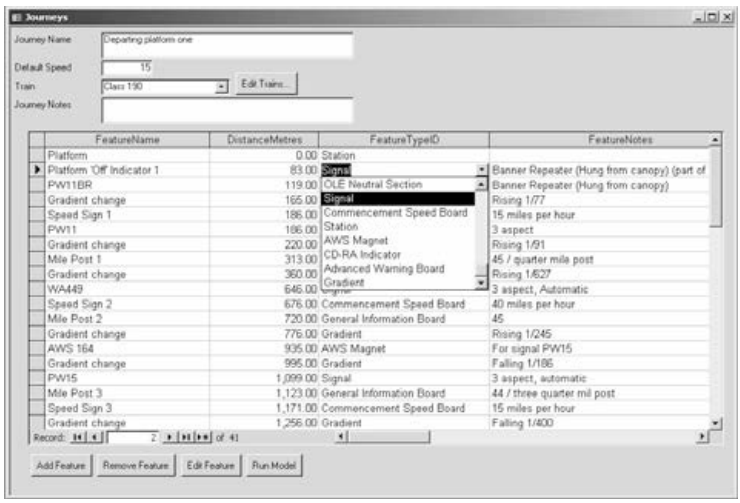


Figure 18.6 The basic RDT screen

- Entering infrastructure features and their distance along a journey, based on a signalling scheme diagram or other source of information
- Running the model to calculate the speed profile and workload rating
- Interpretation of the workload ratings
- Discussion and iteration of the journey, incorporating human factors advice to optimise the driver's workload.

This assessment allows human factors and workload considerations to be made in a timely and efficient manner. The outputs from the technique provide a way of identifying problem areas in drivability terms, and allow for rapid iteration to test the effectiveness of proposed solutions. The RDT has already been used in support of a renewal project and has demonstrated that it is sensitive to differences in the existing and new route designs under consideration, and identified the increased demands on drivers associated with the revised route.

The predictive analysis of route drivability offers a number of advantages:

- Improve the safety, operational efficiency and reliability of routes
- Incorporate human factors early in the project lifecycle, facilitating human factors integration throughout the project cycle
- Objective, quantifiable and diagnostic methodology to support structured expert judgement (SEJ) in decision making.
- Support the GRIP process at stages 3 to 5 (Figure 18.7).

The RDT methodology is also complementary to the design process. For example, new schemes are increasingly being modelled in CAD packages (like SolvSighting) to estimate sighting distances for particular features and establish that they satisfy the required standards. The RDT uses sighting distances to anchor when the driver begins particular tasks, so the output from CAD methods would improve the accuracy of a RDT model of a route.

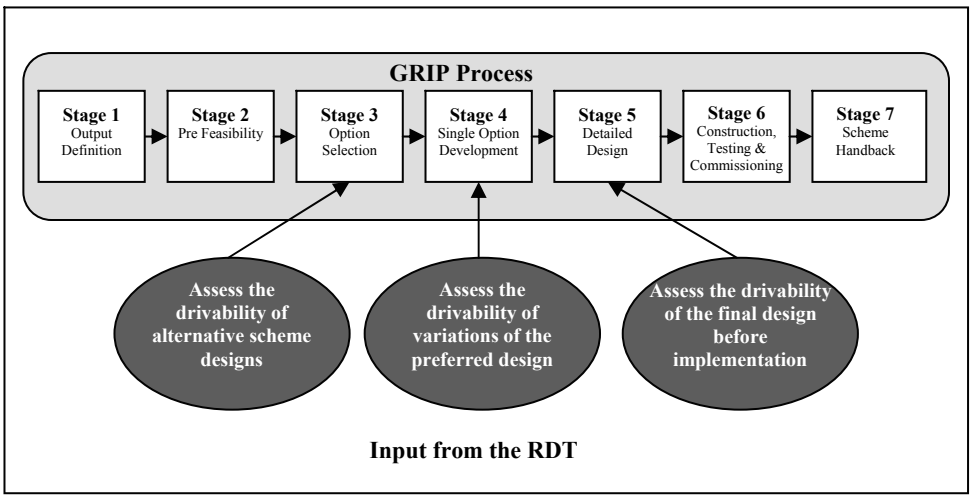


Figure 18.7 Inputs from the RDT to the GRIP stages

Ongoing work

Work on the development of the RDT is continuing with the list of route features, and their associated building blocks, being extended to include such things as signal aspects and route indicators. The extension also includes further development of the workload outputs. In particular, a cognitive conflict parameter is to be added that will provide a clear indication of where excessive workload occurs. This will also identify precisely the route design features that have caused workload problems so that these can be targeted for corrective action. There is also work to develop an engineering procedure to support the application of RDT within projects and to specify competency requirements for its use. This work will be concluded in 2006.

References

Card S.K., Moran T.P., and Newell A., (1983), *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Inc – New Jersey.

Hamilton and Newman (2003), *Cognitive Task Analysis for Signal Sighting Hazards*, HEL/RSSB/03862/RT1, Issue 01, 8th October 2003.

Hamilton, W.I. and Clarke, T. (2003) *Driver Performance Modelling and its Practical Application to Railway Safety.* In Human Factors in Rail Conference, October 2003, York.

Hamilton, Lowe and Blanchard, (2004a). *Validation of ATLAS Workload Prediction Tool for Early Human Factors Analysis*, HEL/NR/03860/RT2, Issue 01, 1st April 2004.

Hamilton, Lowe, Blanchard and Maddock, (2004b), *A Scoping Study for Route Derivability Assessment*, HEL/NR/041022/RT1, Issue 01, 11th August 2004.

Hamilton, Lowe, Blanchard, Hill and Maddock, (2004c) *Early Route Drivability Assessment*, HEL/NR/041024/RT1, Issue 01, 3rd December 2004.

Li, G., Hamilton, W. I. and Clarke, T., (2003) *Driver recognition of railway signs at different speeds – A preliminary study.* In: Contemporary Ergonomics 2003, (ed. P.T. McCabe), London: Taylor & Francis, 373-378

Network Rail Company Specification, 2003a. *Incorporating Ergonomics Within Engineering Design Projects: Requirements,* RT/E/P/24020, Issue 1, August 2003.

Network Rail Company Specification, 2003b. *Incorporating Ergonomics Within Engineering Design Projects: Guidance Note,* RT/E/G/00027, Issue 1, August 2003.
Network Rail, *The Guide to Railway Investment Projects (GRIP),* 2005.

This page intentionally left blank

Chapter 19

Development of a Route Assessment Checklist Tool for Train Operators

Mark Newman, Claire Turner, Ann Mills and John Peters

Introduction

The Human Factors Route Assessment Checklist, or RAC, is based on the Human Factors SPAD Hazard Checklist, a tool that was developed for Network Rail in 2002 to enable a rational, objective and coherent approach to the investigation of human factors causes of signals passed at danger (SPADs). (Human Engineering, 2002; 2004).

In 2004, RSSB and Network Rail sponsored a programme of work to update the content of the SPAD Hazard Checklist (Human Engineering 2004; 2005). Industry SPAD data and research were reviewed to ensure that the checklist contents remained valid, and a number of new hazards were added where supported by evidence. Mitigation measures were also updated to reflect current best practice.

In addition, the sensitivity of the checklist was improved, since research indicates that some hazards, such as driver fatigue, have a greater impact on SPAD risk than others. This was achieved through analysing the responses to a survey of driver attitudes towards the various hazards in the checklist. Every Train and Freight Operating Company (TOC and FOC) working in England, Scotland and Wales was contacted to recruit participants for the survey. In addition, a copy of the questionnaire was hosted on the industry website www.SPADweb.com. Respondents were asked to rate each hazard contained within the checklist on a four point scale:

- This is not a risk
- This is a minor risk
- This is a moderate risk
- This is a severe risk.

Two definitions were provided to support respondents. A minor risk was defined as something that could increase the chances of an accident happening. A severe risk was defined as one that would be very likely to cause an accident.

In all, nearly 400 questionnaires were returned. The results from the survey were used to rate the severity of the hazard by identifying the modal (i.e. the most frequent) response. These results were also supplemented with the data gathered in another survey of driver attitudes to signal sighting hazards (Network Rail, 2003) which used many of the same items and an identical questionnaire format. In the result of a tie between severity ratings, or where the modal rating was only marginally higher than one or more other severity ratings, the hazard was assessed in a workshop by a panel of rail industry SMEs (subject matter experts). This allowed for an unambiguous assessment of the severity of each hazard to be provided. A panel of human factors SMEs also reviewed each item, to verify the severity rating of each before its inclusion in the weighted checklist.

Although the SPAD Hazard Checklist had been designed to identify root causes of SPADs in an incident investigation, many of the hazards it contains can be dealt with proactively as part of the wider SPAD risk management process. TOCs play an important part in this process, e.g. identifying problems with the drivability of routes. The requirement for a formal and consistent method of assessing routes for human factors hazards led to the development of the Route Assessment Checklist.

Aims of the route assessment checklist

The RAC will allow TOCs to:

- Identify those items that, from a human factors perspective, contribute to risk
- Determine which routes are more difficult
- Use this information in training programmes.

The RAC could also be used to extend the risk assessment procedure currently used for signals. The current procedure (Network Rail, 2003) scores existing and proposed track layouts, with the aim of predicting the frequency of collisions and factoring the consequences of these into the assessment of a proposed track design. This process currently uses a list of questions to predict SPAD likelihood; the RAC would therefore complement the process by providing a more sensitive means of SPAD prediction.

Design of the route assessment checklist

The RAC is based, primarily, on the items in the SPAD Hazard Checklist (SHC), and the survey of drivers that was used to weight the items in the checklist. A number of items were dropped from the SHC when it was adapted for use in the RAC. These were typically from the category of Personal issues, (e.g. fatigue); these items are important for SPAD incident investigation, but are unusable within the RAC as they are impossible to predict or identify on the route itself.

The hazards in the RAC are grouped into 7 categories (Table 19.1).

Table 19.1 Route assessment checklist categories

RAC1	Sources of distraction
RAC2	Visibility factors
RAC3	Perceptibility factors
RAC4	Association with line
RAC5	Reading the aspect
RAC6	Interpretation
RAC7	Transitory hazards

These items refer, primarily, to permanent features of the route. When performing a route assessment, the user will concentrate on these features to calculate an overall hazard score and identify those signals where an error is most likely.

However, the RAC also contains a category describing transitory issues, for example *contractors working on the line* and *poor weather*. While these are likely to change on a day-to-day basis, they are also items that can be observed while driving, and so carry some risk. They can therefore be used in a human factors assessment of risks on the route during driver training (Network Rail, 2002).

Process overview

The Route Assessment Checklist has been created in MS Excel for ease of use and dissemination. It can either be printed out or used in its electronic form.

The user is presented with a series of questions in each of the categories in Table 1 above. Questions are designed to identify the presence or absence of a particular hazard on a route and require a yes or no answer. The wording of each question has been checked extensively with SMEs to ensure an unambiguous understanding. Guidance on each question is provided for users. This provides an extended description of each checklist item, giving examples where appropriate and also describing why the item can have an impact on the driver, and what measures are possible to mitigate the hazard.

Use of the RAC is a two stage process: answering then scoring. For each route to be assessed, the user should identify which hazards are present for each signal on the route. A total can then be calculated for how many low, medium and high severity hazards are present for each signal and each route. Where possible, the results of this assessment should be used in the SPAD management programme, as hazards should be mitigated by removal/elimination. Additional training of the driver will often help reduce the risk, but cannot always eliminate it.

Where the identified risks cannot be eliminated, the route assessment score can be used to identify the most challenging routes, and also the points on those routes where a driver is more likely to make a mistake. As it has been suggested that drivers can relax their concentration following a particularly difficult or demanding period of work, this will be *at* or *following* signals (or route sections) with large numbers of identified hazards.

Guidance	
Description	*OVER PROVISION OF SIGNAL AWARENESS AIDS CAUSES HABITUATION /DISTRACTION* Over-provision of signing (e.g., countdown markers or banner repeaters on the approach to signals) causes distraction and habituation. This means that since train drivers are repeatedly presented with the same information day in day out, they can become 'over-used' /insensitive to the information around them. Overuse of reminder aids can misdirect the driver's attention away from signals.
Information needs	Consider: • How many reminder aids are displayed on approach to the signal; • Which reminder aids are used and where.
Information source	Site documentation. Site visit.
Is there evidence which indicates that this is a hazard?	Site visits - signal awareness aids were present at SPADed signals.
Impact of Hazard / Consequence	Vigilance / driver distracted from the signal.
Possible risk reduction measures	Consider reducing the number of signs on approach to multi-SPAD signals if it is considered that this will increase driver concentration. Rule of thumb - there should be no signal awareness aids placed between the AWS magnet and the signal it is associated with.

Figure 19.1 Guidance example

Stage 1: Answering the questions

All the questions in the checklist need to be answered in turn. The guidance provided assists users with understanding what is being asked and why, and what information is required in order to provide the correct response. A site visit and/or cab ride to the signal location may be necessary, or an examination of site documentation/descriptions such as the P'way diagrams. Risk reduction measures are also included, as it is generally preferable to eliminate hazards wherever possible. An example of the guidance is shown in Figure 19.1.

Stage 2: Scoring the questions

For each checklist section the user must count the number of items present in each severity level and record this at the bottom of the page. This process is illustrated in Figure 19.2.

A 'no' answer means that the question is not relevant to a particular signal and should therefore be ignored. An example of how to score the checklist is illustrated in Figure 19.3.

Having answered each question, the user must then add up the 'Page Total' for each part of the checklist and enter this in the box provided on the bottom right hand side of the page. For the example of the checklist shown in Figure 19.3, the Page Total would be low = 1, medium = 2, high = 3. The page totals should then be summed to obtain the overall 'RAC Total'. This number is entered in the box provided on the last page of the checklist.

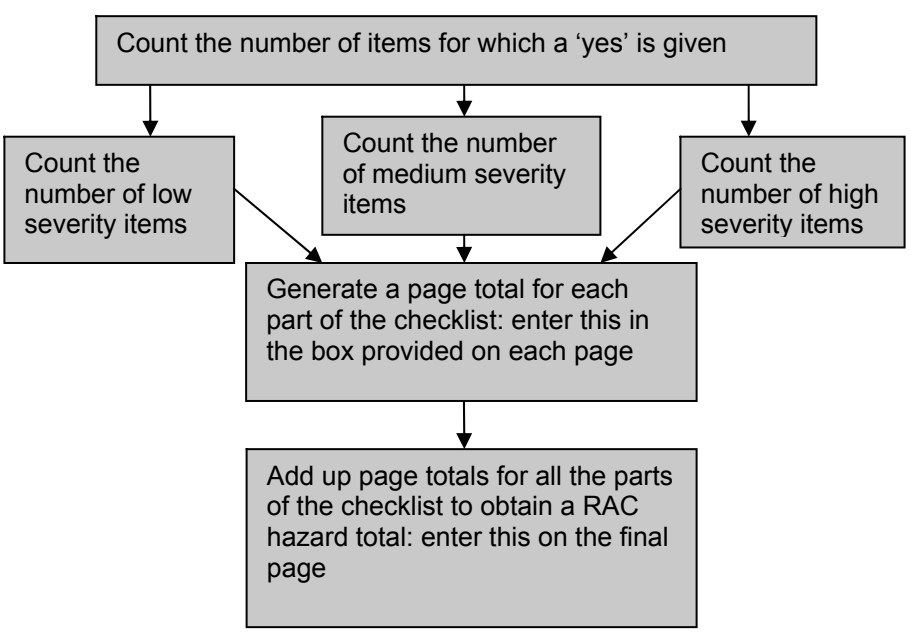


Figure 19.2 Overview of the process for scoring questions

RAC1: Attention/Distraction Factors

Hazard Query		Severity	Answer (yes/ no)
1.	Are multiple reminder aids displayed on the approach to the signal?	low	yes
2.	Is the signal positioned soon after an OLE neutral section or a section gap in the conduction rail?	medium	
3.	Could a complex track layout distract the driver?	medium	
4.	Are there any compelling but non-operational distractions on the lineside (e.g. motorway)?	medium	
5.	Must the driver perform an in-cab activity on approach to the signal, such as arming ATP or use/ set up radio?	high	yes
6.	Is there a falling gradient, which may preoccupy the driver?	medium	
7.	Is there any line-side DOO equipment?	medium	
8.	Does the route change, for example from a fast to a slow line, or vice versa?	low	

Figure 19.3 Checklist scoring example

Interpretation of the output

When used as a tool for route assessment, the RAC will show, for each signal on a route, how many high, medium and low severity hazards there are. A total can be calculated for an entire route by adding together the number of hazards found at each of the signals. There are, at present, no figures to say how many hazards is "too many" on a route. The best use of the route assessment data is therefore to make relative assessments – to identify those signals, those stretches and those routes with a higher than normal number of hazards. These are areas in which errors are more likely.

In addition, there is also a danger that errors may occur immediately after a difficult area (i.e. where there are more hazards than usual). This is caused by the driver relaxing too much after passing the risk area. These danger areas can be flagged up during route learning. It could also be possible to use the total hazard score for a route to determine which routes should be given to which drivers. For example, it might be possible to give routes with low RAC scores to drivers who have recently completed training or who are returning from a period of extended leave (or sickness).

Conclusion

The Route Assessment Checklist allows a more proactive approach to SPAD management by highlighting those routes, and those sections within routes, on which driver error is most likely. This can assist in planning engineering programmes, by highlighting those areas that are most at risk.

The checklist can also be used by TOCs to indicate the more difficult routes on which drivers may require extra training and practice. An added benefit is that it is also possible to look for evidence of a driver under assessment being aware of any hazards encountered, by, for example, driving a route and giving a simultaneous route commentary. Some routes will be more prone to

some of these transitory hazards than others. For example, routes through mountainous areas may be more prone to fog than routes at a lower level, and this could therefore be factored into the route hazard assessment.

The RAC can be also used for rostering purposes, for example by allocating less experienced drivers to easier routes, and can also be used in training programmes as a means of testing whether drivers are aware of the dangers on a route.

The route assessment checklist is likely to undergo a period of field-testing before being made widely available to industry. However, it is anticipated that this will be a cost effective means to support driver training and help reduce the number of SPADs occurring in the future.

References

Human Engineering Ltd. (2002). *Development of a Human Factors SPAD Hazard Checklist*. (Report No. HEL/RT/02719/RT1), Issue 1, 20th September 2002.

Human Engineering Limited (2004). *User Guide for SPAD Hazard Checklist*. (Report No. HEL/RSSB/041123/RTB2). Issue 1, December 2004.

Human Engineering Ltd. (2005). *Overview of Signal Sighting and SPAD Risk Management, Stream B.* (Report No. HEL/RSSB/041123/RTB1) Issue 2, 3rd June 2005.

Network Rail (2002). *Train Driving*. (Railway Group Standard GO/RT/3251). Issue 4, October 2002.

Network Rail (2003). *Prevention and Mitigation of Overruns – Risk Assessment Tools*. (Report No. RS/E/S/14200). Issue 1, April 2003.

Chapter 20

Informing Signage Positioning Rules Through a Human Factors Assessment of Signal/Sign Co-location

Charlotte Hill, Harry Blanchard and Mike Carey

Introduction

A key aspect of the West Coast Route Modernisation (WCRM) project is to upgrade the infrastructure to allow tilt enabled trains to travel at Enhanced Permissible Speeds (EPS). This work includes the implementation of new speed signage for EPS.

When considering the positioning of these new speed signs, it was recognised that placing a sign within reading distance of a signal has the potential to distract the driver from the task of signal reading. In order to reduce this risk, a design rule was developed that placed an 'exclusion zone' for signs around signals. No signs were to be placed within a zone up to 300 yards before the signal, or 200 yards beyond the signal. The application of this rule, however, caused significant problems at some locations where it would result in a major detriment to the line speed profile.

We present a human factors assessment of signal/sign co-location carried out on behalf of Network Rail to examine whether it is acceptable to deviate from the exclusion zone rule. Through modelling of driver perceptual and cognitive behaviour during co-location scenarios it was possible to identify positions within the exclusion zone where it is safe to locate signs, and positions where it is not. In this way, the findings can be used to inform the review and revision of signage positioning rules.

Qualitative assessment of co-location risks

Humans have a finite limit to their information processing resources, hence driver performance will deteriorate under high workload conditions, increasing the likelihood of driver errors. A qualitative assessment of co-location was undertaken to identify the driver errors that could occur during a range of co-location scenarios and their potential consequences. This involved qualitatively examining a selection of co-location scenarios.

This assessment revealed that signal/sign co-location may have the potential to increase the risk of a Signal Passed at Danger (SPAD) or an overspeed. This is best demonstrated by a worked example. Figure 20.1 shows one of the co-location scenarios examined during the assessment, where the driver is approaching a speed board indicating a reduction in linespeed, and a signal showing a double yellow aspect.

In this scenario, high workload conditions are produced by the driver's requirement to time share the tasks of performing the sign-related braking plan and the two reading tasks. This workload may exceed the driver's cognitive capabilities, causing parts of tasks to be delayed, incorrectly performed, or omitted. These performance decrements may result in failure to adequately complete one or more of the following tasks:

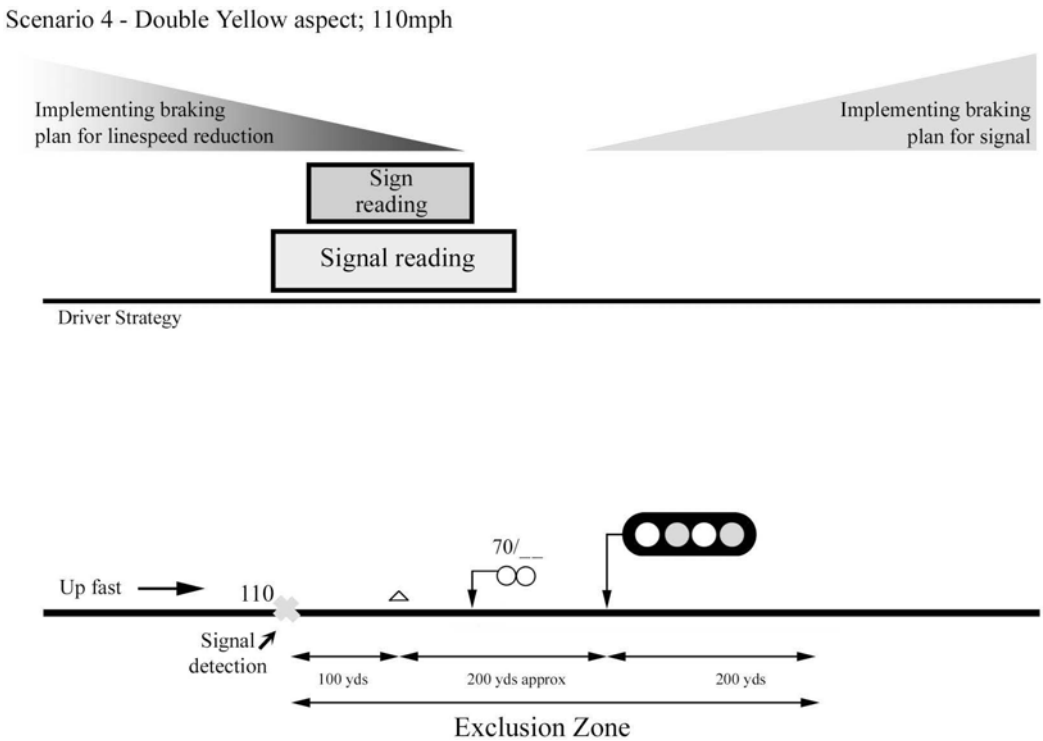


Figure 20.1 Co-location scenario used as worked example of preliminary modelling

- The driver fails to implement the sign-related braking plan correctly.
- The driver fails to read the sign and confirm that the correct speed has been attained.
- The driver fails to read and interpret the double yellow aspect.
- The driver fails to modify the plan after signal reading.

Of the above potential driver errors, three were identified as being associated with potential safety risks. First, failure to implement the braking plan related to the change in line speed may result in an overspeed. Second, failure to read and interpret the signal aspect could lead to a delay in braking, resulting in a subsequent SPAD. Finally, the driver may fail to correctly evaluate and revise current plans. If the driver failed to reduce speed further after reading the signal aspect, perhaps having confused the two braking plans or not having monitored braking adequately, this could lead to a SPAD.

The outcomes of the qualitative assessment revealed that the level of risk associated with potential errors that may occur during a co-location scenario is primarily influenced by the signal aspect being displayed and the type of speed sign.

Development of driver strategy

A model of driver behaviour was developed to allow co-location scenarios to be quantitatively modelled. Based on data collected through driver workshops, cognitive task analyses were developed to represent the following driver tasks:

- Signal reading and response
- Sign reading and response
- Responding to AWS
- Maintaining speed
- Maintaining vigilance
- Checking speedometer.

In order to model driver strategy, an estimate of the time it takes for a driver to complete each task behaviour is required. Performance time estimates used in the co-location study were derived using Perfcalc, a software program that uses a comprehensive set of performance time prediction algorithms drawn from the engineering psychology literature. Perfcalc allows performance times of slow, median and fast operators to be predicted – 'fast man' times represent the fastest times reported for any individual, 'slow man' times represent the slowest , and 'median man' times the average time taken to perform a given task.

The results for fast, median and slow are relatively skewed such that the difference between slow and median man scores is much greater than the difference between fast and median. This difference is far greater than is likely to be true for the driver population working under normal conditions. A modified distribution of times can instead be calculated, in which the population is more evenly distributed. Given that the fast man score is the fastest time, a conservative assumption would be that this time is representative of the top 1% of the population. If the median man score is assumed to be representative of the average of the real driver population, then a 'modified slow man' score, representative of the slowest 1% of the actual driver population, can be derived by a simple process – calculating the difference between fast and median man scores and adding this difference to the median man time.

Throughout the co-location study, median, slow, and modified slow man times have been used when examining the workload associated with co-location scenarios. This allowed the acceptability of co-location to be assessed for a range of performance groups, including the worse case scenario (slow man). The performance timings for each of the CTA behaviours used in the modelling are presented in Table 20.1.

It should be noted that an assumption is made throughout the modelling that drivers both visually acquire and read speed signs. It is recognised that drivers report that they do not need to read speed signs, as they are able to rely on their route knowledge for target speed information. Drivers anticipate the speed change and can use a range of visual cues to identify the location at which a change in speed is required to commence. However, even the most experienced drivers report that they 'sometimes' read signs and it is likely that even if a sign is not 'read', the presence of the sign in the visual field will cause it to be acquired visually. Less experienced drivers on a route are likely to rely on speed signage more. Hence, it was considered appropriate and suitably conservative to assume that signs are both visually acquired and read on approach.

Sequence of signal/sign reading tasks

The sequence of signal and sign reading tasks, and the point at which these tasks commence, was based on the following assumptions regarding driver strategy:

Table 20.1 Performance times for CTAs used to model driver strategy

Task and Associated Behaviours	Median man time (s)	Modified slow man time (s)	Slow man time (s)
SIGNAL READING			
Search	0.54	0.88	1.34
Detect	0.64	1.035	1.61
Read	0.47	0.77	1.24
Process	0.32	0.473	0.527
Monitor	Continuous behaviour until the signal is passed		
Total	1.97	3.158	4.717
SIGN READING			
Search	0.54	0.88	1.34
Detect	0.64	1.035	1.61
Recall	0.32	0.473	0.527
Read	0.47	0.77	1.24
Process	0.32	0.473	0.527
Total	2.29	3.631	5.244
RESPONSE TO AWS			
Detect	0.17	0.24	0.37
Process	0.24	0.38	0.54
Cancel	0.568	0.845	1.03
Total	0.978	1.465	1.94

- The sequence of reading tasks depends on which infrastructure feature is detected first – if the signal is detected first, the signal reading task will begin prior to the sign reading task and visa versa.
- The reading task for the first detected infrastructure feature will commence as soon as that infrastructure feature is detected.

Signal/sign detection points

The detection point for the signal was based on the minimum reading time (MRT), as specified in the Railway Group Standard GE/RT8037 (2003). The standard states that the MRT shall be no less than eight seconds travelling time before the signal. The modelling was therefore based on the assumption that drivers detect the signal eight seconds before they pass it, as this represents the worse case scenario in terms of the amount of time available for the driver to complete the signal reading task. The standard also specifies a requirement to increase MRT in the presence of certain hazards. The MRT assessment form suggests that an additional second is added to MRT when a speed sign is located on approach to a signal. An MRT of nine seconds was therefore used as the detection point for signals when the speed sign is located on approach to the signal.

The detection point for the speed sign was also based on a Railway Group Standard. (1996) GK/RT0033 which specifies that "signs which are required to be read and understood by the driver

of a moving train shall generally be visible for at least 4 seconds". A sign detection point of four seconds before the sign was therefore used for modelling.

Preliminary Quantitative Modelling using ATLAS

As demonstrated by the qualitative assessment of co-location scenarios, the overlap of signal and sign reading tasks is likely to increase workload. This assessment was, however, unable to provide a detailed examination of the workload associated with task overlap, or to provide guidance on whether the workload associated with a given co-location scenario was acceptable or not. The next stage of the project therefore involved quantitative modelling of a selection of co-location scenarios in ATLAS[1] in order to examine the workload associated with the overlap of tasks in more detail.

The selection of co-location scenarios for modelling was based on the findings of the qualitative assessment which suggested that sign position and speed may have the greatest influence on the acceptability of the workload associated with co-location scenarios. Eight sign positions and two train speeds were combined to produce a representative and useful selection of 16 co-location scenarios for examination. The three sections of the exclusion zone were divided into equal sections to create the eight sign positions to be examined, as demonstrated in Figure 20.2.

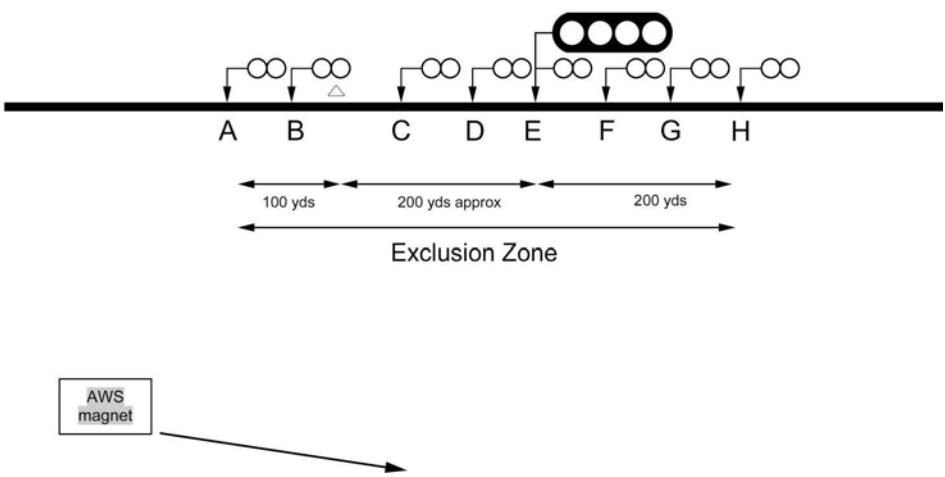


Figure 20.2 Sign positions included in the ATLAS modelling

The distances of each sign position relative to the signal are:

Position A	300 yards before the signal
Position B	250 yards before the signal
Position C	133 yards before the signal
Position D	66 yards before the signal
Position E	At the signal (0 yards before/after the signal)
Position F	66 yards after the signal
Position G	133 yards after the signal
Position H	200 yards after the signal

1 ATLAS is Human Engineering's task performance and workload modelling software. © Human Engineering 1995-2004.

Train speed was modelled at 60mph and 110mph. The outputs from the modelling revealed that signal and sign reading tasks would overlap for some sign positions. Examination of these overlaps revealed a number of behavioural conflicts – instances whereby the driver is required to perform a number of tasks simultaneously that are competing for the same mental resources. Some of these conflicts are considered to represent an unacceptable level of workload. These conflicts are shown in Table 20.2. Co-location scenarios that resulted in overlap of these behaviours were considered to be unacceptable (i.e. the relative position of the signal and sign is unacceptable).

Table 20.2 Unacceptable behavioural conflicts relating to co-location

Conflict between		Nature of conflict
Behaviour 1	Behaviour 2	
Search for signal	Search for sign	Visual
Search for signal	Read sign	Visual
Detect signal	Detect sign	Visual
Detect signal	Read sign	Visual
Read signal	Search for sign	Visual
Read signal	Read sign	Visual
Read signal	Detect sign	Visual
Recognise signal	Recall sign value	Cognitive
Process signal	Recall sign value	Cognitive
Process signal	Process sign	Cognitive
Process AWS	Process sign	Cognitive
Process AWS	Recall sign value	Cognitive
Process AWS	Process signal	Cognitive

Consideration of rescheduling

It may be the case that a driver is able to reschedule one or more tasks in order to mange workload and avoid overlap of conflicting behaviours. Drivers may reschedule by either: task postponement – delaying the start of a task; or task switching - interrupting a task as it is being performed, completing it later. A1 and B1 (on Figure 20.3) indicate the earliest point at which tasks A and B respectively can commence. A2 and B2 indicate the point by which tasks A and B respectively must be completed.

The preliminary quantitative modelling using ATLAS and the possibility of rescheduling taken together suggest that a co-location scenario is acceptable if either none of the primary behaviours overlap, or there is sufficient time available for the driver to reschedule one of the tasks in order to avoid overlap of primary behaviours.

Having defined the above criteria, it was determined that it would be possible to use a spreadsheet program (Microsoft Excel) in order to perform analyses of the acceptability of a wide range of scenarios. The next stage of the project involved quantitative modelling of a larger selection of co-location scenarios using this method.

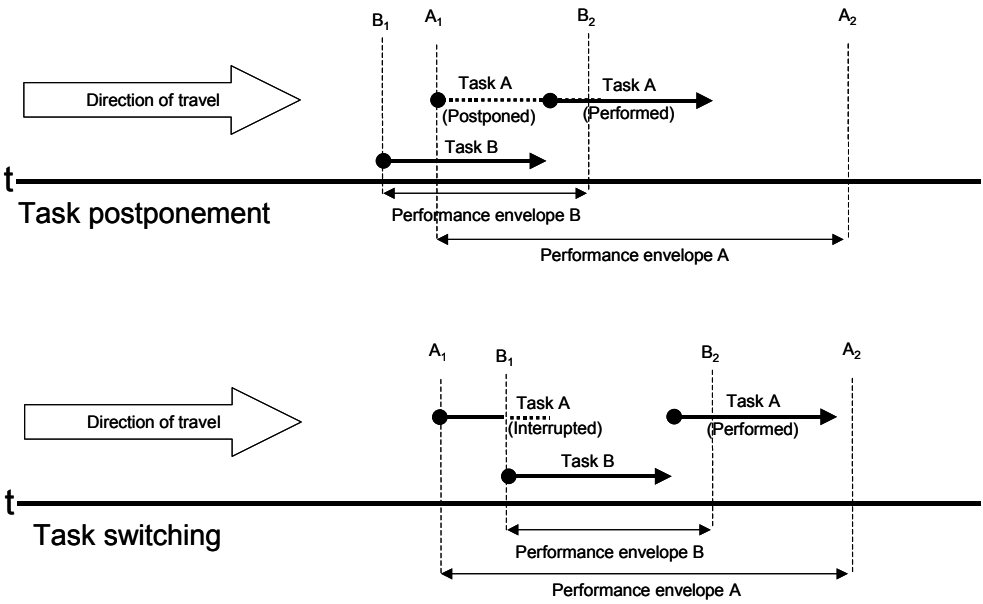


Figure 20.3 Alternative strategies for rescheduling tasks

Further quantitative modelling in Excel

A set of algorithms were written in Excel that allowed the many variations in signal sighting distance, approach speed, sign location and task performance times to be computed simultaneously, to arrive at evaluations of scenario acceptability.

Construction of the model involved selecting the scenarios to be included, calculation of where tasks would occur along the scenario timeline, and evaluation of scenario acceptability.

Selection of scenarios to be modelled

Based on the qualitative modelling, the following variables were combined to form the scenarios to be modelled:

- Location of the speed sign – positions A to H, as shown previously in Figure 20.2.
- Approach speed – between 30mph and 125mph (in 5mph increments)
- Task performance time – median, slow and modified-slow man times

A total of 480 scenarios were modelled.

Calculation of where tasks occur on the scenario timeline

For a particular scenario, the time point for sign sighting, signal sighting, and AWS sounding, are calculated. This calculation is based on the speed of the train, the locations of the infrastructure features, and the following rules:

- Signal reading may only occur once the driver has seen the signal, and must be completed before the driver passes the signal.
- Sign reading may only occur once the driver has seen the speed sign, and must be completed before the driver passes the sign.
- The AWS cancel may only occur when the AWS horn sounds, and must be completed before 3 seconds has elapsed (to avoid an emergency brake application).

For the purposes of modelling, the time that the train passes the signal was fixed at zero seconds. All calculations about time were calculated relative to this reference point. Figure 20.4 presents an example of this method. The three bars show the durations and relative timings of the sign reading task (top), AWS cancel (middle) and signal reading task (bottom).

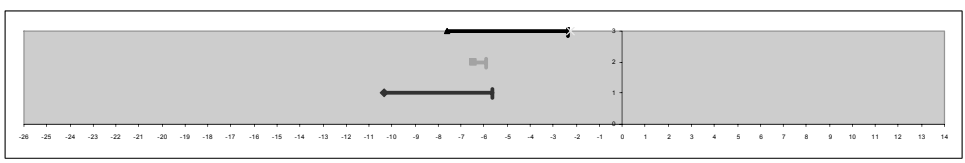


Figure 20.4 Example of a co-location scenario using the Excel modelling method

Evaluation of scenario acceptability

The Excel model was designed to identify task overlap from task times. If the model identifies no overlap of primary behaviours in a scenario, it is classified as "Acceptable" – the driver is able to perform each tasks as soon as the related infrastructure feature is detected without overlap between tasks.

If overlap is detected, the model performs an operation to assess if the driver can reschedule one of the reading tasks in order to avoid task overlap. Rescheduling is only possible if the rescheduled task (a) does not overlap with other tasks, and (b) occurs within the "window of opportunity" for the task. If the model identifies that rescheduling is possible, the scenario is classified as "Acceptable – Reschedule" – the driver is able to avoid overlap of primary behaviours if one of the tasks is rescheduled appropriately. If no such period of time is identified, then the scenario is classified as "Unacceptable" – the driver is unable to complete all the tasks without overlap of primary behaviours.

Results

The main findings of modelling are:

- Signal/sign co-location creates an unacceptable level of workload for some sign positions at particular speeds.
- The acceptability of scenarios is determined by a complex, but predictable, pattern of relationships between tasks, resulting from changes in speed, sign position, and MRT.
- The number of unacceptable scenarios is strongly influenced by the performance time estimates on which the assessment is based.
- Sign positions C and D create the largest number of unacceptable scenarios.
- Sign positions F, G and H create the fewest number of unacceptable scenarios.
- In the worse case (slow man), none of sign positions A-H are acceptable for all speeds.

- Unacceptable sign positions on approach to the signal are only eliminated (at all speeds) if the sign is located at least 310 yards before the signal (based on modified slow man performance times).

Signage positioning results

This work has developed guidance for sign positioning, indicating minimum distances that a sign should be positioned from a signal for different approach speeds. Examples of resulting combinations of speeds and distances (based on median man times) are presented in Table 20.3. Where a minimum distance for speed sign location of 0 yards after the signal is given, this indicates that it may be acceptable to locate a sign at the signal.

Table 20.3 Examples of minimum distances for speed sign location relative to signal

Maximum Linespeed (mph)	Minimum Distance from a Signal for Locating a Speed Sign (yards)	
	Before the Signal	After the Signal
30	0	0
50	250	0
80	250	0
100	250	0
115	250	0
125	250	0

The distances presented in Table 20.3 suggest that the current exclusion zone rule is over-conservative in some cases. Interestingly, positioning a sign at a signal does not create an unacceptable level of driver workload at most speeds – even though the features are physically co-located, drivers can detect and therefore read the signal well in advance of detecting and reading the sign. This demonstrates the fact that it is psychological co-location (i.e. tasks occurring near to each other) that makes a scenario unacceptable, rather than physical co-location (features being positioned near to each other).

The rules presented in Table 20.3 should generally only apply to conditions similar to those represented by the scenarios modelled (i.e. relatively straight sections of track with post-mounted signals), owing to the fact that the outputs may not be sufficiently conservative for more complex design cases. Co-location scenarios within complex infrastructure conditions (e.g. gantry mounted signs, signs with directional boards that apply to diverging routes) may pose other human factors issues in addition to workload and are likely to require individual human factors assessments.

Application of results

The research work has examined in some detail the cognitive demands that can arise from the close location of speed signs and signals. As described in the previous section, results have been

produced that define the minimum distances signs should be placed from signals at various approach speeds.

Two key findings were derived from these results:

- There is a sound basis for imposing a signage exclusion zone around signals, particularly at the higher line speeds that are being implemented on the West Coast. In this respect, the current signage rules are vindicated;
- There is considerable scope for relaxing the signage exclusion zone at lower line speeds (though the analysis only strictly applies to the simplest of plain line applications of speed signage).

Network Rail are currently recommending that the current exclusion zone rule is followed in the first instance, and that ergonomics specialists utilise the findings of this research to address more problematic scenarios. This is because:

- The current exclusion zone rule may be pessimistic but it is simple to understand and apply. The introduction of more complex rules at the current stage of the project is likely to introduce unproductive design change and delay;
- The results are based on a significant number of assumptions about driver behaviour and strictly only apply to the simplest signage design cases. To apply the data sensitively requires a detailed understanding of such limitations, which would further complicate the rules and guidance that can be issued to signalling designers and would risk inappropriate use or extrapolation.

Hence, signalling designers are required to apply the existing rules to locate signs. However, where the location rules would result in a significant decrement to the line speed or generate other forms of confusion, the cases and alternative design proposals are referred to ergonomics specialists within the West Coast project for further consideration and approval. The results of this research can then be applied by these specialists to support the risk assessment of the design proposal.

Conclusions

This research has demonstrated that it is possible to apply knowledge of human cognition in a systematic way to derive practical and empirical design rules for signalling applications. It has provided clear support for existing signage placement rules that were developed at an earlier stage of the West Coast project as 'rules of thumb' (based upon general beliefs regarding driver behaviour). This type of approach has clear potential for wider future use in developing and validating design rules that are applied to lineside signals and signage.

References

Railway Group Standard (1996). Lineside Signs, GK/RT0033. Railway Safety.
Railway Group Standard (2003). Signal Positioning and Visibility, GK/RT0037. Railway Safety.

PART VII
Signalling and Control of the Railway

This page intentionally left blank

Chapter 21

Mental Workload Assessment and the Development of the Operational Demand Evaluation Checklist (ODEC) for Signallers

Laura Pickup and John R. Wilson

Introduction

The subject of mental workload is an emotive topic that has and will continue to create much debate. How best to assess mental workload seems to produce divisions within the human factors community as the approaches to assessment reflect the variability in the interpretation of this concept. The first part of this chapter will summarise the early phases of work completed for Network Rail in reviewing the assessment of mental workload in the context of the railway signaller and developing an assessment approach. The latter part of this chapter will justify and describe the development of a checklist designed to capture the operational demand, one aspect of a signaller's mental workload. It is currently used in the field by Network Rail to provide a profile of a signaller's workstation signifying which entities within the system most impact upon the demand on the signaller.

Mental workload

The literature relating to mental workload has a bias towards presenting the concept as solely relating to theories on information processing and attention. This belief appears to ignore contributions from the cognitive engineering approaches to understanding human behaviour and the relevance of the sociotechnical system, which provides the context that work is completed in. This observation is surprising as both of these areas were suggested as relevant to the topic of mental workload in the 1970's and particularly evident within Moray's collection of chapters (Moray 1979).

This focus on theories of information processing is evident when reviewing the traditional tools used to assess workload (Pickup and Wilson 2002). These typically have been designed for use within simulators and generally do not appear suitable for use in the field or within a live system in real time (Beatty 1982; Wierwille et al. 1985; North and Riley 1989; Meshkati et al. 1990; Army. Research.Laboratory 2002). It may be as a consequence of the focus of workload assessments being required to predict the workload of future designs that typically modelling tools, the most appropriate tools for prediction, are most prominent in the literature. These tools need to consider task difficulty and the compatibility of combined tasks, which relies heavily on theories relating to information processing. It could be suggested that these types of workload assessments, familiar to human factors practitioners, have driven the interpretation of the concept.

The first phase of work for Network Rail included a review the relevant literature with the objectives of presenting the concept of workload in an understandable language to the rail industry and to consider the appropriateness of traditional workload tools for railway signallers. A plethora

of workload definitions were found within the literature but did not appear to encompass all the dimensions relating to information processing, attention, human behaviour and sociotechnical factors that were considered relevant to the mental workload of the railway signaller. Therefore a generic mental workload framework, Figure 21.1, was developed from literature findings and observations in the field of the railway signaller. The development of the framework has been previously described (Pickup et al. 2005a, Pickup et al. 2005b) and will not be elaborated on here.

The framework was intended to ensure the attributes most relevant to mental workload were mapped out, for transparency in discussing each attribute of the concept and clearly illustrating the need for multiple methods of assessment.

The mental workload framework highlighted where tools were required to provide a comprehensive assessment of the overall concept. The product of this work was the development of a 'toolkit' of methods, easily applicable in the field and capable of providing an insight into the level of mental workload and factors directly influencing it. Table 21.1 summarises the tools developed, their function and how each tool is currently used.

Further literature was consulted to consider the concept of measurement itself to explore how to measure such a complex and multidimensional concept. Nunnally (1978) suggests the term 'attribute' as critical to understanding measurement, since measurement always concerns some particular feature of objects and it is not the object that is measured but rather their attributes. This suggests that measurement requires a process that is of abstraction. The key is in understanding the nature of the associated attributes before attempting to measure them. The complexity of the

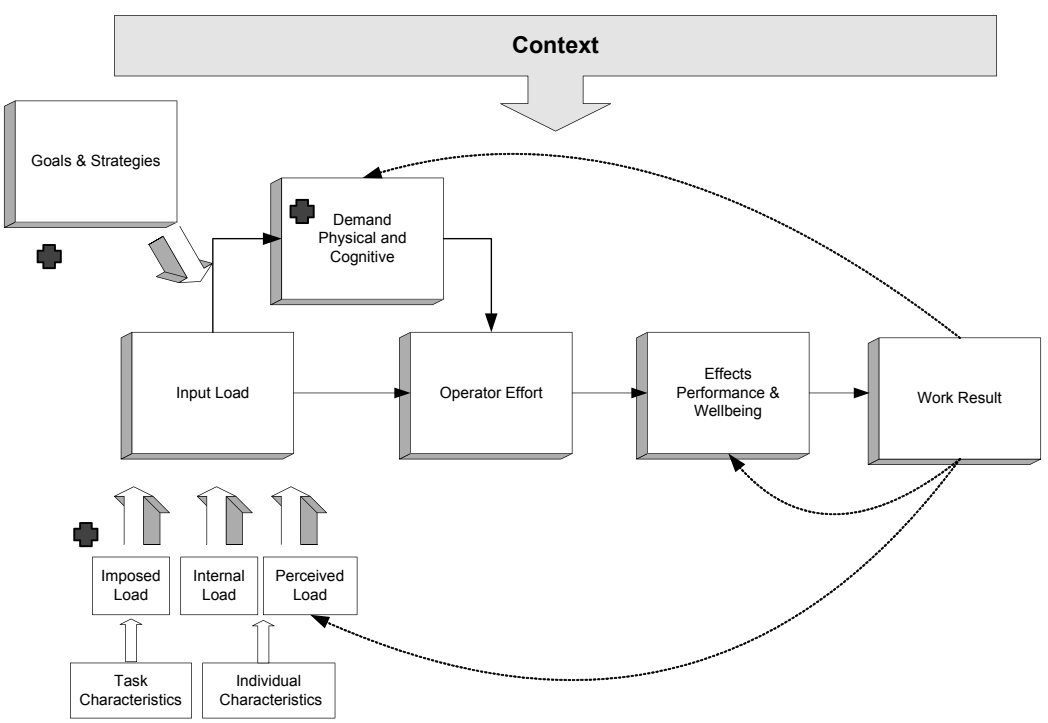


Figure 21.1 Mental workload Framework

Table 21.1 The workload toolkit for field based workload assessments of the railway signaller

Tool Name	Function	Current Use
The Integrated Workload Scale (IWS)	The IWS collects real time perceptions of signaller workload based on a nine-point scale. This tool can be used to identify peaks and troughs in the effort and demand experienced by signallers when responding to dynamically changing work conditions.	The IWS has been applied both in the field and simulated signalling environment. There are currently four options in the interface for applying the IWS tool; pen and paper, Actiwatch, Touch Screen Laptop and a Nine Point Key Pad.
Operational Demand Evaluation Checklist (ODEC)		This tool provides an analytical assessment of the degree of potential loading coming from the work system. ODEC can be used before employing any of the other types of workload tools, to understand and give context to the work of a signaller
The Workload Principles	General ergonomic principles have been applied in the context of understanding signaller workload to provide a set of specific principles that underpin a signaller being able to complete the work they are required to do.	This tool provides a high level overview of the work and workplace of a signaller to identify the presence of any mismatches that may require further investigation during a workload assessment.
Adapted Subjective Workload Assessment Technique (ASWAT)	In the absence of a dedicated railway signaller subjective workload scale, several scales were reviewed in light of the understanding of what the term workload means to signallers. The Subjective Workload Assessment Technique (SWAT) (Reid and Nygren 1988) seems to be both generic in terms of workload dimensions and includes dimensions of workload that were suggested during interviews with signallers as representing their interpretation of the term workload.	This tool can be used within a workload assessment to allow for comparison of signaller workload between two situations or at different times in the day. The application of this tool may therefore be retrospective as a signaller becomes available or as a second signaller can temporarily provide relief from signalling duties.
The Workload Probe	'The 'Workload Probe' facilitates Human Factors personnel to understand the key factors influencing the workload of a signaller. This tool considers both the cognitive and sociotechnical factors that are relevant to the signaller's work. The probe is a series of questions relating to the signaller's work and workplace	This tool is analytical and has recently been developed from a paper based format to electronic due its size and the time required to apply it. The exploratory nature of this tool appears to lend itself more for use during in depth workload investigations or research projects.

workload concept means there is no single physical property or object that can be considered as the 'actual' workload. It is more likely that a combination of characteristics or attributes e.g. complexity of a task or goals which effects the workload will be the focus for workload measurement (Hendy et al. 1993). In conclusion a multidimensional concept, such as mental workload, would struggle to be measured with one scale alone. For clarity, in the future perhaps tools developed for the assessment of this concept should be more specific in what it is they are actually capturing rather than adopting the generic heading of 'workload'.

To understand the most relevant attributes associated with the work of the railway signaller the first two phases of this workload project adopted interviews, observations and talk aloud techniques. This revealed that the signalling system and area under control (these terms are referred to as 'the system' from this point forward) of the signaller were large influences on the mental workload of the signaller.

Overall the demand experienced by the signaller during a shift is heavily influenced by the infrastructure within the area of their control. The characteristics of this area depend upon the

geography and the capacity of the infrastructure at a particular time in relation to the traffic and timetable to be controlled. Finally the technology that provides the interface between the infrastructure and the signaller is also a factor in the demand experienced by the signaller. This determines the level of intervention, vigilance or attention the signaller requires to maintain an awareness of the system's status. These characteristics of a system are unique to every signal box or control centre observed. No one location is a direct replica of another.

The need to capture these attributes in a simple form seemed essential in representing the imposed load upon the signaller that would directly influence the demand they experienced. The term 'operational demand' was adopted to describe this type of demand. The next section of this chapter justifies and describes the development stages of the tool created - the Operational Demand Evaluation Checklist (ODEC).

The development of ODEC

Subject Matter Experts (SMEs) appeared to intuitively judge a signaller's system in terms of the load it did or could impose on the signaller. On questioning them about how they made this judgement it appeared that it was related to a large number of variables or elements of the system. The characteristics of a system also seemed to reflect a shift in the signaller's goals and strategies depending on the operational demand created by these characteristics.

A tool was conceived to assess this operational demand on a signaller to provide a systematic and transparent method of capturing and evaluating the variables that could be used to represent the operational demand created by the area under control. The tool would need to be quick and easy to use, limiting the involvement of the signaller where possible so as not to be intrusive.

A research method was needed that could elicit and explore the knowledge held by rail signalling experts to appreciate the most relevant elements of the signalling system and how each element could be considered as influencing signaller workload. The Repertory Grid Technique was finally settled on as it can provide a structured methodology for extracting knowledge based on relative judgements rather than absolute judgements (Fransella and Bannister 1977). The use of relative judgements was considered favourable based on evidence from Tsang and Vidulich (1994) who suggest that relative techniques encourage retrospective deliberation based on the participant's prototypical memories of all conditions and are also they considered to reflect the overall effect on workload from imposed demands.

The Repertory Grid Technique (RGT)

The RGT is traditionally classified as a structured interview technique that attempts to identify the way in which individuals view their world and understand the importance and relevance of the things that exist within their environment (Fransella and Bannister 1977). The RGT is systematic and an efficient method for considering a large number of elements, which provides a recognised approach to the analysis of the data obtained based on individual perceptions or judgements. The aim of the RGT was to unpack the intuitive judgements that SMEs were making about the demands of a signalling system on a signaller and further explore the interpretation of the concept of workload.

The RGT requires the comparison of the 'constructs' - the attributes or properties perceived by the individual as existing about the concept being considered - to the 'elements' of the world or the context being considered (Fransella and Bannister 1977).

'Constructs', based on Kelly's theory of personal constructs need to be considered as bipolar (Kelly 1955). The theory implies that we judge our world by considering a 'construct' as having a particular attribute whilst simultaneously implying that the opposite of that attribute does not apply.

The bipolar qualities of constructs provide the reference for judging elements against. The RGT presents bipolar constructs, which are considered against each element. The question is posed, is the element more like the positive or the negative pole of the construct?

The application of the RGT was intended to reveal how signalling SMEs understood the construct of workload and which elements of the signalling system reflected the potential for high and low workload situations. Two signalling SMEs with over twenty years of signalling experience and familiar with a range of signalling technology were involved. One SME was more familiar with the South of England and the other the North, providing differing experiences of geographical areas of control.

A two-day workshop was arranged to complete the initial stages of the RGT process. The workshop was video taped to provide a full record of any discussions that resulted in defining or refining the elements, constructs or the RGT process itself.

Elements

Sixty 'elements' representing the signalling system had previously been identified. The researcher had recognised that all of these elements were relevant to the context of the signalling system but did not meet the criteria of falling within the same 'group' (Baber 1996); comparing the number of stations to the number of infrastructure failures was not appropriate as in the eyes of a signaller one represents an object the other an event making it difficult to compare between the two in terms of their relative workloads. To ensure elements were in 'alike' groups the SMEs were asked to sort the elements, written on postcards, into groups that could be considered similar. They were instructed to provide a name for each group that could explain what each set of elements represented within the system. Discussions and justifications for the inclusion of an element into a group assisted in the development of terminology capable of categorising the groups of elements. There were four main groups of elements these were; Operational Infrastructure, Indicators, Process and Service Pattern.

Constructs

For the purpose of the RGT, constructs were required that were considered to be associated with the overall notion of workload but in the context of the elements described and in a terminology understandable by the SMEs. Each group of elements was considered separately to elicit appropriate constructs.

The method for eliciting constructs had to provide a time efficient process that appeared to be natural to the SMEs. This was the 'full context with personal role feature' (Fransella and Bannister 1977). The researcher facilitated the SMEs to consider the groups of elements one by one to justify why each element should be included within the group and what impact each element had upon the work of a signaller. Fourteen constructs were elicited in total from the SMEs with considerable overlap between the groups of elements, therefore all fourteen constructs were considered as having some relevance to all four groups. One construct, effort, that the researcher felt had previously been suggested by other signallers during the exploratory phase of research and was very prominent within the literature relating to workload theories was not elicited within this exercise and was therefore added. The final set of constructs was: activity, ability to do other things, interfere, concentration, association with system, involvement, control, volume of work, load, time available, demand, level of thinking required, effort, frustration and pressure.

Each of the four groups of elements was considered separately and therefore a repertory grid was completed for each group with each SME independently. The process of ranking rather than rating was adopted, as rating an element and making an absolute judgement did not appear to make sense to the SMEs without the context of a particular scenario or event. There are several types of analysis that are possible. The aims of using the RGT were to a) understand which constructs were the most meaningful to SMEs in describing the overall construct of workload and b) extract which elements were most able to reflect the presence or absence of these constructs when judging a signalling system. These aims focused the analysis of the data. The names of the construct and elements analysed at any one time were identified only by letters and numbers throughout the whole process so that the researcher was blind to the identity of each construct and relevant elements until the analysis was complete, avoiding any potential bias.

The first stage of analysis was completed to understand which constructs are most related to one another, and could therefore be considered as collectively describing the SMEs' understanding of the construct of workload. This was completed through a process of rank order correlations and a simple form of cluster analysis (Fransella and Bannister 1977), from which a visual graphical representation of the relationships between constructs is provided.

As this was completed for each SME the outcome differed slightly. Therefore, the constructs that clustered most closely and were common to both SMEs were identified. These were considered by the researcher to be the constructs most capable of representing the overall construct of workload for the group of elements being considered. Table 21.2 illustrates the constructs calculated to be most relevant for each group of elements.

Table 21.2 Constructs most associated to each group of elements

Operational Infrastructure	Indicators	Process	Service pattern
Volume of work	Demand	Frustration	Demand
Involvement	Load	Do other things	Concentration
Effort	Pressure	Concentration	Involvement
Concentration	Concentration	Interfere	Effort
Load	Effort	Time available	
Demand	Do other things		
Requires thinking	Time available		

Whilst three constructs (demand, effort and concentration) appear most consistency across the groups of elements that represent the signalling system there are also considerable differences. This is of interest when considering the historic attempts to provide a single definition for the term workload across work domains, industry and contexts. These findings could suggest that although some key terms are transferable others may exist that are intuitively more relevant to certain contexts even within a single work domain.

The next aim was to understand which elements were most likely to reflect the constructs that had been identified as representing signaller workload. Fransella and Bannister (1977) suggest determining which elements are most associated with each construct is achieved by quantifying the distance between an element and a specific construct. In summary, the analysis involves a

score calculated for each element, this represents the sum of the elements position on the construct when compared to all other constructs. This allows a weighting to be calculated that reflects the relationship of each element to the construct being considered. This method attempts to assess the potential for each element to produce the construct considered. Referring to Fransella and Bannister (1977) for a greater understanding of the analysis process is strongly recommended. Analysis was completed for all of the constructs that had been identified as being most relevant to the concept of workload with each group of elements. The elements were placed in order of which element was most representative of the positive end of the construct spectrum through to those most relevant to the negative end of the spectrum.

The elements that appeared consistently at the top half of the construct tables suggested these were more likely to produce the 'positive' characteristic, higher workload. Due to the nature of constructs and their bipolar qualities it could be considered that those at the bottom half of the tables were more associated to the 'negative' characteristic of the construct, low workload. These data and interpretations were reported to the SMEs to assist in verifying the output from the repertory grids, and the suggestion was that although constructs in the lower half of the table imposed a lower load and subsequent demand on the signaller it was the absence of the elements in the upper part of the constructs table combined with the presence of those in the lower half of the table that would determine situations of low workload.

The next stage was to transfer the information obtained through the RGT process into a tool that could systematically record data on each element of any signalling system. The output of the tool was to provide a profile of a signalling system in terms of having a greater proportion of elements that influence a high or low workload. This would allow documentation (and transparency) of the extent to which the signalling system contributes to the workload of the signaller. Each element was to be scored numerically.

Each element considered within the repertory grid was presented in a checklist format with the elements that had been identified as most likely to create the positive constructs associated with workload highlighted - 14 out of a total of 34 elements. Rather than provide a predetermined single numerical value that could suggested a 'red line' value for each element within a workstation (the parameters for an 'ideal' signalling workstation are not known to the rail industry) categories were created of 'high', 'medium' and 'low' to reflect the values currently existing within functioning signalling systems throughout the country. These categories were based on data collected from over 250 signal boxes. The spread of data for each element was divided to produce three sets of data ranges; as 'high', 'medium' and 'low' categories. Where there were no data for a number of elements, SME judgement was applied to establish ranges of data, which were then tested and adjusted as the tool was applied in the field.

The tool is presented in a Microsoft 'excel' spreadsheet, which facilitated the calculation of the percentage of 'high', 'medium', or 'low' categories that were recorded for each signalling workstation. Elements had been identified from the RGT as more likely to increase the signaller's workload were weighted in the calculation of the final percentages, by doubling the score for each highlighted element. This weighted the distribution of the final percentage of elements within each category. The output from the tool provided a breakdown of percentages for 'high', 'medium' and 'low' categories. The tool aimed to assist in highlighting why and how the signalling system could be interpreted as influencing the load imposed on the signaller and thus contributing to their overall workload. The information provided is intended to further support the human factors experts' understanding and final judgement made towards the overall workload profile presented for a workstation. Both the client and independent ergonomics consultants applied the tool in over twenty signalling centres. As a result of these trials the addition of a non applicable (N/A) column was considered necessary as not all signalling systems appeared to present with all elements, and a low score misrepresented the system.

The tool has been applied on every workload assessment completed by Network Rail over a six-month period, without interference from the lead researcher. Following this the ODEC tool and the empirical data collected were reviewed. Several issues had arisen during this time that were documented and will direct the next phase of the ODEC tool development. These include: clearer definitions of what each element was within the tool itself, and adjustment of the 'high', 'medium' and 'low' categories for all elements to improve the validity of the tool in judging a workstation against a wider spread of data. In view of the large number and variety of signalling systems the development of ODEC has been and continues to be one of iteration.

Summary

This chapter has presented the toolkit currently used by Network Rail to complete field based workload assessments (Table 1). The focus of the chapter has been to present the development of one of these tools. ODEC provides a systematic process to understand and reflect the demands placed by the signalling system on the signaller and is just one piece of evidence that contributes to providing an overall workload profile. The development of this tool has considered the theory behind measurement itself and adopted a transparent method, the repertory grid technique, to sort and understand the relevance of each aspect of the system to the signaller's work and overall workload.

The validity of this tool cannot be demonstrated in the traditional approach of comparing to another previously validated tool as there appears to be no workload tool of this kind and certainly none specific to the railway signaller. Therefore the validity of this tool is proposed through the continual iteration of the tool following field use, workshops with subject matter experts and the integration of empirical data collected from the application of the tool. This process is ongoing, as the tool has been applied more frequently in some types of signalling systems than others.

Acknowledgements

The work reported here was funded by Network Rail. The contribution of Ged Morrisroe, Bob Muffett and John Robinson, Theresa Clarke and Emma Lowe, to development, testing and refinement of ODEC, has been invaluable.

References

Army Research.Laboratory (2002). *IMPRINT Improved Performance Research Integratio*. 2003. Army Research Laboratory, Human Research and Engineering Directorate

Baber, C. (1996). *Repertory grid theory and its application to product evaluation.* In: Jordan, P., Thomas, B. and Weerdmeester, B. (eds) *Usability Evaluation in Industry*. Taylor & Francis, London, pp 157-165

Beatty, J. (1982). *Task-Evoked Pupillary Responses, Processing Load, and the structure of Processing Resources*. Psychological Bulletin 9:276-292

Fransella, F., Bannister, D. (1977) *A Manual for Repertory Grid Technique.* Academic Press Limited, London

Hendy, K.C., Hamilton, K.M. and Landry, L.N. (1993). *Measuring Subjective Workload: When is one Scale Better than many?* Human Factors 35:579-601

Kelly, G.A. (1955). *The Psychology of Personal Constructs*. Vol. 1 and 2. Norton, New York

Meshkati, N., Hancock, P.A., Rahimi, M., and Dawes, S.M. (1990). *Techniques in mental workload assessment* In J.R. Wilson and E.N. Corlett (eds) Evaluation of Human Work: A Practical Ergonomics Methodology. Taylor & Francis, London, pp 749-782

Moray, N. (1979). *Mental Workload Its Theory and Measurement*. Plenum Press, New York

North, R.A., Riley, V.A. (1989). *W/INDEX: A Predictive Model of Operator Workload.* In: McMillan GR, Beevis D, Salas E, Strub MH, Sutton R, Van Breda L (eds) Applications of Human Performance Models to System Design. Plenum Press, New York, pp 81- 89

Nunnally, J.C. (1978). Psychometric Theory. McGraw-Hill, USA

Pickup, L, Wilson, J.R. (2002). *Workload assessment tools.* Institute for Occupational Ergonomics. Centre for Rail Human Factors. Nottingham University, Nottingham, pp 1-33

Pickup, L., Wilson, J. R., Nichols, S., Norris, B., Clarke, T., and Young, M.S. (2005a) *Fundamental Examination of Mental Workload in the Rail Industry*. Theoretical Issues in Ergonomics Science Vol 6,6.

Pickup, L., Wilson, J.R., Sharples, S., Smith, S. (2005b). *A Conceptual Framework of Mental Workload and the Development of a Self Reporting Integrated Workload Scale for Railway Signallers*. In: Wilson JR, Norris BJ (eds) Rail Human Factors. Ashgate, London

Reid, G.B., Nygren, T.E. (1988). *The subjective workload assessment technique: a scaling procedure for measuring mental workload.* In: Hancock, P.A., Meshkati, N. (eds) Human Mental Workload. Elsevier Science Publishers B.V., Amsterdam, pp 185-218.

Tsang, P.S., Vidulich, M.A. (1994). *The Roles of Immediacy and Redundancy in Relative Subjective Workload Assessment*. Human Factors 36:503-513

Wierwille, W.W., Rahimi, M., Casali, J. (1985). *Evaluation of 16 measures of mental workload using a simulated flight taks emphasising meditational activity.* Human Factors 27:489-502

This page intentionally left blank

Chapter 22

Supervision in Signal Boxes

Stuart Dickinson and Emma Lowe

Introduction

Recommendation 43 from Lord Cullen's inquiry into the Ladbroke Grove rail crash in 1999 was that, '*A supervisor should be employed on a continual basis to ensure that the workstations are operated in the most effective way*' (Cullen, 2001).

As a consequence of this Railtrack (now Network Rail) was required to review the supervisory arrangements for all UK signal boxes and to agree a clear policy regarding when and where supervisors should be employed. To that end a method was developed to assess the requirements for supervision in signal boxes and to identify those signal boxes that would most benefit from supervision.

Developing the assessment method

Figure 22.1 presents an overview of the approach used to develop the assessment method.

Literature review

The literature review revealed that supervision of safety critical work is seen as a key factor in reducing accidents and unsafe acts (Allan, 1989; Kraus et al., 1997); it has been suggested that 50% of unsafe acts could be eliminated by the use of effective supervisors (Weber, 1992).

However, there has been little published relating to what supervisors actually do, although it is suggested that in some cases supervision involves a supervisor working in parallel with the worker in a different format to ensure they are doing it correctly (Schmid and Collis, 1999). Large lists of the responsibilities and activities that supervisors have to manage are available. Weber (1992) concluded from these lists that supervisors may have around 74 responsibilities of which only one was to supervise people at work (for example, other responsibilities included providing training, checking tools and investigating accidents).

It is suggested that the role of some rail supervisors has changed due to the introduction of various technological systems, such as TOPS (total operations transport system). As a consequence, it is suggested that supervisors have changed their role from that of traditional directors of operations to 'trouble shooters' who deal with unexpected situations. Technology has not pushed out supervisors, but has resulted in a change in role to be more focussed on information management (Dawson, 1987).

The literature cites a number of factors that contribute to effective supervision in order to ensure safe behaviours. These factors include the following:

- *Communication:* The factor most often found within the literature as being important in supervision is that of effective communication (Lark, 1991; Martens, 1998; Mattila et al.,

1994; Seppala, 1998; Simard et al., 1994). Increased levels of communication, even on non-work related topics, correlate positively with safe working behaviours and generally effective supervision. Key dimensions associated with effectiveness in preventing accidents include the supervisor's involvement in safety matters and the degree of worker involvement they can generate. Hence motivating workers through good communication is also critical.

- *Knowledge of hazards:* In order to be effective, supervisors need a good knowledge of health and safety regulations (Johnson, 1997; Mattila et al., 1994)and the ability to identify and reduce hazards (Simard et al., 1994; Strahlendorf, 1996). In the construction industry, safer sites are those where supervisors ensure a high level of general housekeeping.
- *Safety programs:* It is important that supervision exists within the context of an effective safety management system and an effective safety program (Lark, 1991).
- *Observation and feedback:* Observation should be consistent, though this consistency does not necessarily mean that the supervision should be more time consuming.
- *Accountability:* Holding supervisors accountable for their own performance, via feedback from their subordinates, is also seen as a key factor in ensuring effective supervision.

In addition, it is suggested that having a detailed knowledge of the tasks in question is an important factor in effective supervision (Martens, 1998).

Supervisors often find it difficult to strike an appropriate balance between the pressures of maintaining high levels of production and ensuring safe working practices (ERM, 2002) although effective supervision can result in higher levels of productivity (Weber, 1992). Organisational pressures may be addressed by changing the structure of safety supervision, eg, the use of teams that increase worker participation and increased supervisor commitment to safety issues (Martens, 1998).

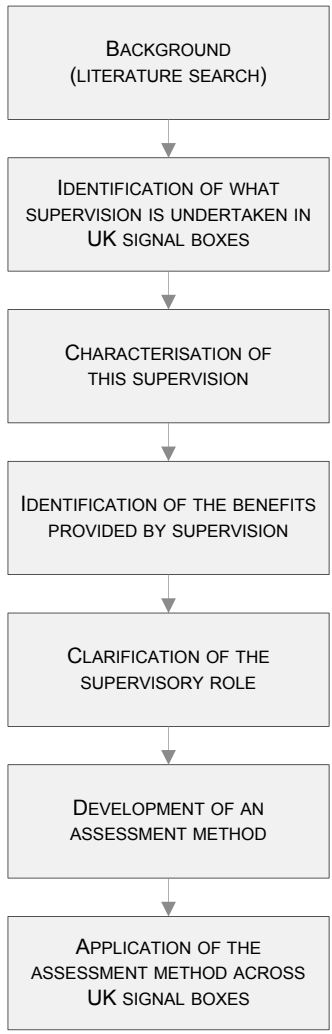


Figure 22.1 Overview of the approach

The literature review provided a useful background to the work that needed to be done to develop an assessment method and highlighted that supervision may encompass a wide range of activities but rarely what is thought of as traditional supervision, ie, close observation and direction.

Understanding the supervisors role

It is not easy to judge whether a signal box will benefit from supervision without a good understanding of the supervisory role. Understanding a box supervisor's role helps illustrate the potential benefits and reduces the likelihood of a mismatch between the expected and actual benefits of a supervisor. Specifically, the aim of this stage of the work was to understand w*hat kind of supervision occurs, which tasks are supervised, and the rationale used to decide which tasks are supervised.*

There was an initial concern that there may be a discrepancy between the expectations associated with a supervisor's role and their actual role. For example, other studies have shown that, '*There is currently very little direct supervision, ie, where the supervisor observes the activity being carried out, checks that the activity has been carried out correctly, and ensures that any necessary corrective action is taken*' (ERM, 2002). That is, the traditional concept of direct supervision is not applicable to the majority of the rail industry. Therefore, adopting supervision in the belief that a supervisor will perform such direct supervision and reduce signaller error may be unrealistic (ie, the intended safety benefit may not be realised in practice).

Given the variability in the signalling equipment, geographical distribution and the different management approaches adopted across UK signal boxes, it is understandable that Railtrack lacked a definitive view of the supervision currently undertaken in signal boxes. In order to gain an improved understanding of the supervision role in signal boxes, a number of signal boxes were visited which operated with supervisors. Interviews were conducted with both supervisors and signalling managers. Interviews with the signalling managers were less detailed but designed to understand their perceptions of the supervision role and to understand how their role interfaced with the supervisory role. The supervisor's were more detailed and included asking them to divide a horizontal bar into three sections, each representing the proportion that the supervisory role is made up of 'management', 'operational' and 'miscellaneous activities'. This was then used as a basis for discussion. In addition to the interviews and box visits, a number of job descriptions were reviewed and task analyses developed. Developing the task analyses helped distil a clear picture of the roles undertaken at each signal box.

The findings from this stage of the work were that there are a wide variety of activities encompassed within the 'box supervisor' role. This means that there are different supervisory roles, depending on the types of activities the supervisor undertook. The difference between supervisors is further illustrated by the variety of job titles, including team leaders, duty shift managers (DSMs), shift manager, and power signal box (PSB) manager that are given to a supervisor in a signal box. A summary of the activities undertaken is provided in Table 22.1 in the form of a checklist of activities.

The perceived proportion of time the supervisory roles involve 'management', 'operational' and 'miscellaneous activities' undertaken by staff at the various boxes is presented in Table 22.2.

The kind of supervision that is carried out in signal boxes is not direct supervision. The supervisor does not spend all their time watching the signaller's actions with the intention of intervening before an error is committed as this would be impractical given the ratio of signallers to supervisors. As one supervisor put it, '*essentially once a signaller has pressed a button to clear a signal there is little that can be done to rectify a regulation error*'. Instead, different kinds of supervision have been identified.

- *Maintaining an overview.* Supervisors will attempt to maintain an understanding of what is going on throughout their area and in some cases beyond their area. The supervisor will intermittently look up at the panel while completing other tasks in what could be described as an unstructured sampling method. The regularity of the sampling is dictated by the amount of information that is required to understand what is going on in their area and the demands of other activities.
- *Focused attention/supervision.* Certain circumstances will attract the supervisor's attention and result in more focussed attention on one panel compared to the others. This may be prompted by something the supervisor has noticed while maintaining an overview, through hearing alarms, noticing urgency in the signaller's voices, a communication from a third party directing their attention to a specific area, or the signaller approaching the supervisor with a request for assistance. Dealing with issues as they arise is a key part of the supervisors

Table 22.1 Checklist of roles and activities the supervisor may perform in each signal box visited

Activity/Role	Box 1	Box 2	Box 3	Box 4a (Shift Manager)	Box 4b (Team Leader)	Box 5	Box 6
Assessing competence of others (competence assessment procedure)	✓	✓	✓	✓	×	×	✓
Observing signallers actions	✓	✓	✓	✓	✓	✓	✓
Observing the track mimic or overview	✓	✓	✓	✓	✓	✓	✓
Monitoring communications (real time)	✓	×	✓	✓	✓	✓	✓
Monitoring communications (recorded)	×	✓	×	✓	×	×	✓
Briefing signallers	×	✓	×	×	×	×	✓
Providing guidance/advice to signallers during operations	✓	✓	✓	✓	✓	✓	✓
Checking delay attribution	✓	×	✓	✓	✓	×	✓
Filling in operational forms (on behalf of other signallers)	×	×	✓	✓	✓	✓	×
Filling in other forms	✓	✓	✓	✓	×	✓	×
Filling in other forms (on behalf of other signallers)	×	✓	×	✓	×	×	×
Responding to signaller's requests for assistance	✓	✓	✓	✓	✓	✓	✓
Answering external calls	✓	✓	✓	✓	✓	✓	✓
Answering operational calls (ie, from drivers etc)	✓	×	✓	✓	✓	×	×
Managing emails/office PC duties, etc.	×	✓	✓	✓	✓	✓	✓
Correcting signaller errors	✓	×	✓	✓	✓	✓	✓
Performance monitoring	✓	×	✓	×	×	✓	✓
Meal break cover	×	×	×	×	✓	×	×
Checking paperwork	✓	✓	✓	✓	✓	×	×
Out of hours rostering	×	✓	✓	✓	✓	✓	✓
Buildings management	×	✓	✓	×	×	✓	✓
Running training	×	✓	×	×	×	×	×
Providing/managing relevant notices/paperwork	✓	✓	✓	✓	✓	✓	✓

Table 22.2 Perception of the supervisory role

Signal box	Job title	Management	Operational	Miscellaneous
Box 1	DSM	20%	70%	10%
Box 2	Team Leader	60%	10%	30%
Box 3	DSM	20%	75%	5%
Box 4a	Shift Manager	10%	50%	40%
Box 4b	Team Leader	5%	60%	35%
Box 5	PSB Manager	10%	88%	2%
Box 6	DSM	40%	60%	0%

role and has been likened to the fire service – '*when the bell goes, the work starts'*.

- *Supervision of signaller's welfare* Generally, supervisors are not involved with additional monitoring and support of signallers, however, their close working relationship and knowledge of the staff means that they are likely to be aware of changes in character/mood that indicate that a signaller may require additional support. In the event of an incident, supervisors have welfare responsibilities and may need to relieve them from the panel and advise signallers of their rights.
- *Supervision of signaller's competence* Arguably, the competency assessment process provides a form of monitoring and supervision. Supervisors play a varying role in the management of signaller competence. All are able to sign off elements that they witness a signaller completing, but some play a more proactive role in completing the competency assessment workbook. Where supervisors are heavily involved in the process it is likely to take up significant time and effort.
- *Supporting activities* Supervisors also play an important role in assisting signallers during incidents and perturbations. Supervisors may make decisions for the signaller, provide guidance/advice, relieve the signaller's workload (eg, form filling, providing an extra pair of eyes) and manage communications from third parties. This supporting role is not traditional 'supervision' but represents a crucial part of the supervisory role.
- *Figure of authority* In some circumstances the supervisor has a role to play as a figure of authority to aid rapid decision making and become something of a figurehead during the management of an incident. In some cases having a managerial level person in the signal box can have advantages when dealing with similar level personnel from other organisations.

Supervisors characterised what they do in various ways, such as:

- responsible for the efficient movement of trains through the area
- the real time managers of the box
- like the fire service, as, when the bell goes, the work starts
- there to ensure the smooth running of the … Signal Box and Station
- overseeing the smooth running of the box
- that of an orchestra conductor… providing guidance for the different areas of the signalling scheme taking into account the bigger picture.

In order to describe the supervisory role in a more objective way that would highlight the safety benefits of the supervisory role, a theoretical model was developed. This model identifies a core set of roles/responsibilities that *all* supervisors undertake. It also proposes two subsets of roles/responsibilities that a supervisor *may* undertake. The use of two subsets reflects the notion that some supervisors are more operationally oriented, while others are more management oriented. The supervisory model included the following components:

The *core* supervisory activities that all of the supervisors perform included:

- Maintaining an awareness of the train running in the area. The level of awareness will depend on the other tasks that the supervisor is required to carry out, and how busy the box is (peak or off peak)
- Assisting the signaller in the event of an incident or failure by providing advice and guidance
- Maintaining an overview perspective and ensuring that the impact of incidents and decisions is considered and minimised where appropriate.

The *managerial* subset included:

- Organising, managing and taking responsibility for the competency assessment process for signallers
- Managing the welfare of signallers, both day-to-day and in the aftermath of an incident/near miss
- Managing the building in terms of its state of repair, visitor access, etc
- Managing the performance of signallers, and the signal box (including delay attribution)
- Managing out of hours rostering

The *operational* subset included:

- Managing the majority of communications for the signaller (except CSR calls and box to box calls)
- Relieving signallers for meal breaks
- Providing guidance on a routine basis, ie, being involved in moment to moment regulation decisions
- Assisting signallers by relieving their workload (filling forms on their behalf, monitoring part of the panel while they deal with a specific issue).

This model proposes a theoretical difference between an operations biased supervisor and a management biased supervisor. It should be noted that this is a theoretical model and supervisors often perform a role that is a blend of these roles.

There appears to be a relationship between the proportion of management activity perceived in the role and the level of involvement in the continuous assessment process. Therefore, a key indicator as to whether a supervisor is a more managerial biased supervisor is whether they are heavily involved in the competency assessment system. This is supported by the view expressed by some signalling managers that a certain 'distance' is desirable between the signaller and their assessor during the competency assessment process.

This concept of two models of a supervisor is supported by the roles performed by the shift manager and the team leader in one of the boxes. The shift manager performs the managerial roles and holds the position of authority, whilst the team leader performs a more operational/signaller support role.

The different types of supervisory role (ie, operational or managerial) will have different benefits. Deciding which of these roles is suitable for a given signal box will depend on the benefits that are most desirable.

Benefits of supervision

Performing the *core* supervisory role is expected to provide the following benefits:

- *Fully informed decision making:* The supervisor provides a focal point during the response to an incident for both staff within the signal box and third parties. Therefore, information will be directed to one person and decisions are more likely to be based on *all* the information available.
- *Reduced number of slips:* Slips are more likely when a signaller's workload is too high. The supervisor's assistance can reduce workload and therefore reduce the likelihood of slips occurring.
- *Catching all available information*: Narrowing of attention occurs when workload is high and a person focuses on certain information at the expense of other information. High workload

and involvement in a problem make this more likely. The supervisor is less involved and therefore is less likely to suffer from attentional narrowing when providing guidance.

- *Reduced signaller stress:* Relieving the signaller's stress by virtue of providing assistance and support; ensuring the signaller does not feel isolated when dealing with incidents and problems.
- *24 hour presence*: The supervisory support role is present in the box 24 hours a day and therefore does not require a detailed briefing to ensure the supervisor is up-to-date with the situation in the event of an incident.
- *More organised/flexible response to manning issues:* Supervisors often manage out of hours rostering, arranging cover when necessary in the signal box. A supervisor can cover panels/workstations in the short term and arrange cover for the rest of the shift. This more flexible approach means that the work can be carried out more efficiently and more safely.
- *Taking account of the train service as a whole:* In maintaining an overview perspective the supervisor has the ability to:
  - provide guidance that takes into account the other panels/areas
  - provide proactive guidance for signallers regarding the effect of incidents in other areas upon their own
  - predict potential problems (eg, a number of late running trains) and provide guidance on this basis.

The *managerial* subset of activities provides the following benefits:

- *Figure of authority:* In the event of disagreements the supervisor can override the signallers and take the final decision, if required. In the event of a time critical decision or a decision that may have consequence in terms of delays the position of authority may facilitate more effective action being taken. The authority is also useful when dealing with third parties, having the authority to make decisions that may affect other parties (eg, TOCs). Actions such as relieving a signaller following an incident, disciplining signallers and providing objective assessment may be better performed by a supervisor with such authority.
- *Reduced workload for signalling managers:* If the supervisor carries out more of the competency assessment, managing for performance and attendance, and attending to the welfare of the signallers, the signalling manager does not have to perform these duties. This relieves some of the signalling manager's workload, which leaves them free to address other issues relating to the safe and efficient running of the signal box.

The *operational* subset of activities provides the following benefits:

- *Reduced workload under normal conditions*: By performing tasks on behalf of the signaller, the supervisor can further reduce the signaller's workload. This may be in the form of managing communications (where a signal box receives a large number of communications, other than drivers and box-to-box communications), filling in forms on behalf of the signaller, operating the panel/area on behalf of the signaller (under their direction).
- *Increased flexibility:* If a signaller is absent, the supervisor can step in and assist the signalling of trains. The way that the cover provided is more flexible if the supervisor is willing to operate as a signaller.

Refocusing of the supervisory role

In order to achieve consensus about what role the supervisor should fulfil and what degree of variability was appropriate a RACI analysis was undertaken involving operational representatives from across the company. The aim of a RACI is to identify, for relevant activities, who is Responsible, who is ultimately Accountable, who must be Consulted or Informed. The power of the RACI technique is primarily in the discussions that are stimulated between the attendees during the RACI workshop and their consequent appreciation of each other's points of view.

A RACI analysis can be completed for an existing ('*as is*') or for a proposed situation, or a hypothetical ideal situation, ie, a '*could be*'. In this case a '*could be'* RACI was used to identify the responsibilities, accountabilities, etc associated with a position intended to encourage the 'professional management of signallers' – the supervisory role. The output of the session was the development two job descriptions, one for a supervisory role (ie, the core activities with a blend of the management and operational supervisory and management characteristics), and another for a signallers assistant (ie, an operational assistant for the signaller).

An assessment approach

It was important to provide a defensible argument as to why supervisory arrangements are the way they are in all signal boxes. However, in practice an in depth assessment is not needed for every box and so a two-stage assessment method was adopted to identify those signal boxes that required supervision.

Stage 1: High level screening: It was felt that the decision not to introduce a 'supervisor' could often be made easily (eg, single manned or single panel boxes), however, defined criteria against which to make these decisions was required. The aim at this stage was to identify signal boxes that do not require further investigation.

Stage 2: In-depth assessment: This was a more in-depth assessment of those signal boxes not previously screened out, with the aim deciding whether or not a 'supervisor' should be introduced, and to provide evidence to support the decision.

The high level screening involved comparison of signal boxes against a set of predefined criteria and a review of all signal boxes by an expert panel.

Although it was generally considered that a single manned signal box did not warrant a supervisory position, there may have been one or two cases where the particular circumstances are considered to make an assistant role justifiable. Use of an expert panel review ensured that those boxes that had extenuating circumstances would be addressed, even if they did not fulfil the criteria previously identified.

The in-depth assessment method was based on the benefits associated with supervisory roles and industry requirements. The steps that were used to complete the assessment are described below; an example of the proforma is also presented in Figure 22.2.

The following steps were carried out when using the proforma to assess a signal box:

1. Consider each of the benefits that may be provided by the introduction of a new role (Column No. 2)
2. For each of the benefits, identify any mechanisms that are currently in place which provide the benefit instead of a 'supervisor' (Column No. 3). Mechanisms include, but are not limited to:

	Column No. 2	Column No. 3	Column No. 4	Column No. 5
No.	Potential benefits of role	Current arrangements that provide the aspect described in column No. 2.	Are you satisfied that all the aspects described in column No. 2 are covered, if not detail mitigations	Need to introduce supervisory role (High/Med/Low)

Figure 22.2 Structure of the signal box assessment proforma

- signallers resident in the box full time to provide support
- signallers being present to provide support when increased workload is likely (eg, planned engineering works, etc)
- local procedures that direct communications away from the signal box
- having a very clear working relationship between signallers and TOCs (including the ability to refuse TOCs requests).

3. From columns 2 and 3, make a judgement as to whether, on balance, the envisaged benefits of the new role are satisfactorily achieved by existing means. If the potential benefits are not provided then detail the steps that are required to provide these benefits. Column 4 effectively provides the argument and justification as to why the introduction of a supervisor is/is not required in the signal box.
4. Consider the benefits of introducing a supervisor and whether they would provide high, medium or low benefit (Column No. 5), where:

   - *High*: gaining the identified benefit would result in *significant* improvement in the safety and performance of signalling operations (over and above those delivered by mechanisms already in place)
   - *Medium*: gaining the identified benefit would result in a modest improvement in the safety and performance of signalling operations (over and above those delivered by mechanisms already in place)
   - *Low*: gaining the identified benefit would result in little improvement in the safety and performance of signalling operations (over and above those delivered by mechanisms already in place).

5. Summarise the number of High, Medium and Low ratings in the table provided.

This process is designed to allow the decision panel to identify the benefits of introducing a supervisory role into that signal box. Importantly it also identified areas in which the signal box might seek to improve without the introduction of a supervisory role.

A judgement was required as to whether the person/team completing the form were satisfied that all the potential benefits are provided by other means. When developing the argument as to

why a benefit is not necessary in the signal box the following issues needed to be considered. It must be noted that this list is not intended to be complete or exhaustive in any way.

- Signaller's workload: The workload that signallers face in the box was a key consideration in the safety argument. If a signaller only deals with a small number of trains per hour then the introduction of a supervisory role is unlikely to be an efficient use of resources.
- *Signalling manager's workload: The supervisory role was to be focused on the 'professional management of signallers' and as such should relieve some of the workload from the signalling manager. If the signalling manager is responsible for the competency assessment of a large number of signallers, then this may be part of the justification for introducing a supervisory role as the ratio of signallers to assessors would be reduced.*
- Other mechanisms: The mechanisms currently in place needed to be taken into account; alternative means of providing the same benefit are acceptable, and may be more suitable in some circumstances.

The method was clearly focussed on what introducing a supervisory role would mean for that signal box, but also recognised the flexibility of working practices and the differences between signal boxes. The signal box assessment approach was taken on and used by Railtrack to assess UK signal boxes.

Study outcomes

The process of completing the signal box assessment proforma highlighted that, of the locations assessed, none actually needed supervisors. However, the method did identify that some boxes would benefit from additional support in the form of a co-ordinator or 'back desk' role that also provided meal relief.

Follow-on work also included the development of the two job descriptions, for a signaller co-ordinator whose purpose is to *operate signalling equipment to ensure that trains are signalled and regulated in accordance with the rules, regulations, instructions and timetable currently in force, to ensure the highest possible standards of safety and train performance* and a duty signalling manager whose purpose is to:

- *ensure that the network is controlled to the highest level of safety and performance by signalling staff, and that infrastructure control systems are operated to work specification.*
- *manage the people involved in signalling and performance reporting within an area of control.*
- *co-ordinate the response to incidents and accidents and ensuring the safe and efficient operation of the signalling centre.*

An additional output of the study was to provide input to the Railtrack policy on supervision in signal boxes.

Conclusions

The supervision in signal boxes assessment method provided a robust and defensible basis for deciding whether or not to provide supervision in signal boxes – unnecessary expenditure

was avoided while maintaining/ improving safety. Furthermore, it was possible to identify methods to improve the working arrangements that did not necessarily require a supervisor to be introduced.

Acknowledgements

ERM wishes to thank the various Signalling Managers and supervisors that provided input to the study.

References

Allan, T. (1989). *People skills: The new tools for supervisors*. Occupational health and safety Canada. Jul/Aug 1989, vol. 5, (4), 46-52.

Dawson, P. (1987). *Computer technology and the job of the first line supervisor*. New Technology, Work and Employment, 2(1), pp 47 - 60.

ERM January 2002 *Supervision of Safety Critical Work: Summary and Recommendations* (Research project completed by ERM for Railway Safety. Ref: 7114).

Johnson, C. (1997). *Beyond belief: Representing Knowledge requirements for the operation of safety critical interfaces*. Human-computer-interaction: INTERACT '97. Edited by S. Howard, J. Hammond and G. Linguard. Chapman & Hall: London.

Kraus, D., Saboda, R., Gramopadhye, A. & Chandler, T. (1997). *Supervisory task analysis: Aircraft maintenance environment.* Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Orlando, FL, Oct. 12-15, 1997, Vol. 2. 1878-1883.

Lark, James (1991). *Leadership in safety*. Professional Safety, 36 (3), 33 - 35.

Linton, S. (1991). *A behavioural workshop for training immediate supervisors: The key to neck and back injuries?* Perceptual and motor skills, 73, 1159 - 1170

Lord Cullen (2001). *Ladbroke Grove Rail Inquiry Part 1 Report* - Health and Safety Commission's Action plan to implement the recommendations, (paragraph 12.17).

Martens, N. (1998). Competence *and skills of the planning supervisor*. Professional Safety. Sep. 1998, vol 45, (9), 30-35.

Mattila, M., Hyttinien, M. & Rantanen, E. (1994). *Effective supervisory behaviour and safety at the building site*. International journal of industrial ergonomics, 13, 85 - 93.

Schmid, F & Collis, L. (1999). *Case studies on human centred design for railways*. In People in Control. Institution of Electrical Engineers, London, pp 44 - 50.

Seppala, A. (1998). *Participative management models and the role of groups and supervision.* In Safety Management: The Challenge of Change, Edited by A. Hale and M. Baram. Pergamon, Oxford, pp 123 - 131.

Simard, Marcel & Marchand, A. (1994). *The behaviour of first line supervisors in accident prevention and effectiveness in occupational safety*. Safety Science, 17, 169 - 185.

Strahlendorf, P. (1996). *What supervisors need to know. Once you have determined that you are a supervisor, what must you do to keep your workers healthy and safe?* OH&S Canada Jan/Feb. 38 - 40.

Weber, J. (1992). *The front-line supervisor's role in safety.* Professional Safety, 37 (5), 34 - 39.

This page intentionally left blank

Chapter 23

Prediction of Signaller Workload

Toby Garner, Mark Newman, Chris Lowe and W. Ian Hamilton

Introduction

The identification of problems early on in the design of a system or product is essential, as this allows changes to be made in a timely and cost effective manner. This is particularly relevant for railway projects such as the re-signalling of lines, as these require significant amounts of capital expenditure. When problems are identified later in a project, or even post-commissioning, the result can be a costly alteration or a restriction on a service that can affect both performance and revenue.

One problem associated with re-signalling is that of identifying the workload that a given combination of signalling scheme and timetable imposes on a signaller. This can affect issues such as the design of the signal panel and decisions on staffing levels. While it is possible to examine workload using simulators, this may be less practicable in some situations (e.g. with Entry/Exit (NX) panels) than others (e.g. VDU based systems), as simulators are costly to build. It is obviously advantageous to be able to predict workload in advance.

Human Engineering Limited was asked by Network Rail to investigate the feasibility of developing a workload model for the role of a signaller using its in-house human performance modelling software, ATLAS. A previous project undertaken by Human Engineering (2003) had investigated the use of ATLAS as a means of predicting train driver behaviour. Subsequently, Network Rail set the objective of investigating the feasibility of predicting the impact of equipment and infrastructure changes on the workload of signallers.

ATLAS

ATLAS is a performance-modelling tool that supports the evaluation of a human-system interface early in the design phase of system development. It takes a scenario-based approach, modelling the activities of the operator. The model of the operator's activities takes the form of a task analysis. This allows an assessment of the operator's workload to be made.

The workload predictions ATLAS makes are objective and traceable, making use of established human factors practice in cognitive task analysis; task performance time estimation; workload assessment; and human error analysis, to generate predictions of the workload profile and performance quality through the course of the scenario.

The signalling task

Signallers undertake operations in accordance with defined procedures and rules, to monitor train traffic and ensure that no potentially dangerous situation arises. Typically this involves ensuring a safe separation between:

- trains and trains; and
- trains and people (including track workers and members of the general public).

When a dangerous situation does arise, the signaller must attempt to resolve the situation, and alert the appropriate authorities, as necessary.

In addition to establishing safe track conditions, the signallers are also responsible for:

- routing trains to meet a desired schedule;
- some aspects of protections and possessions as defined through the Rule Book (Railway Group Standard, 2003).

Essentially the goal of route setting is driven by the aim of executing the flow of rail traffic to a timetable. The train describer provides information relating to the train's class/ speed/ route; the panel provides information regarding its location; and the timetable gives an indication of its schedule and route. With this information, the signaller will make decisions concerning route setting in order to minimise any conflict on shared lines. For example, if a long route is set prematurely for a train, then the route of oncoming train traffic over the shared line may be stalled, leading to the standing of a train and its delay.

Signalling and workload

It is recognised that workload is a factor that contributes to the likelihood of human error occurring. High workload can reduce a signaller's ability to perform certain operations successfully, efficiently and in sufficient time. High workload can also affect a signaller's well-being and job satisfaction; when combined with a high probability for human error, the chances that the error will occur are increased even further.

Different types of workload reflect the different factors that cause them. For example, activities such as route setting, communication via the concentrator, and operation of level crossings, may cause the time pressure experienced by the signaller to increase beyond the point at which performance degrades. This type of workload is distinct from the demand caused by working in a cognitively demanding environment, such as when the signaller is forced to resolve problems caused by disruptions (for example, following a track circuit failure).

The more demanding style of activity can be described as a form of what Rasmussen (1983) termed knowledge-based reasoning. Rasmussen identified 3 different levels of activity:

- Skill-based activity;
- Rule-based activity;
- Knowledge-based activity.

The term skill-based activity typically describes actions that are well practised and can appear to be automatic both to observers, and to the performer of the task. Examples are typically sensori-motor activities such as pressing a button or flipping a switch. Rule-based activities describe action that can be modelled using rules that are triggered, for example, by environmental conditions such as 'if train X is running more than 5 minutes late, then give the junction to train Y'. Knowledge-based activity describes action to resolve situations where analytical processes and stored information are applied to solve a problem.

Routing trains on a typical NX panel is not, generally, considered to involve knowledge-based reasoning when the trains are travelling to schedule (Roth et al., 2001). Instead, activity can be seen to be at the skill- and rule-based level of operation. For example, the appearance of a train on a panel could trigger the rule 'if train X is on time then set route Y up to point Z'.

Cognitive workload increases during periods of degraded operation (such as when there is a track circuit failure), or when there is a train that is not part of the normal timetable (such as a Very

Short Term Plan – VSTP – movement). These incidents can mean that trains, which during routine operation are timetabled to avoid conflict, have to be re-routed. Factors such as routes having different speed restrictions, and trains having different acceleration and braking capabilities, can make re-routing trains a particularly demanding knowledge-based activity. But even in these situations, the Signaller may develop a series of heuristics or rules that are "good plans" for dealing with certain commonly arising degraded situations.

Given that signaller workload is characterised by periods of relatively high and low activity, this means that a battery of measures is required to describe signaller workload:

- Demand models – to show what resources are in use at different points in the scenario.
- Time pressure models – to show how much time signallers have to complete tasks.
- Conflict models – to show where tasks can interfere when performed in parallel.

This battery of measures is available in ATLAS.

Modelling the signaller's task

The decisions taken by signallers are based upon the current situation, the requirement to provide a train service, prediction of what may happen in the future, and how this could affect the flow of train traffic. Many of these activities are cued by the movement of trains, as shown by the panel.

This allows a description to be created of behaviour under idealised circumstances, in which the signaller performs routine activities such as setting routes and monitoring the panel according to the demands of the environment (i.e. the movement of trains) and also experience. The influence of experience is modelled in a series of prioritisation rules such as:

> IF the panel has not been reviewed in the last X seconds THEN review the panel now.
> IF the route ahead of the train is unoccupied THEN set as much of the route as possible, taking account of potential conflict points.

The first of these rules can be easily understood, as it is the requirement to maintain vigilance; if the panel has not been reviewed (for example as part of the activity *set the route*) then do so after a specified period of time has elapsed. The second rule reflects the observed activity of signallers, who frequently set as much of a route as possible for a train, but who may be unable to set a route across a potential conflict point without checking that no other train will require the junction first. Additional tasks can be added to this idealised representation of signalling to represent the consequences of, for example, a track circuit failure or an unanticipated phone call.

The aim of this project was to demonstrate the feasibility and utility of this approach, by modelling a sample of signaller behaviour and contrasting the workload predictions of this model against the observed behaviour and ratings of subjective workload, recorded by the signaller throughout the course of the trial.

Method

Developing a task analysis

The technique of Hierarchical Task Analysis (HTA) was used to describe the core signaller activities. HTA identifies the key steps that a signaller is required to perform in order to complete a task, safely and efficiently, and breaks these tasks down into their component parts. The tasks

included in the HTA cover a wide range of signaller activities, such as route setting, communication via the telephone/ concentrator, and reforming trains; this allows routine operation to be described. Information for the HTA was derived from observation of signallers and from workshops with SMEs.

An extract from the HTA can be used to illustrate how tasks are decomposed into their component tasks. This is shown in Figure 23.1.

Figure 23.1 shows the task *2:Route the train,* and also the sub-tasks which comprise this task such as *2.1:Perform initial assessment* and *2.3:Input route into Signalling System*; the hierarchy can be decomposed even further, to identify the component parts of the sub-tasks such as *2.1.1: Identify new train on panel.* The decomposition is halted when basic tasks such as *hear TD alarm* are reached.

Video data of signallers working at a high-fidelity simulation of an NX signal panel were analysed, to produce a timeline of actions. The panel itself was located in Crewe, and was a high fidelity simulation of a relatively simple track layout with no Automatic Route Setting (ARS). This was controlled by an experienced signaller who was familiar with the panel.

Workload measures

The project collected data that could be used to provide three measures of workload.

1. Predicted signaller workload (as modelled by ATLAS).
2. Actual signaller workload (as modelled by ATLAS).
3. Subjective measure of workload (as modelled by IWS).

Predicted signaller workload (as modelled by ATLAS)

The *predicted* model of workload is based on the task analysis, and incorporates the GOMS (Goals, Operators, Methods, Selection rules) approach to modelling behaviour (Card et al., 1983). GOMS posits that humans will set a series of goals, and by the *principle of rationality,* will act to try to fulfil these via a series of operators and methods. Selection rules determine which of the operators and methods will be used at any one time.

Some rules may even be unique to a signaller, and describe individual strategies and even quirks of operation. Rather than try to identify these context- and individual-specific rules, the

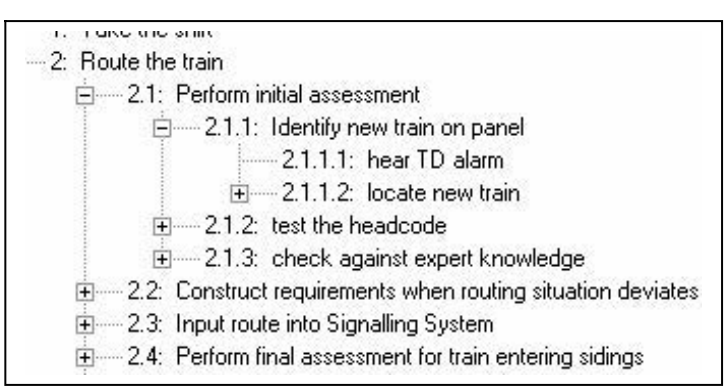


Figure 23.1 Extract from the Signaller Hierarchical Task Analysis

most efficient approach is to accept that individual differences in style of activity will occur, and that these can be accepted as a form of noise in the data, or even ignored.

The combination of the HTA, the timeline, and the prioritisation rules, allowed a model to be built that predicted workload for signallers; this encompassed not only overt activities, such as route setting, but also covert activities, such as decision-making. This was termed the *predicted* model of workload.

Actual signaller workload (as modelled by ATLAS)

Forty minutes of data were analysed using each model, 20 minutes of low workload data, and 20 minutes of high. The timeline was used to drive a workload assessment of *actual* activity, based on the ATLAS method. The observed tasks and task durations were entered into the ATLAS model, and were used to show what workload the signaller was experiencing at different points in the scenario. Since most tasks will be performed in series, the workload models used were those of time pressure and the level of demand on the individual resource dimensions, i.e. T2 and D-VACP.

Subjective measure of workload (as modelled by IWS)

The actual and predictive models were compared against subjective ratings of workload. These ratings were made at intervals of approximately 1 minute throughout the scenario, on the Integrated Workload Scale (IWS) (Pickup et al., 2005).

Results

The following section summarises the major findings for low and high workload scenarios.

Low workload analysis

A comparison of the ATLAS – T2 results for the actual Timeline and the predicted model is shown in Figure 23.2.

Figure 23.2 shows a reasonably close agreement between the actual and predicted workload, as measured by the ATLAS – T2 model, which is an assessment of time-pressure. The most obvious difference is apparent in the second, third and fourth minutes of the scenario, when there are a far greater number of actual behaviours performed than is predicted.

However, inspection of the Timeline shows that the behaviours performed at this time are all visual and cognitive; this can therefore be described as a period of planning ahead in what is in fact an initially unfamiliar scenario. This variation from the predicted workload level can be explained, as the lack of familiarity with the scenario is a breach of one of the assumptions underlying human performance modelling – that people are working in familiar environments. In addition to being opportunistic, this activity is also likely to be interruptible, in that if a second task were to be required, then the planning could be stopped with little effect on performance levels.

There are obvious periods of time in which the pattern of change in the T2 data matches that of the IWS data. Equally, there are periods in which there is a poor match. For example, in the third minute of the trial there is a period of relatively high actual workload, in which the Timeline shows the signaller spending a lot of time using the simplifier, and a low IWS rating. While it is possible that the IWS measure is less sensitive to elective tasks (reading the simplifier at the beginning of the scenario was probably to familiarise the signaller with the forthcoming events – this may not be a normal requirement of the task), such an explanation is not always possible. For example, in

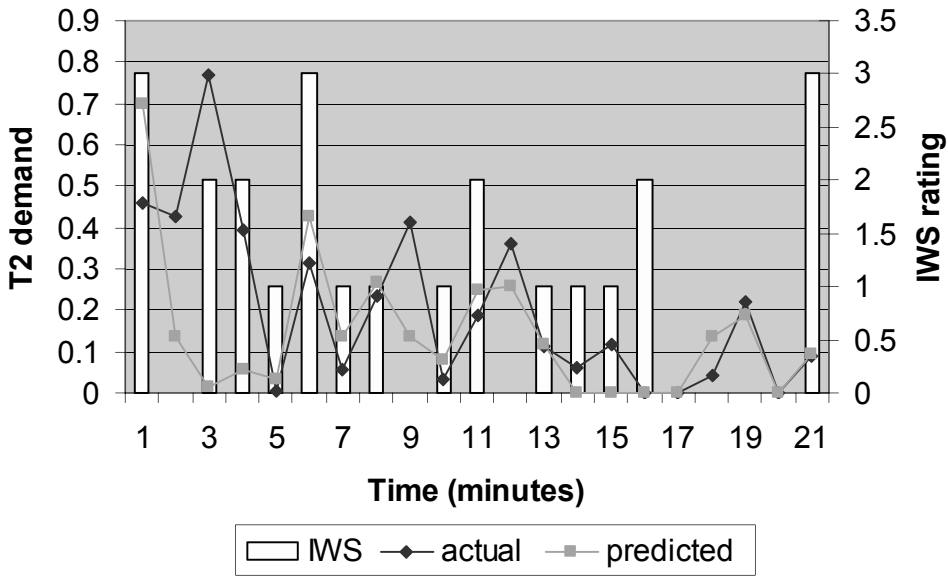


Figure 23.2 Comparison of actual and predicted workload, ATLAS (T2) model

the final minute of the trial (in which the Timeline shows there are just 2 trains on the panel and a single route setting action is performed), the T2 measures show a low level of activity, but IWS shows a higher rating.

High workload analysis

A 20-minute period of high workload, which encompassed 2 faults:

- A problem with a set of points, first noted 9 minutes from the start of the period;
- A Track Circuit Failure (TCF) first noted 20 minutes from the start of the period.

The first analysis conducted used the same method of predicting signaller activity as used in the low load condition. The results of this analysis are shown in Figure 23.3.

There is a significant correlation (0.79: 19df, $p<0.01$ 1 tailed) between the actual and predicted workload data. However, it is also obvious that the predicted workload is much lower that the actual over the course of the period. Examination of the actual scenario data shows a shift in the strategy used by the signaller; during the low load period the signaller had used a relatively consistent strategy of setting a route, then scanning the panel, before going off task. In the high load section the same signaller continuously scanned the signal panel, even during periods in which there were fewer interventions.

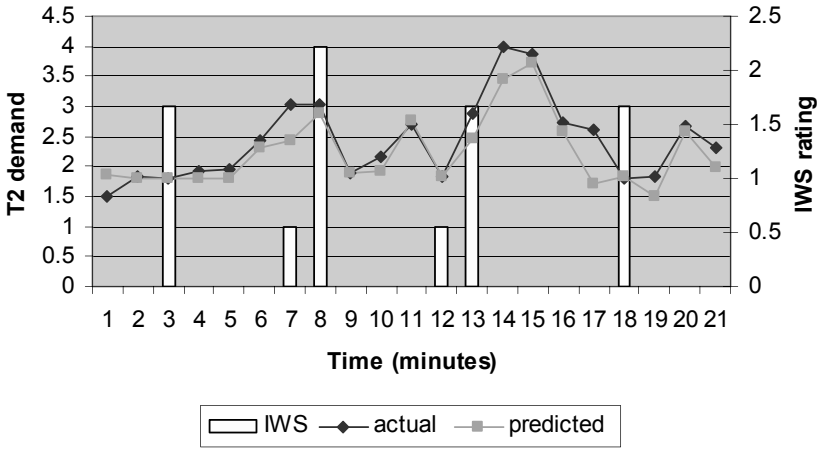


Figure 23.3 Predicted vs. actual workload, high load period

A correlation was calculated for each of the dimensions, for both actual and predicted levels of demand. This is shown in Table 23.1.

Table 23.1 Correlation between D-VACP dimensions and IWS, high workload period

	Visual	Auditory	Cognitive	Psycho-motor	All
Actual vs. IWS	-0.14 (ns)	-0.25 (ns)	-0.49 (19 df, p<0.05, 2 tailed)	0.03 (ns)	-0.17 (ns)
Predicted vs. IWS	-0.18 (ns)	0.05 (ns)	-0.52 (19 df, p<0.05, 2 tailed)	0.04 (ns)	-0.14 (ns)

The strongest (albeit negative) correlation is between the cognitive dimension and the IWS rating. This is the only dimension on which the correlation is statistically significant. However, given that there was no correlation found between the occurrence of periods of cognitive demand in the predicted and actual data, it could be concluded that the cognitive episodes are too few in number to generate meaningful quantitative results.

Conclusion

Is the model accurate?

A model of workload, predicted for a particular panel and timetable, has been constructed, and this has then been contrasted against real world data of signaller activity for the same panel and timetable.

While the correlation between predicted and actual visual, psychomotor activity and responses to auditory stimuli was high, the correlation between predicted and actual cognitive activity was far less successful. This is likely to be the case for two reasons:

Firstly, cognitive activity cannot be observed directly, it can only be inferred from the observational data – this is extremely difficult to do for the occasional checks and decisions that form routine activity.

Secondly, periods of intense cognitive activity are quite rare, and are generally associated with novel situations. None of these occur in the scenario, as even the track circuit and points failure that occur in the scenario required little in the way of problem solving to resolve.

Therefore, it can be concluded that the model can be used to predict most of the activity of the signallers with a reasonable degree of accuracy; however, there is insufficient data to draw a conclusion about the accuracy of models of cognitive activity.

Can the models be used to predict workload?

This model allows workload to be predicted for a timetable, even in the relatively early stages of design.

The number and type of interventions made by the signaller can be determined from the HTA, the set of prioritisation rules, and details of the routes and trains controlled (in terms of train movements and strike-in points, which can either be calculated from route information or estimated by SMEs). This allows for a detailed assessment of workload levels throughout a scenario, which can be used to answer engineering questions, such as whether manning levels are sufficient for the panel design. In addition, the activities of the signaller in response to different timetables can be identified.

It is currently recommended that measures of task load (given as the proportion of time required to the amount of time available) should be no more than 0.8 (Kirwan & Ainsworth, 1992). This needs to be applied cautiously in the context of signalling, given the different styles of activity that may be found. For example, during the high workload period the signaller was always looking at the panel, which, if taken as a task, means that the signaller is always at maximum capacity; performing any activity while involved in a telephone conversation means that the signaller will notionally be at overload.

However, if the standard model of the signaller is used (i.e. the model developed for the analysis of the low workload period in which the signaller is *not* assumed to look at the panel all the time) then an assessment of task requirements can be predicted directly. This will show when the signaller is approaching the critical level of involvement in terms of the requirements of controlling the panel to the timetable.

Application of the models: Analysis of scenario workload

The workload models show the effects of both the requirements of the timetable, and also the impact of time spent reading the simplifier and using the telephone. By excluding the elective activities, such as reading the simulator to plan ahead, it would be possible to identify where the workload peaks caused by the timetable are – and whether this level of demand is acceptable.

The main conclusion from the analysis conducted is that both subjective and objective workload ratings reflect the activity of signallers, but that the overall level of workload is low, and the effect of noise and of elective activities on these data is comparatively high.

The analysis may also be used in comparison of workload on different timetables, to verify that workload remains at an acceptable level under normal operating conditions. It can also be used to

assess the impact of non-normal conditions, such as when the infrastructure is degraded, on the workload of the signaller.

Application of the models: High load and telephone calls

One frequently raised issue is that of taking a telephone call while controlling the signal panel. This is not a straightforward example of a conflict between different tasks, as there are examples of signallers taking a call and simultaneously setting routes on the panel, even during high load situations. This could be due to time pressure, the cognitive involvement in understanding a phone call and the resultant actions, forgetting and interference with the cognitive activities precursing signal activities.

By incorporating details of telephone conversations (start time, duration, priority, etc.) at different times during the scenario, it is possible to predict when the signaller will be able to delay any other activities (because the window of opportunity in which to act will still be open, and possibly even because the signaller can predict when delaying an activity will mean that they will be overloaded at a later time); it will also be possible to predict when the signaller will need to perform activities in parallel.

References

Card, S., Moran, T. and Newell, A, 1983. The Psychology of Human-Computer Interaction. Hillsdale, NJ: Erlbaum.

Human Engineering Ltd, 2003. *Validation of ATLAS Workload Prediction Tool for Early Human Factors Analysis*, (HEL/NR/03860/RT2) Issue 01, 1st April 2003.

Kirwan, B. and Ainsworth, L. 1992. *A guide to task analysis*. London: Taylor and Francis

Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L. and Morrisroe, G. (2005). The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload. *Applied Ergonomics*, 36, 6, 681-693.

Railway Group Standard, 2003. *Rule Book* (GE/RT/8000)

Rasmussen, J. 1983, Skills, rules, knowledge: signals, signs and symbols and other distinctions in human performance models. *IEEE Transactions: Systems, Man and Cybernetics,* SMC-13 257-267

Roth, E.M. Malsch, N., and Multer, J. 2001. *Understanding How Train Dispatchers Manage and Control Trains: Results of a Cognitive Task Analysis.(*DOT/FRA/ORD-01/02) U.S. Department of Transport

This page intentionally left blank

Chapter 24

Collecting Human Factors Attitudes and Opinions from Signallers: Development and Use of REQUEST (The Railway Ergonomics Questionnaire)

Brendan Ryan, John R. Wilson, Sarah Sharples and Eleanor Marshall

Introduction

REQUEST (the Railway Ergonomics Questionnaire) has now been used on five separate occasions since 1998 to collect attitudes and opinions on a range of human factors issues from signallers (and other associated groups of workers) in the railway industry in the United Kingdom. In the first four of these surveys, between 1998 and early in 2004, the survey was conducted with modest samples of respondents, sampling from around 500-600 respondents at a range of signalling and control centres, achieving response rates of between 25% and 50%. The questionnaire has been developed in successive surveys, building a comprehensive tool which was designed to complement other tools which could be used as an audit of human factors issues in the workplace. By early 2004 the questionnaire contained over 300 questions, largely ratings on 10 scales broken down into 27 subscales. In addition there were qualitative questions on problems with the work environment, comments on shiftwork, opinions on the best and worst aspects of the job and opinions on the how the organisation achieved the balance between capacity, performance and safety. However, its strength as a comprehensive tool was perhaps also a weakness as, in its entirety, the questionnaire was now quite lengthy, with a mean completion time of 51 minutes. The length of the questionnaire, in addition to factors such as the lack of opportunity to complete the questionnaire at work and a declining confidence in surveys generally, were felt to have contributed to rather disappointing response rates in its earlier administrations.

Within Network Rail, staff in signalling and control roles (in addition to those in a number of other roles) attend safety briefings on a quarterly basis, receiving updates on a number of rules, regulations or other issues. It was thought that this briefing would provide an ideal opportunity for completion of the survey, giving access to a wide sample of potential respondents, having the opportunity to inform the potential respondents about the purpose of the questionnaire, and crucially, giving respondents time whilst at work to complete the questionnaire. Having received senior management approval during 2004 to use the survey at these briefings the questionnaire was redeveloped substantially in preparation for this administration. Throughout the redevelopment process the researchers were working to a brief that the redeveloped tool should be comprehensive, with coverage of the main scales in the original version of the questionnaire, but a shorter instrument with improvements in design and layout, and which could be administered in less than 20 minutes at the safety briefings. The process of development of the questionnaire, which is outlined in full in a forthcoming journal (Ryan et al., in preparation, a), used a combination of statistical techniques and expert reviews to modify the content of the questionnaire.

In the current chapter, the content of the questionnaire is outlined and the process of administering the questionnaire is described. Examples of findings from the survey are presented.

Overview of the REQUEST questionnaire

The current version of the questionnaire retains a number of core sections which have been developed from established tools within the literature or in previous use by the developers, and an overview of these scales is reported in Wilson et al. (2001). An outline of the main scales and question areas used in the current survey is presented in Table 24.1.

Table 24.1 Main scales from the REQUEST questionnaire and other key question areas which are used in comparisons in the current report

Main scales/ question areas	Breakdown of the scale/content of the question(s)
Job satisfaction	16 questions, largely from Warr et al. (1979), containing extrinsic and intrinsic job satisfaction sub-scales
Teamwork	12 questions developed from work at University of Sheffield, containing sub-scales for supervision and team communication and other teamwork (Wilson and Grey Taylor, 1995)
Job characteristics	Single questions on nine job characteristics (control, feedback, monitoring demand, repetitiveness, problems solving demand, growth, skill utilisation, task significance, responsibility
Workplace quality	10 questions developed within the IOE and used over a 10 year period
Work environment	8 questions developed within the IOE and used over a 10 year period
Usability	15 questions adapted from SUMI scales from Kirakawski (1996) with 10 questions to all signallers and an additional 5 questions to those using VDU/panel systems
Communications	9 questions developed within IOE and Network Rail, with 4 on perceptions of the technical means of communication, person to person communications, outcomes of communication and communication skills, and 5 rating communications with other groups
Workload and demands on time, mental effort and pressure	A question in which respondents are asked whether workload is acceptable or unacceptable in normal service, with a follow up question on times when workload is unacceptable. Three questions on interruptions and the amount of spare time in work, need for extensive mental effort and concentration, and problems causing high level of pressure, based on the subjective workload (SWAT) scales (Reid and Nygren, 1988)
General and safety culture	15 questions developed within the IOE (Farrington-Darby et al., 2005), with sub-scales for communications on safety, conflicts with safety and perceived management attitudes to safety
Working hours, consecutive shifts, fatigue	Questions developed within the IOE after consultation with RSSB on average and maximum weekly hours, consecutive early and night shifts, breaks between shifts, opinions on sleep and breaks, and ratings for fatigue
Physical problems	Single question on experience of physical problems in the last 4 weeks, during or soon after work.
Stress	Based on 20 questions from Cooper et al. (1988), with sub-scales for organisational structure and climate, overload and underload

Improvements in questioning, design and layout of the questionnaire

In addition to reducing the numbers of questions, changes were made in the design and layout of the questionnaire to make completion easier and more satisfying for the respondent. For example, the introductory wording of questions was shortened to enable the respondent to focus more clearly on the content of the item in the scale, without the need to read a lengthy preamble to a question. Sloping response category labels were also introduced, to allow easier selection of the appropriate response categories. These changes were also introduced with the target of fitting the questionnaire into an eight page booklet, so that it was easier for respondents to complete, with less likelihood of missing pages or sections. This target was achieved, and using the services of professional printers to assist with layout and printing the final product was much reduced in length, but also contained enough "white space" to give the perception of a questionnaire which was quick and easy to complete.

Administration of the survey

The approach to administration of the questionnaire was revised in an attempt to improve on previous response rates. Preparatory meetings were held with a member of the organisation who had responsibility for the delivery of the safety briefings. This liaison with a person having a senior role in the briefing process was an important step in developing support from the thirty or so company briefers who delivered the company briefing programme, and who were to encourage respondents to participate in the survey. The briefers were consulted about the questionnaire, with the researchers receiving feedback about the content and presentational aspects of the questionnaire, as well as arrangements for delivery of the questionnaires to the briefers and arrangements for returning the completed questionnaires to the researchers at the University.

Whilst the questionnaires would be given out by the company briefers, it was important to put in place satisfactory arrangements to maintain confidentiality (and perceptions of confidentiality) for the survey. FREEPOST arrangements were made for the briefers to collect and return the questionnaires (which would be placed in individual, sealed envelopes by respondents) in bulk packages, where the respondents were happy to do this. As an alternative, respondents could arrange for personal return of the questionnaire, using the individual envelope which was also supplied with a FREEPOST address.

Materials were given to all briefers to enable return of the completed questionnaires, and these included the FREEPOST response envelopes, postal bags for bulk returns (again with a FREEPOST address), and a sealing sticker for the individual envelopes, endorsed with "University of Nottingham" and the university logo. The sealing sticker was used to give additional confidence to the respondent that the completed questionnaire had been sealed and this was also a reminder that the questionnaires were being sent directly to, and would be analysed by staff at the university. The opening statement of the questionnaire and briefing notes for the briefers also endorsed this aspect of confidentiality.

The survey was conducted between September and November 2004, and the questionnaire was given to all attendees at the safety briefings. Responses were obtained from 3812 employees, a response rate of over 80% of the survey population. The researcher from the IOE attended two briefings to observe how the survey was administered. A brief feedback questionnaire was completed by 26 briefers who attended a subsequent company training day, collecting their perceptions of the administration of the questionnaire at the company briefings. It would seem, on the basis of this response rate and feedback from the briefers, that the procedures which were put in place for the delivery, completion and collection or return of the questionnaires have been successful.

Ratings and opinions have therefore been collected from a very high proportion of attendees at the safety briefings and the target population as a whole. Responses were received from respondents at 758 different work locations, across the eight routes, 18 general manager areas, and 160 signaller manager areas in the organisation. The survey has therefore not only collected information from a large sample of respondents, but also collected information which will be representative of the attitudes and opinions of workers in many different work locations within the organisation.

Preparation of data and analysis of responses

The response rate to the survey was greater than had been anticipated. This was the largest REQUEST survey so far and the numbers of completed questionnaires which needed to be processed were 30 times greater than in previous surveys, posing a significant challenge in the input and processing of data. A number of staff were used for data entry, inputting responses to the questionnaire into a specially prepared Microsoft Excel spreadsheet. Precautions were used in the design of the spreadsheet to minimise errors in data entry, using feedback from the person who had entered data on a previous version of the survey. All data entry staff received written instructions and personal training in data entry to minimise errors. All were asked to adhere to rigorous checking of the data and computer back up requirements. Payment for inputting was based on achievement of accuracy and a satisfactory rate of data entry. 2% of questionnaires for all data entry staff were checked for accuracy, finding very few errors, with no evidence of systematic errors from any of the data entry staff.

The primary method of analysis of the survey is on the basis of a large number of multiple item rating scales. Mean ratings for scales and items in the questionnaire were calculated and used in a series of comparisons, outlined in greater detail below. Other important information has also been obtained from the analysis of free responses to specific questions in the questionnaire. Relationships between factors in the questionnaire have also been investigated.

Examples of findings from the REQUEST questionnaire

The questionnaire has collected a large amount of data on a wide range of issues and detailed findings from the questionnaire are reported in a series of company reports. Reviews of these findings will also be published in the academic literature in due course (Ryan et al., in preparation, b). Only an overview of the findings can be given in this chapter. Table 24.2 outlines the range of analyses which have been conducted and examples of the types of analyses which have been possible are illustrated in a series of Figures.

Comparisons by different locations (route, general manager area and signaller manager area)

The questionnaire was completed anonymously by respondents. Details of work location were requested on the questionnaire to enable comparisons of ratings and responses to the various scales and items by different locations. To protect the anonymity of the respondents, data have been collated so that results are not presented individually for smaller locations, where there may only be a few signallers. Comparisons within the analyses have been made at a number of layers in the organisation – by route, general manager areas and signaller manager areas – and help to identify locations where ratings are good or not so good on a range of human factors criteria. Examples of the types of comparisons which have been possible at these layers of the organisation are illustrated in Figures 24.1 and 24.2.

Table 24.2 Summary of the analyses from the REQUEST survey

Analyses	Details of analyses
Demographic data	Presentation of data on age, grade and experience of respondents, by location and system type
Mean ratings for scales, sub-scales, items	Presentation and interpretation of means, including comparison with previous data and normative data (where available)
	Comparison of means by location (route, general manager area, signaller manager area)
	Comparison of means by system type (lever, panel and VDU systems)
Relationships between factors	Use of scatterplots, correlation and multiple regression models to identify the strongest relationships between factors, and the direction of relationships between the factors
Qualitative analysis of reports	Coding and interpretation of responses to questions, such as times of unacceptable workload, factors affecting adequate sleep and breaks, reasons for fatigue, details of physical problems, other issues and feedback on the survey

Figure 24.1 shows mean ratings for job satisfaction across the eight routes in the organisation. As indicated in Table 24.2, these mean ratings have been interpreted using a range of criteria, including comparisons with other locations in the sample (i.e. comparisons with other routes in the figure), comparisons with data from REQUEST surveys in previous years and comparison with normative data for this scale. These comparisons have indicated that the overall ratings are slightly above the mid point (neutral value of the scale), and suggest that the respondents have some satisfaction with their work. The ratings are broadly comparable with those for Associate Professional and Technical staff in the normative data provided by Mullarkey et al. (1999), though there is some scope for improvement in rating of intrinsic items on the job satisfaction scale (that is features that are integral to the work itself, rather than external to the work). (Note that differentiation between intrinsic and extrinsic items on the scale is not shown in the Figure.) Ratings on one of the routes (Route 6) were particularly low, but relatively high on another (Route 7).

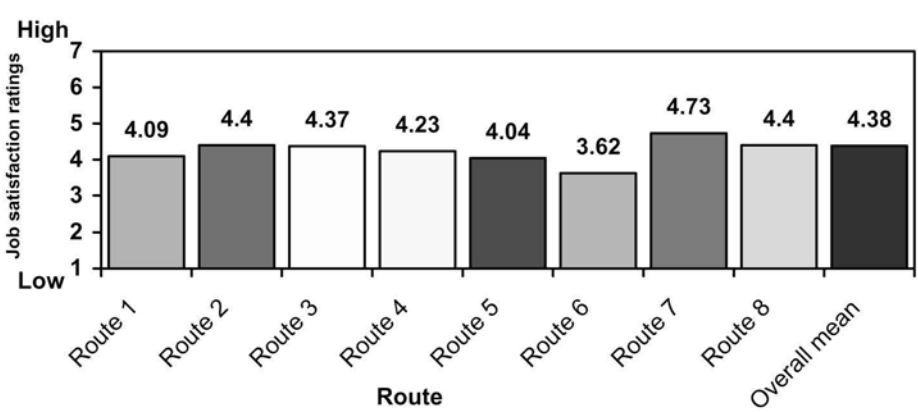


Figure 24.1 Mean ratings of job satisfaction by route

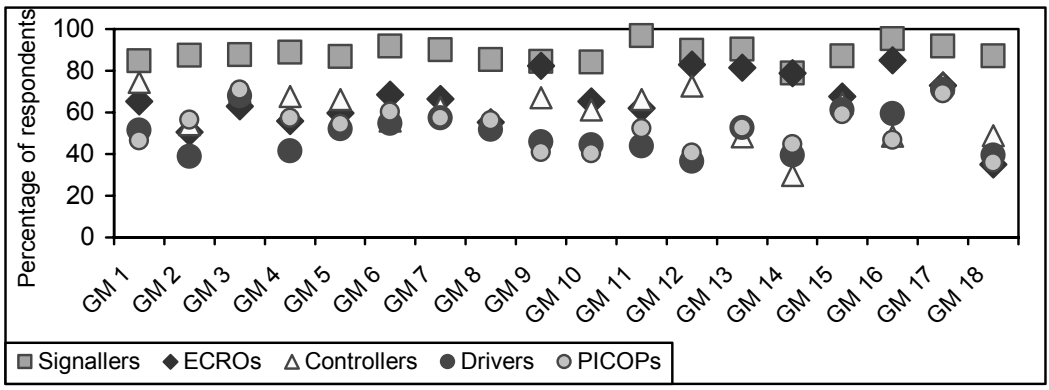


Figure 24.2 Percentage of signallers rating communications with other groups as "good", by general manager areas

Similarly, comparisons have been made between ratings on the range of scales and items within scales by general manager areas and signaller manager areas. As an example of a comparison by general manager area, Figure 24.2 shows the percentage of signallers rating communications with other groups as "good" and indicates that signallers generally perceive that their communications are better with other signallers, than with groups such as drivers or PICOPs. The figure also helps to identify general manager areas where perceptions of communications are better (such as GM area 17) or where perceptions of communications with almost all other groups are poor (such as GM area 18). The Figure shows that in General Manager area 14, few signallers rated communications with controllers as "good". Ratings by signallers for good communications with controllers varied between about 30 and 75% of the respondents across the general manager areas

Comparisons by system type

Differences have been identified between ratings from respondents operating the main types of systems. For example, ratings for job satisfaction were greater from those operating lever systems than those operating panel or VDU systems. Other differences in the ratings between respondents operating different system types and are summarised in Table 24.3.

Relationships between variables in the questionnaire

Relationships have been investigated between pairs of variables in the questionnaire. These have been selected for investigation where there is some basis for a theoretical or pragmatic relationship between the variables. As an example, Figure 24.3 shows as scatterplot of the association between mean ratings for general and safety culture and mean ratings for job satisfaction (r=0.676). This indicates that respondents with high ratings for general and safety culture were also likely to report high ratings for job satisfaction, though of course, this test of association does not give any indication of causation or the direction of the relationship between the factors. Relationships between pairs of factors have been classified as quite strong (r>=0.5), moderate (r>=0.3 to <0.5) or modest (r>0.22 to <0.3) and the factors with the strongest relationships with other factors are summarised in Table 24.4.

Table 24.3 Differences in ratings between respondents operating different system types

Differences between systems	Type of factor affected
Ratings from those operating lever systems are different to ratings from those operating panel or VDU systems	Higher ratings for job satisfaction from lever operators than others Fewer reports of unacceptable workload from lever operators than others More reports of adequate sleep and breaks from lever operators than others Fewer reports of fatigue from lever operators than others Fewer problems with maintenance for lever operators than others
Ratings from those operating panel systems are different to ratings from those operating lever or VDU systems	Lower ratings of workplace quality and work environment from panel operators than others
Ratings from those operating VDU systems are different to ratings from those operating lever or panel systems	Greater problems with learning to operate VDU systems than other systems (on the usability scale)

Table 24.4 Factors with multiple relationships with other factors

Main factors	Factors with strongest relationships with the main factor
Job satisfaction	General and safety culture, Teamwork, Workplace quality, Work environment, Stress, Growth
General and safety culture	Teamwork, Job satisfaction, Stress, Grade
Stress	General and safety culture, Job satisfaction, Fatigue
Teamwork	Job satisfaction, General and safety culture, Stress
Fatigue	Stress, Job satisfaction, Inadequate sleep, Inadequate breaks

A series of multiple regression models have been created to investigate the strength and direction of relationships between the main factors and other factors (and sub-scales) from the questionnaire. It has therefore been possible to investigate the direction of the relationship in our example of job satisfaction and culture ratings. Figure 24.4 illustrates the factors which have been found to affect ratings of general and safety culture using the multiple regression analysis, showing that perceptions of extrinsic items on the job satisfaction scale accounted for 44% of variance in ratings of general and safety culture (using the value for R square change in the regression analysis). Ratings on supervision and team communications (from the teamwork scale) accounted for 8% of variance in the ratings on culture and perceptions of organisational structure and climate on the stress scale had a moderate effect on ratings of culture. Respondent grade was found to have a small effect on culture ratings. In another multiple regression model, in which culture was used as the dependent variable, it is interesting to note that perceptions of culture had very little effect

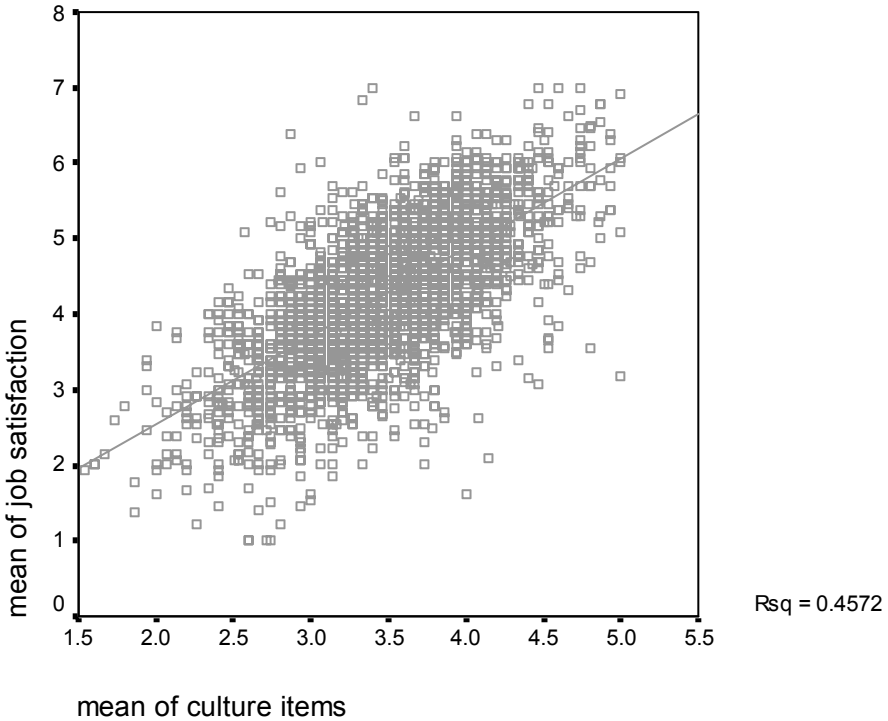


Figure 24.3 Relationship between mean ratings for general and safety culture and mean ratings for job satisfaction

on ratings of job satisfaction. Therefore, it is hypothesised that in the relationship between job satisfaction and culture, perceptions of extrinsic items on the job satisfaction scale are likely to affect the ratings of culture to a greater extent than perceptions of culture will affect ratings of job satisfaction.

Further analyses are being conducted to produce a composite model of the relationships between the factors, to confirm the relationships between multiple dependent variables in the questionnaire.

Qualitative analysis of responses to questions

Responses to a number of questions have been coded and interpreted, and these include questions on times of unacceptable workload, factors affecting adequate sleep and breaks, reasons for fatigue, details of physical problems, other issues and feedback on the survey. As an example, respondents were asked to comment on whether there were particular times when they felt that workload was unacceptable. Comments were received from over a third of the respondents in the sample and these have been classified according to a number of main themes, identifying factors which contribute to workload in these situations or scenarios. The two most common times in which workloads are reported as being unacceptable are during the occurrence of incidents / problems and during possessions or protection of the line. The analysis has identified a number of common factors (including paperwork and interruptions by telephone) which have been reported by respondents as

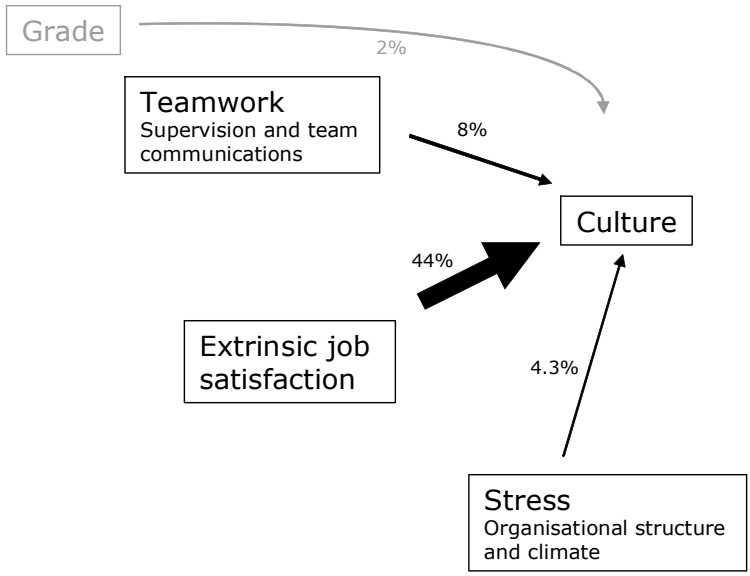


Figure 24.4 Factors affecting ratings of culture, identified through multiple regression analysis

contributing to workload during these times. Outcomes or effects of high workload have also been reported by respondents, such as the inability to carry out other signalling tasks successfully, the lack of time to take necessary breaks, and the increased potential for errors and confusion during signalling operations.

General conclusions

REQUEST achieved excellent response rates, is comprehensive (in terms of coverage), and is one of the largest human factors surveys that we know of. Data have been collected on a wide range of human factors issues. The method of administering the questionnaire during the company safety briefings has been demonstrated to be highly effective, and it is anticipated that the safety briefings will also be a useful means of feeding back findings from the questionnaire to the respondents. Progress is already being made to ensure that this feedback is effective.

The survey has provided a database of findings which is a very useful resource for the organisation. Comparisons have already been made at multiple layers in the organisation on a wide range of issues, identifying locations performing well or not so well, enabling more efficient allocation of resources. It has been possible to benchmark ratings against previous data from the REQUEST surveys, and against data from other organisations on one of the scales (for job satisfaction). Comparisons with data from other audit tools are also on-going (e.g. BASELINE, ODEC). The data are also being integrated with other data held by the organisation, to investigate whether the human factors data from the survey can be used to explain variations in performance statistics and the frequencies of incidents and irregular working in different locations within the organisation.

The survey has also produced a database which is suitable for use in a range of in depth statistical analyses. The investigation of relationships between factors in the questionnaire is currently in progress and can be useful to give insight to important theoretical concepts.

Acknowledgements

The authors would like to thank the many participants at Network Rail who readily participated in the survey, and also thank all those who were involved in the smooth running of the administration of the questionnaire at the safety briefing process.

References

Cooper, C.L., Sloan, S.J. and Williams, S. (1988) Occupational stress indicator: management guide. NFER-Nelson, Windsor, UK.

Farrington-Darby, T., Pickup, L. and Wilson, J.R. (2005) Safety Culture in Railway Maintenance, *Safety Science*, 43, 39-60.

Kirakawski, (1996) The software usability inventory background and usage. In P.W. Jordan, B. Thomas, B.A. Weerdmeester and I.L. McLelland, *Usability evaluation in industry*. London, Taylor and Francis, 169-177.

Mullarkey, S., Wall, T.D., Warr, P.B., Clegg, C.W. and Stride, C.B. (ed) (1999) Measures of job satisfaction, mental health and job-related well-being. A bench-marking manual. Institute of Work Psychology, University of Sheffield, UK.

Reid, G.B. and Nygren,T.E. (1988) The subjective workload assessment technique: A scaling procedure for measuring mental workload. In: P. Hancock and N. Meshkati (Eds), *Human Mental Workload*, Amsterdam, The Netherlands: North Holland.

Ryan, B., Wilson, J.R., Sharples, S., Morrisroe, G. and Clarke, T.
(in preparation, a) Developing a questionnaire (REQUEST) to collect human factors attitudes and opinions from railway signallers.

Ryan, B., Wilson, J.R., Sharples, S., Morrisroe, G. and Clarke, T.
(in preparation, b) Human factors attitudes and opinions of railway signallers: Findings from REQUEST (the Railway Ergonomics Questionnaire).

Warr, P.B., Cook, J.D. and Wall, T.D. (1979) Scales for the measurement of some work attitudes and aspects of psychological well-being. *Journal of Occupational Psychology,* 52, 129-148.

Wilson, J.R. and Grey Taylor, S.M. (1995) Simultaneous engineering for self directed work teams implementation: A case study in the electronics industry, *International Journal of Industrial Ergonomics*, 16, 353-365.

Wilson, J.R., Cordiner, L., Nichols, S., Norton, L., Bristol, N., Clarke, T. and Roberts, S. (2001) On the right track: Systematic implementation of Ergonomics in Railway Network Control, *Cognition, Technology and Work*, 3, 238-253.

Chapter 25

Ergonomics Assessment of Lever Operation in Mechanical Signalling

R.J. Muffett

Introduction

2538 working days were lost to Network Rail in 2003 through injury associated with the operation of levers. These are used within the mechanical form of signalling which was developed during the 19th centaury to control the movement of trains. Although dated, lever frame signalling is still in place in more than 500 signal boxes and is likely to be with us for a good while longer. Mechanical signal boxes have their connection to external devices – such as signals and points – by wire and channel rodding attached to the levers within a metal frame.

The assessment of operability of levers with respect to their manageability is currently measured by reference to the length of the connection, with no consideration given to weight of pull. This does not appear to follow recognised health and safety thinking and guidelines (such as those for manual handling) but the lever systems have been around long before current guidance was developed. Early investigations highlighted that there are many variants of frame and lever types, along with the different release types (triggers and stirrups). The connectivity between lever and the object to be moved gives some clue as to the differences in lever pull weights. However, Rail Group and Line standards give very little insight into appropriate design methods or design criteria for lever based systems.

With so many factors influencing the signaller-lever frame interface, it is apparent that signallers have developed their own methods to carry out their task of operating levers. These have not always come about by good practice or training. The method adopted by an individual is often influenced by body size and frame design, with staff not always aware of the likelihood of injury that may be brought about by their particular approach to accomplishing lever movement. Even having worked as a signaller for many years, the author was not aware himself until recently about the frequency and types of injuries that signallers were suffering while operating lever frames. Few if any signallers realise that there exists a possibility that the physical injuries they are being affected by in their later life may have been caused by the operation of lever frames.

That this is not a new problem is evident from the following quote (and see Figure 25.1):

> Mrs Kathleen Willingham, 36 years old and married with children, worked from 9 a.m. until 10.30 p.m., a 91-hour week, at a level crossing near Colchester. She had to pull a very heavy lever 144 times a day. The LNER provided a stand-in for only seven hours a week 'shopping leave'. Her application for a pay rise to compensate for taking over her husband's duties was declined. She asked to get her hours reduced but the LNER treated her as a part-time employee because she could 'run home between trains'. They argued further that she knew the duties of the job when she accepted employment. Men were allowed to work only eight-hour shifts. *(Excerpt from article on a web site called Signal Box. (Quote and picture in Figure 25.1 appears with permission of Helena Wojtczak (2005)).*

Figure 25.1 Lever boxes required heavy work many years ago and still do so today

Aims of the study

Initial observations of signaller's methods of operating levers raised many questions, the most important being of what is the safest method of pulling levers, and what are safe limits to the weight to be operated. Research was initiated by the Network Rail Ergonomics National Specialist Team to carry out the preliminary research required in order to move towards answers to all of these questions.

The aim of this project was to look at methods of pulling levers from an informed ergonomics viewpoint, gaining an understanding of the risk factors for musculoskeletal disorders for signallers. Of course it was realised that without a substantial longitudinal study it might be difficult to categorically attribute any injury to their work, as activities outside the workplace may have triggered or contributed to the physical problem. Also the lever frame technology is particular to the railway and has been in place for a long time, and is therefore not compliant to modern standards. Therefore the project aims to increase our understanding of the effects of pulling levers, and to identify what methods could be used to limit and manage any risks associated with them. This is to be achieved by analysis of methods of lever operation, and how they are influenced by body size and lever frame design. This should provide data to support recommendations for modifications to levers and frames, training of improved pulling techniques, and a better 'fit' of operators to the task.

Field study approach and methods

A number of approaches to investigate this problem and make recommendations for improvements are possible. These include laboratory experiments with participants drawn from signalling staff or else representing them, to examine postures when pulling on levers with certain forces, and

also computer based modelling of the impact of such activities. However in the first instance it was important to get a better understanding of what forces were involved when signallers were operating levers and what postures they were adopting in doing so, and then to assess guidelines, standards and research evidence available in order to assess both the degree of problem and also any potential solutions. Therefore this chapter is concerned with the first field study that was carried out and with the collation of any relevant guidelines against which to assess the evidence that was found.

Visits have been made by the author and a number of colleagues to various lever frame signal boxes in the UK. This is a rolling and continuing sampling study, and this chapter is concerned only with the visits to the first four sites. The investigation process in each was similar, and consisted of a general discussion with the signallers concerned, and the signalling manager, about operation of the levers and any issues or problems they identified, followed by observation and measurement. The investigator and his colleagues observed the lever pulling technique of the signallers who were on duty and recorded this. Records were made by eye with handwritten notes, via digital photography and also through video taping. In addition a crude first prototype measuring device was applied to the levers. This was a strain gauge attached via a looped strap to a data recorder capable of recording weights up to 100kg through a metal ring. Each lever movement had a measure recorded at three locations through its movement: at rest, during the middle of the pull, and just prior to completion of the pull. The three point measurement gave insight into the variations the signaller was dealing with during the lever operation, and showed great variation across the pull range. The signallers then gave an estimate against the lever pulled in the following categories H = heavy, M= medium, E = easy. There was no prompt for the signaller of the weight actually recorded on the data recorder; it was their impression of the pull on that occasion. In general signallers were rating easy as up to 20kg, a medium rating between 20kg and 50kg, and a heavy pull was rated as above 50kg. Participants involved in tests varied across a spectrum of size and weight, all were men in physically good health.

There are a number of limitations on the measurement of force carried out in these early stages of the study with the existing crude device, not least the fact that it was only able to give discrete point reading of weight rather than dynamically throughout the range of movement. Nonetheless the early recordings provided an excellent first indication of forces, and therefore of the potential for risk of harm to the signaller or of poor performance in lever operation.

The observation study proved extremely useful from a very early stage, in that it was soon identified that there was not one universal method of pulling levers. The postures adopted and then the lever pulling strategies varied considerably, according to the individual signaller preferences, body size and weight of the individual, the space available to pull the lever in the signal box, design of the frame, and the weight and position of the lever itself. These strategies were determined by the investigator to fall into three basic types: swing, drag and shuffle.

1 Swing method of lever operation starts with signaller gripping the lever either with one or two hands, dependant on whether their approach his hindered by other levers already over in the frame. Once the lever is released from the trigger catch in the frame it is often released by the hand/s, there is a short pause in motion but grip is maintained on the lever. The arms are bent with the trunk of the body drawn in close to the lever; the body is then thrown backwards with the feet maintaining their position, and the back kept straight. The arms straighten until the weight of the lever is balanced against the force and weight of the body's backward movement. This movement is very dependant on the signaller's knowledge of the lever frames idiosyncracies.

2 Shuffle method is similar to the swing method with a slight variation in the concluding part of the movement. The body mass cannot be accommodated under the lever's arc and therefore the signaller has to shift while the lever movement is coming to its end. This requires the legs to carry out a shuffling movement backwards in order to prevent collision between body and lever.
3 Drag method starts as with swing and shuffle, but most often there is no pause or trigger release in its early stages. The signaller literally drags the lever through its arc of movement with the whole body moving back in synchronisation. The signaller's trunk gradually bends forward as the movement of the lever brings it towards them, ending with the signaller bent over the lever.

When levers had to be returned to their normal position in the frame, signallers used one common method, which was a reversal of the drag method. The pushing forces required would vary in the same way as the methods of pulling. The body posture would have to accommodate the arc of the lever, bending to get under the angle in order to gain the initial momentum, Swing and shuffle are used but body movement has to be restrained because of the signal box design, with the block shelf situated above the frame at about shoulder height.

The three basic strategies adopted by signaller to move levers should be assessed in detail, but the movements and postures involved do not relate easily to standard handling tasks about which there is good knowledge or risk factors and recommended methods . Drag could be looked at in the context of "pulling", from the explanation given in the HSE Manual Handling Regulation. Swing is a matter of using the body weight to counter the force needed to move the lever; this is the style adopted by smaller (both weight and height) persons. Shuffle is an adaptation of swing, but because the arc of the lever is such that the body cannot be accommodated within the last part of the movement of the lever, they have to extricate the lower half of the body from under the moving lever. This movement is normally associated with persons of a larger girth.

Ergonomics guidelines and considerations

The relevant Rail Industry Guidance states that: "The Maximum recommended distance at which signals can be mechanically operated is 1,000 yards which minimises the effects of temperature change on wire routes" (RT/E/C/19014 Railtrack Code of Practice), and that "Wire operation of mechanical signal should be limited to 1,000metres (1,100 yards)" (Railway Group Standard GK/RT 0039). There is some discrepancy in these documents which make it difficult to determine which distance is most appropriate to use as a reference. Signal boxes tend to be close to points in an attempt to lighten the weight of the pull. Current codes of practice states "200 yards satisfy the Health and Safety Executive" (RT/E/C/19014 Railtrack Code Of Practice) and "Manually worked points should not exceed 320 metres (350 yards) rodding run from lever to furthest point end" (Railway Group Standard GK/RT 0039). The more challenging and restrictive, and therefore least demanding on the operator, is the Railtrack Code of Practice, to which signalling engineers generally work.

As previously stated most of the mechanical frames in use today were implemented at a time when there was a less well developed understanding of the physical demands on the signaller or operator. However one feature of the mechanical system that ties in with some of today's guidelines is the average height for levers at sites visited. With the average length of these being 1010mm this is within the guidelines, based on anthropometric principles and dimensions, of between 825mm and 1231mm.

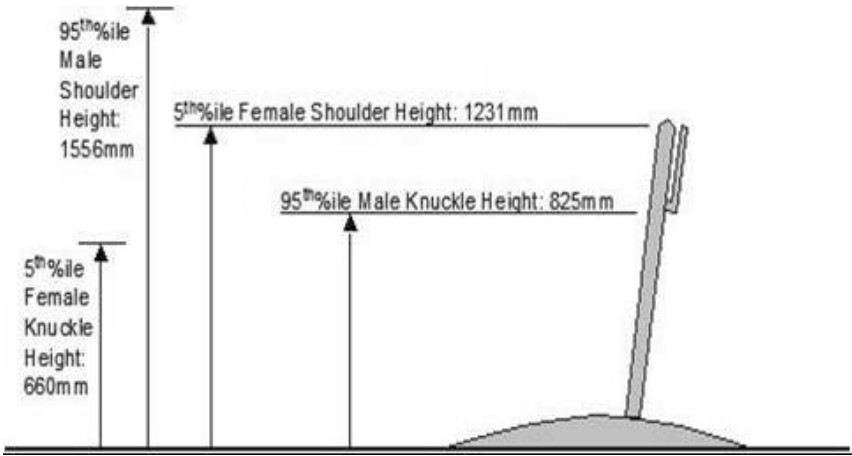


Figure 25.2 Anthropometric measurements (not to scale)

Assuming an action whereby the signaller pulls the lever directly towards the body, using both hands, the difficulties that a 95th percentile adult British male may experience when working with manual signalling will possibly be related to adopting a stooping posture (bent back and neck), and a reduced force application capability due to a non-optimal posture. The difficulties a 5th percentile adult British female may experience when working with manual signalling will possibly be related to shoulder problems and a reduced force application, due to a non-optimal posture (Figure 25.2).

The British Rail Standard Midland style frame design has a higher lever frame base that encourages some signallers to step up, or use one foot on the frame as a fulcrum point and the other on the floor. The lever height, however, is still within 825mm –1231mm (Figure 25.2*)*. This higher frame does not create a different style of pull but forces some operators to stand higher due to their preferred method of pulling. They stand on the frame, moving their feet from the frame to the floor during the movement when the swing and shuffle methods are used. Additional types of injuries could be attributed to this design of frame, because the signaller does not have their feet on the floor in order to maintain balance. The legs should be used as the prime body component in taking the load in this task, but the transfer of the feet off the frame to the floor during the lever movement adds to the risk of lower back injury (Figure 25.3).

In order to reach across the frame from an upright position, the back must be arched. The reach measurement was a minimum of 92cm, which brings into question the loads that can be pulled with the trunk bending forward.“The amount of force a user can exert on a control depends on posture, position, and the length of time the force is applied” (Galer 1989).

The maximum pulling and pushing forces of the hand during standing work have been studied by Rohmert (1966) and Rohmert and Jenik (1972) (both quoted in Kroemer and Grandjean, 2000). Pulling forces are greatest in the vertical plane and lowest in the horizontal plane. Pulling and pushing forces are of the same magnitude whether the arms are held out sideways or forwards (in the sagittal plane).

The Health and Safety Executive website describes in detail how back pain can be aggravated: “Physical activities that can aggravate back pain, or at least be more difficult when you have back pain, include: stooping, bending over or crouching, including work at PCs (poor posture); lifting objects which are too heavy or bulky, carrying objects awkwardly; pushing, pulling or dragging excessive loads; working beyond normal abilities and limits, and when physically overtired; using

Figure 25.3 Melton station with British Rail Standard Midland frame

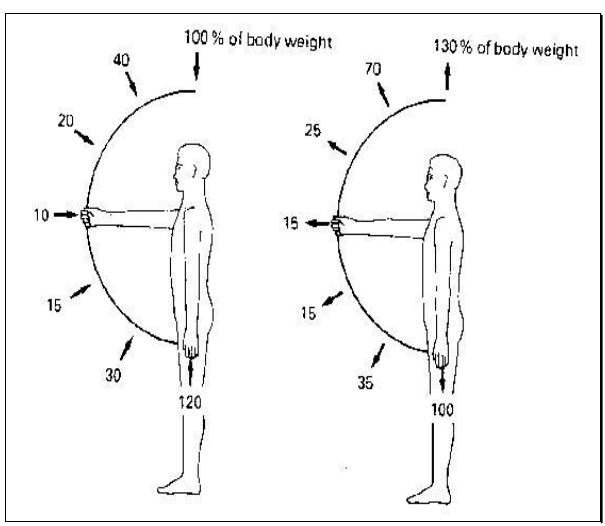


Figure 25.4 Maximum force of pulling (left) and pushing (right) for a man, feet 300 mm apart. The values at various angles are given as a percentage of the body weight. (simplified, after Rohmert, 1966 – in Kroemer and Grandjean, 2000)

poor lifting techniques (or where you are unable to use lifting aids due to the working environment); stretching, twisting and reaching; prolonged periods in one position, leading to postural strain; situations where the whole body is subjected to vibration, jolting and jarring. As with many ill health conditions, some people are more susceptible to back pain than others" (Excerpt from HSE website: www.hse.gov.uk).

HSE Manual Handling Guidelines under the heading "Pulling pushing" recommends

Action	**Men**	**Women**
Force to stop or to start a load	20kg	15kg
Sustained force to keep the load in motion	10kg	7kg

These guidelines also show that were "twisting more than 45% the load should be reduced by 10%". NIOSH guidelines refer to a recommended lifting weight of 23 kg

The question that arises is; what is a safe weight or effort that an operator can exert given the nature of the task, frequency and duration? The Work Safe Western Australia Commission's 2000 Manual Handling Code of Practice (Appendix H) provides the following guidance concerning the handling of heavy objects:

> The risk of injury increases as the weight of the load increases.
> Evaluating the risk of weight of the object needs to take into account:
>
> - how long the load is handled
> - how often the load is handled.
>
> As a guide, the risk of back injury increases when loads over 4.5 kg are handled from a seated position or when loads over 16 kg are handled from positions other than seated. As weight increases, the percentage of healthy adults who can safely lift, lower or carry decreases.
> Generally, no single person should be required to lift, lower or carry loads over 55 kg. THIS LIMIT WOULD ONLY APPLY, HOWEVER, WHEN THE LOAD IS WITHIN THE PERSON'S CAPABILITIES AND NO OTHER RISK FACTORS ARE PRESENT (eg. no bending or twisting is required to pick up the load; the load is compact and easy to grasp; it is held close to the trunk and not carried frequently or for long distances).
> On occasions, objects over 55 kg may be moved but not lifted, eg. Rolling a 200-litre drum.

The amount of force a user can exert on a control depends on posture, position, and the length of time the force is applied. (Galer, 1989) 'Jerk' forces used to loosen tight controls may be two or three times larger than the corresponding maximum steady forces; forces which must be applied over a long period should, it has been suggested, not exceed 15% of the corresponding maximum force over 5 seconds, although some continuous application may last for an hour or more without rest (Galer, 1989).

HSE Manual Handling Guidelines do not now give a maximum weight to be lifted, but recognise that the task itself and other factors can affect the performance limits.

Discussion

For the field study there was no existing tool or method of measurement that could be used on a dynamic movement such as lever pulling. The initial research carried out concentrated on analyses of posture (static) and the pressure exerted on the various body parts against the time that the

position is maintained. Further research will be required to enable a better understanding of body stresses while weights / forces are being managed through a dynamic movement of lever pulling. Also, a further search will be made for guidelines drawn from the same or very similar work; there is little at present although some transfers and interpretations can be made. For instance, the work done in the movement of trolleys (HSE Manual Handling Regulations) has some connection, in that the movement from the static position and again in stopping the trolley may be loosely comparable to lever movement.

The only comparable research to this investigation was carried out in the USA (Page, McMahan, and Brown 1990) on the movement of levers at point switches. However, this is not applicable here, because the measured action starts with the lever parallel to the ground and is more a lifting movement from a static position with no connectivity through wires or rodding. This US work also concentrates upon hand forces exerted and not on whole body movement. These point switches are similar to point operation within a UK marshalling yard but with an upright lever. There is no lever frame or other levers in close proximity to inhibit access.

The load guidance given in the (HSE Manual Handling Regulations) for lifting is 20kg; if you took this load in the context of the swing method of pulling levers, it would fall within the context of a snatch weight method of operation (Galer, 1989). This method allows for doubling or even tripling the pull weight, taking the total load to possibly 40 or 60 kg. Interestingly when the tests were carried out, and cross comparisons with signaller's subjective ratings were made, it was in the 50kg area that signallers started saying the pull weights were "heavy". Such criteria could also be applied to the shuffle method of pull, but the implications of the movement of the lower half body while dealing with this load would have to be evaluated. There would also be a need to evaluate what upper or ceiling weight might make the use of the shuffle and drag methods inadvisable. Western Australia Commission's 2000 Manual Handling Code of Practice (Appendix H) says "On occasions, objects over 55 kg may be moved but not lifted, eg. rolling a 200-litre drum".

No method of measuring the pushing forces required to return levers into the frame has been tested yet. The advisable push weight is also 20 kg (HSE Manual Handling Regulations) and so the conclusions arrived at with pulling would apply, apart from not having the snatch element to assist with getting heavy weights moving. It is also important to recognise that starting the movement with the lever at the full extent of its arc creates the problems of approach stance, virtually having to get under the lever. If the back is to be kept straight then the 50th–95th percentile male would need to bend his knees in order to accommodate a position whereby the signaller could start to move the lever. The signaller would have to be able to move forward with the lever. The signaller will, with heavy pushes, be arching his back toward the conclusion of the movement, as their feet are static. Signal stop control levers will mostly return to normal in the frame with little assistance, whereas distant signal levers do require assistance but not with the same vigour as pulling them. Point levers and facing point lock bars do not change in their characteristics from pulling to pushing.

Further work

It has become clear throughout the project that operation of levers in lever frame signal boxes is a potential health and safety risk, with several factors influencing the degree of risk, but that there is currently little if any relevant research upon which to draw.

Before any firm recommendations can be made regarding strategies, weights and forces associated with the operation of levers, additional work must be carried out:

- More detailed measurements of the lever forces by developing a more sophisticated measuring system.

- Further analysis of the effects of the lever type in conjunction with the signallers' size and stature would also allow for a greater understanding of how these two factors influence the signallers posture and perception of the forces applied.
- A more detailed analysis of the number and type of injuries and also if they are more prevalent with specific lever types
- Computer modelling of the postures and forces involved within typical (constrained) workspaces, using Jack or similar ergonomics modelling system.

Work on all of these is now on-going. If any recommendations are made in the future regarding maximum forces, lever and frame design they would be applied only to replacement systems. Reduction of risk, if necessary, in all current lever systems would have to be managed through changes in process and training.

References

Galer, I. (1989) Applied Ergonomics Handbook. (2nd Ed). Butterworths

Health and Safety Executive (2004) Manual Handling Operations 1992 (as amended 2002) (3rd Ed) HSE Books.

Kroemer and Grandjean, 2000, Fitting the Task to the Human (Fifth Edition): A Text book of Occupational Ergonomics. Taylor and Francis

Network Rail Worksafe Manual Legal and Health Safety Responsibilities.

Network Rail Professional Signallers Handbook (2004)

Network Rail Department of Human Resources (2005)

Page, G.B; McMahan, P.B. and Brown, R.T. (1990). The development of ergonomic guidelines for railroad hand switch operations.

SMIS Safety Management Information System

The National Institute for Occupational Safety and Health (NIOSH)

The Work Safe Western Australia Commission's 2000 Manual Handling Code of Practice (Appendix H).

Wojtczak, H. Railwaywomen: Exploitation, Betrayal and Triumph in the Workplace. Hastings Press, 2005.

This page intentionally left blank

Chapter 26

From the Horse's Mouth: The Contribution of Subject Matter Experts (SMEs) to Study of Rail Work Systems

Gemma Cox, Trudi Farrington-Darby and Richard Bye

Introduction

The involvement of SMEs (subject matter experts) in rail research has increased in parallel with the interest and associated research programmes in Rail Human Factors (Rail HF) over the last decade. The nature of HF investigations is such that humans form the focus of our enquiry. This coupled with the complex nature of a domain with a long history that is also subject to regular changes creates an environment where the role of SMEs is crucial.

Our experience of working with SMEs in several functional areas, on different projects and where SMEs have been involved in different capacities, has resulted in observations which are commented on here.

Definitions and the role of SMEs

There are several definitions of SMEs and these are largely related to the reason for involving the SME in an investigation and also the nature of the project in which they are assisting. On the whole SMEs are considered to be individuals who exhibit the highest level of expertise in performing a specialised job, task, or skill within the organisation.

The term SME has been widely used in the field of software development and the development of expert systems. SMEs were people who were considered to hold knowledge in their heads that could be captured and represented using technologies such as expert systems and artificial intelligence, usually though a series of rules linking knowledge items. More recently SMEs have become viewed as more than a head full of knowledge that is there to be captured. SMEs are viewed as "content specialist who advises or assists the designer" (Ibstpi, 2005).

Rail HF research has, like other areas of research such as psychology, seen a shift away from the dominant experimental design approach using subjects in sterile laboratories performing isolated and often simple and meaningless tasks to study microcognition. This shift has been towards more naturalistic methodologies where the SME are not so much subjects but sources of information. Traditionally SMEs' technical knowledge has been the primary source of data but recently they have taken on a slightly different role. They are the domain guides and, gatekeepers to the settings to be investigated. This change of role in research has accompanied a general change from the specific definition of SMEs as advisors in design to something looser and sometimes more informal.

In a review of the subject of expertise (Farrington-Darby and Wilson, 2005) it is suggested that expertise is neither static nor something that an individual possesses for every aspects of a job. There has been a move to thinking of SMEs as people who contribute more than one type of

knowledge about specific areas of a domain. It may be that different SMEs are needed for different aspects work or activities. Accessing knowledge of the social structure of a team or system may be as important as obtaining knowledge about technical aspects of a system. More recently web definitions of the term tend to be informal, steering away from specific technical knowledge and towards domain knowledge. An SME is described as "an individual who has extensive experience and knowledge in a domain and who is recognized as an authoritative source of information about the domain." (Process Impact, 2005).

In the rail industry SMEs are operational or ex operational staff who have generally worked within the industry for a number of years, building up their knowledge and skill by moving up through the ranks to positions with increasing authority. Through this process SMEs gain a large amount of knowledge, much of it tacit, not just of the specifics of the job, but on how best to go about performing a job. They typically have a great deal of social expertise along with deep understanding of the organisation and how it operates in reality. Different individuals will vary in what they offer in their role as an SME. They may differ with respect to the degree of social versus technological expertise, for example.

The contribution of SMEs to the research process

The SME may contribute in roles that reflect the traditional view of the term. They may work with software engineers to develop expert or knowledge support systems through the elicitation of their specialist knowledge. They also continue to work with designers to provide input on the development or evaluation of design ideas for new and modernised systems in rail – for instance displays for ERTMS. This involvement of SMEs is in keeping with the philosophy of a participatory approach used in ergonomics. This approach ensures that potential users of new designs and upgrades, be they technologies, work systems or organisational structures, are consulted and sometimes drive the development, implementation and testing of these systems. Further to these roles SMEs are working with researchers, consultants, and in-house ergonomists to assist in:

Establishing investigation background and familiarisation

Ergonomics investigators new to rail, a particular function in rail or just a specific environment spend some part of their project time becoming familiar with the domain. This happens to greater or lesser degrees dependent on the nature of the project and the knowledge and experience of the investigator. Without satisfactory background, investigation run the risk of being unable to ground findings sufficiently, failing to set the correct boundaries on studies and understanding the interactions within the system as a whole. The investigator may also run into issues with credibility when approaching rail staff in the field or reporting on the findings at the end of the process. Familiarisation is more than technical data on how technology works and what people are supposed to do. SMEs provide important data on the reality of the system's performance. This enables intelligent lines of questioning ands saves time and money.

Enabling access and developing contacts

Gaining access to data sources in the form of people and environments is challenging because of the pressures of work requirements and the dangers of "research fatigue". The challenges include where to look, who to speak to, and how to make contact. SMEs are expert in their subject but their years of experience have also taught them how the organisation is structured and operates. They will have formed working and personal relationships with the people around them. This knowledge

and experience is not something that can be obtained from a manual or organisation chart. An SME can find out who the researcher needs to speak with and how they can be contacted. They will be able to make personal introductions to the relevant people to a point where any researcher will be able to make the most of the access they have been given.

Providing primary data

The contribution of SMEs to the research process can provide primary data which supplements data from other sources such as observation and archival data. From a practical perspective, information that may require long periods of time for collection and interpretation in a direct observation setting may become apparent over a conversation with an SME but that is not to say that any methodology employing SMEs is based around casual conversations only. There are a number of formalised methods involving SMEs. Shadbolt (2005) provides a review of knowledge elicitation methods. Many methods relying on SMEs have developed from the move towards studying human behaviour and performance in a more naturalistic way. Critical Decision Method (Klein et al., 1989) interviews make use of accounts from individual SMEs about critical events to probe for various cognitive behaviours and information use. Data acquired from interaction with SMEs using formal and informal methods can ensure the context and depth that provide connections to the real world and make data more meaningful.

Verifying existing data

SMEs are consulted throughout the research and design process. They contribute to content, as well as providing criticism and verifying data. Verification is concerned with whether the findings are "really" about what they appear to be about. Respondent verification is the practice of taking research findings back to the people being studied, such that the researcher may be more confident that they are representative. Outputs from SME interviews may also be corroborated by other SMEs from different operational areas. It is important in both quantitative and qualitative approaches to research that the interpretation of data following its management (reduction, analysis and representation) is verified. The findings become less useful and more obvious if they cannot be verified by an SME as a true reflection of what occurs.

Examples of work with SMEs in the rail domain

Each of the authors of this chapter have been involved in rail projects where SME input has proved crucial to the success of the work. The projects span the key rail functions of maintenance, control, signalling and driving.

Maintenance

During a study of track workers' safety behaviour (Farrington-Darby et al. 2005a) SMEs were consulted at the start of the investigation and again through the analysis phase. The research involved a series of semi-structured interviews with front line operators, supervisors and managers. The topic was sensitive, the interviews were fast moving and context rich. In order for the independent researchers to keep pace with the group interview conversation, follow up the interviewee statements with intelligent questions, and interpret data during analysis, technical background knowledge was required. Consultation with SMEs before the interview stage was arranged to ensure this background knowledge was acquired in the most efficient way possible. Many of the participants

interviewed were SMEs in their technical areas and they were approached when the time came for the development of recommendations in the form of a strategy with practical measures. This study made use of technical knowledge of SMEs as a source of data but also for explanations and accounts of real world situations. When it came to the subject matter of the "safety behaviour of track workers" there were no better experts than the track workers themselves. They were seen as SMEs in their own behaviour and in the safety culture at the time.

Control

A series of studies examining the work and expertise of railway controllers (Farrington-Darby et al., 2006) provides another example of SME involvement as primary data sources. Controllers, duty control managers, trust delay attribution clerks and data quality managers have all contributed to developing a picture of the function of control and the various roles. SMEs have been interviewed using critical decision method (Klein et al. 1989) and task diagram (Militello and Hutton 1998) interview techniques. The value of SMEs in work analysis and investigating expertise is that they bring data that goes beyond the accepted and well documented procedures and practices that can be accessed through manuals and rule books. SMEs are able to reflect on practical situations and provide data on knowledge and skills developed over time that have important implications for safety and performance of the systems with which operators interact.

Driving

In safety critical environments, where the occurrence of situations of interest to the researcher is unpredictable and access to environments is restricted, SMEs' retrospective accounts are often the best source of data regarding operator behaviour. Train driving is a useful example of a rail function where access can be more difficult. It is possible to observe driver behaviour in train simulators but where more naturalistic methods are considered more appropriate cab rides can be arranged. In a study of train drivers' use of line side permanent speed restriction signage (Farrington-Darby et al. 2005b) a combination of interviews and cab rides was used to develop a representation of drivers' use of signage as their experience progressed. Drivers in real driving situations are not able, for safety reasons, to communicate with a researcher travelling in the cab. SMEs such as instructors and assessors are often consulted when studying skilled human behaviour at work because of their expertise and ability to verbalise details behind overt behaviour. Cab rides where the researcher and driver are accompanied by an experienced instructor or assessor provide useful data about actions as they occur in unfolding situations in the real world but where the operator cannot be interrupted or distracted.

A second piece of work with train drivers as part of the RRUK project revealed that using novice drivers as SMEs is also extremely useful to understand the systematic differences between experts and novices. Using novices gives another layer of context and helps to answer the 'why' questions. Recent work on train cab refurbishment invited an expert driver to participate in the redesign process.

Signalling

Whilst investigating signaller expertise in the RRUK project, SMEs were consulted. SMEs provided a rich data source as well as clarifying, explaining, and adding meaning to the familiarisation process. The SMEs were able to give access for the researcher to signalling centres and the signallers themselves and thus allowing the researcher to build up a database of SMEs in different locations

with different and varying levels of expertise. In this particular case the breadth of research was important so a cross section of SMEs was chosen. The SMEs had varying signalling experience covering all types of signalling equipment. Interviews with a range of SMEs revealed their differing working practices along with a number of core elements that contributed to the "expert signaller".

Limitations and problems

The involvement of SMEs as data sources in research is not without its problems and limitations. The term SME is made up of two parts, "subject matter" refers to knowledge and "expert" refers to the person with the greatest depth and breadth of knowledge. Work in the rail domain has demonstrated that different SMEs contribute to the research process in different ways. In order to understand a system it is often necessary to consult with more than one SME. Thus the limitation is that one SME may not provide all the data required and that the researcher needs to be aware of this in order to ensure comprehensive coverage.

When selecting SMEs it may be helpful to substitute experience for expertise as it is difficult to define an expert whereas more experienced drivers will have spent more time in 'deliberate practice' (e.g. Ericsson and Lehmann, 1996) and experienced more unusual events.

When considering train drivers from different routes, driving different trains, working in different company cultures and all working within a larger system it is necessary to consider if breadth is important; then it may be necessary to work with a cross section of SMEs. Alternatively if the researcher is interested in a particular subject matter (e.g. a particular train operating company) then the SMEs chosen will be limited to that particular subject area.

Data from SMEs is open to the criticisms that have been directed at qualitative research and in particular the use of self report techniques. Many of the methods used with SMEs also involve retrospective accounts. There are still questions over whether an SME can provide accurate data about their own cognitive work. There are also questions over the validity of data from accounts of incidents that have occurred in the past. There are flaws in most methods and techniques and it is the researcher's responsibility to manage those issues. Retrospective accounts can be compared in principle to reality or other similar accounts. As for SMEs' abilities to report on their own cognitive processes, it is rare that it is this level of cognition we are interested when we involve SMEs. Many studies where SMEs contribute are at the macrocognition level of investigation (Klein et al., 2003). For example, instead of studying time to recognise a lineside speed restriction sign, which is a study SMEs may be involved in, the researcher would use the SMEs accounts of using speed signs to understand what they use them for in the process of regulating train speed.

The research findings are only ever as good as the SME and the researcher. SMEs can be an unknown quantity. It is also difficult to define what makes a good SME. Often the success of SME involvement may rely on the relationship between the SME and the researcher. Those that stand firmly in the repeatable experimental design camp find this a difficult concept to accept in empirical research. The reality is SMEs and researchers have individual knowledge and skills, motives, biases and personal issues. Some of the more formalised methods reduce the effects of these factors. It is also important for the researcher to recognise these elements in the SMEs with which they are working. Much like the non-response to a questionnaire these can be clues to culture and the nature of the environment being studied. More informal methods can also help in making an unknown quantity less so.

There is an inherent problem with SMEs that often work with researchers and designers. They spend so much of their time outside their operational setting and their role there that they are described as "going native". This is much the same problem as researchers who become part of the community they study. SMEs' skills and knowledge may become outdated the longer they

are not involved directly in frontline operational work. In these circumstances the naïve and uncontaminated responses from SMEs unaware of the research agenda can be lost. These two changes may vary in how problematic they are depending on the nature of the study.

The researcher is sometimes in the position to decide who to involve from a pool of SMEs. The behaviour of the researcher and way in which the research is approached has an important influence on the SMEs' willingness to take part. Protocols such as informed consent and confidentiality help to develop a situation where SMEs speak freely abut sensitive issues. It is helpful to select SMEs that are comfortable and confident when communicating their knowledge and their views. Articulate SMEs are invaluable in studies where methods include exploratory thought and open ended questions but also where deep analysis is required. This may seem like limiting an already small pool of experts but a consultation where no rapport is established and responses are strained can result in an unfavourable experience for both researcher and SME, one that is best avoided if at all possible. It is the researcher's responsibility to determine how many different SMEs are consulted on a given issue to ensure that the conclusions are meaningful and broadly representative of generally held views

There is also the ever present limitation of researcher time and resources, SMEs that are comfortable and confident, able to talk freely and openly, will ultimately save the researcher valuable time. It may be possible to extract very useful information from SMEs that are not as well practised in expressing their views and recounting their experiences but these situations typically involve a greater investment of researcher time. The time required to prepare and facilitate an open researcher-SME interaction will be greater in these situations but may still lead to a very productive interaction. It's not sufficient for the SME to have the requisite experience; they must also be able to communicate it. It is therefore essential that the issue of interest is structured and presented in a form that is meaningful for the SME. Work with drivers has found that using terms such as 'cognition' and 'information requirements' is not appropriate since although drivers understand what is meant by such terms they do not exist in the natural language of train driving. Researchers must think like the SME and frame questions appropriately, otherwise the SME is forced to approach issues from the researcher's perspective and much of the value may be lost.

Guidance for work with SMEs

Some guidance on working with SMEs is implicit in the preceding section. Other guidance is explained here.

There is a sense of pride that can be seen in much of the work performed by rail operational staff. Their willingness to give freely their time, knowledge and experience as an SME is a manifestation of the personal interest railway men and women have in the industry in which they work. Further to this SMEs are willing to provide researchers with their vote of confidence risking their own credibility and reputations when providing researchers details of contacts and introducing investigators to friends and colleagues.

The researcher-SME relationship is more than investigator-subject. SMEs are actively involved in the issues and work being investigated. They are not a knowledge base to be elicited and captured. The researcher-SME process must be two-way, it is important to avoid the exploitation of SMEs. In these situations the SMEs supply information which is used by the researcher and the SME gains nothing or little from the process. The researcher must be prepared to give something back to the SME.

There are several ways in which we have tried to ensure this is the case:

- In setting up the logistics of SME based work their needs are always put first. The venue, time, duration of every interview is set by the SME. Like any good interview practice

the interviewee must feel comfortable. The SME sessions should not interfere with work shift patterns or mean the SME has to travel further than is necessary, even if it means the researcher working in a local depot late at night in the middle of winter!

- Confidentiality is always respected. Trust is a central theme to our SME studies. This is more than ensuring a face is blurred on a picture. It means amongst other things, protecting data with passwords, making contingency arrangements for what to do if potentially difficult issues are reported to the researcher that should be reported to the safety groups in that organisation and ensuring the researcher is prepared to take notes if SMEs do not wish to be recorded. Previous work with SMEs has demonstrated the importance of meeting the SME in an appropriate environment. If the meeting takes place within earshot of the SME's colleagues it may inhibit his/her responses and effect his/her concentration. The SME may speak more freely if not overheard by colleagues, line managers etc.
- Ethical and respectful treatment of SMEs extends to informed consent. Not just to be sure SMEs are aware of the study but to share with them what is being studied and the research process. Experience shows that many SMEs are not only very interested in the research projects to which they are contributing to but that they often have useful tips and advice for the project.
- The researcher must always explain the purpose and nature of the study as well as explaining why the SME was chosen. In addition to this it is beneficial for the researcher to tell the SME something about themselves, which allows for a more balanced process where both sides are giving information about themselves. If the researcher is able to trigger interest in the research project it will make it easier to ensure that the SMEs gain from the research process.
- The theme of respect is carried through into the interview itself. There is often an element of openness to the nature of the research with SMEs. There will usually be a central focus to interviews and observation exercises with SMEs. Experience shows that if SMEs are left to add in their own ideas and issues through conversation this indicates that the researcher respects what they have to say and what they are concerned about. This has often led to important and interesting findings that have embellished studies. Great care must be taken to establish an equal non-hierarchical relationship with the SME, the relationship must be two-way, with associated costs and benefits for both the SME and the researcher. The research process should be viewed as a partnership between the researcher and the SME. In order to ensure a productive meeting the researcher must listen attentively to what is being said, giving the SME their full attention. People react well when they feel they are being listened to. This may mean that on occasion the researcher listens to points raised by the SME with no direct relation to the study.
- The researcher can offer or agree when asked to pay the SME or their organisation for their time. This payment can be financial (although this offer is rarely taken up) or in kind. One author has provided human factors introductory sessions for train operators who have provided SMEs for work analysis.
- Another important contribution from the researcher can be made in the form of feedback. This can occur either as the project progresses or at the end of the project. Feedback emails, posters and symposiums have all been used in the process of SME based research by the authors.
- Verification is another means to deliver feedback, maintain SME involvement and ensure robust and meaningful findings. Taking representations of the interpreted and analysed data to an SME illustrates the researcher's willingness to accept that he or she is not the expert and that there is a commitment to portraying the SME's views accurately.

First contact with SMEs is very important and care must be taken to ensure the right impression is made. In some cases it may be useful to learn about the SME beforehand, i.e. job title, job responsibilities, number of years with company, and number of years in professional field. Although knowing too much may arouse suspicion, reduce cooperation and engender feelings of insecurity in the SME e.g. "Who is this person? How do they know me? What do they *really* want? I'm telling them nothing". Preparation for a meeting with an SME is important, even it will be relatively informal the researcher must have an idea of what he/she wants to achieve. Preparing in advance gives confidence and allows the researcher to make the most of the meeting. It is important to gain trust with the SME, building up this trust and establishing a rapport may be a slow process, relaxed informal initial conversation is important to engage the SME and make them feel at ease. Mutual understanding is imperative in formulating a successful researcher-SME relationship.

Personal appearance can be an important consideration, impressions that may pose an obstacle to the formation of the researcher-SME relationship must be avoided at all costs. Important items to consider include appearance, speech and conduct. Different types of SMEs require subtle differences in the above and it is important that the researcher is sensitive to this. The researcher must be socially competent. Individuals who are socially competent are able to elicit positive responses from others. Of particular importance when working with SMEs is communicative competence; ideas must be communicated clearly and confidently, making sure non-verbal features such as body language and tone of voice don't cause a problem in communication. Previous work has demonstrated the importance of not being afraid to look stupid. It is important for the researcher to request clarification, even of basic issues, and to minimise the chance of being misled.

References

Ericsson, K.A. and Lehmann, A.C. (1996). *Expert and exceptional performance: evidence on maximal adaptations on task constraints*. Annual Review of Psychology, 47, 273-305.

Farrington-Darby, T. and Wilson, J.R. (2006) *The Nature of Expertise*. Fundamental Review. Special Issue. Applied Ergonomics 37, 17-32.

Farrington-Darby, T. Pickup, L. and Wilson, John. R (2005a) *Safety Culture in Railway Maintenance*. Safety Science 43 p. 49-60.

Farrington-Darby, T. Wilson, John. R., Norris, B. J. and Clarke, T. C. (2006) A Naturalistic Study of Railway Controllers. Accepted to appear in Ergonomics.

Farrington-Darby, T., Wilson, J.R. and Norris, B.J. (2005b). *Investigating train driver behaviour: the use of lineside information when regulating speed.* In: J.R. Wilson and B.J. Norris (eds), Rail Human Factors. London: Ashgate, 60-69.

Isbtpi (2005) www.ibstpi.org/glossary.htm (accessed 12/09/05)

Klein GA, Calderwood R, and MacGregor D. (1989). *Critical decision method for eliciting knowledge*. IEEE Transactions on systems, man and cybernetics 19(3): 462-472.

Klein, G., Ross, K.G, Moon, B.M., Klein, D.E., Hoffman, R.R. and Hollnagel, E. (2003). *Macrocognition.* IEEE Intelligent Systems, 18(3): 81-85.

Millitello, L. and Hutton, R. (1998). *Applied cognitive task analysis (ACTA): practitioners toolkit for understanding cognitive demands*. Ergonomics 4(11): 1618-1641

Process Impact (2005) www.processimpact.com/UC/Module_3/Export/data/downloads/glossary.html (accessed 12/09/05).

Shadbolt, N. (2005). *Eliciting Expertise*. In J.R. Wilson and E.N. Corlett (eds.), Evaluation of Human Work: A Practical Ergonomics Methodology, 3rd Edition: 185-218. CRC Press, Boca Raton.

Chapter 27

Work Analysis and Distributed Cognition Representation of Integrated Rail Operations

R. Bye, T. Farrington-Darby, G. Cox, G.R.J. Hockey, J.R. Wilson and T. Clarke

Introduction to the rail human factors research context

Research in the UK rail industry has considered HF issues for several decades, but such analysis has not always been systematic and rail has been seen to lag behind domains where HF research is well established such as aviation, the process industries and the military. That said, since the early 1990s systematic programmes of fundamental and applied work in the UK and abroad have facilitated substantial growth in rail HF. A wide network of HF specialists, including in-house practitioners and supporters as well as external consultants and academic researchers, now bring their knowledge and experience to the rail industry. It is important to ensure that the various approaches contribute to a coherent stream of research, evaluation and practical guidance, but equally important is that the understanding developed in individual rail HF projects is achieved within the context of an integrated system.

Although the boom in rail HF has provided significant opportunities for an improved understanding of rail operations it has also presented challenges for managing the scope and organisation of the work. Such challenges are evident at both strategic and practical levels, and have surfaced from shared experiences of working on rail HF projects and in reviews of existing research (e.g. Hockey et al. 2005; Wilson & Norris, 2005). Some of the challenges we have identified are outlined in the following section.

Shared terminology

One problem that has become apparent in the growing body of HF work in rail (and in other industries) is that individual practitioners inevitably favour particular methods and approaches, frame problems in unique ways and use terminology that may differ from that of colleagues with varied perspectives and experiences. This can result in a divergent set of HF methods, findings and recommendations for a single issue. If varied approaches continue to be adopted it becomes increasingly important to clearly define the terms of reference for a project and to identify an appropriate place to commence the analysis.

Establishing suitable commencement points

For any given problem knowing where to start can be challenging, particularly if the specialist is new to the domain or moves between rail operations and locations. To fully appreciate the working environment, HF practitioners need to recognise the various social, organisational and technical artefacts that combine to satisfy the work goals, however this level of familiarisation can be difficult

to achieve thereby adding cost to projects and taking time from already tight schedules. Domain familiarisation activities often require access to physical environments, people and situations. The key to successful familiarisation exercises and the challenge to the investigator is selection of an approach and method that is sensitive to safety, intrusiveness and research fatigue.

Loss of rail HF expertise

Once projects are complete there is a danger that a lot of the knowledge acquired through in-depth study leaves with the investigator. Research reports and other deliverables can be catalogued for later retrieval online or on CD Rom (e.g. RSSB, 2004), but much of the lost wisdom may not be contained within a final report.

Developing an integrated picture

Thus across the recent and current projects of rail HF, pockets of knowledge and understanding develop as the need for them emerges. What we do not currently have is a place where our understanding of the rail industry as a joint cognitive system is represented. This chapter reports on the progress of a rigorous long-term research project that aims to develop a HF representation of the contemporary context for rail research and to provide a practical tool for HF professionals. The intent is not to review the research landscape but to review the rail domain itself and provide a macrocognitive HF model of integrated rail operations.

Building a macrocognitive representation of rail

> "The concept of macrocognition is a way of describing cognitive work as it naturally occurs"
> (Klein, Ross, Moon, Klein, Hoffman & Hollnagel 2003, p81)

The development of the proposed model of integrated train operations was driven by the identification of difficulties in understanding the complex and interacting nature of the systems that operationalise rail industry goals. The model was therefore designed to assist HF specialists through the provision of context and a description of the human performance at the heart of work in the rail industry.

Capturing and representing work in the rail industry is difficult even when the focus is limited to the interactions between a few core roles – in this case driving, signalling and control. To deliver rail services, work is divided between a range of cognitive, physical, technical and social artefacts. These human and mechanical artefacts are distributed in time and space, and interact as a joint cognitive system to control the dynamic state and constrain possible action trajectories. The cognitive, integrated and distributed nature of the work presents opportunities, challenges and additional layers of *complexity* for the human operators and also for HF specialists who must understand individual functions and the factors that mediate task performance in the rail system as a whole. HF specialists need to consider the *context* of local and distal situations in order to make meaningful recommendations.

A useful representation of rail operations must therefore capture the *cognitive* elements of the work and the *distribution* of functions among multiple agents and locations. From a cognitive systems engineering (CSE) perspective (e.g. Hollnagel & Woods, 1983; 2005) the representation must also account for agent-environment mutuality (Flach et al., 1995; Woods, 1988). Activities are only meaningful in relation to their functional context and the opportunities and constraints presented by the work domain. As a provider of transport services to the public the rail industry also connects with life outside of the work domain. Thus the boundaries of the current analysis

are extended beyond the human-machine interface to include features of the wider environment as additional layers of context.

The model takes a macrocognitive perspective (e.g. Cacciabue & Hollnagel, 1995) as we are concerned with exploring the *results* of cognitive processes and resulting human performance in the context of rail service delivery. This contrasts with microcognitive approaches that study cognition *per se*, as a context-free mental process of inputs and outputs (e.g., Newell & Simon, 1972) or as embodied action (e.g. Clarke, 1997) which is contextually grounded but bounded to inner mental processes.

A description of the model

Our model of integrated rail operations will describe and show: (a) the skills, knowledge and general expertise of key rail functions, (b) the management of information distributed between train drivers, signallers and controllers, and (c) the interaction of these operations within the wider rail context. The approach used to produce the model of integrated rail operations is guided by a framework of cognitive work analysis (CWA; Rasmussen, Pejtersen, & Goodstein, 1994; Vicente, 1999), in so far as the work domain is decomposed into a number of levels of abstraction beginning with and giving primacy to the constraining features of the external environment. The aim of the work analysis is to provide a contemporary HF model of rail operations rather than to support revolutionary design, thus it deviates from the CWA approach that is primarily concerned with enabling *"...workers* [to] *adapt to the unexpected and changing demands of their jobs"* (Vicente, 1999, p.XIV).

For analysis of any system there are a number of candidate levels of granularity. The work analysis was structured to produce progressively detailed representations, starting with the features and components of rail as a domain, and graduating through levels describing rail functions, work activities, tasks and processes (Figure 27.1).

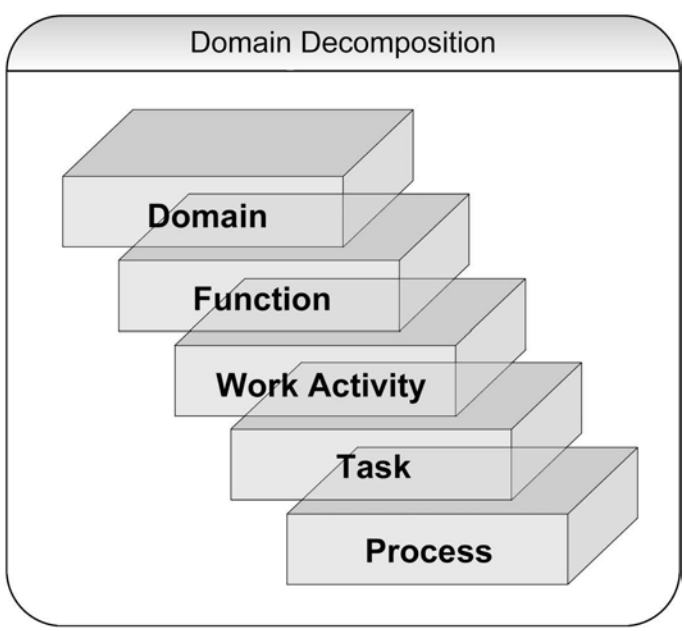


Figure 27.1 Levels of domain decomposition

This form of representation allows the user to move through different levels of decomposition to view a vertical slice of the rail industry system. For example, the rail *domain* is concerned with operating safe and efficient train services whilst at a *function* level, signalling systems control the movement of traffic in a given area. At the *work activity* level signallers (and associated systems) maintain the safe separation of trains and regulate traffic as an example of a common *task*. The *task* level shows outcomes mandated by particular higher-level goals and the *process* describes the particular actions and processes required e.g. a signaller may detect a change in the system state and plan a series of actions to perform traffic regulation. This is an example of a vertical slice through the model. Definitions and typical questions addressed at each level are shown in Table 27.1.

Table 27.1 Questions reflecting the information sought at each level of decomposition

Level of Decomposition	Definition	Typical Questions
DOMAIN	The system being considered in terms of core business, key players and physical nature.	What is the key business for UK rail? What is the basic structure of the industry, and who are the governing bodies and component organisations? Who are the key players? What are the components and features of the rail system?
FUNCTION	The role of the different operational functions in the domain.	What are the roles of each function in rail operations? How are functions organised to satisfy top level goals and contribute to rail service delivery? What are the features of each function's work systems e.g. staff, environments, work settings, technologies etc?
WORK ACTIVITY	The high level categories of activities performed within a functional role.	What are the main activities for a given functional role? What is the nature of the work in a function? Which activities will provide focus for task level analysis? What interactions/collaborations/conflicts occur?
TASK Task Element	Tasks that are performed to satisfy work activity. Task elements	What are the main tasks that are performed? What interactions, collaborations and conflicts occur? What are the key steps (3-6) that make up each task? What is the operator doing in each step and why? What information and knowledge is required? How do operators know what information they need, where it is located and how it can be accessed? Where are the lines of responsibility for each of the tasks?
PROCESS	What are the processes that underlie the tasks.	Can fundamental HF research be used to describe the essence of what is occurring in the task elements? Are there any common or generic processes occurring? Which aspects of the process are most challenging?

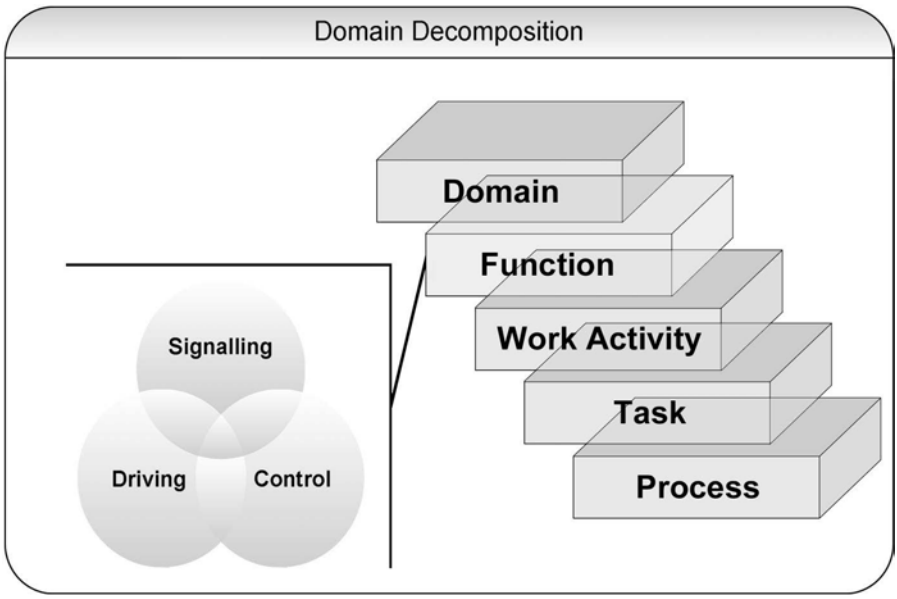


Figure 27.2 Further descriptions within levels of the domain decomposition

Figure 27.2 shows that the representation also affords horizontal analysis within any given layer so that for example in *function*, the nature of information transfer (e.g. what, how, why, when etc.) between drivers, signallers and controllers can be displayed. Within the practical resource constraints of this project at this time the analysis of the full range of features at any given level is prevented. For *function* the focus is on driving, signalling and control since these were identified as three of the core operations in rail service delivery that could be investigated within the resources of the project at this time. Interfaces with staff involved in maintenance, platform management and other work are explored but only in the light of their interactions (or conflicts) with the core functions. These limitations of the model will be made explicit to the user.

The choice of representational technique and the underpinning constructs were informed by desk-based research (e.g. Jackson, 2005) and practical considerations including a user requirements analysis. A key feature of the representational technique adopted is the ability to portray the dynamic links between system artefacts and the availability of multiple descriptions, levels and viewpoints which were designed to reflect the need to examine rail operations from different perspectives. In summary the model is a descriptive (not diagnostic nor predictive) account of integrated rail operations. It embraces and includes the core values of distribution, meaningfulness, context, complexity, interaction, collaboration and macrocognition, and above all aims to represent the sociotechnical system beyond the isolated microcognitive performance of the individual.

Model development and actualisation

With the initial features and framework of the model identified the next challenge lay in the methodological approach used to populate the model. The approach taken in representing work and populating the HF model of integrated rail operations used methods and concepts found in cognitive

task analysis (CTA; e.g. Schraagen, Chipman, & Shalin, 2000) and CSE, including ethnographic field observations (Hutchins, 1995) and structured interview techniques. The key methods used for data capture are all based in a tradition of naturalistic enquiry. This is essential in investigating real world behaviour where context and meaning are key. Observation and interview methods were used extensively and great care was taken to analyse the data in a systematic and traceable fashion. Following each iteration of data collection and analysis the conclusions were discussed with the original operators who provided feedback, modification and verification. The methods and techniques listed in Table 27.2 yielded varying degrees of data for each level of abstraction in the model.

Table 27.2 Methods used for data capture

Method	Key Reference
Direct observation	(Gomm, 2004)
Video captured observation	(Heath & Luff, 2000)
Verbal protocol analysis	(Ericsson & Simon, 1993)
Focus groups	(Sinclair, 2005)
Web searches and document analysis	(Bisantz & Drury, 2005)
Questionnaires. Informal and semi-structured interviews	(Oppenheim, 1966)
Critical decision method interviews	(Klein et al., 1989)
Task diagram interviews	(Militello & Hutton 1998)

The ethnographic approach involved long periods of familiarisation and observation in the field but was essential in enabling the investigators to understand the operational context. It was particularly effective in facilitating an appreciation of the informal social networks, highly coded language, 'tricks of the trade' and other minutiae that underpin driving, signalling and control work in the UK rail industry.

Although data collection and representation were distinct activities both were concerned with the organisation of the representative framework and the difficulty of where to set the boundaries for the limits of analysis. Within the current theoretical context it is argued that there are many overlapping joint cognitive systems since each layer of analysis is subsumed within another, and at any given level the scope can be broadened to include ever more peripheral features. The approach was to collect progressively more detailed data until additional information failed to add value to the scenario or context of interest. As a general heuristic the analysis went no deeper than cognition at the level of problem detection, planning and action to manage uncertainty, and no wider than substantive interactions with peripheral artefacts.

Bringing the model to life

Establishing appropriate methods, and performing the data collection and analysis, is challenging, but represents only half of the task of building a computer-based model of integrated rail operations. The additional component of the project is to develop a web-based software representation, to produce a useful and accessible tool for rail industry practitioners and potentially wider audiences. It is hoped that the representation will provide a rich source of information, thereby allowing

users to understand the roles and interactions of individual functions during normal operations and also perhaps for illustrative scenarios. The interface will display which of the five levels of decomposition is in focus to support an awareness of whole-system issues and facilitate fluid transfer between levels. In addition the user will be able to apply filters to show particular features of an operating context of interest, such as an operator's information requirements, display equipment, and communication strategy.

User requirements analysis

Preliminary user requirements interviews were conducted with industry stakeholders from Network Rail and the Railway Safety Standards Board (RSSB) at a very early stage in the project lifecycle. This analysis was designed to provide the first clues to the required content and functionality of the model so that data collection efforts could be focused appropriately.

A programme of interviews and workshops with other stakeholders will refine design iterations of the user interface and ensure that the software is targeted at end users. It is necessary to have a clear understanding of potential user requirements so that we can orient design decisions towards meeting real user needs and focus development objectives on meeting criteria for quality in use (ISO, 2004).

Model outputs

The model was developed to include features that would address issues such as the need for: (a) shared approaches and terminology, (b) an integrated, whole-system view of rail HF, (c) a repository of HF rail expertise, and (d) transparent commencement points for projects. Thus the key attributes of the model include:

- A generic description of rail operations (perhaps including illustrative scenarios) showing the features that interact to deliver rail services in the UK.
- A representation of the wider features constraining localised activities in the rail system that may not be obvious to individuals with limited domain experience.
- A tool designed to accelerate exploratory work at the start of HF projects so that decisions to focus on narrow elements of the system can be substantiated.
- A resource to promote awareness and facilitate a deeper understanding of HF issues among engineers, operators, project managers and other rail personnel.
- A knowledge management system displaying completed work as well as work in progress and target areas of investigation, perhaps with links to existing reports.

The specific properties of the representation will change as the model is refined but it is hoped that with the provision of the features listed above it will be of practical use to the rail HF community. To satisfy these goals much work is still to be done and a number of challenges must be overcome.

Questions, considerations and challenges

A challenge facing the development and subsequent delivery of the model is the demonstration of robustness. It is extremely difficult for the model to be validated in the traditional sense since it is founded on subjective data from specific real-world contexts taken at a specific point in time. That said, the methods used to collect the data have received extensive use and verification in a range of

domains, and the results of our investigations are subject to both acceptance by the target operators and corroboration by independent subject matter experts.

A related problem concerns the ability to accurately represent elicited knowledge, especially when *what to collect* and *where to stop* depends on the individual researcher. In an attempt to reach a consensus, and control for idiosyncrasies and variability, three researchers shared data to produce collaborative findings – any issues, concerns or unexplained variability were jointly investigated.

Additional problems will come into play over time as new systems are introduced and the model becomes increasingly out-dated. Model maintenance is a very real issue but it is hoped that our approach in describing goals, key functions, generic steps and underlying processes for tasks and activities, will reduce and simplify the necessary revisions. The representation will be delivered with explicit caveats that clearly state its limitations as well as what it does, and does not, offer. The model does not list rail HF issues, is not prescriptive, diagnostic, or complete (at this time) and it is not an expert system.

Before the model is delivered, data verification workshops, additional iterations of data collection and analysis, and scenario-driven simulation workshops are planned. Work on the design of the computer-based representation will be extended through rapid prototyping, usability testing and analysis. We will also continue to solicit comments, criticisms and suggestions as part of the design process. Development of the representation following delivery will require industry support to maintain, revise and extend the model so that it remains useful in the face of developments such as ERTMS and GSM-R. As an online web-based system, resources will also be required to run the website and control access so that inappropriate information is not made available to the general public.

Early signs show that this work analysis approach is well suited for HF research in rail, and that it can deliver the quality of information required to populate a macrocognitive model of integrated rail operations. The next stage is to produce the prototype computer-based model to see whether the data can be meaningfully represented. It is hoped that these early prototypes will be well received by target users from the rail HF community who will play an important role in shaping its development and refinement, and ultimately its acceptance.

Acknowledgements

We are grateful to those who have supported our work through the provision of time, information and advice. Special thanks go to the subject matter experts at Network Rail, RSSB and GNER, without whom this research would not have been possible. The research was conducted for RRUK project B3 and funded by EPSRC with additional funding from Network Rail.

References

Bisantz, A.M. and C.G. Drury (2005). Applications of Archival and Observational Data. In: JR. Wilson and E.N. Corlett, (Eds.) *Evaluation of Human Work*. London: Taylor and Francis: 61-82.

Cacciabue, P.C. and Hollnagel, E. (1995). Simulation of Cognition: Applications. In: J.M. Hoc, P.C. Cacciabue, and E. Hollnagel (Eds.) *Expertise and Technology: Cognition and Human-Computer Cooperation*. Lawrence Erlbaum Associates: 55–73.

Clarke, A. (1997). *Being there: putting brain, body and world together again*. Cambridge, MA: MIT Press.

Ericsson, K.A. and Simon, H.A. (1993). *Protocol Analysis: Verbal Reports as Data*. Cambridge, MA: MIT Press.

Flach, J.M., Hancock, P., Caird, J. and Vicente, K.J. (1995). *Global perspectives on the ecology of human-machine systems*. Hillsdale, NJ: Erlbaum.

Gomm, R. (2004). Observing naturally occurring events. *Social Research Methodology: a critical introduction*. Palgrave: MacMillan: 216-243.

Heath, C. and P. Luff (2000). *Technology in Action*. Cambridge: Cambridge University Press.

Hockey, G.R.J., Bye, R., Farrington-Darby, T., and Wilson, J.R. (2005). *The understanding and integration of human factors across the railway network*. Technical report B2/1 for RRUK.

Hollnagel, E. and Woods, D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. Boca Raton, FL: Taylor & Francis.

Hollnagel E, and Woods D. (1983). Cognitive systems engineering: new wine in new bottles. *International Journal of Man-Machine Studies*, 18(6):583-600.

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.

International Organisation for Standardisation (2004). *Software engineering - Product quality - Part 4: Quality in use metrics*. Geneva (ISO 9126-4).

Jackson, S. (2005). *Preliminary modelling of the work of drivers, signallers, planners and maintenance engineers within the rail industry*. Unpublished technical report for Rail Research UK, B3. Centre for Rail Human Factors, University of Nottingham.

Klein GA, Calderwood R, and MacGregor D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on systems, man and cybernetics*, 19(3): 462-472.

Klein, G., Ross, K.G, Moon, B.M., Klein, D.E., Hoffman, R.R. and Hollnagel, E. (2003) Macrocognition. *IEEE Intelligent Systems*. 18(3): 81-85.

Millitello, L. and Hutton, R. (1998). Applied cognitive task analysis (ACTA): a practitioners toolkit for understanding cognitive demands. *Ergonomics*, 4(11): 1618-1641

Newell, A. and Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ.: Prentice-Hall.

Oppenheim, A.N. (1966). *Questionnaire design and attitude measurement*. London: Heinemann.

Rasmussen, J., Pejtersen, A.M., and Goodstein, L.P. (1994). *Cognitive systems engineering*. New York: Wiley.

RSSB (2004). *Human Factors Research Catalogue* (CD Rom). London: Railway Safety and Standards Board.

RSSB (2003). *Hierarchical task analysis of the UK train driving task*. Prepared by Human Engineering for RSSB.

Schraagen, J.M., Chipman, S.F., and Shalin, V.L. (2000). Introduction to Cognitive Task Analysis. In: J.M. Schraagen, S.F. Chipman, and V.L. Shalin, (Eds.) *Cognitive task analysis*. Mahwah, NJ: Lawrence Erlbaum Associates: 3–23.

Sinclair, M.A. (2005). Participative Assessment. In: JR. Wilson and E.N. Corlett (Eds.) *Evaluation of Human Work*. London: Taylor and Francis: 83-111.

Vicente, K. J. (1999). *Cognitive work analysis. Toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wilson, J.R. and Norris, B.J. (2005). *Rail human factors: past present and future. Applied Ergonomics*, 36 (6):649-660.

Woods, D (1988). Coping with complexity: the psychology of human behaviour in complex systems. In L.P. Goodstein H.B. Andersen and S.E. Olsen (Eds.) *Tasks, errors, and mental models*. London: Taylor & Francis: 128-148.

Woods, D.D., Roth, E.M. and Bennett, K.B. (1990). Explorations in Joint Human-Machine Cognitive Systems. In: W. Zachary and S. Robertson (Eds.), *Cognition, Computing and Cooperation*. Norwood, NJ: Ablex Publishing: 123-158.

This page intentionally left blank

PART VIII
Planning for the Railway

This page intentionally left blank

Chapter 28

Planning, Reasoning and Patterns of Inferences: An Empirical Study into the Reasoning of Staff Planners in The Netherlands Railways

René Jorna and Derk Jan Kiewiet

Introduction

This chapter has a theoretical purpose: putting the knowledge system methodology common-KADS, to an empirical test. It also has a practical purpose: looking at what kind of reasoning strategies staff planners in the Netherlands Railways use. Concerning the first purpose we want to know to what extent theoretical models for generic tasks, such as planning or scheduling, that have been developed by KADS (Schreiber et al., 2000), fit the empirical reality of reasoning. The KADS-methodology has largely been developed for the implementation of knowledge systems. The models behind KADS are mostly theory-driven. Therefore, from an empirical perspective it is also interesting to look at the actual reasoning activities of planners. There is a related strategical issue here. The Netherlands Railways (Nederlandse Spoorwegen, NS) are in a process of redesigning their overall planning processes. Knowing how planners reason and solve planning problems will help in the mutual accommodation of planner and supporting software.

Planning is an activity that takes place in every organization. Planning is done by human planners. Our research question concerns knowing how planners reason and what plans (solutions) they prefer. Although software vendors are claiming that planning can and should be done by software, practice shows that in planning and scheduling situations human planners and not software are leading. Because software support for planning is increasing, it is important to have a better empirical look at what human planners in the real world do. We will discuss some general characteristics of planning in section 2. In this section we will also explain what planning in the NS means.

The KADS-methodology was developed in the 1980's to fasten and to structure the development and implementation of knowledge systems projects (Breuker & vandeVelde, 1994). Sometimes KADS is used as a knowledge-structuring tool after knowledge elicitation is accomplished. From our planning research the most important notions in the KADS methodology are task and inference. KADS, its important concepts and methodology, will be discussed later.

This study is quasi-experimental and exploratory. We will discuss the assignment the planners had to make and the procedure we followed. The data concerns a) the solution the planners choose, b) the time it took to reach the solution and c) the reasoning patterns behind the solutions. We discuss the results of the various kinds of data and give conclusions followed by a short discussion.

Planning in general and planning in the Netherlands Railways (NS)

Planning as a task

Planning is not a well-ordered or well-defined subject (Jorna et al. 1996). Various scientific disciplines study planning, e.g., economics, management and organization studies (Jorna et al., 1996), operations research (Johnson & Montgomery, 1974), mathematics, artificial intelligence (Rich & Knight, 1991) and (cognitive) psychology (Hayes-Roth & Hayes-Roth, 1979). In mathematics and operations research (Sierksma, 1996), the focus of planning research is on formulating sophisticated algorithms that can solve planning problems.

Concerning the nature of planning, we will use the following definition in our research: "Planning or scheduling is attuning (assigning) different kinds of entities (object types and tokens), taking into account different kinds of constraints and working towards minimizing or maximizing various goal functions" (Jorna et al. 1996). For example, assigning (attuning) engine drivers to tasks on railway lines implies assigning "engine driver" to "task" (object types) or driver 1, 2, 3, etc. to different trajectories and hours and days of a week (object tokens).

Despite the importance of the organizational context and the support of algorithms, ultimately a human planner has to assess, finish or adjust the plans him or her self (MacCarthy & Wilson, 2001). From a cognitive perspective, planners are information processing systems (Newell & Simon, 1972). We argue that from a cognitive perspective, not withstanding individual differences in experience, age and training, the reasoning steps of planners are comparable. This is in line with the generic task approach that says that inference sets in generic task execution are similar, despite the different knowledge domains people work with (Kiewiet, Jorna & van Wezel, 2005). We will discuss more details of the generic task approach.

The general planning structure in the NS

The Netherlands Railway (NS) transports one million passengers daily. Passengers are transported in 2700 rail carriages that contribute to approximately 5000 train services each day. The trains run between 384 stations in the Netherlands. NS-Passengers (NS-P) is responsible for the transportation of all passengers and has three departments: Production, Marketing and Staff & Organization. Within the department of Production, logistics is responsible for all planning and scheduling within the NS.

Within the NS, four kinds of planning partitions can be distinguished. The first is between timetables and all other plans or schedules. The second concerns the partition between planning rolling stock and planning staff. The third concerns the partition between local planning and central planning (of stock and staff) and the last concerns the distinction between year plan (long term) and day plan (short term), in relation to rolling stock and staff. Overall, approximately 350 planners are continuously involved in making plans and schedules.

The overall planning process is as follows. The first step is making a timetable. This timetable contains departure and arrival times, the stop-off stations and the type of trains that run between the stations, such as intercity trains, local trains, etc. The timetable is the basis for the rolling stock plan, indicating which kind of trains and how many carriages are put into service. Then planning staff follows, in an abstract sense. Tasks are scheduled, not real train guards or engine drivers, and these tasks are allocated to shifts. No plans are constructed from scratch; existing plans are always used as the basis for new plans and schedules. We conducted the research in the units Year Plan and Day Plan.

Planning rolling staff within the NS

The research was carried out in the department for rolling staff planning at central headquarters: the main office in Utrecht. Within rolling staff planning there are two planning units: Year Plan (YP) and Day Plan (DP). Rolling staff planning refers to planning all train tasks into shifts. An example is the shift of an engine driver. Starting at station A, the driver goes to station B, then back to station A and from there to station C (on the same or different train types). This shift takes 8 hours and has various constraints, such as familiarity of the trajectory, changing trains, etc. For planners in both units, the main "object" they have to plan is a "train task", i.e., an activity on a train running from A to B performed by, for example, a ticket collector. If the time to travel between A and B is two hours, theoretically there are four of these tasks within one shift of eight hours. However, regulations and constraints prevent these kinds of shifts. Work has to satisfy variations in trajectories, types of trains, and time of the day. The difference between Year Plan and Day Plan is related to the time horizon (the planning period) of adjustments to existing plans.

Tasks, inferences and cognition

A task is a set of activities in order to reach a (set of) goal(s) taking into account (a set of) constraints (Waern, 1989). One such a task is planning or scheduling. We follow KADS in saying that the planning task is a generic task.

KADS stands for Knowledge Acquisition and Design System (Schreiber et al., 2000) and was developed over the years as a counterpart for mostly technology-driven software development in the area of knowledge technology. KADS was meant to start with an organization, user and system role perspective. KADS does not restrict the perspective of knowledge engineering to "mining from the expert's head" (Schreiber et al., 2000, p. 15). KADS can be used as a knowledge elicitation instrument, but the methodology is especially developed for other kinds of knowledge gathering. Knowledge can also be found in documents, handbooks and proven instruments. The overall structure of KADS consists of 6 models: the organization, agent, task, knowledge, communication and design model (see Figure 28.1, after Schreiber et al., 2000). No layering is involved in the models, except that the upper three relate to the context, the middle two to conceptualizations and

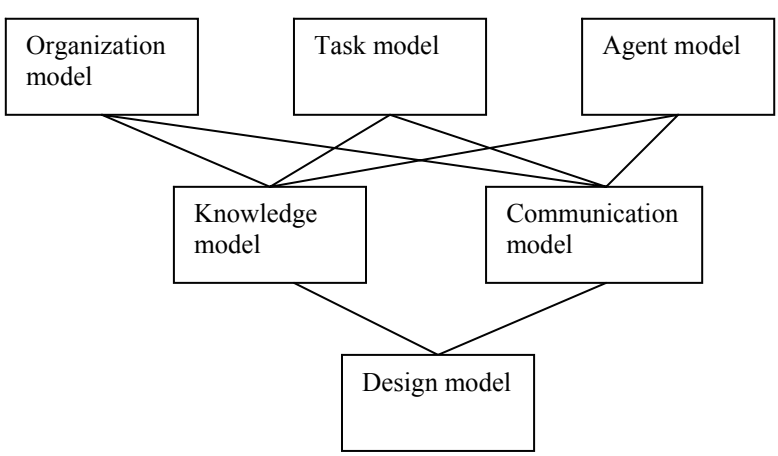


Figure 28.1 Model suite in the CommonKADS methodology

the lower one to the artifact, the knowledge system that has to be developed. KADS main rationale is the structured implementation of a knowledge system. However, before this can be realized the various models have to be created. In this research we only focus on the task and knowledge models.

In CommonKADS a task not only relates to a model, a task is also imbedded in knowledge. A knowledge model consists of three components: a domain, a reasoning strategy and a number of smallest reasoning steps. The knowledge domain concerns the facts and structures of parts of reality. In the NS the knowledge domain concerns facts about trains, tracks, engine drivers or stations. It also concerns structured knowledge, for example that tracks 4 and 5 have a combined entrance to station X. A reasoning activity in KADS consists of goals, decomposition structures and ways to control the task progress.

The third important knowledge component is the smallest reasoning step. In the terminology of KADS this is an inference. "An inference carries out a primitive reasoning step" (Schreiber et al., 2000, p. 104). In KADS this reasoning step is not the same as mental reasoning. "An inference is fully described through a declarative specification of its input and output. The internal process of the inference is a black box, and is considered not of interest for knowledge engineering" (Schreiber et al., 2000, p. 105). The black box is taken for granted; in it something dynamic takes place, which changes an input into an output. From a knowledge engineering perspective this is satisfactory. If the activity inside the block box does the trick, that is enough. However, from a cognitive psychological or ergonomics perspective the inside of the black box is intriguing.

A human actor uses knowledge to fulfill a task. This knowledge consists of two main parts: static knowledge and dynamic knowledge. Static knowledge concerns domain knowledge, whereas dynamic knowledge concerns inferences, tasks and strategies. Inferences are organized in well-structured patterns and are called tasks, such as the diagnosis, the configuration or the planning task. However, a human actor also has the ability to combine tasks or to go beyond the established patterns. This more general reasoning activity is called a strategy. Strategies, tasks and inferences apply to domains.

The empirical research focuses on the planning task or, in terms of KADS, on the fixed inferences patterns related to planning. In the perspective of KADS, planning as a generic task and differences in domain knowledge do not affect the more general task knowledge. A NS planner works with a completely different domain compared to a planner in staff scheduling in hospitals. Their domains have little in common, however their inference patterns resulting in the task and even in a strategy, are supposed to be quite similar. KADS discusses 12 examples of these generic tasks: Classification, Diagnosis, Prediction, Design, Planning, Scheduling and others. For all these generic tasks KADS developed what they call templates. "A task template forms a common type of a reusable combination of model elements" (Schreiber et al., 2000, p. 124). An example of the structure of a KADS inference (Select) is as follows (Figure 28.2).

Select takes Plan specification as input and gives an Action as output (see Tansley & Hayball, 1993, p. 353). KADS formulates various inferences for the different planning tasks. In the situation of the NS we do not think a clear defined planning task exists. Besides planning, the NS planning task also includes replanning, information gathering and evaluation. For this reason we analyzed the set of inferences for our planning and scheduling task in the NS and we ended with 13 inferences. We

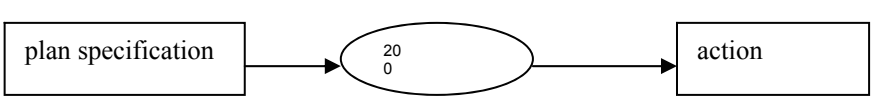


Figure 28.2 Example of a part of the inference structure for planning

also slightly changed some of the names of the inferences. A 14th inference indicates that sometimes planners describe what they see. We named this activity Observe/Describe. The inferences used in our research are: Count, Conclude/Deduce/Critique (from cause to effect), Classify, Evaluate, Hypothesize/Abduce (from effect to probable cause), Identify, Introduce, Decompose, Select, Sort, Combine, Compare and Delete. The choice for this set is theoretical; it is model-driven. The question is whether they match reasoning in reality. In relation to this we look at the resulting plans and at the time it took to reach an answer.

The four assignments in the experiment have different characteristics. Assignment 1 is the most simple and assignment 4 is most complex one. Here we only discuss the results of assignment 2. We have no space for the other assignments.

Based on KADS and our own experiences in planning research we have various expectations concerning the occurrences of inferences. Because counting (addition and subtraction) is an essential part of planning we expect that the inference Count is often mentioned. Based on common knowledge, we also expect inexperienced planners to count more than experienced planners and also more in Day plan than in Year plan. Experienced is defined as more than 2,5 years of experience with planning within the NS.

Because planning in the assignments, in correspondence with practice, often is replanning, we expect that the inference Hypothesize/Abduce is more often mentioned than the inference Conclude. Hypothesize argues from effect to cause and when one has to change an existing plan, one considers the plan as effect and has to trace back what concerning a cause must be changed to integrate the alteration. This holds for all groups of planners. Furthermore, in the KADS list of inferences for planning the inferences Select and Sort play a dominant role. We expect this also to be the case for our groups, whether inexperienced or experienced planners and Year or Day plan.

If replanning is an important aspect of planning in the NS, planners have to decompose an existing plan into its components. Then including the alterations in the assignments, the planners have to attune again the changed object types or instantiations. From an inference perspective, one might expect that the following combinations of inferences show up: Select and Decompose, Decompose and Identify and Compare and Combine.

The assignments are constructed in such a way that more solutions are possible. This is according to planning practice. We classified the solutions for the different assignments in categories such as the most efficient or the most simple. We also measured the solution times of every assignment for every planner. The measurements are rough data in the sense that the precision of response latencies in laboratory settings cannot be accomplished. In accordance with general knowledge of novices and experts, we expect experienced planners to use less time than inexperienced planners. We expect the difference between Year and Day plan to result in faster solution times for Day plan. In the planning practice of the NS time pressure is stronger on Day plan than on Year plan.

Assignment, subjects and procedure

The assignment (only number 2)

In designing the assignments the following criteria are used. First, an assignment has to be representative for both units. If Year plan has more knowledge of a certain domain aspect than Day plan, the assignment is adjusted for this information. Because real names of station might influence the solution, we worked with fictitious names. Secondly, the size of the assignment has to be limited. This is due to the availability of planners and protocol transcription. In theory, the assignments could be solved within 10 minutes. Thirdly, solving the planning problem should not involve information system support. Therefore, the assignments can be made by hand and with

paper and pencil. Fourthly, because we are interested in the possible solutions, each assignment should contain at least two solutions. An important aspect in assignment 2 is whether the planners are willing to violate constraints. Furthermore, every assignment has a time horizon as a starting point. That is to say that a change lasts one day. Because we were interested in the interpretation by the planners of the assignment with regard to the time horizon, we asked planners after their answer for one day, whether they would choose a different solution if the change would last for one month or would be permanent.

The assignment (2):

> Within a month an extra train will be put in for section F-G. The rolling stock for this train is not used further on. The tasks for the chief train guard/engine driver that starts at 9.23 AM has still to be planned. Below, in Figure 28.1, the situation with the new task is depicted. Besides the new task three existing shifts are represented. Your assignment consists in adjusting the shifts to the alteration. If necessary, you may adjust all shifts. We ask you to answer the question for three scenarios: alternation for only one day; alteration for one month; alteration permanently.

In this assignment the question is to put in an extra task using three available shifts. Possible solutions are categorized according to the following. In this assignment, solution 3 is the most efficient and solution 1 the most simple (Figure 28.3).

1. The extra task can be put into the first shift. The problem with this solution is that the minimal transition time of 25 minutes is not met. The transition time is 23 minutes. The transition time has a 2 minutes deficit.
2. The extra task can also be rearranged in the third shift, instead of a passenger's task. In contrast to solutions 1 and 3, the shift is longer.
3. The extra task can also be rearranged into shift 2 after the second task of shift 2 is replaced to shift 1. This means that no shift is lasting longer as well as that the minimal transition time is recognized.

Units and subjects

Two units participated in the empirical research. Unit Year Plan (YP) changes existing plans up until 8 weeks before realization. This unit consists of 12 planners plus 4 (supporting) managers. Unit Day Plan (DP) makes the plans for rolling staff from about 8 weeks up until 36 hours before execution. Their work is comparable to YP, with the exception that YP uses a very precise software program called CREWS, whereas the program for DP is the older and less sophisticated VPT (Vervoer Per Trein). Furthermore, because of the near execution deadline, DP holds more shifts in reserve (the "800-shifts"), because DP has less degrees of freedom and more time pressure to finish the plan. The DP-unit consists of 17 planners plus 4 (supporting) managers.

In both units, the planners are more or less similar in terms of skills and experience. There are no systematic differences between the planners. Most had worked for the NS for more than 15 years. No special training exists to become a planner; skills are acquired on the job. Planners also work alone most of the time. 34 planners participated in the research. One YP planner did not want to make the assignments, and 2 planners of YP did not come to a solution. 12 planners were inexperienced and 22 experienced.

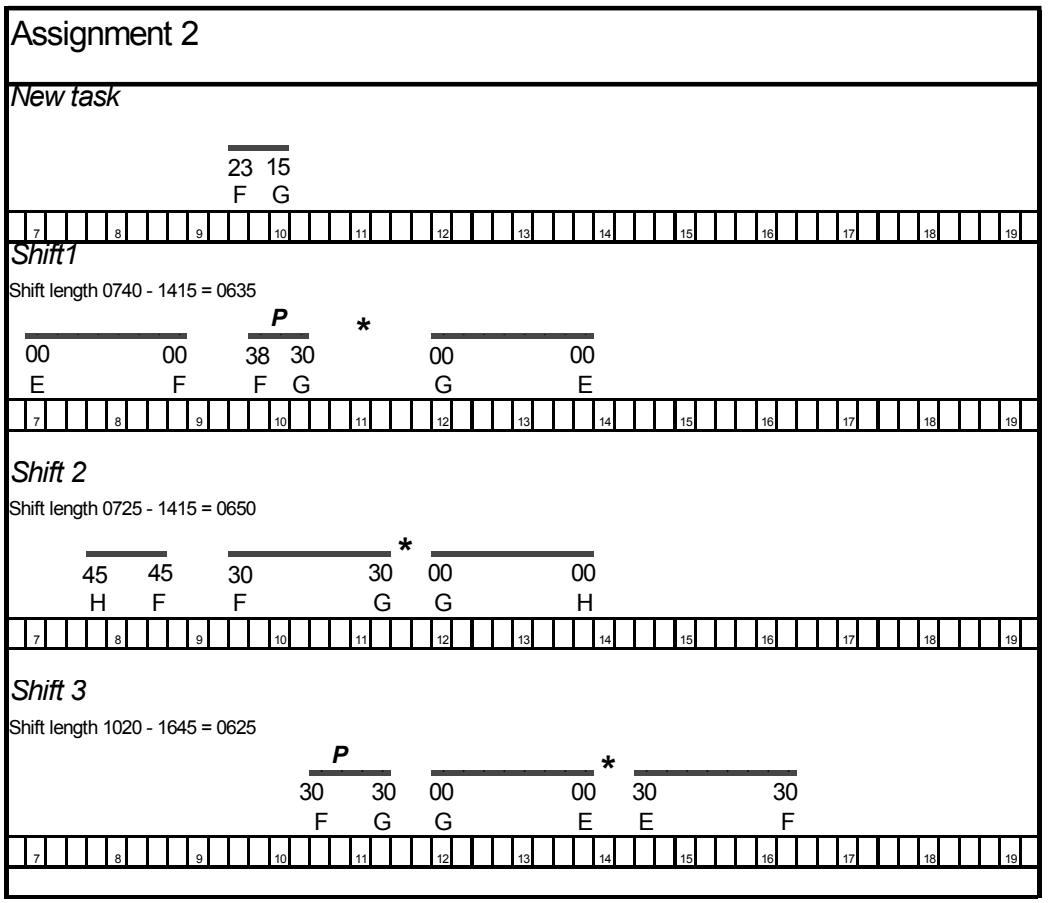


Figure 28.3 Assignment 2, new task to be inserted into three shifts (P=Passengers task; * = lunch)

Procedure

Every planner received the four assignments on chapter with a cover page and a short explanation. To exclude a possible order effect, half of the planners received the assignment order 1 to 4 and the other half 4 to 1. Before starting the assignment the think aloud protocols were practiced with the planners by using a dummy assignment. From the start of the assignments everything was taped on video and audio. The transcribed protocols were used for inference analyses per assignment and per planner. Solutions were noted, for every assignment, for every planner and for the three time horizons of day, month and permanent. Time was recorded from the start of the assignment until the answer to the question about the last (third) time horizon.

Results

Solution times

The first analysis involved the solution times of assignment 2 (see Table 28.1). Table 1 gives the means and standard deviations of these times for the different groups in this study. As we expected, the experienced planners needed less time to find a solution than the inexperienced planners did. Also, experienced planners showed less variance in their solution times than inexperienced planners (sd = 66; sd = 152). There is also a clear distinction in means between YP and DP: 193 seconds versus 132 seconds. As noticed before, this difference was expected due to the observation that under normal circumstances DP is under a larger time pressure than YP and therefore will find a solution faster.

Table 28.1 Solution times

Group	Mean solution time (seconds)	Standard deviation (seconds)
All (n=34)	157	108
Experienced planners (n=22)	133	66
Inexperienced planners (n=12)	201	152
Year Plan (n=14)	193	115
Day Plan (n=20)	132	98
Solution 1 (n=22)	144	122
Solution 2 (n=2)	185	57
Solution 3 (n=7)	184	97

Table 28.2 Solution choice

Group		Solution		
		1	2	3
All	n	22	2	7
	%	71	6.5	22.5
Experienced planners	n	14	1	5
	%	70	5	25
Inexperienced planners	n	8	1	2
	%	72.7	9.1	18.2
Year Plan	**n**	5	2	4
	%	45.5	18.2	36.3
Day Plan	**n**	17	0	3
	%	85	0	15

We also looked at solution times for different solutions. We deduced earlier that solution 1 was the simplest solution, and we see indeed that this solution was found faster than the two other solutions. The mean solution time and standard deviation for the second solution should be interpreted with caution: only two planners gave this solution.

Solution choice

The solutions chosen by the planners are given in Table 28.2. Solution 1 was preferred (71% of the planners, as opposed to 6.5% for solution 2 and 22.5% for solution 3). The choice for a solution was almost the same for experienced as for inexperienced planners. However, YP and DP were quite different in their solutions: where solutions 1 and 3 were almost equally chosen by YP, solution 1 was by far the preferred solution for DP.

Inferences and tasks

Only the inferences for day horizon were analyzed. First, we transcribed the protocols for each planner and each assignment. Then each assignment was scored in terms of occurring inferences for all planners by one researcher. Four researchers accomplished this task. The researchers discussed and trained themselves in scoring the inferences beforehand. Scoring inferences was in line with the sequence of the protocol statements (Ericcson & Simon, 1984). A list of inferences resulted for assignment 2. Then we calculated the number of planners that used an inference. The results of this analysis are in Table 3 and Figure 28.4. In Table 3 we see 8 columns. Column 1 gives the frequencies for all. Columns 2 and 3 give the frequencies for experienced and inexperienced; columns 4 and 5 for YP and DP and columns 6, 7 and 8 for solutions 1, 2 and 3.

Table 28.3 Frequency of planners that uses inferences

Inference	All (n=34)	Exp. (n=22)	Inexp. (n=12)	YP (n=14)	DP (n=20)	Sol.1 (n=22)	Sol.2 (n=2)	Sol.3 (n=7)
Count	22	13	9	12	10	10	2	7
Deduce	33	21	12	14	19	21	2	7
Classify	10	6	4	3	7	5	0	4
Evaluate	7	4	3	4	3	4	1	2
Abduce	24	14	10	12	12	14	2	6
Identify	26	17	9	12	14	16	1	7
Introduce	1	1	0	1	0	0	0	1
Decompose	14	8	6	6	8	7	0	6
Select	-	-	-	-	-	-	-	-
Sort	-	-	-	-	-	-	-	-
Combine	27	17	10	12	15	17	1	6
Compare	14	10	4	6	8	11	0	2
Delete	2	1	1	1	1	1	0	1
Observe	31	19	12	13	18	19	2	7

We see that the inferences Select and Sort were not used by any of the planners. This is contrary to our expectation, based on KADS, that Select and Sort would have a prominent role in the inferences.

The other expectation we had (Abduce is more often mentioned than Deduce) is also contradicted by the results: Deduce is observed within 97.1% of the inference patterns, whereas Abduce is observed within 70.6% of the patterns (see Figure 28.4-A). These results hold when the whole group is split up into subgroups in several ways (Figure 28.4-B, C and D). There were some striking differences between YP and DP: the inference Count was used less often by DP than by YP (50% versus 85.7%), as was Abduce (85.7% versus 60%). Also, differences in inferences used were observed for the various solutions. Count was used less for solution 1 than for solution 3, as were Classify, Abduce and Decompose (see Figure 28.4-D).

Next, we looked at the occurrences of two subsequent inferences in the protocols. As with the single inferences, we calculated the number of planners that used a certain pair of inferences at least once. Table 28.4 and Figure 28.5 give the results for inferences that were used by 10 or more planners (i.e. by one third of the planners).

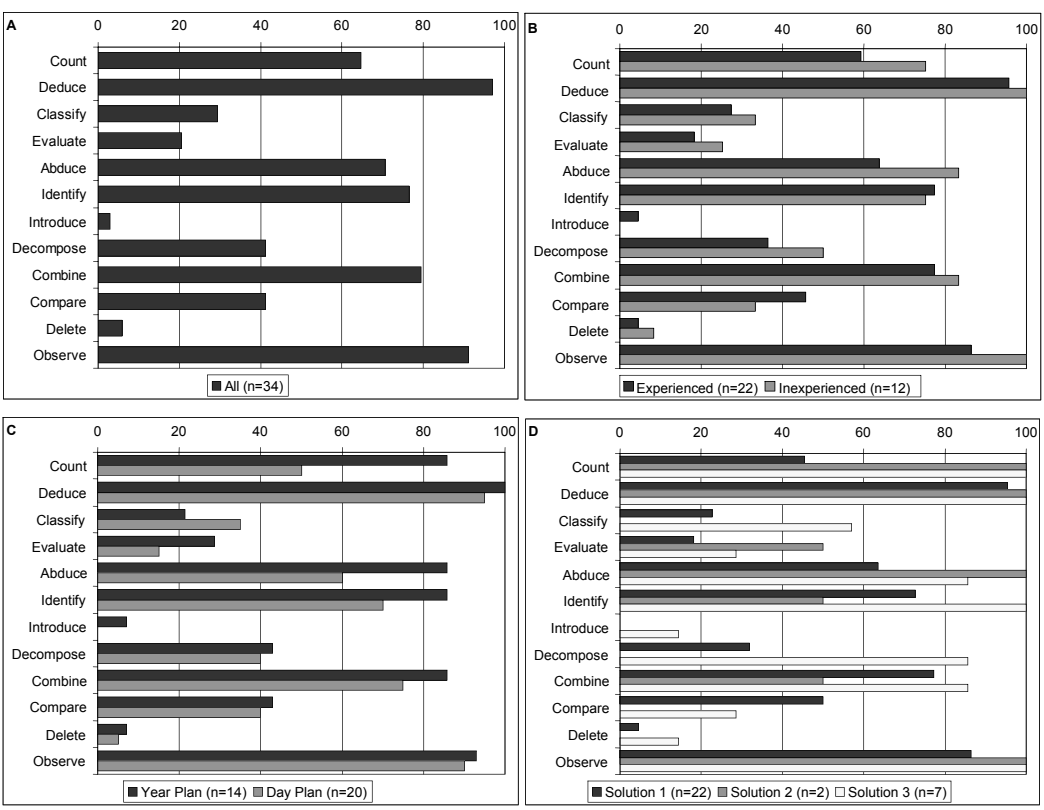


Figure 28.4A–D Percentage of planners that uses inferences

Table 28.4 Frequency of planners that uses the sequence of inferences (only sequences that were used by 10 or more planners are shown)

Sequence of inferences	All (n=34)	Exp. (n=22)	Inexp. (n=12)	YP (n=14)	DP (n=20)	Sol.1 (n=22)	Sol.2 (n=2)	Sol.3 (n=7)
Count/Deduce	17	12	5	9	8	9	2	5
Deduce/Deduce	11	5	6	6	5	7	1	3
Deduce/Abduce	12	7	5	7	5	8	1	3
Deduce/Identify	11	7	4	6	5	5	0	5
Deduce/Combine	10	7	3	6	4	6	0	3
Deduce/Observe	24	13	11	12	12	13	2	6
Abduce/Deduce	11	6	5	7	4	6	2	3
Decompose/Combine	13	7	6	6	7	6	0	6
Combine/Count	10	6	4	4	6	4	0	5
Combine/Deduce	12	8	4	7	5	5	1	3
Observe/Deduce	18	11	7	10	8	10	1	6
Observe/Abduce	12	6	6	6	6	7	2	2
Observe/Observe	11	7	4	6	5	5	1	4

All other possible combinations of inferences (with frequencies <10) were left out of consideration. We see that overall the combinations Count-Deduce, Deduce-Observe and Observe-Deduce are used the most (at least by 40% of the planners). We also found differences between experienced and inexperienced planners (Figure 28.5-B) for the inferences Deduce-Observe (a difference in usage of 33%); for YP and DP (Figure 5-C) for the inferences Deduce-Observe (26%), Observe-Deduce (31%) and Count-Deduce (24%); and for the different solutions (Error! Reference source not found.-D) for Count-Deduce (31%), Deduce-Identify (49%), Deduce-Observe (27%), Decompose-Combine (58%), Combine-Count (53%) and Observe-Deduce (40%).

Conclusion and discussion

Empirical study of the reasoning of planners requires a conceptual framework that can be applied. KADS is such a framework. In line with the expectations, we found that experienced planners use less solution time than inexperienced planners and that Day plan uses less time than Year plan. For solution outcomes the simple, constraint-violating, outcome is preferred to the more complex correct outcome. This is more the case for Day plan than for Year plan. For the inferences we often found Count, but neither Sort nor Select showed up as we predicted from the framework of KADS. Because we found Deduce alone or in combination more often than Abduce we conclude from this that the planners saw the assignments as planning and not as replanning. To our

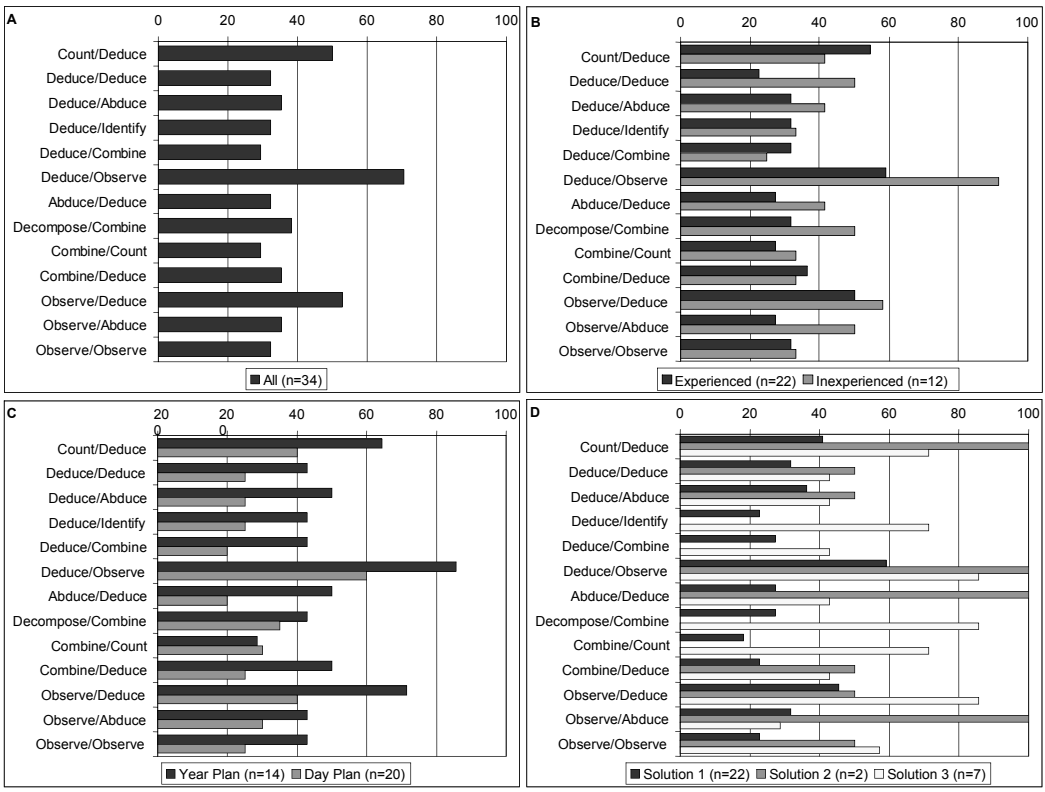


Figure 28.5 A–D Percentage of planners that uses the sequence of inferences (only sequences that were used by 10 or more planners are shown)

disappointment many of the other inferences – reasoning steps – that we expected to take place, did not systematically show up in the protocols.

The results show us three things. First, the best solution may be different for various planners. Software support and algorithms should take that into account. Secondly, the organizational context influences the solution outcome, times and reported inferences. Year plan and Day plan, although working on the same plan, reason differently in a systematic way. Thirdly, a better understanding of inference patterns opens the way for interactive planning support systems, i.e., systems that adjust to the reasoning strategies of humans. The elicitation of inferences of planners is a necessary step in developing such software.

Acknowledgements

We like to thank Erwin Abbink (NS), Thomas de Boer (RuG), Luit Gazedam (RuG), Geert Helder (RuG) and Harrie de Vaan (RuG) for their efforts in practice in realizing this research project.

References

Breuker, J. and Velde, W. van de (eds) (1994). *CommonKADS library for expertise modeling: reusable problem solving components.* Amsterdam: IOS Press.

Ericsson, K.A. and Simon, H.A. (1984). *Protocol Analysis: Verbal Reports As Data.* Cambridge: MIT Press.

Hayes-Roth, B. and Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science, 3*, 275-310.

Johnson, L. A. and Montgomery, D.C., 1974. *Operations research in production planning, scheduling, and inventory control*. New York/London: Wiley.

Jorna, R.J., Gazendam, H., Heesen H.C.and Wezel W. van (1996). *Plannen en Roosteren: Taak-gericht analyseren, ontwerpen en ondersteunen*. Lansa: Leidschendam.

Kiewiet D.J., Jorna, R.J. and Wezel, W. van (2005). Planners are more different than we think: an analysis of domain representations using Multi Dimensional Scaling. *Journal of Applied Ergonomics*. (in press).

MacCarthy, B.L., and Wilson, J.R. (eds). (2001). *Human performance in planning and scheduling*. London: Taylor & Francis.

Newell, A. and Simon, H.A. (1972). *Human Problem Solving*. Englewood Cliffs, New Jersey: Prentice-Hall.

Rich, E. and Knight, K., 1991. *Artificial Intelligence*. New York: McGraw-Hill.

Schreiber, A.T. et al. (2000). *Knowledge Engineering and Management: The CommonKADS Methodology.* Cambridge: The MIT Press.

Sierksma, G., (1996). *Linear and integer programming: theory and practic*e. New York: Marcel Dekker.

Tansley, D.S.W and Hayball, C.C. (1993). *Knowledge-based systems analysis and design: a KADS developer's handbook.* New York: Prentice Hall.

Waern, Y. (1989). *Cognitive Aspects of Computer Supported Tasks*. Wiley & Sons: Chichester.

This page intentionally left blank

Chapter 29

Implementation of a Test System for Evaluation of New Concepts in Rail Traffic Planning and Control

Bengt Sandblad, Arne W. Andersson, Arvid Kauppi and Johan Wikström

Background

Tomorrow's train traffic, with higher speed, more frequent traffic, and many independent traffic operation companies, requires new principles and technical solutions for efficient train traffic control. Today's control systems are often designed to support the operator's possibilities to react on alarms, conflicts and disturbances and to solve problems and conflicts that occur. However, in order to perform efficiently, operators should be able to follow the dynamic development of the traffic system over time and prevent disturbances. In order to achieve this, we must change the control paradigm from low level technical control tasks into higher level traffic planning tasks. This is done by replacing the more traditional control tasks by real-time re-planning tasks [Sandblad et al. 2002].

Advanced laboratory prototypes have already been implemented and tested. By connecting user interface prototypes to a train traffic simulator [Sandblad et al. 2000] it is possible to perform experiments with the design of new user interfaces and decision support tools, and to test new control strategies for the train traffic control operators. In a laboratory at Uppsala University a control room environment has been set up, where it is possible to evaluate the prototypes. It is also possible to perform more limited experiments on site, at local traffic control centres, using a portable computerized simulation and evaluation tool.

Earlier research studies

Our research has been based on a very detailed description and analysis of how train traffic is controlled today, the mental models of the dispatchers and the strategies they use for decisions and control actions. The research has consisted of mainly the following steps:

- Observations and interviews with dispatchers and other professionals at the traffic control centres. Analysis of the findings and identification of problems and development areas.
- Seminars with experienced and responsible professionals from the national rail and traffic control administrations. Here the visions and restrictions for future development of control systems have been specified.
- Iterative specifications and evaluations with the help of a working group consisting of experienced operative traffic control professionals.
- Tests and evaluations in a laboratory control room environment using a train traffic simulator system.

In order to support real-time planning of train traffic we provide the traffic controller with an interactive computerized time-distance graph. Prototypes of new user interfaces that support the new control strategy have been designed, implemented and preliminary tested in the laboratory environment at Uppsala University. The interface is designed to integrate all decision relevant information into one unified interface and to support continuous awareness of the dynamic development of the traffic process [Kauppi et al. 2005].

The computer based time-distance graph is designed in such a way that it visually supports the operators' situation awareness of the current status and the projection into the future [Endsley 1996]. The user interface, with its planning view, can support early detection of upcoming conflicts, identify possible re-planning alternatives and their predicted effects. The new control strategy has a potential to better support the traffic controller's ability to handle continuous re-planning, with the goal to always have a functional traffic plan at hand. This plan can be automatically executed except when a technical malfunction hinders this. Automatic functions that support execution of the traffic plan must be predictable and easy to understand. The automatic functions must never change the controllers' traffic plan but are only allowed to strictly execute what are specified in the traffic plans. The traffic plans mainly consist of time table and track usage information. Obvious interface design, easy to interpret concerning the automation, helps to keep the human in-the-loop and to avoid automation surprises [Bainbridge 1983]. By re-planning, the operator is in control of what the automatic function will do and when. Hence, the operator is continuously in full and active control.

We have also evaluated different approaches to design of decision support systems in operative train traffic control [Hellström et al. 1998, Kvist et al. 2002]. We have found that more advanced automated decision support systems are today not a realistic alternative for different reasons. More research and development of methods are needed in this field. We have decided to focus our efforts on supporting the controllers through better presentation of information, improved information observability and quality, help with early detection of conflicts and disturbances, identification of possibilities and limitations for re-planning and evaluation of effects of re-planning alternatives.

Towards an operative test implementation

Experiments with the new control strategy, operator interfaces, decision support systems and automatic execution functions have been performed in our laboratory environment with satisfying results. Many important aspects can be investigated in the simulated environment at Uppsala University, but some issues must be evaluated in a real operative environment. To work in a laboratory environment, and to control a simulated traffic system, will always mean that we have simplified the situation. The real traffic system is more complex and stochastic compared to our laboratory models. The work tasks of the controllers are also more complex and diversified then what we can create in the laboratory. It will never be possible to evaluate all relevant aspects of the new control system in a pure simulated environment.

We also face large practical and economic problems when the laboratory prototypes will be implemented and deployed as a part of the real train traffic system. It will not be possible to develop a complete traffic management and control system only for test purposes, but we must implement the new control functionalities on top of the existing basic control infrastructure. Our prototype system has a focus on planning, re-planning and automatic execution of control commands and all the other tasks of the traffic controllers are not supported by the functionality in our new system. When malfunctions in the infrastructure hinder the automatic execution the controllers must go back to the old control functionality. Thus, the new control system, control by re-planning, must be implemented as a complementary module.

The questions for the STEG project are:

- Does the new control paradigm, principles, tools and interfaces really contribute to more efficient traffic control and a better work environment for the traffic controllers?
- Is it possible to implement the new control principles and tools as an integrated part of the already existing infrastructure? What of the original ideas must be changed in order to make the implementation possible and economically realistic?
- How can our research prototype environment support requirement specifications and evaluation for real implementation and development?

Through the implementation and evaluation of an operative test system we hope to reach several different goals. We can perform tests over longer time periods, and in this way evaluate the system when the users have become experienced. We hope to detect eventual difficulties e.g. concerning planning activities, interface details or design of automatic functions. We also hope to learn more about the practical limitations that the present systems, infrastructure, traditions and culture mean for the transformation of the control paradigm and for implementation of the new functionality. Finally, we also need more knowledge and experiences about how safety aspects are influenced by the introduction of the new control systems.

The STEG project

Project organisation

The STEG project has been divided into several different steps or phases. On a high level the following main steps have been identified:

- Benefit-cost analysis.
- Risk analysis and assessment of the project as such, including backing procedures if certain parts of the project fail.
- Identification of test site. A test site was selected that fulfilled a number of requirements, e.g. availability of different track structures, single track, double track, mixed traffic types, more complex stations, connections to other traffic control regions, availability of input data for track diagrams, technical specifications etc.
- Requirement specification for the test system.
- Several different technical investigations concerning compatibility, availability of input data, technical platform, technical performance, safety, security, communication etc.
- Specification and test of control algorithms, e.g. for the automatic execution of control commands from the traffic plans.
- Technical development according to specified development model, including a user centred process.
- Implementation.
- Test and evaluation.

Need for complementary research to support STEG

From our earlier research and experimental studies in the laboratory, we have a more or less complete solution for the new proposed control paradigm, control by re-planning. When the

operative test system is being specified and developed, it is not possible to implement the full prototype system. Some parts are not relevant to the operative test environment, other parts are not possible to implement because of limitations in the existing infrastructure etc. During the specification and development phases we must identify problems and questions concerning the detailed functionality and design of the system that must be solved. Some of the more important, and from a research point of view most interesting, problems are:

Early detection of disturbances and conflicts

In order to obtain enough precision in the conflict detection and presentation we need a good model of the interlocking system, train running times and dynamic relations between train routes. Today, the models and the input data have a rather low quality. Our solution is to simply show conflicts in the interface and let the traffic controller analyse the degree of conflict. In the future, when the models have a higher quality, we can develop more advanced methods for conflict predictions. There are different reasons for the poor quality. One is the problem to convince the railway companies to deliver high precision data about their transport to the control centres.

Automatic execution functions

The purpose of the automatic execution functions is to generate and deliver control command sequences to the underlying control system in time. Because of the lacking quality in traffic predictions, the algorithms must have large margins. This results in a non optimal performance. Measures to improve precision in data are most important, since this can significantly improve the total performance. The traffic controllers could be allowed to update the traffic plans closer to real-time. The actions of the automatic functions must be clearly shown in the user interface in order to support good situation awareness and avoid automation surprises.

Notations in the graph

It is necessary to be able to make notations in the dynamic planning graph. The notations will define all restrictions for how tracks can be used in the re-planning activities, i.e. the notations will document all non train reservations of track capacity. Things which will occupy track segments can be maintenance or extraordinary vehicle movements. All such events must be noted and made observable to other traffic controllers. Notations must be easy to make, identify and observe.

Detailed track diagrams

Today it is unclear how much track diagram information is needed for the controllers. The presentation must be detailed enough to support the understanding of conflicts, status of the infrastructure, restrictions and degrees of freedom in the re-planning activities etc. Different levels of detail in the presentation will be tested.

Traffic control in complex stations

The design of support systems, e.g. interface elements and decision support functions, for traffic control in more complex stations, has not been investigated enough. In our operative test environment we will not cover complete traffic regions, and because of that we do not need advanced solutions in this respect. On the other hand we will not be able to evaluate the total performance.

Work environment and design of the workplace

We have a detailed picture of what kind of presentation system that is needed for optimal performance and a good work environment. This should require very large presentation areas with high resolution and quality and without disturbing frames. Because of economic and practical reasons, we will not be able to implement an ultimate technology. The exact lay-out of the work place will be made so that the individual traffic controller has an optimal solution concerning ergonomics and information overview. The design of the interface has been thoroughly evaluated in the laboratory environment. It is also important to design the total work place so that efficient cooperation between traffic controllers is supported. The solutions must also be adjustable to different traffic situations and to different times of the day.

Remaining design and implementation questions

There are some important problems concerning the available technical solutions that we will not be able to solve within the STEG project. This will restrict our possibilities to develop efficient support systems and to evaluate the new control paradigm. The two most severe limitations in this respect are:

- Train speed and position. There are no technical solutions available that now allow us to observe train speed and position with high precision. Today the best precision is the identity of the block section. We will investigate possibilities to improve the precision by using running time predictions. It is important to not implement presentation of data that do not fulfil security requirements. In the future positioning systems with a high precision will be available.
- Communication with train drivers. Today we are not able to automatically communicate new traffic plans to the train drivers. This means that the train drivers will drive according to old and obsolete plans. By doing so they will not be able to perform optimally. In very urgent cases the controllers can phone the train drivers to inform them about changes in traffic plans, train stops etc. It is also not possible for the train drivers to easily inform the traffic controllers about late departures, speed restrictions caused by machine problems etc. In the future we will have efficient communications links for such purposes, with ERTMS/ETCS.

Test environments and procedures

In our previous research activities we have developed laboratory, software prototypes of the new train traffic control systems. Three basic configurations of the test environment are in use today.

First there is the stationary workplace prototype environment at Uppsala University, used mainly to test different workplace setups, interface prototypes and to perform presentation of the research. Secondly, the portable test environment used for performing case studies of different graphical user interface setups and interaction strategies. Thirdly there is an online test environment that is being used by traffic controllers' for online-training prior to case studies, using the portable test environment. The online test and training environment (Simson-online) is available at http://www.it.uu.se/research/project/ftts/.

To start investigating whether or not re-planning in a computerized time-distance graph results in reasonable solutions, experiments were conducted. As an example of this, 12 professional train dispatchers were presented with the same preconditions and they all got the assignment to solve the same disturbed traffic situation. Eight out of the 12 dispatchers used the computerized graph to re-formulate a new valid traffic plan. The other four dispatchers performed, as they are used to, the re-planning task in a paper based time-distance graph. On a logical level there was no

significant difference between the solutions produced with the computerized or with the paper based graph. Basically two different variants of solutions were produced. The majority of the train dispatchers chose an efficient but risky solution, resulting in minimized delays. The others chose a softer approach in solving the problem that resulted in much more delays, but that probably could be considered to be a more robust solution. Initial indications were that there is no significant difference in solutions produced through planning with a computerized graph in comparison to the paper based graph used today. In connection to this experiment, by interviewing and using questionnaires, we also got an opportunity to see how the train dispatchers experienced the basic concepts of the new control strategy. Reactions were very positive to the concepts and ideas, even though the experiment only allowed for limited interaction (changing the duration of train stops) in the computer based time-distance graph. Further tests in the laboratory environment are continuously being performed.

It has been found that the experimental test environment is a very powerful tool also in supporting communication between STEG project members and future users of the operative system. The test environment can visualise the requirements and help traffic controllers to understand what future work will contain and look like. They are able to contribute better to the specification and evaluation activities.

The STEG project has specified a plan for the evaluation of the total operative test process. A base-line registration of important aspects on performance and subjective experiences are being performed prior to the implementation of the new system.

There are several important aspects concerning benefits, efficiency, functionality and work environment that we will not be able to evaluate from the operative tests. We are only implementing our new control system in one traffic segment. The full benefits of the new control principles and the implementation will only be achieved when a more complete traffic area is involved. We will also not be able to implement some basic technical prerequisites for a full evaluation. Examples of the more important things which will still be missing are an accurate system for indication of train position and speed and a fully functioning communication channel between train drivers and the traffic controllers.

Conclusions

We have since several years developed and evaluated new principles, interfaces and decision support systems for future train traffic control. When an operative test system now is being specified and developed, there are many restrictions that must be taken into consideration. In order to support the specification of requirements and to follow the process with development, deployment and evaluation of the test system, supportive research questions have been formulated. We have found that the traffic control laboratory environment can help us to perform studies that can guide good design decisions.

The specification and planning phase for STEG is now completed. The work to develop the system has been initiated. The research activities that are needed in order to evaluate design decisions and necessary exceptions from the original and evaluated research design are continuously being specified and effectuated. The STEG project as such is monitored and evaluated through an action research procedure.

Acknowledgement

This project has been financially supported by the Swedish National Rail Administration. We especially thank all professional traffic controllers who have been engaged in the research activities.

References

Andersson A.W., Sandblad B., Hellström P., Frej I., Gideon A. (1997). A systems analysis approach to modelling train traffic control. *Proceedings of WCRR'97,* Florence, Italy 1997.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19,* 775-779.

Billings C.E. (1991) Human-centered aircraft automation: A concept and guidelines. *NASA Technical memorandum 103885*. Moffet Field, CA: NASA Ames Research Center.

Endsley M.R. (1996). Automation and situation awareness. In R. Parasuraman and M. Mouloua (Eds), *Automation and Human performance: Theory and applications* (pp. 163-181). Mahwah, NJ: Lawrence Erlbaum.

Hellström, P., Sandblad, B., Frej, I., Gideon, A. (1998). An evaluation of algorithms and systems for Computer-Aided Train Dispatching, *Computers in Railways VI,* Wessex Institute of Technology, 1998.

Kauppi A., Wikström J., Hellström P., Sandblad B., Andersson A. W., (2005). Future train traffic control, control by re-planning. In J.R. Wilson, B. Norris, T. Clarke, and A Mills (Eds), *Rail Human Factors supporting the integrated railway* (pp. 296-305). Ashgate Publ. Ltd. 2005.

Kvist T., Hellström P., Sandblad B., Byström J., (2002). Decision support in the train dispatching process. *Computers in Railways VIII*, Wessex Institute of Technology, 2002.

Sandblad B, Andersson AW, Byström J, Kauppi A. (2002). New control strategies and user interfaces for train traffic control. *Computers in Railways VIII*, Wessex Institute of Technology, 2002.

Sandblad B. et al. (2000). A train traffic operation and planning simulator. *Computers in Railways VII*, Wessex Institute of Technology, 2000.

Wikström J., Kauppi A., Hellström P., Andersson A., Sandblad B. (2004) Train traffic control by re-planning in real-time. *Computers in Railways IX*, Wessex Institute of Technology, 2004.

This page intentionally left blank

Chapter 30

Task Oriented Support for Train Shunting Planning

Wout van Wezel

Introduction

Train shunting scheduling for the Netherlands Railways is a complex planning problem that is performed by about 130 full-time planners. It concerns the planning of day-to-day shunting operations at the large stations in the railway network. Input for this planning is based on the long term schedule of arrivals and departures at a station, rolling stock and track maintenance schedules, and specific circumstances that have to be taken into account when preparing the plan for a particular day or night. Currently, the planners mainly work manually. The plans are made and revised on paper first, after which the outcome is put in the computer. In our study, the central question is: how can these planners be supported in their task with an advanced planning system?

Advanced planning systems offer many advantages. Graphical manipulation of the plan, real time constraint checking, plan evaluation, real time data exchange, and automatic plan generation are functions that are commonly offered by advanced planning systems. Unfortunately, train shunting planning gets very little attention in literature. From a somewhat abstract perspective, shunting can be depicted as a combination of inventory location planning, routing, and staff scheduling. This, however, does not do justice to the specific aspects of train shunting planning. As a consequence, it is not trivial how these functionalities could be applied for shunting planning. To overcome the lack of knowledge about how to support shunting planning, we apply a task oriented approach to resolve our research question. In a task oriented approach, an analysis of the task as it is performed by the planners is input for the design of the system. In the chapter, we describe the results of the task analysis and the developed prototype.

Advanced planning systems

Advanced planning systems are tools that help in creating detailed plans and schedules. Such systems are basically like any other kind of decision support system consisting of a database, a user-interface, a model base with algorithms, and links to external systems. The database stores all information, the user-interface shows the stored information and provides possibilities to update and manipulate the information, and the model base contains generation algorithms. Scheduling specific functionalities are a constraint checker, a goal evaluator, and a blackboard where multiple solutions and partial solutions can be stored for later retrieval. Figure 30.1 shows a typical architecture using these components. Literature contains many descriptions from different perspectives (e.g., empirical approaches, system engineering approaches, cognitive approaches) for the various components. The two parts that interact with the human planner are the user interface and the algorithms.

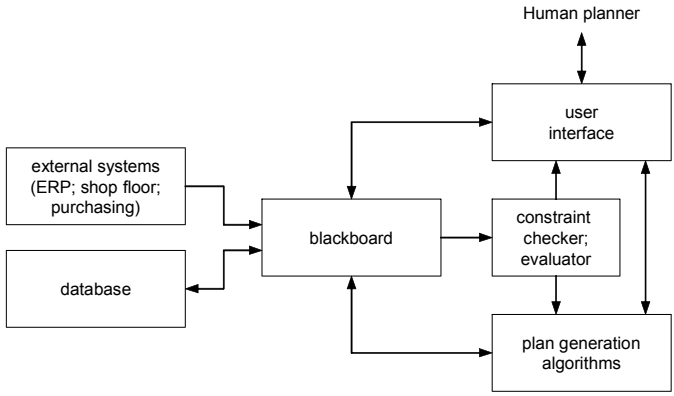


Figure 30.1 Scheduling system architecture

Literature provides little theory on user interfaces for planning and scheduling. Most advanced planning systems are created for production planning, staff scheduling, or routing. Scheduling specific user interface elements in such systems are for example a Gantt chart, a machine dispatch list, a throughput diagram, a performance evaluation interface, and a resource calendar (Pinedo & Yen, 1997, p. 374). Literature on graphical user interfaces for train shunting planning is not available. For the design of our system (which will be described later), we have therefore taken the current representation used by the planners as a starting point.

In sharp contrast with the amount of literature on user interfaces for scheduling, there is much literature on scheduling algorithms. There are two approaches in the development of such algorithms. In the first approach, the developer of the algorithm describes the characteristics of the scheduling problem in variables in a mathematical model. Then, an algorithm is devised that can solve the model, i.e., the algorithm finds values for the variables such that a good solution for the problem is found. In the second approach, the developer looks at the way in which the planners solve the scheduling problems. In contrast with the first approach, the primary aim is not to create an algorithm that solves as much of the scheduling problem as possible. Instead, algorithms are explicitly designed to set up a collaboration between the planner and the system. This latter approach very well fits with the task oriented approach we applied.

Train shunting planning

Trains that arrive at a station do not necessarily leave in the same configuration. A train can be split in its individual coaches and these coaches can be combined again in a train. For example, the train from Groningen to Zwolle consists of two coaches and the train from Leeuwarden to Zwolle consists of one coach. In Zwolle the trains are connected, after which they leave as one train to Den Haag. During the night the passenger trains stay at the station. The ‘storage’ capacity, however, is limited. In addition to changing the configurations during the night, all trains must be washed at a track that contains the washing equipment. The task of the planner is to plan the movements of the trains and coaches and to decide on what track trains can stay. To plan the movements, the planner must assign train drivers, train shunters (the ones who connect and disconnect coaches), and routes to move trains from one track to another. The major bottleneck in the shunting yard that

we investigated is the washing track. Since the time it takes to wash a train is largely independent of the size of the train, it is advantageous to combine carriages in trains before they are washed. Therefore, the nightly transition can not be solved in one stage, since trains are shunted during the night for washing as well.

Because the main time table is stable, planning mainly means adjusting existing plans. Changes are made on the basis of events, for example a track that needs maintenance, an extra train that will pass through the station because of a pop concert, etc. The plans are made manually on paper, after which the result is entered in the computer. Figure 30.2 shows an example of a part of a shunting plan. The computer can check the validity of the plan and send the data to traffic control.

Approximately 130 planners from five major stations in The Netherlands plan the shunting operations for all stations. The Netherlands Railways is in the process of renewing all logistical computer systems. Because of the specific characteristics of the shunting planning problem, and because of the many planners that would have to use the tool for shunting planning, four research projects were started to test two approaches to shunting planning support. In three of those research projects, the emphasis was on developing mathematical algorithms from different perspectives. In a fourth research project, the emphasis was on a task analysis of the shunting planners. In a follow-up fifth project, we have combined the mathematical and task oriented approaches by designing algorithms in a task oriented fashion. In the following section, we describe the prototype that was the result of this approach.

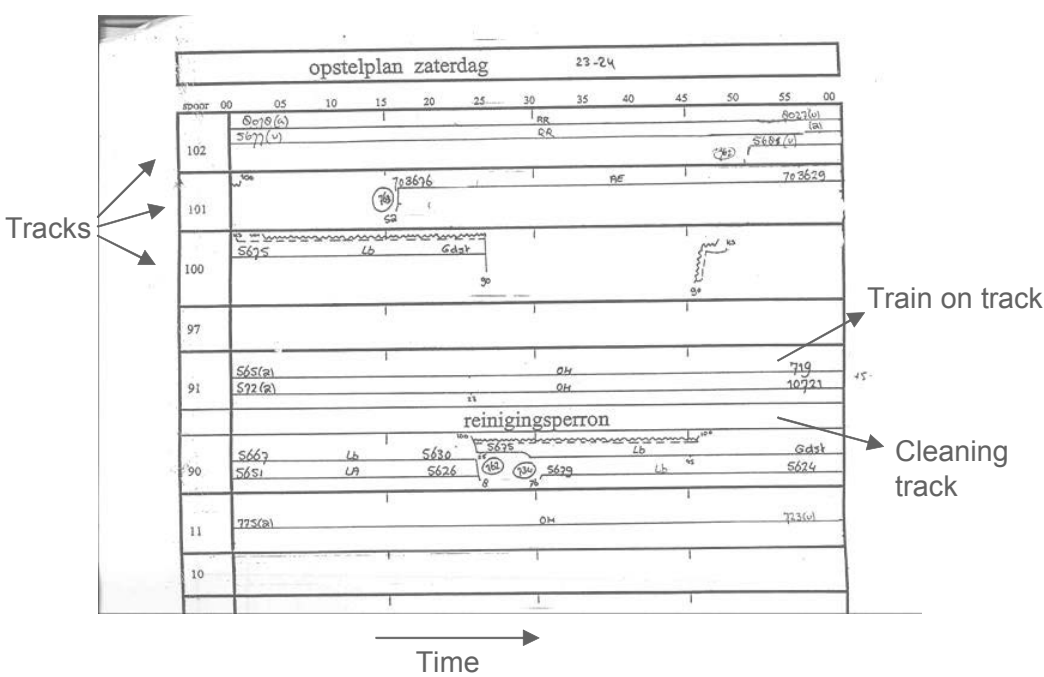


Figure 30.2 Example of a train shunting plan

Prototype description

Here we elaborate on the aspects in which the user is involved: the user interface and the algorithms. First, however, we will give a concise overview of the results of the task analysis that provided the input for the development of the system.

Task analysis

An extensive task analysis with two planners revealed a hierarchical task structure. The activities of the planners contain the following high level subtasks:

1. Determine nightly transitions. Several trains of the same type stay overnight at the station. Coaches of the same type are exchangeable, so the planner must decide what incoming coach will become what outgoing coach.
2. Determine tracks. From the national timetable, the planner knows the amount, arrival/departure times, and tracks of both incoming and outgoing trains. After subtask 1, the planner can search for tracks that the coaches can stay on during the night. Because there are many more coaches than shunting tracks, multiple coaches are put on one track. Therefore, the dominant constraint is that a coach should not be blocked by trains to the left and right of it at the time it must be moved to the departure track.
3. Determine routes. When subtask 2 is finished, the planner must make sure that a movement is indeed possible by determining the route the train should use.
4. Assign shunting staff. The route that is used determines the amount of time that is needed to move the train and the time that is needed for shunting staff to walk from one task to another. Consequentially, assigning tasks to shunting staff is the last planning subtask.

The structure is hierarchical because each decision constrains the lower decisions. For example, the possible routes are determined by the tracks that the train will be on. These tasks are performed in this order no matter the kind of event that triggered the task, but always for a few coaches at a time. The planner does not perform subtasks sequentially for all trains, but only for a limited set of related trains at a time.

Although the task order is fixed, the kind of event determines what task is started with. For example, when a track is scheduled for maintenance, the planner will start with subtask 2 and try to find new tracks for the trains that are scheduled for the track on which the maintenance will take place. Only if he can not find tracks (subtask 2 is over constrained), he will change the nightly transitions and from there perform subtasks 2, 3 and 4. The consequence of changing a nightly transition will be that a part of the existing schedule becomes obsolete and that subtasks 2, 3, and 4 have to be done again for many of the trains. This means much work and therefore the planner tries to avoid this.

In the prototype, we provide support for subtasks 2, 3, and 4. The reason for this is that in the task analyses and experiments we have focused on the day-planners, for whom subtask 1 is not a frequently performed task.

Algorithms

Traditionally, scheduling algorithms tend to neglect the human as a decision maker; literature about human interaction with scheduling algorithms is scarce. Moreover, the literature that pays attention to human/algorithm interaction reasons from the perspective of the algorithm, e.g.

1. The planner can choose from a number of alternative solutions that are generated by algorithms (Lauer et al. 1994; Ulusoy & Özdamar 1996).
2. The planner may specify weights on goal functions (Smed et al. 2000; Gabrel & Vanderpooten 2002), after which the algorithm generates a schedule.
3. The planner can steer the backtracking process of the algorithm (Bernardo & Lin 1994).
4. The planner can specify parameters for the algorithm (Ulusoy & Özdamar 1996; Dockx et al. 1997; Oddi & Cesta 2000; Meyers et al. 2002).

Such algorithms, however, are based on domain analyses. In other words, when creating an algorithm, the developer looks at the characteristics of the "objects" that must be attuned such as machines, jobs, staff, shifts, vehicles, etc. In general, there are three phases in using such algorithms:

1. Translation of the domain to the model. This means a formalization of the objects that must be attuned, including their characteristics, constraints, and goals. Some examples of types of models are Linear Programming (LP) models, Traveling Salesman Problem (TSP) models, Assignment models, and Constraint Satisfaction Programming (CSP) models.
2. Solving the model. A model contains a number of variables, constraints, and goal functions. The variables must be assigned values such that the constraints are not violated and the goal functions get the highest possible values. Often, the kind of model that is chosen goes hand in hand with the solution technique. For example, in order to use the Simplex method, the model must be formulated as a LP model.
3. Translation of the solution to the domain. Once a solution has been found the values of the variables must be translated back to the domain.

The paradigm in algorithm development is often that algorithms can solve scheduling and planning problems better than humans. This is true, but only when we look within step 2. When a model has been formulated, the computer is better and faster at finding the best conflict-free solution or a solution that violates as few constraints as possible. However, we must take into account that (a) information might be lost during the translation, and (b) the solution must be interpreted and understood by the human planner in the context of the actual situation. In other words, the algorithm is only a part of the planning process that is for the rest driven by human decision making.

To overcome these problems, we based the algorithms on the subtasks that are performed instead of on the domain. The design of such algorithms requires a different model-building process. From the field of cognitive science it is known that a decision maker implicitly makes a trade-off between the (cognitive) cost of applying a decision aid (efforts to understand and employ the model and process the information) and the expected benefits (increased quality and speed of obtaining a solution). Decision aids (such as a planning tool or an algorithm) can be improved by applying the following three steps:

- Decompose the planning problem into sub problems and obtain estimates for the efforts (costs) to manually find solutions to these sub problems.
- Identify the sub problems with a high potential of effort (cost) reduction for the decision maker and identify a decision aid that reduces the total effort to find and use a solution for such a sub problem.
- Incorporate specific features for automating storage, retrieval, and computational tasks in the decision aids to manipulate the cognitive effort associated with using these decision aids (Benbasat & Todd, 1996).

In order to investigate whether this approach will also be worthwhile for scheduling and planning, we have developed a task-oriented scheduling system prototype for the shunting planners at the Netherlands Railways. This prototype implements the idea that algorithms should be created for subcomponents of the task strategy to support the problem solving process. The focus is on the level at which the system and the user communicate (Newell, 1981). In other words, human and system should be problem solving in coincident problem spaces (Prietula et al., 1994, p. 660). In our research, we try to find such coincident problem spaces by looking at the subcomponents of the task strategy of human planners.

Looking at the subtasks on the task structure, we can discern four basic assignment tasks: find a free track or combination of tracks, match incoming to outgoing coaches, route a train, and assign tasks to train drivers/shunters. By implementing algorithms for these basic assignment tasks, all steps in the task structure can be supported algorithmically.

1. Track-finding algorithm. Finding a track for a train that is available during a specific time window by varying the time window and constraints. Criteria that will affect the decision to what track the train should be moved are amongst others: the length of the time interval it can stay at this track, the routing distance (i.e. number of direction changes and total mileage) to this track, the previous activities of driver and/or shunter, and the consequences for future actions with this train (i.e., internal cleaning, external cleaning, routing to the track from which it has to leave in the morning, etcetera). The problem the algorithm has to solve is defined as finding a sequence of partially overlapping time intervals from the moment of the actual move to the moment of departure. Sometimes, an additional feature of the sequence is that the cleaning track must be visited somewhere over time. Therefore, the algorithm will have to include the possibility of stating a set of intermittent nodes (i.e., intervals on the cleaning track) from which at least one has to be included in the final sequence before the departure track is reached. Finally, it has to be possible for the planner to block several tracks that may not be included at all in the final sequence, as they have to be reserved for other purposes such as maintenance or trains running through the station. The track finding problem is solved using a K-shortest path algorithm (Van Wezel & Riezebos, 2005).

2. Train unit matching algorithm. Determine how the coaches that enter the station are matched to the coaches that leave. This is performed by a mixed integer programming algorithm that matches arriving to departing train units. The algorithm is described by Freling et al. (2002). This algorithm can also be used on a high hierarchical level by deleting large parts of an existing plan and matching train units again.

3. Routing algorithm. Given the current plan and infrastructure information, the inputs for this subtask are the source and destination tracks for a train. The output of the subtask is a list with the shortest feasible routes for shunting the train to the proposed track. A modified version of the undirected K-shortest path algorithm of Shier (1976) is used to determine the K shortest paths from the source track to the destination track. Modification of the algorithm of Shier was necessary in order to determine the occurrence of direction changes in a route, which is the primary optimization criterion in the weighing of routing alternatives.

4. Driver/shunter assignment. The input for this task is the moment of train movements. In Zwolle, there are at night 6 train drivers available. After a movement, the train driver must walk to the track where he must move the next train. The main criterion is minimizing the overall walking

distance, but a trade-off must be made with the time that drivers must wait at a track (which means a walking distance of zero) for the next train. Waiting too long means they would go to the canteen, which increases the walking distance. A shortest augmenting path algorithm is used that combines shunting activities in cycles in such a way that within a cycle the walking distance is minimized (Jonker & Volgenant, 1987). Multiple cycles are assigned to a shift. Longer cycles means less overall walking distance but at the same time longer waiting times. The balance is found by calculating multiple solutions with different cycle lengths and combining the cycles to shifts such that waiting time is minimized without violating constraints.

As mentioned, the algorithms are made for the basic assignment subtasks. This means that there is not an algorithm that creates the whole plan. Rather, the plan is created by making a number of consecutive planning decisions where each decision is either taken by the human or by the algorithm. This allows the planning process to be highly interactive, because the human planner and the algorithm can both participate in each step of the problem solving process. For this, however, the algorithms must be integrated in the user interface. In the next subsection we explain how this was accomplished for one of the algorithms.

Graphical user interface

We described generic functionality of advanced planning systems. We stated that there is not much literature on user interfaces for planning systems in general, and no literature at all on user interfaces for shunting planning. For that reason, we have taken the paper versions of the plans as a starting point, and added interactivity. For our purpose, the following requirements had to be met:

- The planning screens should follow the hierarchical task structure.
- The screens should allow graphical manipulation (dragging and dropping).
- Screens should be linked real-time. For example, if the planner changes the departure track of a train in one screen, all other screens showing the departure track should be updated immediately.
- The input for and outcome of the algorithms must be highly transparent.

Figure 30.3, Figure 30.4, and Figure 30.5 show examples of the screens in the prototype that conform to these requirements. Figure 30.3 shows the main planning screen which is the graphical representation of the paper version shown in Figure 30.2. All kinds of manipulations are possible, for example, moving trains in time, changing the track, splitting a train in individual coaches, etc. Constraint violations are shown by colors in the bars, and the movements of individual coaches can be tracked easily.

Figure 30.3 gives a time/place representation of the plan. In Figure 30.4, the shunting yard is shown. It shows for a specific time what coach is on what track. In this screen, the tracks the coaches are on can be changed by dragging and dropping. Additionally, the route of the selected train is shown; colour coded tracks show the route from the previous track to the current track.

In Figure 30.4 and Figure 30.5, an example of one of the interactive algorithms is shown. If a train is dragged to another track, the routing algorithm automatically determines a new route. If the planner does not want the default route, he can choose one of the alternatives as shown in Figure 30.5. Furthermore, the planner can determine the route himself by clicking on the tracks in Figure 30.4. Each time he clicks on a track, that track is added to the 'via'-list, meaning that the routing algorithm will include that track in the route. Immediately after adding a track to the 'via'-list, a new route is calculated and shown in the user interface.

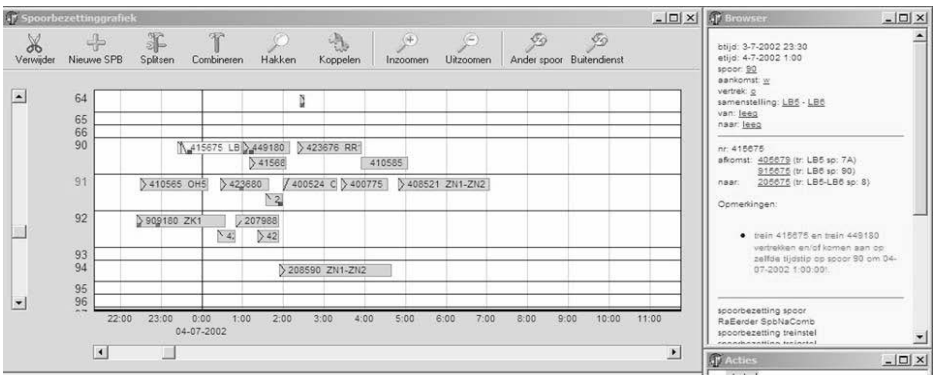


Figure 30.3 Shunting plan

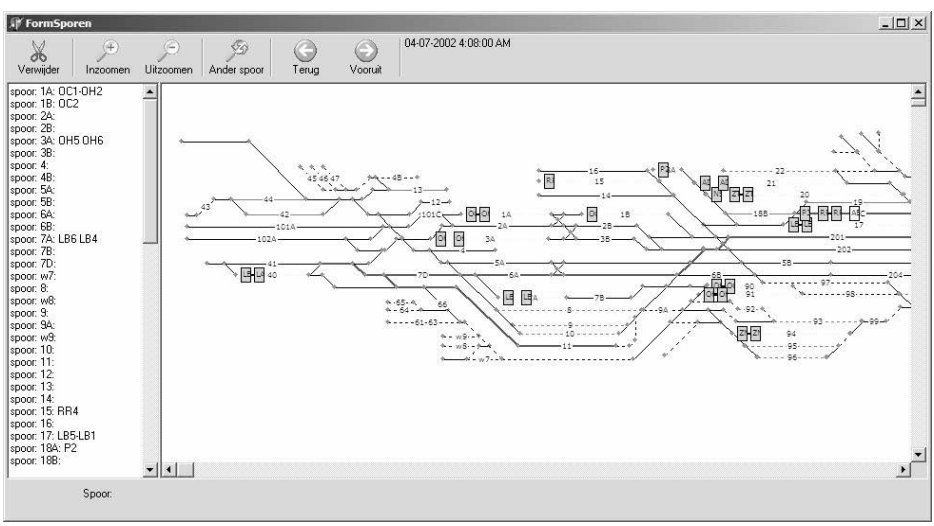


Figure 30.4 Graphical overview of the shunting yard

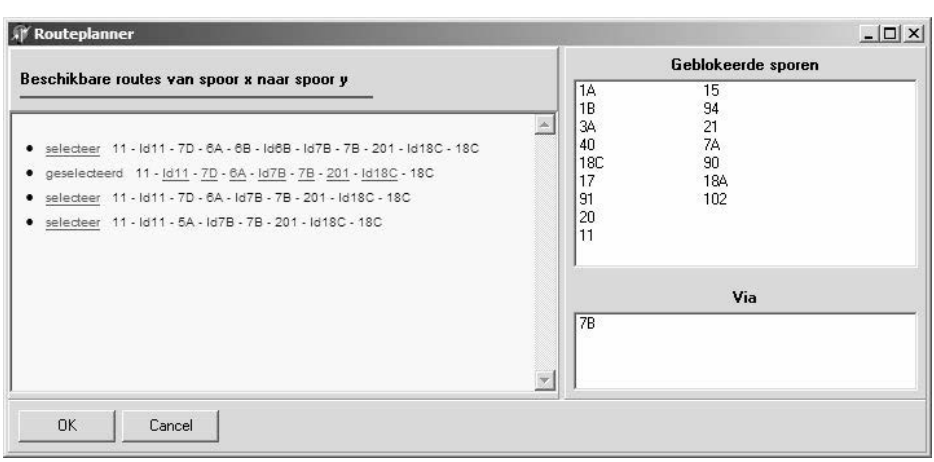


Figure 30.5 Routing alternatives

Conclusions

Once, there might have been the belief that computers and clever algorithms would soon render human planners obsolete. This belief is obsolete itself nowadays, as it is widely recognized that both human planners and computer systems have their own distinctive role in planning. We have tried to contribute to this debate by combining methods from different scientific fields in the design of a shunting planning support system.

The goal of the shunting planning project was to create task support, including algorithms, for shunting planners. In-depth task analyses with planners resulted in a number of task structures. By using these task structures to design planning support we combined cognitive science (for the task analysis) and operations research (for algorithms) to design a prototype planning system. The bottom-up approach of algorithmic design provides an extension to the mixed initiative scheduling approaches found in literature. By describing the creation of the plan as a hierarchic problem solving process, we place the discussion about human versus computer at the level where it belongs: an algorithm should not be created for a planning problem, but for a planner's subtask. Because of the bottom-up approach that was taken, the algorithms are very flexible in their use.

After the algorithms were implemented and integrated in the prototype, the efficiency of the algorithms could be evaluated by comparing the time it takes to solve problems of several levels of complexity. However, the concept of planner-oriented design of algorithms is much more difficult to evaluate. Whether the algorithms effectively support the task of the planner is currently being researched by empirically testing several kinds of support with a large number of shunting planners.

Acknowledgements

This study has been supported by the Netherlands Railways. We gratefully acknowledge the management and planners of this company. Specifically, we want to thank Dr. L. Kroon and the planners of the Netherlands Railways, location Zwolle, for their willingness to co-operate.

References

Benbasat, I. and Todd, P. (1996). The effects of decision support and task contingencies on model formulation: A cognitive perspective. *Decision Support Systems*, vol. 17, pp. 241-252.

Bernardo J.J., L Lin, K.S. (1994). An interactive procedure for bi-criteria production scheduling. *Computers and Operations Research*. vol. 21, no. 6, pp. 677-688

Dockx K., Boeck, Y. de and Meert, K. (1997). Interactive scheduling in the chemical process industry. Computers and chemical engineering. *An international journal*, 21, 9, pp. 925-946

Freling, R., Lentink, R. M., Kroon, L. G., and Huisman, D. (2002). *Shunting of passenger train units in a railway station*. Econometric Institute Report Erasmus University Rotterdam no. EI2002-26. Note: Accepted for publication in: Transportation Science.

Gabrel, V. and Vanderpooten, D. (2002). Enumeration and interactive selection of efficient paths in a multiple criteria graph for scheduling an earth observing satellite. *European Journal of Operational Research*, 139, 3, pp. 533-542

Jonker, R. and A. Volgenant (1987). A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38, pp. 325–340.

Lauer J., Jacobs, L.W., Brusco, M.J. and Bechtold, S.E. (1994). An Interactive, Optimization-based

Decision Support System for Scheduling Part-time, Computer Lab Attendants. *Omega : the international journal of management science*, vol. 22, no. 6, pp. 613-626

Myers, K.L., W.M. Tyson, M.J. Wolverton, P.A. Jarvis, T.J. Lee and M. desJardins, M. (2002). PASSAT: A User-centricPlanning Framework. In: *Proceedings of the 3rd Intl. NASA Workshop on Planning and Scheduling for Space*, Houston, TX

Newell A (1981). The knowledge level, AI Magazine, vol. 2, no. 2, pp. 1-20.

Oddi, A. and A. Cesta A (2000). Toward interactive scheduling systems for managing medical resources. *Artificial Intelligence in Medicine*. vol. 20, no. 2, pp. 113-138

Pinedo, M. and Yen, B.P.-C. (1997). On the design and development of object-oriented scheduling systems. *Annals of Operations Research*, 70, p. 359-378.

Prietula MJ, Hsu W, Ow PS (1994). MacMerl: Mixed-Initiative Scheduling with Coincident Problem Spaces. In: Zweben, M, Fox MS Intelligent Scheduling. San Francisco: Morgan Kaufman.

Shier, D.R. (1976). Iterative methods for determining the k shortest paths in a network. *Networks*, vol. 6, pp. 205-230.

Smed J., Johtela, T., Johnsson, M., Puranen, M. and Nevalainen, O. (2000). An Interactive System for Scheduling Jobs in Electronic Assembly. *The international journal of advanced manufacturing technology*. vol. 16, no. 6, pp. 450-459

Van Wezel, W. and J. Riezebos (2005). Algorithmic support for human rail shunting planners, *Proceedings of the 18th International Conference on Production Research (ICPR18)*, Salerno, Italy

Ulusoy, G. and Özdamar, L. (1996). A framework for an interactive project scheduling system under limited resources. *European Journal of Operations Research*. vol. 90, no. 2, pp. 362-375

PART IX
Engineering Work and Maintenance

This page intentionally left blank

Chapter 31

Understanding the Underlying Causes of Procedures Violations and Developing Effective Preventative Strategies

David Embrey

Introduction

It is widely recognised that procedures have a vital role to play in the operation of all safety critical systems. Many accident and incident reports cite 'failure to comply with procedures' as a central cause. In some respects this is a truism, since the actions that lead to an accident are unlikely to involve compliance with the documented procedures. However, a deeper analysis of the way in which procedures are perceived and used by the workforce indicates that there are major differences between these perceptions and those assumed by the Safety Management System. This disparity of vision between the end users of procedures and those that frame their content is one of the major reasons for the widespread non-compliance that is the reality in most organisations. This issue will be explored in detail in subsequent sections of this chapter.

The usual definition of a violation or non-compliance is as follows:

> Where a person deliberately performs a task in a manner which is different from the documented, official procedures (even though they are aware of the content of these procedures).

However non-compliance can be more usefully defined as defined as:

> A deliberate deviation from a previously agreed and documented Best Practice (if such a best practice exists and is known to the person involved).

From this perspective a deviation from a poor formal procedure may in fact be a positive response if it reduces risk (or correctly balances safety, production and other risks). In an enlightened organisation, such a deviation should not normally result in disciplinary actions against the person involved. However, once an agreed best practice has been developed, in consultation with the workforce, which both manages the risk and is practical in terms of the day to day realities of performing the job, it follows that reckless deviation from this agreed practice should attract disciplinary action unless justified by exceptional circumstances. Any process which involves the most experienced workers in a creative discussion regarding the best way to do things is likely to be beneficial and in the long term will be a powerful means for reducing violations. A practical process for achieving such a consensus will be described later.

Causes of non-compliance

A number of research results provide insights into the causes of non-compliances (see for example Health and Safety Executive, 1995). In the nuclear industry, the Advisory Committee for Safety

in Nuclear Installations (ACSNI, 1993) suggested the following as common reasons for non-compliance:

- Procedures may be impractical or extremely difficult to implement
- Procedures may be perceived as being primarily to 'protect the management's back'
- There may be an inadequate understanding of the risk or consequences of non-compliance
- There may be pressure (explicit or implicit, real or perceived) to trade production needs against strict compliance
- Procedures may be known but not fully understood

Discussions with personnel in several safety critical industries have indicated that the following issues have a significant impact on procedure usage:

- Lack of a clear policy on procedure use (i.e. are procedures meant to be followed to the letter or are they essentially guidelines, which need to be interpreted in the light of the specific conditions of use?)
- Too many and too complex procedures (this comment reflects a common concern that many procedures are unnecessary for simple tasks carried out frequently by experienced operators)
- Poor presentation (procedures which are easy to use and which confer a real benefit are likely to be complied with)
- Procedures are out-of-date or irrelevant (these characteristics reinforce the lack of credibility of many procedures)
- Time pressure (this reflects the fact that procedures are often written for use under normal conditions, and may be impractical in abnormal or difficult situations (e.g. extreme production pressures)
- Lack of management commitment (this refers to management pressures to 'cut corners' or use alternative methods if production pressures mean that the documented procedures cannot be applied)
- Lack of resources (this refers to time or other constraints that prevent compliance)
- Individual preferences (differences of opinion exist with regard to how procedures should be interpreted, leading to variations in working methods)
- Lack of knowledge that they exist (this indicates that the situations under which procedures need to be referred to when carrying out the task, are not clearly distinguished from those in which knowledge and experience are deemed to be sufficient)
- Procedures depersonalise the job (this relates to the extent to which personnel should be allowed to introduce variations into the way in which tasks are performed in order to increase their feelings of autonomy and therefore job satisfaction).

The Procedures Culture Survey: A case study

Prior to beginning a programme of work to minimise non-compliances in an organisation, it is desirable to measure the perceptions of people regarding the incidence of such problems and the factors that are believed to influence their likelihood. Based on some of the issues identified in the preceding sections, together with other data gathered from the research literature, a survey tool was developed to address these requirements, called the Procedures Culture Survey, (PCS). The application of the PCS is illustrated in the following case study.

In response to a previous assessment that indicated a procedures compliance problem in a large multinational organisation, the PCS was distributed to nearly 1000 operators and managers. The following sections summarise the results of that survey.

Extent of usage of procedures

The first set of questions related to the extent that procedures were actually used for different categories of task. For tasks perceived to be safety or quality critical, the use of procedures was high (75% and 80% respectively) but not universal. For problem diagnosis (regardless of whether a system was safety critical or not) only 30% of the respondents used procedures. For routine tasks, only 10% of the respondents said they used procedures.

The results of the survey indicated that, even in tasks where procedures were said to be used, only 58% of the respondents had them open in front of them when carrying out work. These figures imply that the actual average 'on-line' usage of procedures for safety/quality critical, problem diagnosis and routine tasks is quite low, i.e. 43%, 17% and 6% respectively.

Use of standardised working methods

Procedures provide the basis for standardised working practices, which, if followed, should ensure that the objectives of the task are achieved. One of the items in the survey concerned the use of 'Black Books' i.e. personal sets of notes held by individuals as informal job aids. The results indicated a very high usage of Black Books by both operators and managers (56% and 51% respectively). The existence of such books suggests that existing operating procedures do not contain information which is necessary to support the day to day performance of tasks.

The PCS also assessed the extent to which procedures should be regarded as being guidelines, or needed to be followed 'to the letter'. Although there was considerable agreement that safety and quality instructions should be followed to the letter (90% and 75% respectively) for most other categories of task about 50% of respondents believed that they were primarily guidelines. This came as a considerable surprise to the management of the companies included in the survey.

Strategies for improvements

The final part of the survey considered the question of why procedures were not used. Following prior discussions with technicians, seven factors were investigated with regard to their impact on procedure usage. These are set out in Figure 31.1.

Figure 31.1 indicates a high level of agreement with most of the suggested reasons for lack of usage of procedures. Another part of the survey asked people to indicate the five main reasons that procedures were not used, and the five changes that would be most effective in improving the quality of procedures and their use. The most highly ranked reasons for procedures not being used were as follows:

- If followed to the letter the job wouldn't get done
- People are not aware that a procedure exists
- People prefer to rely on their skills and experience
- People assume they know what is in the procedure
-

'Procedures are not always used because…' (percent agreeing)		
Accuracy	…they are inaccurate …they are out-of-date	(21) (45)
Practicality	…they are unworkable in practice …they make it more difficult to do the work …they are too restrictive …too time consuming …if they were followed 'to the letter' the job couldn't get done in time	(40) (42) (48) (44) (62)
Optimisation	…people usually find a better way of doing the job …they do not describe the best way to carry out the work	(42) (48)
Presentation	…it is difficult to know which is the right procedure …they are too complex and difficult to use …it is difficult to find the information you need within the procedure	(32) (42) (48)
Accessibility	…it is difficult to locate the right procedure …people are not aware that a procedure exists for the job they are doing	(50) (57)
Policy	…people do not understand why they are necessary …no clear policy on when they should be used	(40) (37)
Usage	…experienced people don't need them …people resent being told how to do their job …people prefer to rely on their own skills and experience …people assume they know what is in the procedure	(19) (34) (72) (70)

Figure 31.1 Reasons for non-usage of procedures

The most highly ranked strategies for improvements were:

- Involving users in the design of procedures
- Writing procedures in plain English
- Updating procedures when plant and working practices change
- Ensuring that procedures always reflect current working practices

There were no significant differences between the reasons for lack of procedure usage, but 'involving users in the design of procedures' was rated significantly higher than any of the other approaches to improvements.

One definition of Best Practice is the performance of a task in a manner which achieves its required objectives whilst minimising any SHE (Safety, Health and Environment), economic and quality risks. The PCS survey demonstrated that, in many cases, formal written procedures are not representative of Best Practice and may not correspond to the way in which the work is actually done. There are two main reasons for this disparity. Firstly, procedures are often written by technical specialists without a hands-on experience of the work environment and a lack of awareness of a task's practical constraints. Secondly, there is rarely a system in place for ensuring that procedures are modified to take into account organisational learning and gradual changes in working practices. These problems gradually erode the credibility of the official procedures, and can give rise to the use of informal undocumented methods that may or may not be effective in controlling risks.

Conclusions from the Survey

In this industry, the majority of tasks were performed without the on-line use of step by step written procedures. There were also variations in the ways in which tasks were performed, between both shifts and individuals, which sometimes differed significantly from the 'official' procedures. People will not follow procedures if they feel they are impractical, and they will not routinely use written procedures if they believe they have sufficient skill and experience to get the job done. However, the existence of 'Black Books' indicates that there is a definite need for some form of on-line support, which is not provided by the existing procedures systems.

What is the appropriate level of detail for procedures: the role of tacit knowledge

In general, cognitive scientists have been disparaging about the usefulness of conventional, step by step written procedures. They are usually regarded as grossly simplified representations of the complex and adaptive ways that we interact with an uncertain world. Dekker (2004) has suggested that procedures are resources for action, but cannot be prescriptive for all situations that might arise. Although this is clearly true up to a point, in real high risk systems, nobody believes that a set of written procedures will be applicable to all situations that might arise. However, this is not an argument against developing procedures for the many situations that might reasonably be anticipated and using this information to improve the likelihood of successfully managing the situation. There is often a comprehensive knowledge base available regarding how to handle safety critical situations in an organisation. If this knowledge can be systematically gathered and represented in a manner that can be readily accessed in a complex situation (usually referred to as a job aid, e.g. flow diagrams, or checklists), this will relieve operators of the burden of having to maintain these strategies in their memory. This in turn will reduce the mental workload and free up processing capacity to develop innovative solutions if the scenario is outside the expertise possessed by the operator or represented by the procedures.

The ability to use more compact job aids is based on the assumption that the operator possesses the necessary tacit knowledge and skills to fill in the activities not specified in the procedure. In this context, 'tacit knowledge' means both physical and mental skills that have been acquired either by means of formal training or by experience on the job, including informal mentoring by more experienced workers. Unfortunately, tacit knowledge is rarely documented, even in a well organised training system.

Another argument for capturing undocumented tacit knowledge relates to the retention of the cumulative skills and knowledge acquired by people over their working lives. The retention of this knowledge is what defines learning organisations, as opposed to those who are doomed to continually relearn the lessons of the past due to corporate amnesia. Unfortunately when companies downsize, it is usually the older, more experienced workers who are the first candidates for early retirement. Without a formal process for capturing the tacit knowledge acquired over a number of years, the organisation is likely to be less safe and less efficient following the loss of experienced personnel.

Although procedures have their limitations, they are nevertheless an important method for supporting operators in performing both routine tasks (such as maintenance) or handling more complex situations such as disturbances. However, procedures must be seen as complementing rather than replacing skills and knowledge. In later sections a simple process will be described for deciding on the type of procedural support that is appropriate given various characteristics of the situation and the task. Because the tacit knowledge which is an essential underpinning of a safe and efficient organisation is usually invisible, it is important that processes are applied to

acquire and codify this knowledge. In particular, these processes must provide efficient means for identifying tacit methods for performing tasks that have evolved over time, which manage the risks, and which are acceptable to the workforce. These methods or 'Best Practices' need to be developed by consensus if they are to become the accepted way of doing things in the real world of day to day operations.

Developing a consensus based procedures culture using CARMAN

The key to maximising compliance to procedures lies in developing a culture where the preferred practices are identical to the documented Best Practices, which in turn incorporate all the available knowledge concerning the best ways to operate the system. In this context, 'Best' means that working methods not only have to control risks; they also need to be acceptable to the workforce in terms of credibility and practicality. These 'Reference Task Descriptions (RTD)' provide the basis for training content and competency assessment, and also guide the development of the most appropriate form of on-line support. In order to achieve these objectives, a consensus process is required which evaluates the hazards associated with existing working practices and enables agreement to be reached about how these practices need to be changed to control any unacceptable risks. It should be emphasised that such a process must not only include shop floor personnel, but also technical specialists who may have insights into why a task should be performed in a particular way.

Human Reliability Associates have developed a process called the Consensus based Approach to Risk Management (CARMAN) to achieve these objectives. The original impetus for CARMAN came from a number of procedures improvement projects where the main focus was on improving the usability of procedures by applying ergonomics design standards to issues such as readability, layout and formatting. However, it was found that even when the usability of procedures was considerably improved, their level of usage was sometimes still low, and procedural violations still occurred. This led to work aimed at understanding the causes of procedural non-compliance, and the development of the CARMAN approach that combined insights from task and risk analysis, group processes, and work on organisational learning. This approach was gradually refined by being applied to a number of organisations.

In CARMAN, the working practices which are actually used by operators and technicians are examined, using a participative process, which documents the variations that exist, and then evaluates the practices from the point of view of whether they control the risks. Best Practices are then developed and documented, which take into account the preferences and insights of the people who actually perform the tasks, whilst ensuring that risks are adequately controlled.

The process seeks to create a neutral forum for the exchange of information about differing working practices (e.g. between shift teams) and also to allow insights to be gained into the risks associated with different ways of carrying out tasks. Technical specialists contribute to this information exchange process, but do not dominate it. This is because it is essential to ensure that the developers of the revised procedures have a shared sense of ownership.

The CARMAN process

CARMAN comprises two stages: the development and documentation of Best Practices, and the production of job aids, competency standards and training programmes based upon the documented Best Practices.

The overall structure of Stage 1 is shown in Figure 31.2. Prior to commencing the steps shown in Figure 31.2, it is first essential to appoint a facilitator, and to provide training in the tools and

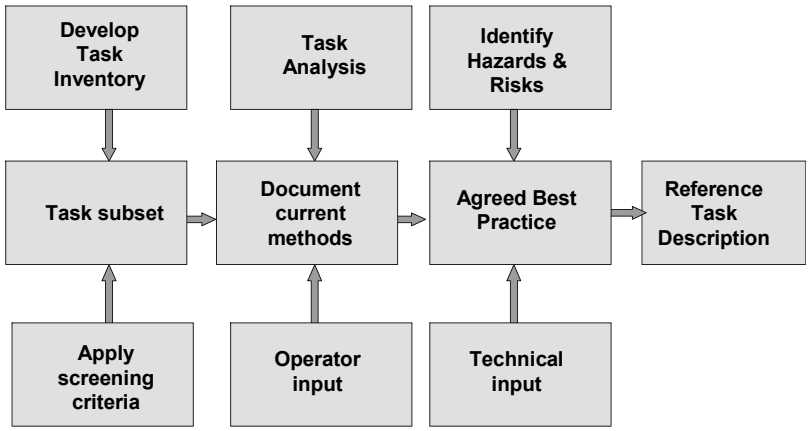


Figure 31.2 Stage 1 of CARMAN

philosophy of CARMAN. His or her role is to collect information from the various stakeholders, and to assist in the development of consensus regarding Best Practice. It is essential that personnel respect the facilitator, and that he or she has good communication skills. It is also desirable to provide some awareness training for the technicians, and also basic training in task analysis, which is an important technique in the process.

The first step of Stage 1 is to list the tasks that exist in the system. This list is called a Task Inventory, and is intended to ensure that no important tasks are omitted. Following the development of the Task Inventory, a screening analysis is conducted based on assigning a measure of risk (likelihood of failure weighted by severity of consequences) to identify the subset of tasks that are considered to be critical, and the current practices for these tasks are then documented using Hierarchical Task Analysis (HTA). Usually there will be differences between individuals and shift teams regarding how the tasks should be performed. These are compiled by the facilitator, and then resolved by convening consensus groups, which examine the similarities and differences between methods. These groups also evaluate the consequences associated with various types of error, and on the basis of these risk assessments and the discussions, consensus is reached on the Best Practice. At this stage, technical specialists are invited to the consensus sessions to comment on the draft Best Practices. Unless the specialists provide specific reasons for modifying the Best Practice, this is then appended to the database in the form of a Reference Procedure together with information concerning the possible hazards and consequences.

In Stage 2 of CARMAN, shown in Figure 31.3, the Reference Task Descriptions in the Best Practice database are used to develop competency specifications, training programmes and supporting job aids, based upon the level of on-line support required for each task. The primary factors that are considered when determining the level of on-line support are the severity of consequences if the task fails, the frequency with which the task is performed and its complexity. The more severe the consequences, the lower the frequency of task performance and the greater the complexity, the more elaborate the level of support that is provided.

An example of a decision rule is shown in Figure 31.4. It can be seen that the majority of tasks will be performed without written instructions. As the tasks become more critical, complex and infrequent, the level of support increases. However, overall, less than ten percent of the tasks require step by step instructions for on-line support. This is in sharp contrast to many organisations, where nearly all cells in the matrix would specify the use of step by step instructions.

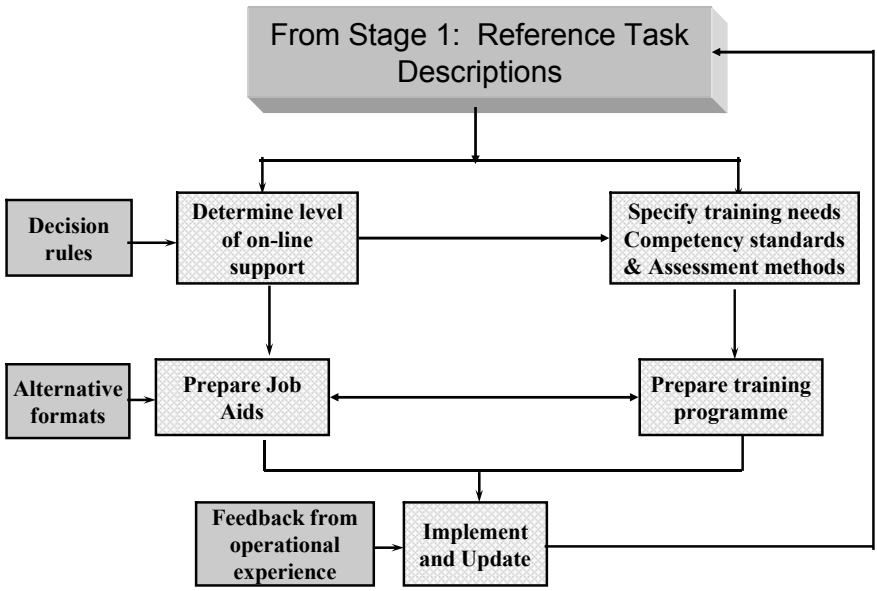


Figure 31.3 Stage 2 of CARMAN

Task Criticality	High			Medium			Low		
Task Familiarity	Freq	Infreq	Rare	Freq	Infreq	Rare	Freq	Infreq	Rare
Task Complexity									
Low	NWI	NWI	JA	NWI	NWI	JA	NWI	NWI	NWI
Medium	NWI	JA	SBS	NWI	NWI	JA	NWI	NWI	NWI
High	JA	JA	SBS	NWI	JA	SBS	NWI	NWI	JA

No Written Instruction required (NWI)
Job Aid required e.g. checklist/memory aid (JA)
Step by Step instruction required (SBS)

Figure 31.4 Decision aid for choosing level of Job Aid Support

Conclusions

This chapter has described an approach to procedures which is based on the concept of developing agreed methods of working with the active involvement of the workforce. The CARMAN approach provides an open environment where the variability in the ways in which safety critical tasks are performed can be reviewed and improved working methods developed which recognise the realities

of the environment. In addition, CARMAN has considerable benefits in developing an open culture with regard to information sharing, which is an essential requirement for a learning organisation. CARMAN has been successfully applied in several multinational organisations over the last few years, and has been shown to produce measurable benefits in terms of both safety and financial criteria. It provides a framework to control safety, quality and other risks by the development of a participative culture, and the integration of training and procedures. It also provides a logical method for selecting the type and degree of on-line support required to complement the skills acquired through training and experience. The Procedures Culture Survey tool provides a valuable insights into the current effectiveness of the procedures system in an organisation, with specific indications with regard to where improvements will be most effective, It is recommended that it is used as a before and after measure when applying an improvement process such as CARMAN, as a means of measuring the effectiveness of the intervention.

References

Advisory Committee on the Safety of Nuclear Installations (ACSNI)—Study Group on Human Factors (1993) *Organising for Safety* London: Health and Safety Commission.

Dekker, S (2004) *To engineer is to err* In: 'Human Factors for Engineers' Sandom, C, and Harvey, S. (Editors) London: Institution of Electrical Engineers.

Health and Safety Executive (1995) *Improving Compliance with Safety Procedures: Reducing Industrial Violations.* London: Health and Safety Executive (Human Factors in Reliability Group).

This page intentionally left blank

Chapter 32

Human Factors in the Management of Engineering Possessions: The Roles of the Engineering Supervisor and PICOP

Brendan Ryan, John R. Wilson, Alex Schock, Emma Lowe and Fiona Kenvyn

Introduction

Engineering possessions are important features of railway maintenance and renewals activities, allowing workers to get access to the railway in the absence of traffic. The rule book (GO/RT3000) allows for a range of different types of possessions (for example, for short periods of time or extended periods of time, on a single line or all lines in a section of track) and specifies the duties of key staff in setting up the possessions and the delivery of safe systems of work within these possessions.

The Institute for Occupational Ergonomics (IOE) at the University of Nottingham has supported staff at Network Rail from the West Coast Route Modernisation (WCRM) Safety Group and the Ergonomics Group in a programme of research after analysis of data held by the organisation raised some concerns about the numbers of incidents or possession irregularities which were occurring on the West Coast Route. Whilst they had not resulted in serious injuries or events, approximately 40% of these irregularities could have resulted in a fatality given different combinations of circumstances. In parallel, Network Rail have begun to take a fundamental look at possession planning and management, in order to improve performance such as time on track and minimising engineering train movements.

This chapter presents an overview of the work which has been carried out in this programme to understand why irregularities were occurring in WCRM possessions, investigating human factors issues in the management of engineering possessions. This work has been carried out in two phases. The first phase concentrated on the work of the Engineering Supervisor, who plays a key role in setting up a work site within a possession and the management of access for work in the work site. A second phase of work is currently in progress, and this is looking at the practicalities of a series of proposed changes to possession rules. This is considering, in particular, the wider effects that these changes could have on the performance of key staff in the management of the work in a possession (such as the PICOP (Person in Charge of the Possession) and including new roles which have been introduced in these proposed changes). General overviews of the roles of the Engineering Supervisor and the PICOP are presented, and different work situations are described which can have an effect on these.

Phase 1 – Field study investigating the role of the Engineering Supervisor

Role of the Engineering Supervisor

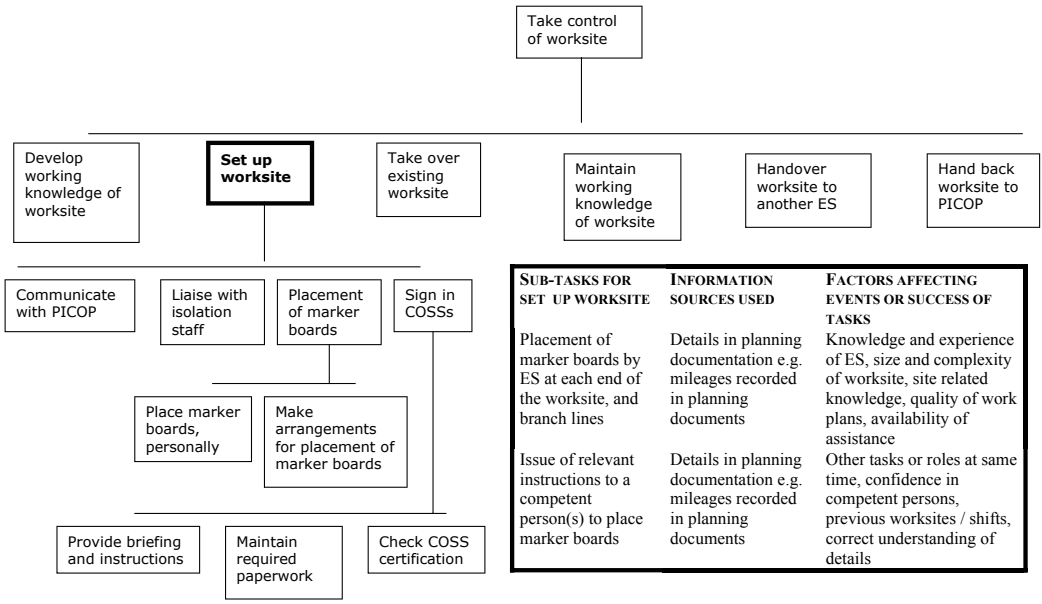
The Engineering Supervisor is defined in the Railway Group Standard GO/RT3261 as “a person certified as competent and appointed to be in charge of each separate work site within a possession”.

Duties of the Engineering Supervisor are recorded as "establishing, maintaining and giving up work sites", and "authorising the movement of trains into or within work sites". In practice, the role of the Engineering Supervisor is dependent upon a number of factors, and these include the complexity of the track layout, the nature of the work in the work site (or even the neighbouring work sites), the size of the work site, the type of equipment and plant being used, isolation requirements, and the duration of the work. For example, major track relaying works will be likely to require multiple contractors and groups of workers in different parts of a large work site, extensive provision of plant and equipment for a variety of tasks, and the movement of many trains into the work site, within the work site and out of the work site (for supplies of materials, equipment, tamping and removal of scrap). The work will be likely to run over a number of shifts and the Engineering Supervisor may have to hand over control of the work site to a colleague. In contrast, a smaller maintenance job might involve a single gang of workers in a shorter work site in which work is carried out over a part of a shift. This type of work may require the use of a limited amount of equipment, and there may be no movements or limited movements of trains or road rail vehicles. As a further variation, work on overhead line equipment (e.g. adjustment and minor repairs) will be likely to require a number of road rail vehicles working in a large number of different locations within the work site, with frequent movements throughout the shift. Each of these different types of work will place different demands on the Engineering Supervisor: in setting up the work site correctly and placing the marker boards appropriately and in a timely manner; signing in and briefing workers effectively; assessing and controlling the risks from multiple groups of workers on the site; controlling the movements of trains and road rail vehicles; and handing back the work site in a condition which is safe for the passage of trains.

Approach to the investigation

The primary method of collecting information for this investigation was through observations in the field, though a number of preliminary steps were taken to develop knowledge of the domain and provide a focus for the field investigations. Firstly, relevant documentation was examined (e.g. Group and Company Standards). Basic descriptions of the role were interpreted using prior knowledge gained in previous projects and used to prepare an interview protocol (e.g. Farrington-Darby et al, 2005). A detailed interview was then carried out with an experienced Engineering Supervisor. The interview was initially unstructured, using general prompts to collect descriptions of the role, knowledge used in the job and experiences in the job. More specific prompts and questions from the interview protocol were used, where necessary, to collect information on the range and frequencies of tasks in the role, details of experience of different types of possessions, perceptions of communications, briefings, types of records and documentation used, involvement in planning of work, shift changeover procedures, and awareness of errors and irregularities. A second interview (a group interview) was conducted with two experienced Engineering Supervisors and a construction manager. General descriptions and opinions from the interviewees were collected and topics from the interview protocol were discussed. Through careful facilitation of the interview it was possible to use the experiences from each of the members of the group to prompt recollections from other members of the group, thereby enriching the discussion. These interviewees had also agreed to participate in the later part of the field study (i.e. observations in the field), so a second aim of this meeting was to begin to develop a rapport with the participants so that they would be more comfortable with being accompanied and observed during the field study.

Information from these preliminary investigations was used to develop hierarchical and tabular task analyses for the role of the Engineering Supervisor and parts of these task analyses



SUB-TASKS FOR SET UP WORKSITE	INFORMATION SOURCES USED	FACTORS AFFECTING EVENTS OR SUCCESS OF TASKS
Placement of marker boards by ES at each end of the worksite, and branch lines	Details in planning documentation e.g. mileages recorded in planning documents	Knowledge and experience of ES, size and complexity of worksite, site related knowledge, quality of work plans, availability of assistance
Issue of relevant instructions to a competent person(s) to place marker boards	Details in planning documentation e.g. mileages recorded in planning documents	Other tasks or roles at same time, confidence in competent persons, previous worksites / shifts, correct understanding of details

Figure 32.1 Part of the hierarchical task analysis and a relevant extract from the tabular task analysis for the work of the Engineering Supervisor

are illustrated in Figure 32.1. These were initially constructed from self reports of workers and the information from documentation such as the Group and Company Standards. Initial steps have been taken to validate the task analyses through observations at sites in the current programme of research, though further validation and development will be necessary through additional observations in the field.

Three researchers were allocated to different parts of the West Coast route and carried out observations during different types of engineering works in the field (e.g. track relaying, overhead line work). Where possible, the researchers attended pre-site meetings to develop an in-depth understanding of the type of work which would be observed, and how this work fitted in with the range of other works which would be in progress on the network. For example, an overview of how items of work were selected from the work bank and scheduled into a programme of works was provided by a company planner during a face to face meeting. At a PICOP meeting, two to three days before the proposed works in a possession, it was observed how representatives from various companies (such as Engineering Supervisors or other planning staff) confirmed their final arrangements for the work site limits and the types of works which had been scheduled within the work sites in the possession. The PICOP agreed access arrangements and contact details for the sites, and details of likely train or vehicle movements within or between work sites in the possession were discussed. At a company meeting of Engineering Supervisors it was observed how important information from the PICOP meeting was relayed to Engineering Supervisors and isolation staff who would be covering a number of shifts during the course of the works in the possession. Further opportunities were taken during these meetings to build relationships with staff, thereby increasing the likelihood that observed behaviours in the field would be natural.

Observations were made at nine different work sites over 13 observation periods, collecting general observations of the activities and tasks of the Engineering Supervisor. Attention was focused, in particular, on the following: the length and content of briefings, the numbers of

COSSs signing onto the work site with the Engineering Supervisor, the numbers and types of communications to and from the Engineering Supervisor. Whilst the primary method of collecting data was through unobtrusive observation in the field, the researchers took opportunities to seek clarification from the Engineering Supervisor where necessary, such as in relation to the content of telephone conversations (only one side of which could be over heard whilst shadowing the Engineering Supervisor), or the reason behind some activities which were observed. Such clarification was sought in such a way that it did not interfere with the work of the Engineering Supervisor. At periods when the Engineering Supervisor had more time available, a variety of issues were discussed in greater depth.

The researchers were PTS (personal track safety) qualified, but were accompanied by experienced staff from Network Rail or other participating companies during all parts of the field work. This was essential for reasons of personal safety, but this also enabled easy access to sites via access points which can be difficult to find without local knowledge and experience. The researchers received appropriate briefings on safety and wore standard high visibility clothing and personal protective equipment on all work sites.

Findings from the field observations

Pen and paper records were maintained in the field, using a simple matrix to note down the times and details of activities, communications or events. Field notes were transcribed and used to produce a range of figures to illustrate the findings. One such figure is included by way of example, below. Figure 32.2 shows a timeline for communications to and from the Engineering Supervisor in the period of two hours after he arrived at the site. The figure helps to illustrate the range of people who were in contact with the Engineering Supervisor during this period. The majority of

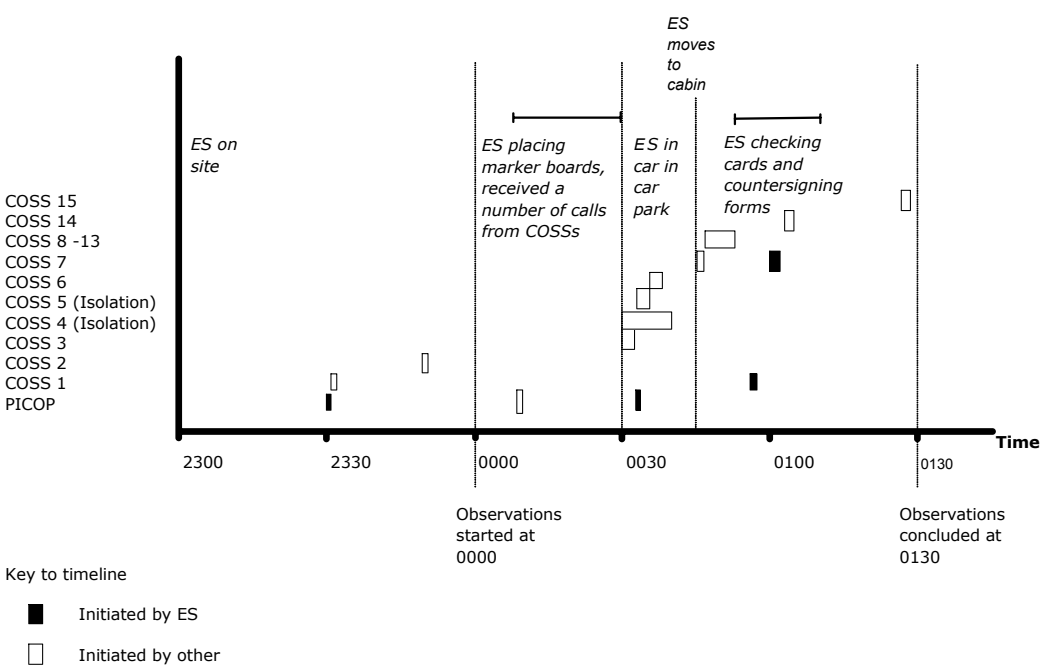


Figure 32.2 Timeline showing communication to and from the ES

the conversations were short, of only one to two minutes duration, but were often consecutive, with one individual queuing up after another to speak with the Engineering Supervisor, particularly in the period between 0030 and 0100. During one of the longer conversations (with COSS No.4 who had limited experience in the role), the Engineering Supervisor was helping the COSS to complete a form and the conversation was interrupted by other people who needed to speak with the Engineering Supervisor. The Engineering Supervisor also broke off the conversion at one point to contact the PICOP about a related matter.

Other figures (not shown) have been used to illustrate the frequencies of communications with various individuals and the location, and illustrate the area (mileages) of the work site where different contractors or work groups were working, indicating the potential for overlap in the working areas of these work groups.

Details from observations from the fieldwork, interviews, discussions and reviews of documentary evidence were tabulated, and through a series of qualitative comparisons, common themes have been identified. Table 32.1 provides a summary of the range of issues which have been considered in this part of the project.

Table 32.1 Themes on which information has been collected in fieldwork, interviews and review of documentation

Main themes	Breakdown of the main themes
Organisation and planning	Timings of possessions • including problems with the coordination of timings for possessions and isolations Coordination of work through planning • within companies • between companies on site or in neighbouring work sites and successive work sites • between functions, such as the effect which tamping can have on previous overhead line work Variability of workload of the ES • by type of work site • type of shift e.g. single shift or consecutive shift • higher workloads at the start and end of the shift Availability of resources on site • including the volume and quality of paperwork Suitability and competence of personnel • including expertise and experience • local knowledge • access to off site support in the event of problems

Continued

Main themes	Breakdown of the main themes
Briefings	Quality of briefings • content • complacency and degree of attention of participants • numbers attending (which was seen to be in excess of 50 in some work sites) • location of briefing (size and the suitability of on site cabins) • interruptions • use of visual aids / plans
Communications	Workload arising from communications • frequency of communications and details of common contacts • Quality of communications • communication protocols Methods of communication • mobile telephone, radio or other • factors affecting choices of communication method
Other issues	Placement of marker boards • including communications and instruction of competent person Movements of trains and road rail vehicles Role of the ES in supervision of work in the worksite

Methodological issues arising from the field observations

The Engineering Supervisors were happy to participate in the study and appeared to be very comfortable with being observed. They were keen to answer questions and readily participated in more in depth discussions of the role. The nature of the work which was observed during the field study indicated that the Engineering Supervisor is likely to have periods of high intensity work (particularly at the beginning and end of shifts) and this presented some difficulties in terms of seeking clarification on some activities or communications. In particular, the researcher's ability to hear only one side of telephone conversations was often a problem, though it was usually possible to back track during quieter periods to discover the source of any relevant communications and the content of these conversations.

Periods of high intensity work were found to be at the beginning and end of a shift, so subsequent observations were scheduled to capture relevant data in an efficient way. Through assistance from the participating companies, a schedule was prepared which allowed observations at several sites during a single night shift, with staggered start and finish times. Observations were carried out within the hour leading up to the start of the work site (when the Engineering Supervisor was on site and making early contact with the PICOP) and then during the first hour of work at the work site, when the majority of staff were signing into the work site and receiving briefings. Similar observations were then made at the start of a nearby work site, before moving to a final site to see the completion and hand back of that site at the completion of the works.

There were some occasions when an Engineering Supervisor could not be observed. On one occasion an Engineering Supervisor had to go to another part of a site with a colleague and

arrangements for transport for the researcher were not available (though generally the researchers were able to accompany the Engineering Supervisor during visits to different parts of the work site). Occasionally, the researcher was involved in discussions with other on-site staff when an activity involving the Engineering Supervisor occurred. Sometimes, it appeared that the Engineering Supervisor just needed a short break from being observed and retired to a car or other part of the site. The desire for this type of privacy is understandable and could be accommodated with the research. Where it was felt that periods of observation had been missed the researcher used a variant of a verbal diary system (e.g. Van der Beek et al., 1994) to collect any missing observations. The Engineering Supervisor was asked, for example, what tasks he had done, who he had spoken to, or what calls he had received during the fifteen minutes or so that he had not been observed. The use of diaries is a legitimate and reliable tool in human factors research and it is felt that such questions were effective within the current research, enabling the collection of important information, without endangering the strong relationship which had been built between the researcher and participant.

Phase 2 – Evaluating proposed changes to possession rules

The study of the Engineering Supervisors described above met the needs of the Network Rail Ergonomics Group to begin to understand more about the socio-cognitive human factors of track work (e.g. briefings, communications, potential for error) and the West Coast Route Modernisation Safety Group to reduce the potential for error in possessions. Both of these groups, together with the Network Rail Engineering and Operations functions (at a very high level) recognised that much more might be investigated and much more developed in relation to new ways of planning, managing and running possessions. Thus the work broadened into a more general study of possessions and of ways to maximise time in carrying out engineering works.

The IOE and the Network Rail Ergonomics Group provided human factors support in an on-going evaluation of proposed changes for the management of engineering possessions. These proposed changes introduce two new roles and the changes in working arrangements could also impact on the role of the PICOP. However, it must be stressed that these are purely hypothetical changes for now, with much more assessment and consultation to take place before any concrete proposals are made.

Proposed roles of the site supervisor and site movements controller, and the role of the PICOP

Draft rule changes have outlined new functional roles to take on many of the existing duties of the Engineering Supervisor, including a site supervisor (to take control of the work site) and a site movements controller (to control rail vehicle movements within the work site). The proposed changes also include changes to the sizes of work sites, the relative locations of work sites, and the nature of work and responsibility for the work within work sites. The changes do not include any specific changes to the role of the PICOP (who is defined in the Railway Group Standard GO/RT 3261 as "a person certified as competent and appointed to establish, maintain and give up a possession of a line or lines, and authorise the movement of trains over sections of line or lines identified as being under the control of the PICOP") though it is envisaged that the proposed use of a larger number of smaller, more focused worksites with one principal contractor, could have implications for the workload of the PICOP. These proposals have been discussed with interested parties from a wide range of organisations and working arrangements have been tested in two desk top exercises. Details of the human factors evaluation of the most recent of these exercises is outlined below.

Approach to the evaluation

A two day workshop was arranged at the Network Rail training centre in Watford and was attended by planners, possession assurance managers, PICOPs, Engineering Supervisors and isolation staff. Proposed changes in working arrangements were explained and a desk top simulation of a possession was run. The scenario was based on a script which specified work in five worksites in the possession. All supporting documentation for the worksites was provided and the track layout for the possession was illustrated on a large plan. Individuals who played key roles in the scenario received detailed briefings of the nature of the work which would be under their control within the scenario and were directed to different rooms within the training centre for the purposes of the role playing scenario. Communication between individuals was by way of a telephone system and standard safety critical communications were to be used for all communications.

Attendees were allocated roles to play within the scenario which complemented their usual roles within the industry. For example the PICOP role was played by an experienced PICOP, isolation staff took on isolation roles in the scenario, an experienced signaller was used for the signalling role and the new roles of the site supervisor and site movements controller were played by experienced Engineering Supervisors. Other parts in the scenario (e.g. train drivers, other competent persons) were played by attendees who were familiar with the roles and the type of dialogue which would be used.

Staff from WCRM at Network Rail observed the progress of the scenario, considering the practicalities of the rule changes. Four human factors specialists participated in the human factors evaluation, in observations of the work of the PICOP, site supervisor, site supervisor of access, site movements controller and the nominated person for isolation. Detailed notes on the frequency and content of communications to and from individuals were collected using a matrix which had been prepared prior to the study. After completion of the desk top scenario, feedback was obtained from the participants, collecting their perceptions of the desk top trial and the likely effectiveness of the proposed rule changes.

Findings from the observations of the desk top exercise

Notes on observations of the desk top exercise were transcribed and used in producing a series of figures to summarise the findings of the trial. Figure 32.3 is included below by way of example, and illustrates the range of activities undertaken by the PICOP. Most of the communications were to or from the site supervisors (the majority of whom were also acting as site movements controllers in the scenario) and related to the movements of trains or road rail vehicles. The analysis of the data has indicated that communications to move a road rail vehicle from one work site to another work site could vary between three and seven telephone calls. As the proposed changes in possession rules introduce greater numbers of smaller work sites (and the use of a specified access point for those in a number of work sites), there is clearly the potential to greatly increase the numbers of movements of vehicles between work sites. Consequently, the numbers of communications required to move vehicles between and through adjoining worksites will also be likely to increase, thereby adding to the workloads of the PICOP, site supervisor and site movements controller. Further analysis is now necessary of the nature of the communications required in the movement of vehicles within and between work sites. Preliminary analysis of the communications in the current study has indicated that there are likely to be differences in the nature of communications for the movements of trains and road rail vehicles, with greater numbers of communications being likely with the movement of trains. Regrettably, in-depth analysis of these communications has not been possible because of the loss of data when the voice recording telephone system failed during the

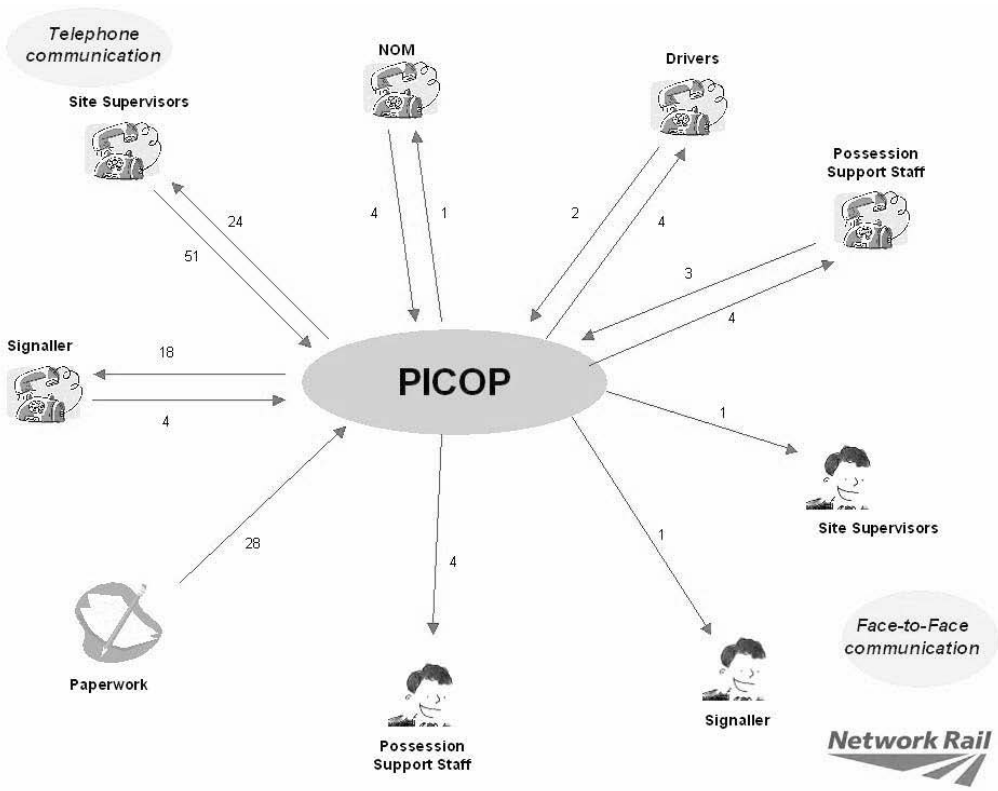


Figure 32.3 Activity analysis of the PICOP

desk top scenario. It has been suggested that there may be scope for streamlining the numbers of telephone calls which are needed during the movement of vehicles between work sites, though a detailed analysis of the risk involved in any reduction in the numbers of such communications will be necessary.

Methodological issues arising from the field observations

The attendees at the workshop participated fully in the desktop scenario. A number of the participants had taken part in a preliminary exercise which was run some months previous to this, and so were familiar with this type of role play experience. Both the positive approach of the participants to the role playing and the previous experience of those participating were strengths and contributed to the quality of the exercise and its findings.

A number of the participants commented on the lack of complexity of the scenario. The scenario which was used was straightforward, with a limited number of train and road rail vehicle movements, both within and between neighbouring work sites. There were no complications arising in the possession. The exercise provided valuable information on the likely communications for a sample of tasks or activities within the scenario, but this will need to be extended to test the arrangements in more complex track layouts and with larger work activities, also testing the types of communications which would be needed between the site supervisors and site movements

controllers. The exercise has so far provided a good opportunity to collect early feedback on the perceptions of those in the roles which could be affected by the changes, though again, feedback from more complex scenarios will be essential.

It had been intended that a specialised telephone system would be used for recording all communications between participants during the course of the trial, but due to technical power problems this system had to be abandoned after a few minutes of the trial and participants used their mobile telephones for subsequent communications. Feedback from the participants indicated that they welcomed this change as the specialised telephone system did not seem to be intuitive to use and the participants were more comfortable using their own mobile telephones. Communications during the trial were therefore simplified, but the loss of the specialist telephone system had the unfortunate consequence of the loss of ability to record the communications. The problem for the researchers of only being able to hear one side of a telephone conversation (which had been experienced in the field study with Engineering Supervisor) was experienced again in this desk top exercise. This effect was perhaps greater in this current exercise because of the high frequency of communications involving the PICOP, with little opportunity to seek clarification of the source and content of the communications.

Observations at the desk top scenario gave some insight into the numbers of communications between key staff, and it was this aspect of the work which was the primary focus of observations in this part of the study, giving an early indication of likely effects of the changes in arrangements on workload. It must be borne in mind that the time of the possession (which would have been an eight hour possession) was compressed into a four hour period for the purpose of this desk top exercise. This was designed in this way to save time during the exercise, and realistic gaps were not included for the completion of the simulated work activities which were included in the scripts as there would have been few communications between key roles during this time. Therefore, the frequencies of communications may be informative but the gaps between the communications will be shorter than would be anticipated in relation to real work. Further work is in progress to develop a suitable means of measuring workload in key roles, such as the PICOP and site supervisor.

The exercise has also given some insight into the frequency of use of paperwork (such as completion of the mandatory forms, reference to plans or work site documentation). However, it has not been possible to obtain accurate records of the frequencies with which different items of paperwork were completed or used as points of reference. It was often difficult during the real time observations, particularly when the observer was also recording details of conversations which could be conducted concurrently with the paperwork task, to determine which documents were being used by the PICOP, or how many different documents were looked at to identify relevant information, or for what purpose they were used. Therefore, revisions to the study methods would be required in further studies to obtain accurate information on this type of data.

General conclusions

The research outlined in this chapter has given an indication of the high numbers of communications involving the Engineering Supervisor, at least in parts of the shift and in large worksites. Factors which are likely to contribute to the high numbers of communications and the potential for high workload for the Engineering Supervisor have been identified. Observations within a range of worksites have raised questions about the extent to which the role includes elements of direct supervision of the work in the worksite, though such supervision has been found to be unlikely where the work site is large and multiple work activities are carried out by multiple work groups or contractors.

Initial proposals have been made within the industry to assess implications of introducing a new role of site supervisor, in place of the Engineering Supervisor, with wider supervisory responsibilities, but having responsibility over a smaller geographical area and with fewer items of work within the work site. However, it has been established in the current work that the division of responsibilities within engineering possessions is complex and such changes in responsibility could shift problems with workload to other roles. Concerns have been raised during the desk top simulation of a straightforward possession about the frequencies of communications involving the PICOP, who would have considerably more site supervisors reporting to him in the proposed method of working than Engineering Supervisors reporting to him under the existing method of working. Additional options for changes may include support for the PICOP or increasing the numbers of PICOPs, but this could again shift the workload problem to the signaller. Findings from another project (the REQUEST survey of attitudes and opinions of signallers, Ryan et al, 2005) suggest that many signallers may already be dissatisfied with workload arising from possession related work. Options for changes will be evaluated in future work.

Human factors input has been included at an early stage of proposals to review possession arrangements and has received support from senior managers within the organisation. In this project a range of human factors methods have been applied and the successes and limitations of the methodology have been discussed. A range of data have been collected and used effectively to evaluate both existing and the proposed changes to working arrangements.

Acknowledgements

The authors would like to thank the managers and staff at a number of companies who gave their time willingly for this piece of research and the members of senior management at WCRM who have driven and supported this work. The authors would also like to acknowledge the work of Anthony Slamen, formerly of Network Rail, for his role as a researcher in the Engineering Supervisor project.

References

Farrington-Darby, T., Pickup, L. and Wilson, J.R. (2005) Safety Culture in Railway Maintenance, *Safety Science*, 43, 39-60.

GO/RT3000 The master rule book, Railway Group Standard.

GO/RT3261 Persons in Charge of Possession (PICOP) and Engineering Supervisors (ES), Issue 2, December 1999

Ryan, B., Wilson, J.R., Sharples, S., Marshall, E., Morrisroe, G. and Clarke, T. (2005) Collecting attitudes and opinions from signallers: Development and use of REQUEST (the Railway Ergonomics Questionnaire), Second Rail Human Factors Conference, London, UK.

Van der Beek, A.J., Braam, I.T.J., Douwes, M., Bongers, P.M. Frings-Dresen, M.H.W., Verbeek, J.H.A.M. and Luyts, S. (1994) Validity of a diary estimating exposure to tasks, activities, and postures of the trunk, *International Archives of Occupational and Environmental Health*, 66, 173-178.

This page intentionally left blank

Chapter 33

Human Factors Integration for the Computerised Track Access Control System

Chris Lowe, Dan Lock, Barnaby Annan, Paul Thompson and Paul Raistrick

Background to project – Track Access Control

The Track Access Control (TAC) Department provides a service to all London Underground lines, supporting safe and timely access to track area working sites by suitably prepared and qualified personnel. A safety related aspect of this service is to ensure that the electrical current remains switched off at the relevant traction current rail sections while workers are located on that particular section of the track. Essentially this is achieved by recording the location of track workers when they begin work, and then the departure of the track workers from the work site when they call to report that they have left the work area. It is only after the work site is reported as being clear that the current can be re-charged.

The roles associated with the Track Access Control function, and the other interacting roles within London Underground can be seen in Figure 33.1.

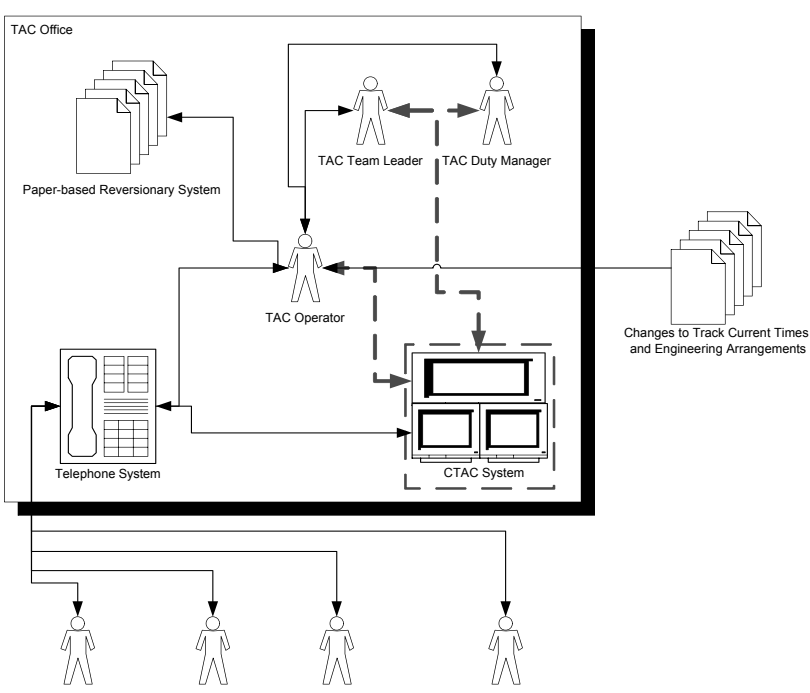


Figure 33.1 TAC department roles, equipment and interactions

The central role within the Track Access Control process is that of the TAC Operator. The primary role of the TAC is to operate the telephone and computer system to grant safe access to the track and to ensure that individuals are suitably qualified and informed. A top-level task analysis can be seen in Figure 33.2.

At the start of the shift, around 22:00, the TAC will prepare for Engineering Hours. This involves reviewing the appropriate Engineering Notices and Current Arrangements and noting any changes to the times at which track circuit sections are usually re-charged. After 23:00, the work crews will begin to call the TAC Department to request access to the track. Each track-based work crew is under the control of a Protection Master, who must call (by telephone) the TAC Department to book on, reporting the number of persons and various other critical data about the work and state of the track during the work, and the sections of track affected. This information is then verified by the TAC. The TAC must use their experience, knowledge of the applicable published engineering notices, and a computer system called TACMIS to verify the competency and safety licences of the Protection Master. The validity and nature of the work, which will have been previously submitted to LU by the contractor and approved, is also checked against a computerized database.

If the details of the work and the licences of the Protection Master are in order, the TAC will record the work location that the Protection Master requires, and provide the Protection Master with a unique reference and a call-back time. The call-back time is the time by which the Protection Master and associated workers must have left the track and reported this fact to the TAC Department. When a Protection Master calls back after leaving the track, the TAC must identify the correct working party and book them off the system. If a Protection Master fails to book off by the call-back time, and cannot be contacted directly by the TAC office, then procedural measures are employed and a section or sections of track may have to be searched to ensure that the track is clear. Once all Protection Masters have reported that they have left the track and that the line is safe to run trains, the TAC can call the appropriate power-switching centre and report that the track circuit sections are clear. The Power Control Room Operator will then re-charge the associated track circuit sections.

The TAC Department currently utilise a part paper-based, part computerised approach. While a computerised database is used to check Protection Master competencies and work booking details, the recording of a work party at a particular work site is presently recorded on a paper form. This form can involve the duplication of information from one part of the form to another, and relies on the professionalism and vigilance of the TAC to avoid errors.

The purpose of the Computerised Track Access Control System (CTAC) is to provide a computerised system that will replace the existing paper-based system used by the TAC Operators. This system will integrate with the telephone system, and transfer some of the forms and activities performed by the TAC Operators into software.

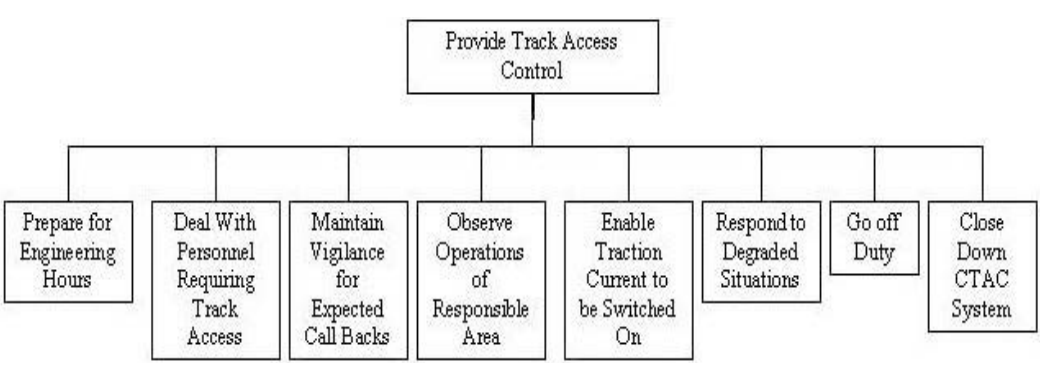


Figure 33.2 Top level goals of Track Access Control operator

Therefore, the CTAC Human Factors Integration (HFI) programme is principally concerned with the design, development and introduction of the CTAC Software System, and the way in which this interacts with the users within the TAC Department. The HFI programme is intended to provide assurance on the safe and efficient interaction between the TAC Operators, Team Leaders and Managers, and the CTAC system.

Background to project – Human Factors Integration

The CTAC HFI programme is being conducted in accordance with the LU standards 2-01018-003 (Integration of Human Factors into Systems Development) and M1035 (Good Practice in Human Factors). The relevant parts of the LU standards 2-01018-004 (Universal Human Computer Interaction) and 2-01018-002 (The System Engineering Process Applied to Projects) are also addressed through the HFI programme, and certain other human factors engineering standards are being used for guidance purposes.

Much has been written on the subject of 'integration' in relation to human factors (Booher, 2003). Integration is often seen as central to the successful conduct of human factors within engineering projects, but very little guidance is ever provided on how to conduct 'integration', or even how the term can be defined. Our strategy for human factors integration on the CTAC project had five components:

Coherent view of human factors

We realised that before human factors can be integrated with other processes, there must be a coherent view within the human factors programme itself. The scope of human factors relevant to the project, and the relationships between the types of human factors, needs to be defined. This is often achieved through the use of the 'human factors domain' concept. For CTAC, we found that the EHFA output was sufficient to provide coherence to the HFI programme (see later).

Goals and values

We thought that we would increase our likelihood of successfully integrating human factors into the wider engineering processes if the human factors goals were aligned with the overall goals of the CTAC project. Fortunately, the Operational Concept and User Requirements Specification provided a common point of interest for both the human factors and other engineering domains.

Process and methods

For the CTAC project, we tried to ensure that the HFI programme 'spoke the language' of the wider software and safety engineering. This meant that we attempted to carry out integration at the process level, and opportunities were taken to address human factors within software and safety assessment methods and techniques.

Timescales

At the start of the project, we realised that timeliness would be an important aspect of integration. Conducting human factors engineering activities at the correct time would ensure that important human factors and engineering activities are occurring when the issues have common relevance within the overall project.

Outputs

The fifth type of integration we attempted was integration of deliverables. This refers to the requirement for human factors and other project outputs to be mutually supporting. We stressed the need for information to be shared in a format that was of common benefit.

CTAC human factors goals

In accordance with 2-01018-003, an assurance-based approach to the HFI programme was attempted. To support the assurance-based approach, it was recognised that there was a need for project-specific top-level 'requirements' that would distil and articulate the wishes of the client and user group with regard to human factors. To address this need, we used the concept of a 'human factors goal'. A 'human factors goal' is a type of derived requirement or recommendation.

The human factors goals were derived during the Early Human Factors Analysis (EHFA) phase of the project. Table 33.1 presents the high-level Human Factors Goals for the CTAC project.

Table 33.1 High-level CTAC Human Factors Goals

Goal	Source	Addressed Through
The CTAC system shall not incorrectly advise the TAC that the line is clear	User Requirements Specification	Human Error Analysis, Safety Engineering Integration, User Interface Design and Testing
The CTAC system must be able to support a flexible method of managing access requests	User Requirements Specification	Human Error Analysis, Safety Engineering Integration, User Interface Design and Assessment, Organisational Change and Training Plan
The new CTAC system should support a method of working that serves to reduce or 'level' the amount of time a Protection Master would spend in a telephone queue.	Operational Concept, User Requirements Specification	User Interface Design and Assessment, Workload Analysis
The CTAC system should not introduce any significant new sources of human error, and appropriately manage the sources that are common with the paper-based system.	Human Reliability Assessment of Track Access Control Centre	Human Error and Workload Analysis, Safety Engineering Integration
The development of the CTAC system should demonstrate compliance with the applicable LUL Category 1 Human Factors Engineering Standards	User Requirements Specification	All work packages

Through the associated work packages, these human factors goals have been further decomposed, associated with measurement methods, and operability criteria have been developed where relevant.

The purpose of these 'human factors goals' was to provide a small number of clear objectives, traceable to previous analyses and feasibility studies, that could be used to provide focus to the HFI programme. It was hoped that this would ensure that the HFI programme related to the wider project objectives, and also had a set of top-level objectives under which the remainder of the HFI programme could be organised in a fashion that supported an assurance-based approach, rather than a process or methodology-centred approach.

We found that having this type of high-level derived requirement was useful in capturing significant high-level human factors issues that were relevant to the CTAC project. Using this approach, we were able to address concerns such as the need to recognise the transition from an existing paper-based system to a computer-based system, and the need to provide assurance that potential hazards associated with human reliability were identified and appropriately managed.

Overview of CTAC HFI programme

The overall structure of the HFI programme, in relation to the wider system, software and safety engineering activities, can be seen in Figure 33.3. The human factors activities were organised into a number of themes (shown in the left-most column of Figure 33.3). Activities within each theme were then aligned with the overall system lifecycle model.

The CTAC HFI programme began in August 2004. It is scheduled to run beyond the design and installation phase of the CTAC System, into the first year of operation with the new system.

	Concept and Feasibility	Requirements Definition	Design and Implementation	Installation and Handover	Managed Service
WP1 Human Factors Integration Management	HFI Plan HF Issues Database	Update, Mitigate and Resolve HF Issues			
WP2 Early Human Factors Analysis	Interviews and Observations Review of Operational Concept				
WP3 Task, User and Context Assessment	Predictive Task, User and Context Model				
WP4 Human Error and Workload Assessment	Human Error Assessment Workload Assessment				
WP5 Safety Engineering Support	HAZOP Attendance	Safety Case Integration			
WP6 User Interface Testing	HCI Style Guide	Usability Performance Criteria	Usability Assessment		
WP7 Change and Training Support				Change and Training Plan	
WP8 Ongoing HFI Maintenance				Update or review HFI deliverables	

Figure 33.3 Alignment of HFI work packages with project lifecycle

Human Factors Integration Database

After the HFI structure had been established, and the top-level human factors goals derived through EHFA, the individual work packages were developed and executed. To support the CTAC HFI Programme, and in particular the execution of the human factors engineering work packages, we used a 'Human Factors Integration Manager' database tool. This tool is a Microsoft Access database that assists with the management of human factors integration projects. It assists the analyst in the following activities:

- Management of a Human Factors Issues Register
- Representation of Hierarchical Task Analysis information
- Performing a number of human factors analyses, including: #
- Human Performance Risk Analysis
- Human Error Identification
- Workload Analysis
- Training Needs Analysis
- Generation of reports associated with the human factors analyses and reports.

An overview of the Human Factors Integration Manager Tool can be seen in Figure 33.4.

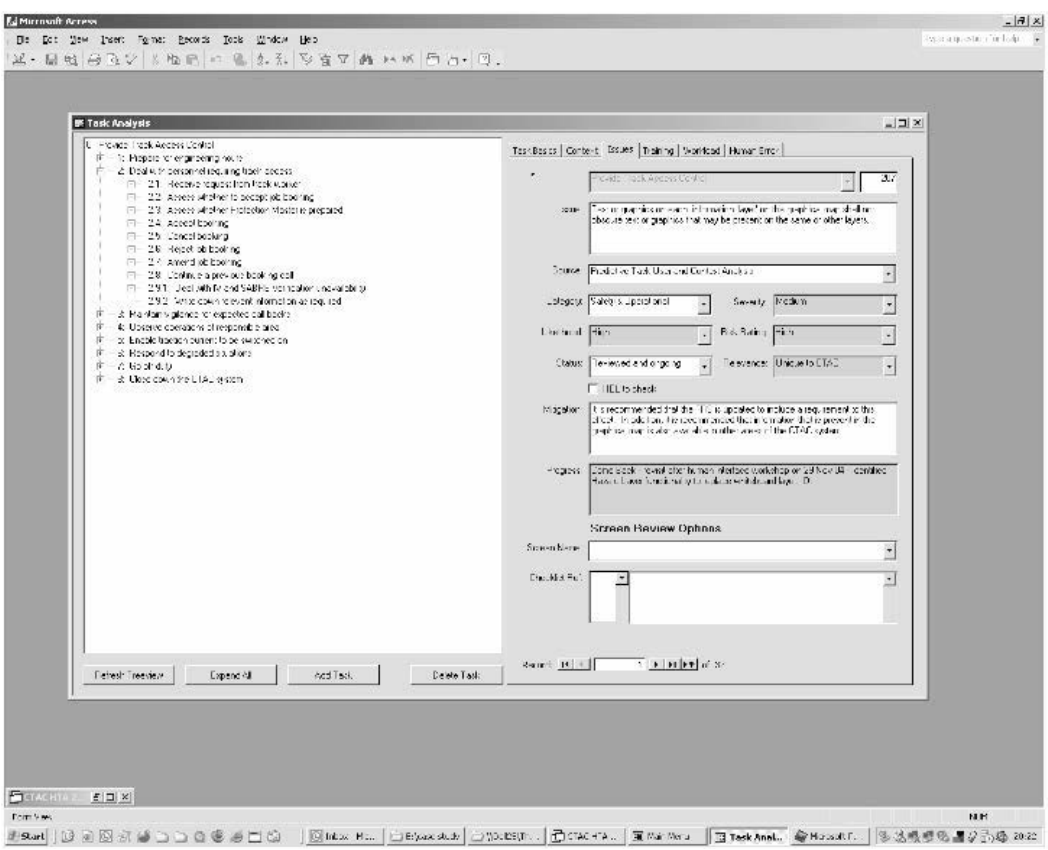


Figure 33.4 Human Factors Integration Manager issues log

Identification of human factors issues

An early human factors engineering work package for the CTAC project was to conduct a predictive task analysis of the new system. The purpose of this activity was to construct a predictive model of user interaction with the proposed CTAC System, in accordance with the Functional Requirements Specification, prototype CTAC user interface, and CTAC Operational Concept. These task data were represented within the HFI Manager tool.

The hierarchical task analysis (HTA) functionality within the tool presents the various levels of task description within a 'tree view' format that can be expanded to reveal the tasks and sub-tasks that contribute towards the goal. The tool also supports the recording of plans. 'Drag and drop' can be used to move parts of the task analysis around within the tree.

The database tool also provides a way of associating contextual information, such as photographs, with tasks in the task analysis. We found that this provided a useful means of storing photographs and video footage that was recorded during the interviews and observations at the TAC Office that were used to construct the task analysis. Showing the photographs in context with the task communicated a common understanding of the work environment to other project team members.

Based on the predictive task analysis, other types of analysis were performed in order to identify potential issues with the CTAC concept design. In particular, a qualitative workload and human error screening exercise was performed on each task in the database. The Workload and Human Error Analysis was designed to answer questions that related to the top-level human factors goals. These questions were:

- What are the credible sources of human error, and the associated underlying cognitive error mechanisms, associated with the new system?
- What is the predicted workload that TACs will experience while working with CTAC, especially in relation to the booking-on process?

The CTAC predictive task analysis was screened for potential human error and workload issues. The human error screening generated a number of potential human errors of omission and commission associated with the CTAC system. Following a workshop review with TAC Operators and DS&S and LU CTAC project team members, a set of high-priority or safety-related human errors were identified. It was these tasks and potential errors that were taken forward for further analysis.

The workload screening results were reviewed at the same workshop. It was agreed that further scenario-based workload assessments would be undertaken. Normal and abnormal scenarios were constructed to cover the main phases of CTAC interaction over the shift. The detailed human error analysis further examined the underlying psychological error mechanisms associated with each error, and put forward mitigations based on addressing these root causes. The detailed workload analysis used ATLAS, Human Engineering's human performance modelling software.

As a result of this work, human error and workload issues were identified with the proposed CTAC system design. All of these issues were associated with plans for resolution, and tracked through the Human Factors Integration Manager tool.

Design to mitigate human factors issues

Once the analysis phase of the CTAC concept design was complete, the emphasis of the HFI programme turned towards design activities. In this phase of the project, we attempted to derive practical solutions to the identified human factors issues. This phase relied on working closely with the software development team and safety assessment specialists.

A collaborative design approach was taken to areas of the user interface that had been identified as requiring further development. A user interface workshop was held at the TAC Office, attended by representatives from Human Engineering, DS&S and LUL CTAC Project Teams, and TAC Operators. The purpose of the user interface workshop was to discuss and brainstorm user interface solutions and options to the identified issues.

Training design was also used to address certain classes of human factors issue. This involved two steps: a training needs analysis (TNA) process in order to identify tasks where training support would be beneficial and a training material design phase where course materials and a user guide were developed to address in particular the tasks with human performance risks.

The HFI Manager database includes functionality necessary to perform a training needs analysis (TNA). Using pre-defined definitions, tasks from the task analysis can be set a difficulty, importance and frequency rating. The performance risk and the TNA ratings are automatically calculated from the information in the difficulty, importance and frequency fields. Consideration was given to the extent that a particular CTAC task introduces new knowledge, skills or attitudes (KSA) required from the TAC User. In instances where the CTAC task introduces new KSA attributes (i.e., the CTAC task is novel, and not just a new implementation of a task common to the existing paper-based system), this was noted in the analysis and the Training Objective and Recommendation adjusted accordingly.

Assurance that human factors issues have been addressed

The third phase of the CTAC HFI Programme is concerned with providing assurance that the human factors goals, and issues identified during the analysis phase, have been appropriately managed. In relation to the human factors goals and issues, the assurance strategy is based on three elements: demonstration of process sufficiency, demonstration of achievement of criteria, and demonstration that risks have been reduced as low as reasonably practicable. In order to demonstrate that the human factors integration process has been adequate, traceability from the requirements of 2-01018-003 (Integration of Human Factors into Systems Design) to the execution of human factors work packages can be demonstrated. This provides basic assurance that the conduct of the HFI programme has been carried out in accordance with the HFI Plan, by competent individuals.

For criteria-based requirements, a range of compliance activities are planned. The type of compliance demonstration activity that is required relates to the human factors goal or issue against which assurance must be provided. For example, a compliance checklist is being developed as part of the HFI programme, in order to extract acceptance criteria from the relevant standards and identified human factors issues. This checklist will be used to show traceability between the relevant standards and the delivered CTAC System.

Conclusions

Although the CTAC HFI programme is on going, the project has enabled us to draw a small number of conclusions. Given that the LU category 1 human factors standards attempt to provide an ‘assurance-based’ framework for HFI, we found that this places a great deal of importance on an ‘early human factors analysis’ stage. During this stage on the CTAC project we were able to perform a type of ‘project characterisation’ where we were able to assess the nature and scope of human factors input that was required, and how this affected both the approach to integration in a management sense, and the focus of the human factors engineering work packages. Most contemporary human factors integration process standards recognise the requirement for some

form of project characterisation and we found this activity to be very useful in ensuring that the HFI programme had a focus and only addressed the significant risks within the CTAC system.

Using a multi-faceted approach to integration with other project areas has also been relatively successful. Attempting to align with the wider project phases, and adapting the human factors engineering methods accordingly, provided the HFI project with a means of influencing the wider project decisions. This was because the human factors outputs were timely, and relevant to the wider questions of the project.

The Human Factors Integration Manager database tool also proved to be useful during the project. We found that the main use of the tool was in organising the large amount of information produced by the human factors engineering analyses, and providing an audit trail for identified issues. This allowed the human factors analysts to concentrate on assisting the wider project, rather than concentrating on human factors methodology. We believe that the CTAC HFI Programme has demonstrated that the category 1 LU human factors standards provide a reasonable framework within which HFI can be conducted.

References

Booher, H. R. (2003). *Handbook of Human Systems Integration,* John Wiley & Sons Inc.

LUL, July (2003). *M1035 – Manual of Good Practice – Good Practice in Human Factors.* London Underground Limited Director of Engineering.

LUL, December (2003). *2-01018-003 Engineering Standard – Integration of Human Factors into Systems Development.* London Underground Limited Director of Engineering.

LUL December (2003). *2-01018-002 Engineering Standard – The System Engineering Process Applied to Projects.* London Underground Limited Director of Engineering.

LUL February (2004). *2-01018-004 Engineering Standard – Universal Human Computer Interaction.* London Underground Limited Director of Engineering.

This page intentionally left blank

PART X
Level Crossings

This page intentionally left blank

Chapter 34

The Impact of Aggressive Driving on the Design of Level Crossing Safety Measures

Andrej Godec and Zdravko Toš

Introduction

There have been significant endeavours in the field of level crossing safety, not only on the railway networks of Croatia and Slovenia, but also on the road network, in contributions from various professionals. In general, the safety level is adequate and socially acceptable, but society and responsible institutions are continuously looking for cost-effective improvements.

Several years ago there was an "open" Slovenian conference devoted to the problem of safety at level crossings. The conference proceedings show an adequate level of knowledge, good recognition of problems, wide coverage, presence of professional potential, etc. The only sad fact was a lack of money for all the suggested improvements. The reality is that "...the society has as much safety as it wants to have..". The society does not want to spend millions of Euros where there is a risk of just ten fatalities, if there is a possibility to save hundreds of lives (e.g. in hospitals, with better equipment) in the same period and for the same amount of money.

To further investigate the level crossing problem, the authors performed a system analysis as the first step in finding the principal issues in the present situation, taking into account a wide spectrum of influences. The study was set up in the light of the over 400 previous relevant publications found by the authors, for instance covering risk assessment and aggressive driving. Our work was to contribute to knowledge of design of the level crossing geometry, road quality, visibility triangle, etc. (Stipetić (2002) or Evans (1991)), while a specific effort was made to recognise and evaluate improvements on the entire network and in cases with a high incidence of rule violations on roads.

Systems approach in the analysis of risk at level crossings

As a first step in the research, a systems analysis was made comprising:

- Relationship to the external world
- Relationship to road traffic in connection with an actual crossing
- Relationship to railway traffic in connection with an actual crossing
- Micro-space around the crossing in the strict sense of the word, influenced by various spatial functional segments, devices, procedures and system relationships characterising the behaviour of participants and devices.

This structure first appeared in Jelaska (1999), and its importance can be briefly explained as follows. The external world has a decisive role in that society would support effective spending of money with the purpose of minimising risks in all of its activities, but the infrastructure owner

prefers to minimise the risk in its own infrastructure. These two divergent goals should be balanced, as can be understood from Litman (2003). The car driver is a part of the road subsystem. At the same time he is the weakest link within all the subsystems. His behaviour strongly influences the micro design of the crossing. The railway subsystem operates trains whose mass, speed, and other parameters require absolute priority in the crossing. The crossing in its strict sense involves a number of problems concerning risk and adequate safety protection. The car driver behaviour belongs to the road subsystem, while the signalling and safety equipment is linked with both the railway and the road subsystems. All these represent several relationships, which must be examined within each subsystem and integrated in the interaction with other parts in the total system. The efficiency of such system abstraction is in its analytical and integrative power. It allows an analysis within each subsystem, and at the same time the essential relations can be elaborated at the level of the whole system (Figure 34.1).

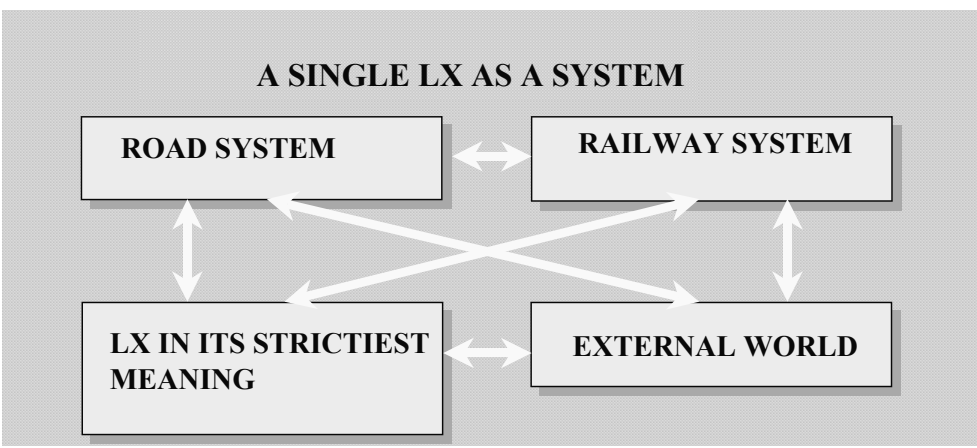


Figure 34.1 The level crossing (LX) problem in the system approach

Set of level crossings versus single level crossing

The model framework

The system of level crossings is modelled here in a specific way by following the three main dimensions:

- individual fatality risk,
- time, and
- critical points.

The *individual fatality risk* per year is taken as the most suitable variable. Compared to the fatality, other indicators, such as fuel consumption, travel time and distance, may significantly differ from source to source and they are not so precisely defined as the fatality. The fatality is often used as an index in other fields like health and safety at work, or city planning as well as transport, thus enabling comparison with safety levels taken from different sources. Figure 34.2 shows a block diagram indicating the composition of the risk indicator.

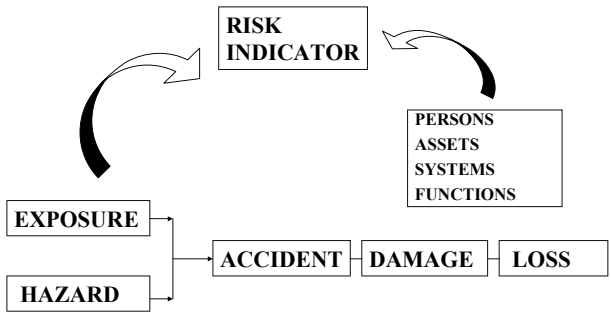


Figure 34.2 A diagram of risk indicator calculation

The next important dimension is *time – usually given in years*. Large economies, for instance, often use a 30-50 year period. Yearly values are suitable, because many statistics, plans, and investments are well defined on a yearly basis, as are fatality and damage data.

The third dimension is a set of *critical points*. The set comprises all critical points on a railway network. This set can be easily enumerated, much more easily on the railway than on the road network, which is often more complex. We use the term "points" because crossings are traditionally observed as points on a network, in contrast to sections of road and railway lines. One advantage of such a data record is that it is a compressed inventory of data, and another advantage is flexibility. With the same input set the crossings can be easily modelled also as:

- redesigned,
- with changed safety class,,
- closed
- eliminated,
- redefined for pedestrians only, etc.

Thanks to such adaptability, in further research one can include, if necessary, points of interaction of neighbouring road and rail line sections, prohibited but trespassed crossings, etc.

Investment scenaris comparisons

Each level crossing defined in the input set of observed critical points on the railway network has all the attributes necessary for the model calculations. The safety classes (type of protection) can be defined according to the goals of research. Each point belongs to only one safety class at a time. After a change, e.g., after an investment in safety improvement, the modelled critical point can be simply reassigned from one safety class to another from that moment on. Among input data per crossing, the improvement is defined by the expected cost of investment, coefficients for lower expected fatality risk, etc.

A typical analytical approach is possible at the single level crossing. Improvements in safety are treated as investments. Various safety classes for implemented measures are defined:

- St Andrew's cross,
- flashing red light,
- half-barriers,
- full-barriers,

- crossing closure,
- level separation, and
- combinations of these with medians (central barriers).

Relevant investments and maintenance costs are considered as costs. Savings are described as equivalents in the statistical value of life. Savings are based on the already mentioned risk calculations.

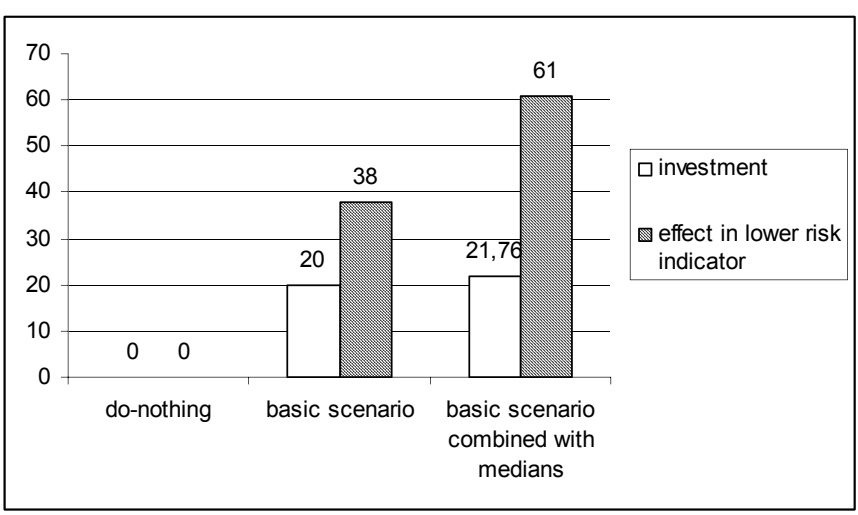
At the railway network level, due to an inherent lack of money, owners tend to avoid expensive solutions like level separation etc. From the standpoint of the infrastructure owner, an evaluation must cover the entire railway network instead of evaluating one single level crossing (e.g. Nichelson, 2000). If the owner's budget is limited, one expensive level separation may postpone many inexpensive improvements so that the overall result over a longer period could be negative. In other words, it is possible to find potentially better scenarios regarding an overall risk reduction.

Figure 34.3 presents one example of output results. The aim is merely to show the completeness of this approach. It can actually cover many different scenarios over a long period, with various safety classes and types of traffic. Among the more than twenty input parameters, the proportion of aggressive drivers in a particular road vehicle flow can also be found.

Improvements in modelling the impact of aggressiveness

Hourly variations of road vehicle flow

There is a certain lack of knowledge while looking for a good "network approach". One problem, often disregarded, is in the impact of road flow variations during a 24-hour period. The road owners usually apply just the Average Annual Daily Traffic (AADT) value, but this should be improved



Legend: x axis – scenarios
y axis – present values in million € of cost and effects for each scenario (medians' effect on aggressive driving reflects in the scenario combined with medians, contributing to the effect increase from 38 to 61 million €).

Figure 34.3 An example of comparison of three trial scenarios

when used in level crossing studies. Better involvement of flow variations is important since more dangerous violations may be expected in a dense flow, or within specific hours, or during a nervous drive to work or to school. Additional convenience may be expected with cheaper automatic traffic counting making the relevant data available.

A road vehicle flow with a high share of rule violations

An ever-present problem is of drivers disregarding rules while driving a road vehicle. The authors applied separate calculations of risk reduction for normal driving (characterised just by unintentional disregard for rules) and for aggressive driving (characterised by intentional and unintentional disregard for rules).

A specific effort to find a better approach in risk analysis at level crossings came with the idea to split the road flow of vehicles into

- the part of flow driving normally, and
- the part of flow driving aggressively,

each having a representative driver in the model.

Some experience from previous research in Slovenia shows heavy problems in road traffic, especially in the cases of some rural roads and numerous drivers making obvious rule violations. Several publications show the evidence of a high proportion of speed limit violations. Novak (1999) found from 80% of drivers violating low speed limits to 10% at 130 km/h, within the violating regions in Slovenia.

Consequently, aggressive driving can also be presumed at level crossings. Some observations performed in Eastern Slovenia confirm the existence of such rule violations. These violations appear at flashing red lights and observers report frequent slalom driving. Level crossings on rural roads with traffic flows in the range from 1500 to 10000 AADT were observed. However, considerable differences among statistical data from level crossings at various locations remained unexplained. Hypothetically they may be caused by various reasons, e.g., by local behavioural or cultural differences.

The concept of a reference level crossing

Road users cause a significant majority (approximately 97%) of level crossing accidents by disregarding at least one safety aspect. In contrast, the railways have built their safety systems carefully, with redundancy, so that only after a series of mistakes or errors resulting from long-term neglect can an accident be caused in the railway system.

Among road user problems, "risky" driving attracted a lot of research attention. This problem can be found in Witte (2000) followed by a question of what is the share of drivers showing risky behaviour at each particular level crossing. Interesting facts can be found in Corbett (1999) and Pearce (2002) and also in Meekr (1997), Hellman (2004) and Vijendran (2004). In our concept, the aggressive driving at level crossings can be determined by rule violations in spite of a clearly understood warning for punishment.

Based on these ideas, a particular vehicle flow can be modelled as split into two parts: non-aggressive and aggressive driving. The road flow modelling in general is based on Akcelik (1994). With the intention to measure relative variations in the model, the concept of a reference level crossing is used. |The reference level crossing flow was defined at 3653 AADT, for the present situation in Slovenia, for the set of level crossings with passive protection.

Pearce (2002) looked for causes in fatality accidents on roads in general. The study analysed court decisions. The causes can be divided into two groups: 56% unintentional (inattention or wrong judgement) and 44% intentional (rule violations or aggressive driving) causes. Combined with Corbett's (1999) findings, where 1.44% drivers intentionally violate rules, it seems that proportionally such drivers risk a many times higher fatality rate.

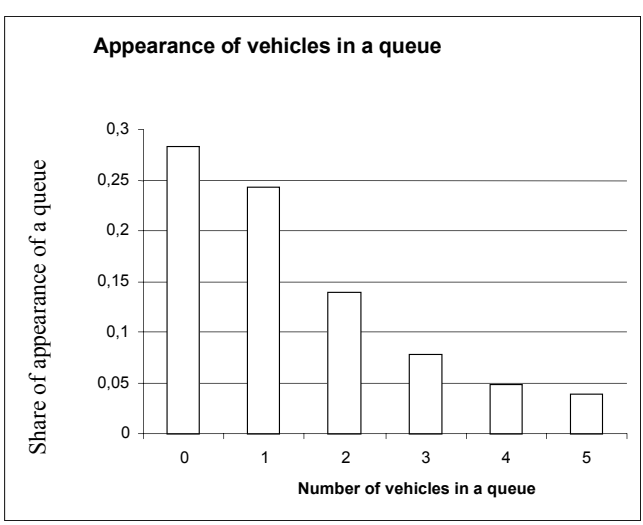
Based on these facts and the average traffic flow of 3653 AADT for a level crossing with passive protection the aggressive drivers will risk 1 fatality on 7 billion passages and normal drivers will risk 1 fatality on 385 billion passages a ratio of 1: 54,57. A reference level crossing helps to measure the impact of changes for various input data, such as: safety measures, variations in AADT, share of aggressive drivers, types of railway or road traffic, train length and speed, etc.

Efficiency of investment in medians

In the literature there are data to predict changes in risk when new safety measures are implemented. Compared to the St. Andrews cross an additional stop sign will half the risk. On average, video surveillance will reduce the risk approximately 4,4 times, flashing lights 3,5 times, and barriers 7,5 times. Full barriers with the four quadrants technique, video surveillance, and prosecution will reduce the risk 50 times.

In practice, actual results deviate a lot from these average and predicted factors. Therefore it is useful to investigate how these factors vary at various locations. There is not only a question of regional driving culture each crossing flow may show considerable deviations in response to the applied class of safety measure.

Within the scope of a single level crossing, the exposure of a road vehicle is calculated by a model. The model enables queue parameter calculation, so it is possible to estimate the additional effect and necessary length of medians. The authors made specific efforts to include and to evaluate the risk reduction at level crossings with a high share of rule violations, if possible by added medians (in a trial presented in Figure 34.3 – to obstruct the 2nd, 3rd, 4th or 5th vehicle with an aggressive driver in a queue to overtake). Such diagrams (as the one presented in Figure 34.4)



Legend: x axis – number of vehicles in a queue
y axis – share of occurrences of a queue of x-length

Figure 34.4 An example of the calculation of queues

depend strongly on input data. Many input parameters can be changed in calculations of a particular case: e.g., train flow characteristics, length and speed of trains, share of passenger trains, road flow characteristics, geometry of crossing, etc. The results from this and similar analyses of certain level crossings with significant aggressive road vehicle driving, can be inserted back into the overall model for the railway network as it is presented in the investment scenarios comparison.

Conclusions

The work presented here makes a contribution to the application of safety measures on a set of level crossings by the use of cost-effectiveness evaluation. The calculation of the risk of individual fatality has a central role due to its overall importance both for society and for an individual. Other losses may also be included through their equivalent values. This enables an evaluation of:

- risk indicator reduction as the effect of investment, and
- the relationship between aggressive driver behaviour and the design of safety measures.

If an observed road flow shows higher than average rule violations, it is efficient to combine medians with the existing or proposed safety equipment at this particular level crossing. The main advantage lies in the prevention of overtaking by drivers slaloming around the waiting vehicles.

Consequently, there are two improvements proposed in the chapter:

1. at the level of a single level crossing, the length of medians for a given risk reduction can be determined, depending on the proportion of aggressive drivers within the road flow, and
2. such risk reductions can be inserted in a model for evaluation of the investment scenarios on a set of level crossings on the railway network.

References

Akcelik, R., Chung, E. (1994). Calibration of the bunched exponential distribution of arrival headways. Road and Transport Research 3 (1), pp 42-59.

Corbett, C., Simon, F. (1999): The effects of speed cameras: How drivers respond. Centre for criminal justice research Department of law. Brunel, UK, Brunel University, Feb. 1999, 118 pp.?

Evans, L. (1991): Traffic Safety and the Driver, LC 90-44233, ISBN 0-4421-00163-0. Van Nostrand Reinhold, 1991, 404 pp.

Hellman, A., Caroll, A. (2004): Lessons Learned From The Evaluation of a Four-Quadrant Gate Grade Crossing With Obstruction Detection. Presented at the 8th International Level Crossing Symposium, CD-ROM Proceedings, The University of Sheffield, Sheffield, UK, 14-16 April, 2004, 15 pp.

Jelaska, M., Godec, A., (1999): The system analysis of road-railway level crossings, (in Slovene). 6th International scientific and professional colloquium: Traffic control, Maribor, 8 Dec. 1999. Maribor, Slovenia, University of Maribor, 1999, sep.pag. 5 pp.

Litman, T. (2003) Integrating Public Health Objectives in Transportation Decision-Making. American Journal of Health Promotion, Vol. 18, No. 1, 2003, pp. 103-108.

Meeker, F. Fox, D., Weber, C. (1997): A Comparison of Driver Behaviour at Railroad Grade Crossings with Two Different Protection Systems. Accident Analysis and Prevention, No 1, Vol.29 (2000). pp. 11-16.

Nichelson, R., (2000): Grade separations – when do we separate? 1999 Highway-rail grade crossing safety conference, Grade crossing technologies – The new millennium, Conference proceedings, Texas transportation institute, College Station, 18 Oct 1999. College Station, US: Texas A&M University, April 2000, pp. 60-73.

Novak, Š., Kolenc, J., Medeot, M. (1999): Impact of traffic sign and speed limits on the road traffic safety, (in Slovenian). 6th International scientific and professional colloquium: Traffic control, Maribor, 8 Dec. 1999. Maribor, Slovenia, University of Maribor, 1999, sep.pag. 6 pp.

Pearce, L.M., et al. (2002): Dangerous Driving and the Law (No.26). DTRL, PO Box 236, Wetherby, West Yorkshire LS23 7NB, 7. Feb. 2002, 230 pp.

Stipetić, A., Toš, Z., Janjić, D. (2002): Safety Elements at the Road and Railway Traffic Conflict Points, Promet-Traffic-Traffico, Vol. 14, 2002, No. 4, pp. 209-212.

Vijendran, M., Little, A.D. (2004): Red Light Violations at Level Crossings. Presented at the 8th International Level Crossing Symposium, CD-ROM Proceedings, The University of Sheffield, Sheffield, UK, 14-16 April, 2004, 14 pp.

Chapter 35

Human Factors Issues at Level Crossings: A Reference Tool for Inspectors

Hayley Dixon, Andy Baker and Claire Dickinson

Introduction

Level crossings are locations where the railway is crossed by a road or path at the same level and where vehicles, pedestrians and other users cross the railway. There are approximately 8000 level crossings, of about 12 different types on Britain's railway infrastructure, each with a variety of risk control measures depending upon the characteristics of the crossing users, the train traffic, the road or path and the surrounding environment.

Level crossings currently present the largest risk of a multi-fatality incident on the railway network. HMRI, a division of the HSE, has a role in the approval, inspection and investigation of incidents involving level crossings. To ensure risks are better controlled, HMRI were seeking to develop their understanding of HF issues at level crossings.

Davis Associates was commissioned by HSE to achieve the following objectives:

- Define a list of human factors issues at level crossings
- Capture the findings so as to support the development of a tool for Inspectors
- Develop a tool for use by Inspectors that supports their day-to-day work methods

This chapter describes the methods undertaken for gathering the HF issues and for developing a suitable tool to support Inspectors in identifying HF issues.

Method

Four different sources were used to gather the HF issues at level crossings. A literature review was initially carried out, followed by site visits to level crossings across the UK, visits to two signal boxes and interviews with a range of stakeholders.

Literature review

105 documents were reviewed from rail and road, transport, and to do with user risk perception and behaviour. These documents covered both UK and worldwide level crossing research and incidents, involving both pedestrian users and vehicle drivers.

Level crossing visits

Visits to 45 UK level crossing sites helped build a greater understanding of the issues first-hand. While at each site, observations were made of pedestrians and vehicle drivers using the

crossing. An assessment of HF issues present at each of the crossings was undertaken supported by the information acquired from the literature review. Two examples are shown in Figures 35.1 and 35.2.

Figure 35.1 shows a passenger being dropped-off close to the level crossing. Vehicle drivers behind moved on to the other side of the road and into the path of oncoming cars, causing conflicts of right of way between cars while on the crossing. Observations at other sites also identified similar problems brought about by vehicle drivers dropping passengers off for stations located next to crossings.

Figure 35.1 Drop-off locations

Figure 35.2 shows a line of queuing vehicles over the crossing. A busy right hand turn ahead of the crossing resulted in the traffic queuing across the crossing. The Highway Code requires vehicle drivers to move over the crossing only once the exit is clear. Observations at crossings identified a preference for queuing across the level crossing while a vehicle ahead of them waited to cross the path of oncoming vehicles or held up for other reasons.

Figure 35.2 Queues over the crossing

Signal box visits

To understand the role of the signaller and the sequence for opening and closing crossings, visits to four signal boxes were undertaken. The signallers provided a detailed explanation of the sequence for operating the crossings and an account of the behaviours of users.

Interviews

Interviews with 22 stakeholders with various levels of experience and areas of expertise of crossings and user behaviour further supported the identification of HF issues. Stakeholders, such as rail and road industry experts and those representing the interests of passengers, provided a wide perspective of issues relating to human behaviour.

Correctly defining the boundaries of each interview was key to ensuring that the discussion generated HF issues and the reasons behind user behaviour, and was not focused on 'fault-finding' with individual users, duty-holders or others industries. Each interviewee was provided with a set of guidelines prior to the interview, detailing examples of HF issues. These provided a framework for the discussion and facilitated users to talk around the issues, ensuring maximum output from the interviews.

Database

Commonalities of human factors issues from each of the above stages were drawn together and recorded in a database, created for the purpose of this project. A database was used as the most efficient way of recording the human factors issues because:

- The findings were not necessarily unique to one piece of research or respondent and the database allowed this to be conveyed easily and without duplication;
- The database allowed identification of many variables associated with a single issue, such as the level crossing type or the specific user;
- The database was to be the primary source of information for developing an Inspectors' tool, providing a traceable source of information; and
- The database allows for easy searches and updates to be made, by choosing the required data entry fields (e.g., level crossing type or user details) or using keywords.

Each database entry has a set of data fields (Figure 35.3) that permitted a range of information to be recorded against each HF issue. This included both a title and description of the HF issue, the behaviour of the level crossing user as a result of the issue and the type of user and crossing it was relevant to. Where appropriate, an example of the HF issue was provided with a brief description and a photograph of a level crossing or area relevant to the issue.

To help maintain traceability of information, a set of checkboxes identified the source of the HF issue, such as the employer or organisation of those interviewed, which level crossing visit or document type reviewed.

Level crossings: communication and user behaviour issues

Ref 12 | Creation date 15.06.2004 | Last modified 02.02.2005 | davis

Issue

Regularity of trains

1. Level Crossing type

1.1 Protected	1.2 Unprotected
☒ AOCL ☐ MCBcctv	☒ UWC ☐ Undefined
☒ AOCR ☐ Undefined	☒ UWC/T
☐ ABCL	☒ UWC/MWL
☒ AHB	☒ OC
☒ MG	☒ FP/MWL
☐ MCB	☒ FC

2. User details

2.1 Level crossing user type	2.2 Age	2.3 On own
☒ Pedestrian ☐ Motorcyclist	☐ 0-10 ☐ 60+	☐ Individual
☐ Farmer ☐ Passenger	☐ 11-15 ☒ Undefined	☐ Group
☒ Cyclist ☒ Horserider	☐ 16-20	☒ Undefined
☒ Car driver ☐ Train driver	☐ 20-35	
☐ Van driver ☐ Other	☐ 35-50	
☐ HGV driver ☐ Undefined	☐ 50-60	

Issue at a level crossing

3. Description of issue / design feature

Variations in train schedules, such as engineering works, unexpected delays to train services, and line speed restrictions etc., all contribute to fluctuations in trains passing a point at a supposedly 'known' time.

4. User behaviour

The supposed regularity of trains passing a point at a known time is being used by level crossing users to make judgments of when to cross. Users are assuming trains to pass a crossing at a particular point in time, thus believing the crossing to be safe at certain times. Many users also believe the timetable to be 'set in stone', thus allowing them to make an accurate assessment of when they should and should not cross. However, users fail to consider the variations in train schedules and that many trains, such as freight are not scheduled under passenger timetables.

The research refers to a landowner who retains a train timetable within his tractor cab to allow him to make 'safe' assessments of when he should cross. Other research also identified 4% of users considered a crossing to be 'safe' because they were knowledgeable of the train timetable.

GI/GN7611 indicates that the regularity of trains to be a factor in posing a high risk to users, due to "the rarity of them encountering a train and the reduced vigilance that they may therefore demonstrate in crossing".

5. Error type

☒ Error
☒ Violation

6. Sources of information

6.1 Interview source	6.2 Level crossing visit - areas	6.3 Document source	6.4 Document details
☒ RAIL INDUSTRY ☐ ROAD INDUSTRY	☐ Milton Keynes	☒ RAIL INDUSTRY ☐ GOVERNMENT ☐ TRL (road)	ar/us
☐ Network Rail ☐ AA	☐ Doncaster	☐ HSE/HMRI ☐ DfT (rail) ☒ Independent	GI/GN7611
☐ RSSB ☐ Bus drivers	☐ Hertford	☐ Network Rail ☐ DfT (road) ☐ Non-UK	
☐ HSE/HMRI	☐ Peterborough	☐ RSSB ☐ Other dept.	
☐ RPC ☐ OTHERS	☐ Bedford	☒ Railway GS	
☒ BTP	☐ Redcar	☐ Others ☒ OTHER	
		☐ TRL (rail)	

Figure 35.3 A database entry with all of the data fields

How HMRI address these identified issues, in terms of the appropriate risk control measure, (to ensure the risk is as low as reasonably practicable) is dependent upon knowing the type of failure may be committed by the user. The human failure type was identified for each issue and for the purposes of this work they were simplified into the following categories:

- Violation behaviour: a deliberate deviation from the correct procedure. The user perceives the benefits (i.e., saving time) of undertaking the violation outweighs the risks of committing it (i.e., risk of being hit by a train). This type of behaviour is most appropriately dealt with through emphasising the consequences.
- Error behaviour: an unintentional behaviour that deviates from the correct procedure. The user carries out a task (i.e., crossing the railway) but fails to take account of the correct crossing procedure (e.g., does not comprehend change in audible warning tone). The applications of appropriate measures that design out the error best address this type of behaviour.

Findings

The diversity of human risk perception, the environments around level crossings and the types of crossings themselves, all contribute to many human factors issues. A total of 104 human factors issues were identified. Some of these issues were recognised as being outside the control of HSE and the rail industry in general and came under the responsibility of other organisations, such as the Highway Authorities. However, they were maintained within the database as they continued to impact on the behaviour of users and were considered important subject areas for Inspectors to be aware of and learn about.

A range of recurring HF themes were evident from the findings. These themes briefly describe the reasons for the user behaviour as a result of the HF issues.

Competence

The competence theme relates to the crossing user not being aware of or failing to fully understand the correct rules and procedures for using the level crossing safely. An example of a database issue that demonstrates this competence theme is the awareness and application of the Highway Code. The Highway Code currently contains 275 rules for vehicle drivers. Due to the current method for learning the code and its depth and complexity of legal requirements, not all vehicle drivers are fully aware of the exact procedures for responding correctly to the information at different types of level crossings.

Distraction

Distraction has also been identified as an HF theme throughout the database. This may also result in the user failing to take account of the correct crossing rules and procedures. An example of an identified HF issue is: ‘Noisy surroundings may impair the performance of the users to detect trains at level crossings’. Another distraction issue is related to group behaviour: ‘People in groups may undertake more risky behaviour than when on their own’.

Inadequate design

In many areas of the UK, level crossings have been established for some time. The urban development of areas to include more homes, industry and larger road infrastructures has to be accommodated

through changes to the level of protection at crossings. The inadequacy of level crossing design is a feature of some of the HF issues. For example: 'Housing developments: increases road traffic and level crossing use'.

Individual perceived control

The level of perceived risk can change dependent upon the user's situation. Violations can be explained by an understanding of how people assess the perceived risks. For example, one of the HF issues relates to the time of day: 'Risk-taking behaviour at level crossings increases during rush-hours, at midday and at the beginning and end of the school day'. A vehicle driver calculates the perceived risk of crossing illegally instead of waiting at the crossing, compared with the risks of other things, such as being late for an appointment. If this risk, such as being late, outweighs their assessment of the potential risks associated with being hit by train, then the user is likely to cross.

Risk compensation

An individual's propensity to take risks is influenced by their own experience and that of others. A risk compensation model assumes that the degree to which we take risks varies from one individual to another. An individual's target risk level changes dependent upon the positive (i.e. saving in time) or negative (i.e. injury) impact which risky behaviour achieves. Therefore, if a person's target risk level is low and their perception of risk is high, then that person will behave in a cautious manner. However, if the target level remains the same, but the perception of risk is also low, then it is suggested that the person will behave in a more risky manner.

Recent RSSB unprotected crossing research discussed the issue of risk compensation. Crossing users aware of poor sight times at some crossings, perceive the crossing as dangerous and therefore compensate by being extra vigilant. However, where sight times are good, user risk perception is low and they may adapt their behaviour, acting less cautiously. The research does not suggest that poor sighting times should be permitted but: 'Good sighting distance should indicate the level crossing as high risk'.

Familiarity

A person's familiarity with a task can also influence their behaviour. Habits form over time to help people cope with regular situations and environments, through applying behaviours that require minimal thinking. This enables people to undertake particular tasks (and many tasks at once) without having to concern themselves with the finer details of how that task is actually formed, thus allowing them to concentrate less.

These finer details that people gradually fail to take into account when undertaking a regular task include all the information presented to us from the environment, which we process to determine the most suitable behavioural response. However, by not thinking about these finer details of a task, users may miss the external cues from the environment that would inform them whether the behaviour they are undertaking is appropriate (Burgess, 2002).

Familiarity presents other problems. Continued implementation of a task which does not present dangers on a regular basis can lead users to have a lower perception of risk, therefore behaving less cautiously under these circumstances. An example HF issue is: 'Crossings with low frequency of trains are likely to increase the risk-taking behaviour of regular users'. A vehicle driver using a crossing that has only a few trains passing each day may reduce their level of caution. The combination of continued use of the crossing and only ever seeing a few or possibly no trains

during this time, removes the individual's "ability to think logically and rationally about their behaviour" (Burgess, 2002) when using the crossing.

Complacency

Known as the "influence of attributional biases on people's comparative risk evaluation" (Weyman & Kelly, 1999), people will perceive themselves to be at a lesser risk than others. This is often related to judging themselves as being more skilled, and therefore leading to reduced levels of caution. Another way of describing it is to suggest that users take a "it won't happen to me" approach while crossing.

Mental models

People form mental models of situations to help them make sense of and put structure into the world around them. However, if the model does not contain all the correct information or they have misrepresentations of the environment, the user may perceive themselves as safe when undertaking tasks, when in fact they are exposing themselves to danger (Weyman and Kelly, 1999).

In the context of level crossings, users can form an incorrect mental model of train speed and distance. An example issue from the database is 'train speed and distance is underestimated by users, which may result in increased decision making errors by level crossing users'. The model formed by level crossing users may be based upon a more common and easily understood model, such as road vehicle speed, which exposes them to an increased risk when judging train movement.

Identifying HF issues at level crossings

To assess the feasibility of identifying potential HF issues at level crossings, a trial exercise was carried out. Visits were made to six level crossings located in close proximity. HF issues relevant to the crossing type were assigned to each crossing using the following three methods.

Physical presence

This involved verifying the presence of HF issues based on any physical evidence at the crossing. For example, a nearby 'road junction', 'type of trains' using the crossing, 'farm land' in the vicinity of the crossing or a 'traffic calming system' on the approach are all issues that may influence the behaviour of people when using a crossing. An assessment of the crossing from a user's perspective, both on foot and from a vehicle was also made to ensure all physical evidence was confirmed.

Observations

This involved spending time at all of the crossings and observing the actual behaviour of users, checking this against the issues relevant to the type of crossing.

Interviews

Interviews were carried out with the local British Transport Police, a local elected politician, school head teacher, signallers and staff from a local bus company. The use of local knowledge was a valuable method for gaining insights into the behaviour of level crossing users. As in the approach taken with stakeholder interviews, each interviewee was provided with a set of guidelines

prior to the interview, which detailed some examples of HF issues. This defined the boundaries of the interview, ensuring the discussion identified HF issues at the particular crossing and the reasons behind user behaviour, and did not focus on 'fault-finding'. The interviews were central to validating many of the issues at each crossing.

Developing an Inspectors' tool

The next stage of the project set about developing the database of HF issues into an appropriate HF reference tool for use by HMRI Inspectors. Interviews held with Inspectors aimed to understand their job role and methods of work for carrying out their job. Inspectors with a range of different experience and varying roles within HMRI were included to ensure a broad range of methods were captured and considered in the development of the HF tool. In addition to the interviews, a review of internal HSE documents and the HSE website with references to work methods, procedures and guidance was also undertaken.

Main findings from interviews

The Inspectors' jobs covered a wide range of methods, from their day-to-day tasks to using a range of reference material to support their role. A set of key findings was evident from all of the interviews.

- The main objective of an Inspector's work is to assess level crossings against current standards and legislation.
- While in-field, Inspectors have limited carrying capacity, which dictates they carry only the most relevant legislative documentation.
- Inspectors prefer to use self-documented, high-level aides-memoire to support them in exploring all issues at level crossings.
- The integrity of an Inspector's position is of utmost importance, therefore reference to lengthy checklists or in-depth notes is inappropriate.
- In preparation before and after site visits, Inspectors refer to various sources of information, in particular the knowledge of other Inspectors, but also in-house reports, databases and the Internet.
- Inspectors also have individual preferences for working methods.

The Inspectors stressed the importance of developing a reference tool that was easy to use and that added value to their work to avoid it being an unused resource.

In-house database

After assessing the working requirements of Inspectors, it was apparent that the most appropriate format for providing them with the HF issues at level crossings was through an in-house database system.

A set of basic requirements were defined prior to developing the HF tool. These covered the following areas:

- End-users – the main users were operational inspectors, Rail Inspectorate contact officers and HMRI human factors personnel. However, it will be designed, in terms of terminology and dialogue, so as not to exclude other potential end-users.

- Information content – to match that available within the HF database. However the format will also allow for ease of updating with new HF issues when required.
- Format for delivery – the format will be focused on delivering an in-house tool, however it will also be delivered with the flexibility to provide summaries of HF issues for use in-field. This takes account of the individual working methods of Inspectors and the potential for changes to how work may be undertaken in the future.

A review of best practice guidelines for presenting information on screen also supported the development of a HF tool that met with both the job requirements of Inspectors and was easy to use.

HF reference tool

A prototype HF tool consisting of an introduction, help and search page was developed using a database software program. The tool met the requirements defined above, by providing three viewing formats, each with a different quantity of information, thus supporting the different individual working methods of Inspectors.

A presentation of a prototype reference tool to two groups of HSE/HMRI Inspectors and staff provided an opportunity for gathering feedback about the overall look and feel of the tool, use of language and the methods of searching for HF issues. Detailed usability testing of the tool with Inspectors on a one-to-one basis was not possible due to restrictions on their availability. However, in-house tests with users that had not been involved with the development of the tool helped to further critique the tool's interface.

A final version of the HF tool was provided to HSE/HMRI. This has subsequently been made available via the HSE intranet and is being used by Inspectors. The tool also permits the administrator of the system to make updates to the content for including new HF issues when required.

Acknowledgements

The work was funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.

The HF issues are available from the HSE website at the following address: www.hse.gov.uk/research/rrpdf/rr359.pdf.

References

Burgess, C. (2002), *Risk behaviour* (www.ex.ac.uk).

Health and Safety Executive (2005), *Level crossings: Summary of findings and key human factors issues*, Research report 359, HSE Books, ISBN 0-7176-6132-6.

Weyman, A.K. & Kelly, C.J. (1999), *Risk perception and risk communication: a review of literature*, Health and Safety Laboratory.

Chapter 36

Human Factors in the Upgrading of Railway Control Equipment

John Wood, Mark Brunt, Claire Fix, Andy Harding and Theresa Clarke

Introduction and background

Selected level crossings are monitored remotely by signallers or other staff with CCTV cameras mounted adjacent to the track. This allows a number of sites to be controlled by a single operator at a dedicated workstation. The original specification for these CCTV systems was based around black and white images. The pictures sent to the level crossing operator are used to check whether the crossing is 'clear' before allowing the train to continue its approach. The existing displays used in signal boxes are black and white cathode ray tubes (CRT) which are getting increasingly less reliable and costly to maintain.

Network Rail has decided to replace trackside black and white cameras with colour equivalents. The unit costs of 'flat screens' has been steadily decreasing whilst the quality of images has been improving. The CCD human factors team were asked to determine whether there would be any significant degradation in the images presented to the operator if the CRT displays currently used in signal boxes were replaced with their flat-screen equivalents.

Preparatory work and experimental design

It was decided early on that actual level crossing pictures would be used – preferably taken at a number of different types of crossing – and that judgements would be made by experienced level crossing operators. It was also agreed that the object of the experiment was to compare the performance of the existing CRT displays with flat screens – it was not to establish what the smallest target might be that could be identified through the CCTV system. Effectively it was a comparison between current and future display technologies.

As a first step the ergonomics team undertook a pilot study. This involved getting some sample video material from level crossings and playing it back through the alternative displays. Getting hold of sample video from existing cameras proved to be more difficult than expected. The original concept was to use the spare camera at each of the selected sites and replace it with a colour camera. The advantage of this approach was that both the existing camera (black and white) and the temporary replacement (colour) would be mounted on existing infrastructure and would be focussed on the same scene under exactly the same weather and lighting conditions. Despite several attempts this failed, as did tapping usable records off the existing terminals. The logistics involved in replacing equipment safely – on a live line – was also complex and time consuming. This was solved using a special purpose van, see below.

Pilot trials were conducted at CCD offices using the edited video material. These trials allowed experimental protocols to be tested and refined. A progress meeting was held on completion of the pilot studies and the conclusions reviewed. It was agreed that the following variables would be examined – with appropriate recordings being made at selected crossings:

Factor	Variables to be examined
Crossing type	Single v multiple track
Lighting conditions	Daylight v dusk v nightime
'Target'	Wheelchair v child v sheep
Image type	B&W v colour presentation

Both Network Rail and the ergonomics team reconfirmed that the primary objective was to identify whether there would be any significant degradation in operator performance when a black and white CRT was replaced by a flat-screen.

The existing requirement to 'check that the crossing is clear' sets neither an upper or lower limit on the nature of the object to be seen. In deciding targets to be used the ergonomics team selected ones which were at the smaller end of the spectrum of possible targets and also of such a size that they *would* induce errors – thus allowing for comparisons between displays to be made. Two targets were finally selected a) an empty wheelchair and b) a full-scale paper-mache model of a 3 year old child (Figure 36.1).

As previously mentioned the proposal to use the existing CCTV infrastructure for the recordings was abandoned in favour of hiring a specially adapted vehicle with a retractable mast (Figure 36.2).

The van included video recording equipment which allowed for records to be made independently from the NR CCTV infrastructure. Although using the van did not always allow for the mast to be positioned in an identical location to the Network Rail existing mounting pole this was not

Figure 36.1 Targets used for trials

considered to be a significant drawback as far as the study was concerned.

At each site a series of video records was taken with each type of target and under different lighting conditions. Figure 36.3 presents the locations used for target location.

The recording of the raw video clips took was spread over 3 days including one overnight session. The video recordings were edited into fifty pairs of 10 second clips in both monochrome and colour. The fifty clips included a mix of all targets (including some without a target present), every level crossing type, all lighting conditions and all test locations on the crossing, Figure 36.3. In selecting the kit to be used to run the trials portability was a significant consideration. The decision to use a large number of experienced signallers for the trials meant that the experiment would need to be set up in a signalling centre – with all the constraints that this imposed.

Figure 36.2 Mobile camera van

Experimental set-up and conduct of trials

The experiments were carried out in the Peterborough signal box which maximised the number of available, experienced level crossing operators with CCTV experience. These trials were run over 2 shifts involving a total of 21 signalling staff.

Following a standardised briefing session subjects were given some sample tests for familiarization. Subjects were presented with 2 screens – a black and white CRT monitor and a flat-screen which could be run with either colour or black and white images – and a small keypad for recording their decisions.

Once having familiarized themselves with the experimental procedure subjects were asked to look at a sequence of images and determine whether they were clear or not of targets. Where a target was noted subjects had to indicate what the target type was on a colour-coded keypad. Where no button was depressed, and thus no target judged to be present, the video sequence automatically moved onto the next scene after a set time. The experimental protocol followed the procedure in Figure 36.4. Errors and response times were automatically recorded.

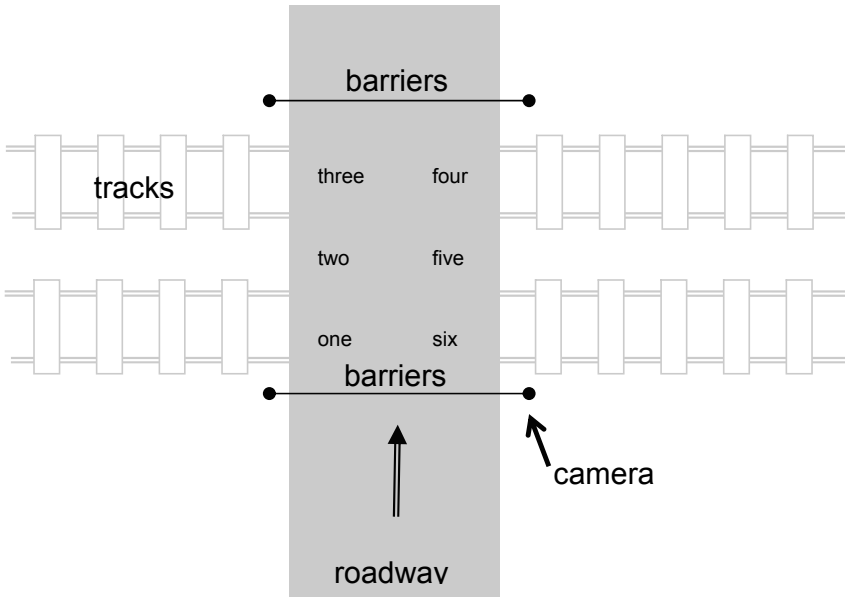


Figure 36.3 Locations used on crossings for test targets

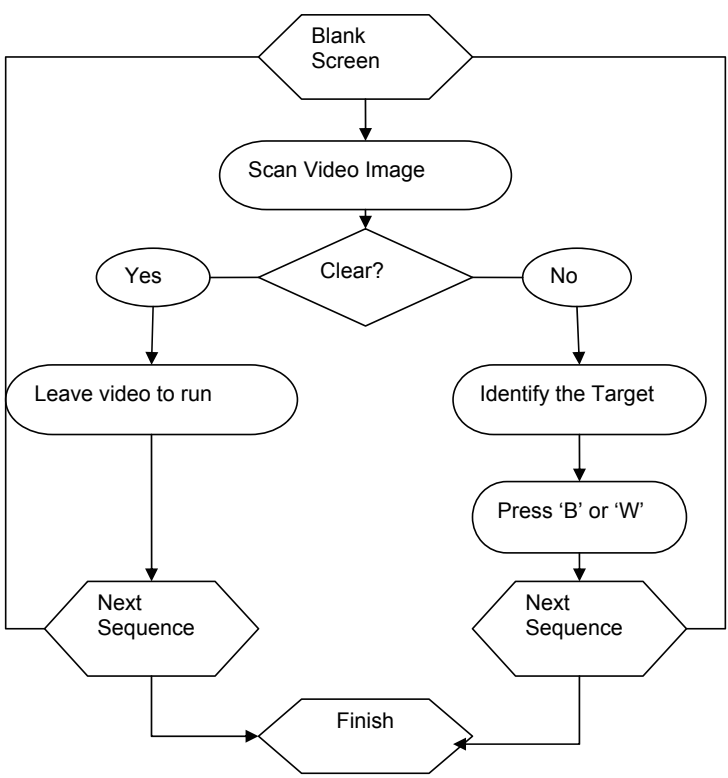


Figure 36.4 Experimental protocol used during trials

Subjects continued until they had completed the entire sequence of 50 tests following which 2 more repeats were undertaken so that presentations on CRT (black and white), flat-screen (black and white) and flat-screen (colour) had been completed. At the end of the trials they were asked for their subjective views about the alternative screen technologies and a more general open-ended question about any other comments.

Results

The data collected were subjected to a battery of statistical tests including repeated-measures ANOVA, Friedman's test, Bonferroni Post-hoc test, Wilcoxon Signed Ranks test and 'T' test. The main conclusion was that there was no significant difference between the different displays and thus the replacement of the CRT with flat-screens would not result in any significant degradation in signaller performance. The key results are presented in Figures 36.5 and 36.6.

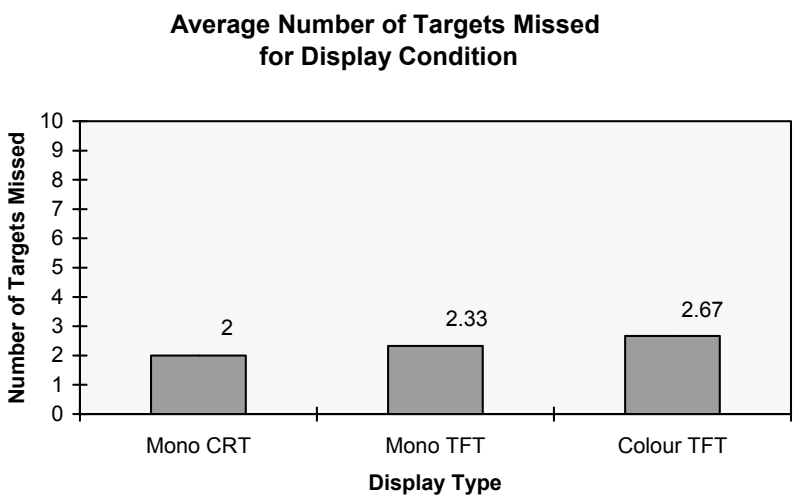


Figure 36.5 Targets missed v display type and image

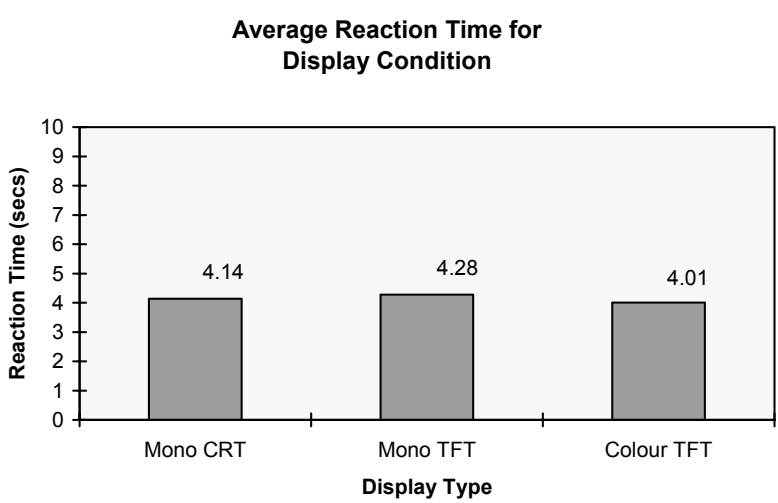


Figure 36.6 Reaction time v display type and image

The targets missed results were tested statistically. Although the graph shows a difference between the error scores, this was not statistically significant. The number of targets missed, in general, was not affected by display type.

The reaction time performance with the mono TFT is marginally worse than the existing condition, mono CRT, and the full colour condition is marginally better. Statistical tests showed the differences to be significant. However, in real world terms a reaction time difference of a few tens of milliseconds would not be an important differentiator since the operator is not under time pressure.

In general, participants felt that the colour clips presented on the TFT monitor added better definition to the images.

Conclusions and discussion of results

The overall conclusion was that there would be no sacrifice in safety by replacing the existing CRT monitors with flat-screens showing either black and white images or colour ones.

When performance against type of targets were assessed it was noted that there was a difference between the 'boy' and the 'wheelchair'. No boy targets were missed. The wheelchair was difficult to see in some positions where its visibility was marginal. Using targets on the threshold of visibility raises wider issues about the minimum size of target that a signaller might be expected to see.

Examining the numbers of errors against the crossing type a startling pattern emerged – most of the errors could be associated with one crossing. Significantly more targets were missed at Maxey level crossing than the other level crossings, Figure 36.7. Also, reaction time for detecting targets at Maxey was significantly longer than reaction time for the other level crossings under all three display conditions. Nearly all participants commented that they found the far corners of Maxey level crossing the most difficult to see. It was also noted that participants tended to move closer to the display monitors to look at the Maxey level crossing.

The evidence suggests that creating a satisfactory image on CCTV for a large crossing might not be that easily achievable. It may be useful to examine further whether there is a limit on the maximum size of level crossing that can be examined using current CCTV systems.

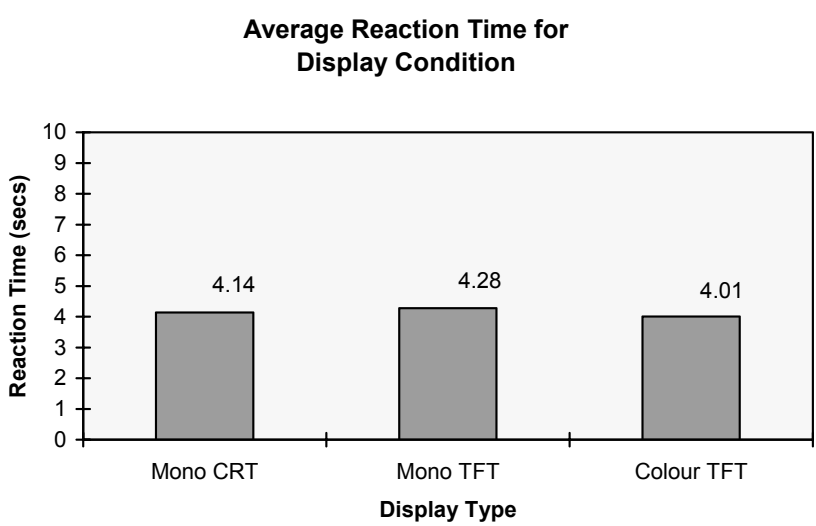


Figure 36.7 View of 'Maxey' level crossing

Acknowledgement

CCD and NR would like to thank David Jones, and the staff at Peterborough signal box, for their assistance in recording video clips and running the trials.

This page intentionally left blank

Chapter 37

The Effect of a Level Crossing Upgrade on Signaller Workload and Staffing Requirements: A Case Study

Amanda J. Widdowson

Introduction

Background to case study

This chapter describes work conducted in response to a recommendation for a detailed workload assessment of two signalling panels (numbers 4 and 5) at Motherwell Signalling Centre (SC) during the night shift, 10th-11th May 2005. The recommendation, from a previous report (Stapley, 2002), was prompted by the proposed modification of the Cleghorn Automatic Half Barrier (AHB) Level Crossing (LX) on Panel 5, to a Manually Controlled Barrier (MCB) CCTV level crossing. The upgrade was necessitated by the high level of risk associated with the level crossing.

Currently, one operator is responsible for Panels 4 and 5 during the eight-hour night shift; two signallers operate these panels during the early and back shifts. The introduction of the new CCTV LX is likely to increase the signaller's operational demand. It was therefore necessary to determine the extent to which this increase is likely to have a negative impact on safety and efficiency, and the resulting impact on staffing levels; in other words, whether one or two signallers will be required to operate Panels 4 and 5 during the night shift. The case study is described in full in the assessment report (Widdowson, 2005).

Motherwell Signalling Centre – an overview

Motherwell SC contains seven N-X (Entry-Exit) panels in total, covering the permanent way from the Rotherglen Exclusive to Cove on the West Coast Mainline (WCML). Panels 4 and 5 are standing workstations. Panel 4 contains an existing MCB CCTV LX (Logan's Road). The two panels that were the focus of the assessment are illustrated in Figure 37.1.

Method

Network Rail's suite of assessment tools were applied to Panels 4 and 5 at Motherwell SC by an ergonomist, (the author), and two Subject Matter Experts (SME) from Network Rail, during the eight-hour night shift 10th-11th May 2005. This case study will focus only on the workload and operational demand tools applied, namely:

- SME Commentary;
- Integrated Workload Scale (IWS);
- Operational Demand Evaluation Checklist (ODEC);
- Baseline Analysis Model.

Figure 37.1 Panel 4 (Left) and 5 (Right)

These tools are described in the following sections.

SME commentary

This refers to a technique whereby a SME records the activities being conducted and the number of trains on the panel, at five-minute intervals.

Integrated Workload Scale (IWS)

The IWS is a simple 9-point scale produced by University of Nottingham on behalf of Network Rail for estimating perceived effort and demand (Pickup et al, 2005). 'Work too demanding' is at one end of the scale, and 'Not demanding', at the other (each number on the scale is defined and presented to signallers). Signallers are prompted to provide a rating every five minutes. The IWS and SME Commentary were used to produce a timeline analysis. The results are presented in the 'Results' section of this chapter.

Operational Demand Evaluation Checklist (ODEC)

ODEC is a Microsoft Excel-based software tool that was also developed by University of Nottingham on behalf of Network Rail. It was specially designed to generate an operational demand score for the operation of signalling panels. Specific data collected includes the number of trains, signals (automated and controlled), station platforms, junctions, depots/sidings, level-crossings, ground frames, Hot Axle Box Detectors (HABD), Patrolman's Lock Out Devices (PLOD), special box instructions, points traversed, and incidents on the panel. Definitions of each of the data items are incorporated. The tool automatically calculates whether the number the user inputs to the checklist is 'high', 'medium' or 'low', (or not applicable), according to pre-defined judgements. The number in each of the three categories is weighted and totalled, to give a final ODEC percentage score.

The ODEC tool was applied to each panel (4 and 5) individually, and then the results combined in a third ODEC checklist combining the results of the individual panels.

Baseline Analysis Model

The Baseline Analysis Model, produced by CCD on behalf of Network Rail, is another Microsoft Excel-based software tool. It addresses 'ergonomics effectiveness', as well as operational demand. This chapter is only concerned with the operational demand section. Essentially the model takes the

form of a checklist of questions that were asked of the Shift Supervisor. The questions addressed the number of controlled signals, sidings, trains, stations and level crossings on the panels; the frequency of use and type of level crossing; the track complexity (calculated via the number of point ends and track miles on the panel); and how many other tasks, (such as communications and paperwork), the signaller is required to perform at the same time as the signalling activities.

The baseline checklist questions were answered as they pertained to the night shift (22.00-06.00, Monday to Friday) and 14.00 Saturday until 07.00 Monday. The results for this, and all the other tools applied, are presented in the next section of this chapter.

Results

Timeline and SME commentary results

The activities recorded by the SME were categorised and the results, presented in a histogram (Figure 37.2). In the figure, 'T1' represents the first time interval at 23.01 where activities and traffic numbers were recorded and the IWS rating, provided. The last time interval shown is 'T25', at 00.59.

Analysis of Figure 37.2 reveals that most of the signaller's time was spent viewing or scanning the panels, as one might expect. A significant amount of time was also spent setting routes. The peak reported workload during the timeline analysis is represented in the Figure 37.2 as T14-T20. Isolation activities defined as reference to isolation diagrams and lists, placement of reminders, but not paperwork such as form filling, or telephone communications pertaining to planned isolations were taking place during this time. A reasonable amount of time was spent using the CCTV (existing and simulated[1]), including during isolation activities. Interestingly, a lot of paperwork was performed during this period of peak reported workload. This includes activities such as viewing simplifiers, and form filling.

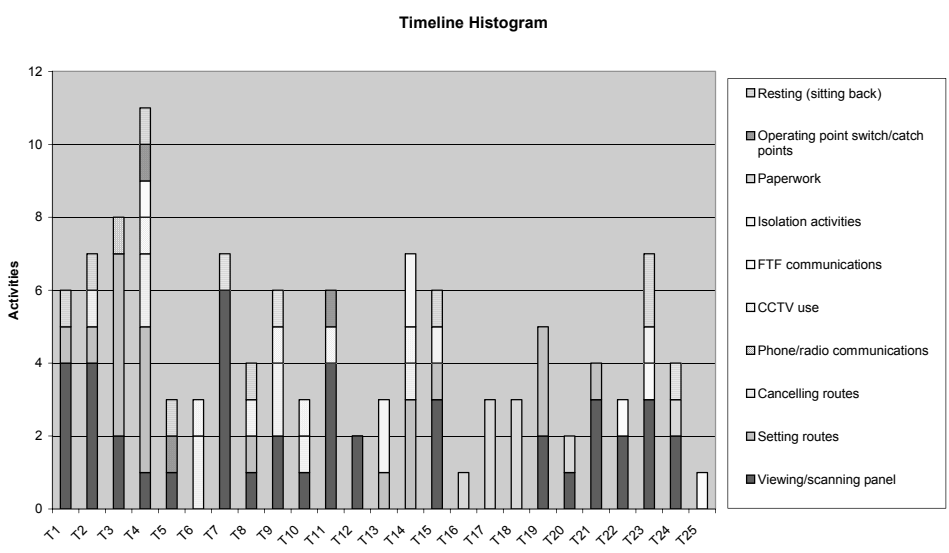


Figure 37.2 Timeline histogram (23.01-00.59)

1 During the SME commentary, the operator was asked to simulate operation of the future Cleghorn CCTV LX when trains passed through the relevant section of Panel 5.

IWS Results

The majority of the Integrated Workload Scale (IWS) ratings were '*moderate effort'*, (scores of 5 or 6). The IWS analysis therefore demonstrated that the operator only experienced moderate workload during their busiest time period. In fact, there were no ratings higher than '*moderate pressure*'.

The highest reported workload occurred between 00.05 and 00.35 (illustrated in Figure 2 as T14-T20). There were two to three trains on the panel at this time; a relatively low number (the highest was seven, at 23.05), so it does not appear that high traffic levels were the cause of the high reported workload. Therefore, it is proposed that the number and/or type of activities the operator was performing was the cause of the higher workload. According to the SME commentary, the operator was performing a higher *number* of activities at this time, but more importantly, he was dealing with planned isolations (defined in the previous section) for engineering work at this time. It seems very likely that the tasks associated with isolation activities and track protection were the cause of the signaller's reported higher workload.

During the SME commentary, the operator was asked to simulate operation of the future Cleghorn CCTV LX when trains passed through the relevant section of Panel 5. There were three simulated uses of the Cleghorn LX during the forty-minute peak of reported workload, and one telephone call to the former barrier attendant at the existing Logans Road LX. However since the workload, even at this peak time, was only reported as "moderate pressure" (IWS score 5), this did not appear to adversely affect the operator's workload and therefore efficiency. However, when the new Cleghorn MCB CCTV is incorporated in Panel 5, there may be occasions where the signaller is required to operate one of the LX at same time as performing another activity such as talking on the phone, especially during isolations. This may result in delays to the services and reduced time spent checking the CCTV monitors, causing a safety risk. However, in conclusion, the reported workload was found to be low enough even at peak time, such that the likely increase in operational demand caused by the introduction of the new MCB LX is not anticipated to have a significant detrimental effect on performance efficiency or safety. In fact, the introduction of an additional signaller to Panels 4 and 5 during the night shift may reduce the workload to a level that is too low to sustain satisfactory vigilance levels.

The lowest IWS ratings ('minimal effort') were received when there were only one or two trains on the panel and the operator was scanning the panels, checking paperwork, and talking to the shift manager regarding isolation and block information. It is assumed that the operator was performing the latter two activities because he was not busy. These scores occurred towards the end of the assessment, between 00.40 and 01.00.

It should be noted that only one operator, on one night, was tested. More valid results may be generated by testing several operators during the same time period on different nights. However, the results in this analysis would seem to suggest that the Panel 4/5 signaller's operational demands are low enough to allow them to take on the additional tasks necessitated by modification of the Cleghorn level crossing.

ODEC results

The ODEC scores for each panel are listed in Figure 37.3.

It is likely that the proposed introduction of the MCB LX monitor to Panel 5 is likely to increase the operational demand associated with that panel. The current ODEC score for both the panels operated simultaneously by a single operator, as is the case during the Night shift, is 60%. This is already slightly high, so the introduction of the additional tasks associated with the LX may be detrimental to efficiency and performance The ODEC results therefore suggest that two signallers may be required to operate Panels 4 and 5 during the night shift.

Panel Number	ODEC Score (%)
4	51
5	47
4 and 5 combined	60

Figure 37.3 ODEC scores

However, some of the data collected by the ODEC tool is based on number of times used per week (eg. ground frames), and averages over the last week (eg. PLOD activations and HABDs), rather than the number per night shift. The ODEC results may therefore be slightly inaccurate in this respect. Furthermore, the current staffing requirement for Panels 4 and 5 during the night shift is only one operator, because the workload is perceived to be sufficiently low. The ODEC scores should not be considered in isolation, but rather in conjunction with the findings from the SME commentary, IWS and baseline operational demand checklist.

Baseline operational demand results

The Baseline Analysis Model contains measures of both ergonomics effectiveness and operational demand, as discussed previously. The operational demand score was 2.40, where a score of 1 refers to low demand and 3, to high demand. The baseline analysis therefore found the operational demand of the Panel 4 and 5 signaller's task to be high. This score is based on readings taken from the panel such as the number of point ends, track miles, controlled signals, trains, level crossings and stations controlled, as well as the frequency that communications, alarm processing and paperwork activities are performed.

Traffic regulation (ie. number of trains) was found to be high (receiving a score of 3) on the two panels during the night shift. Communications also accounted for a large proportion of the operational demand.

Conclusions and recommendations

The conclusions generated by the assessment are summarised in this section.

a) The IWS analysis demonstrated that the operator experienced a moderate workload during their busiest reported time period (23.00-01.00).
b) The timeline analysis illustrated that most of the signaller's time was spent viewing or scanning the panels. A significant amount of time was also spent setting routes.
c) The timeline analysis found that the reported workload was low enough even at peak time, such that the likely increase in operational demand caused by the introduction of the new Cleghorn MCB CCTV LX is not anticipated to have a significant detrimental effect on performance efficiency or safety. In fact, the introduction of an additional signaller to Panels 4 and 5 during the night shift may reduce the workload to a level that is too low to sustain satisfactory vigilance levels.
d) It should be noted that only one operator, on one night, was observed. More valid results may be generated by testing several operators during the same time period on different nights. However, the results in this timeline analysis would seem to suggest that the Panel 4/5 signaller's operational demands are low enough to allow them to take on the additional

tasks necessitated by modification of the Cleghorn level crossing.

e) The ODEC results found slightly high operational demand (60%) for current night shift operation of Panels 4 and 5, by a single signaller.
f) The baseline analysis model demonstrated high operational demand on both panels (2.40). (Although practice has shown that this level of workload can be satisfactorily managed by one operator).
g) In summary, the signaller's moderate reported workload (as measured by the IWS in the SME commentary during the busiest period), and the reasonable ODEC score, contradict the high baseline operational demand results. This may be because the individual operator tested was experienced enough to cope with the high operational demand. This individual may also possess good coping strategies for stress. The baseline analysis model 'People and Organisation' section showed that only two of the forty signallers at Motherwell had less than one year's previous signalling experience. This would seem to support the conclusion that the contradiction between the operational demand and workload measures was caused by experience.
h) This case study illustrates the use of Network Rail's suite of workload and operational demand assessment tools, and demonstrates how Network Rail assesses staffing requirements.

Recommendations

The following recommendations were generated by the case study:

1. The assessment showed that the night shift operation of Panels 4 and 5 can currently be performed by one *experienced* signaller, despite the high operational demand associated with these panels. The further increase in operational demand and workload likely to be incurred by the future addition of the Cleghorn MCB CCTV LX to Panel 5 is expected to be low enough for one *experienced* signaller to cope with. The recommended staffing requirement for the Motherwell Panel 4 and 5 Night-shift operator, bearing in mind the future additional MCB, is therefore one *experienced* signaller.
2. More precise results may be obtained by further workload assessments of different Panel 4 and 5 signallers, taken during the same time period (23.00-01.00), over different nights. However, the results generated by this assessment should provide sufficient indication of workload and operational demand to facilitate the staffing recommendation.

References

CCD on behalf of Network Rail, '*Baseline ergonomics assessment programme: User guide, A guide to the Network Rail method for ergonomics assessment of signal boxes*', Ref: NR/Tser089/User Guidev1.

Pickup, L., Wilson, J.R., Norris, B.J., Mitchell, L. & Morrisroe, G. (2005). The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload. Applied Ergonomics, 36, 6, 681-693.

Stapley, N, Human Engineering Limited, for Railtrack, 5th July 2002, '*Ergonomics Assessment of the Motherwell Signal Centre*', Ref: HEL/RT/02739/TC1, Issue 02.

Widdowson, A, Vectra Group Limited, for Network Rail, 16th June 2005, '*Motherwell Signalling Centre Workload Assessment (Night Shift)*' Ref: WCE-NR-HF-REP-016, W054-242-HF-REP-005000, Issue A02.

PART XI
Accidents and Safety

This page intentionally left blank

Chapter 38

Maximising Information: An Effective Interview Technique for Investigators in the Rail Industry

Emma Lowe and Claire Turner

Introduction

Information provided by the driver often provides a critical component of the SPAD investigation process. Thus, the ability of driver managers, who usually undertake the initial interview with the driver, to obtain accurate and detailed information is of great importance. However, this is one area where little attention has been paid in terms of the process or indeed the skills and knowledge of those undertaking the interview. Following the Ladbroke Grove accident in 1999, the rail industry, and in particular Network Rail, was criticised over the quality of SPAD investigations, especially with regard to identifying any contributory human factors. As a result of the work undertaken to improve consideration of human factors in investigations, it became apparent that the issue was not always that investigators failed to understand what the human factors issues might have been, but that complete and accurate information was difficult to obtain, or was not forthcoming. This chapter describes the work undertaken to identify good post-incident interviewing practice that could be adopted by the rail industry.

Current investigation practice

Network Rail is responsible for ensuring that all category 'A'[1] SPADs that occur on its infrastructure are investigated thoroughly. Following a SPAD event the driver will talk to a signaller and together they will complete the Signal Passed at Danger Form RT3189. Details of the incident will then be communicated to the Route Operations Control who will contact the driver's company and initiate evidence collection, such as testing of the signalling equipment and arranging for the data from the on-train monitoring equipment to be downloaded. The driver's manager or someone from the train operating company's on-call team will interview the driver within 2 hours of the SPAD incident. This first interview is primarily about assessing fitness to continue work, if necessary, and explaining the investigation process that will occur. A second interview will take anywhere between 24 hours and 3 days after the incident to obtain information from the driver about what happened. As the driver is often the only party involved in a SPAD, their testimony about what

1 Signals Passed at Danger is the term used to describe an incident when any part of a train has passed a stop signal at danger without authority. SPADs are categorised and a category A SPAD is when the train has passed the stop signal when the stop aspect was displayed correctly and there was sufficient time for the train to be stopped safely at the signal.

happened is critical and Network Rail accident investigators are very dependent on the evidence obtained in an interview between the driver manager and the driver.

However, both the investigators and the driver managers (from here on in known as the interviewers) reported specific difficulties with obtaining information from drivers. Firstly, drivers can be defensive and uncooperative. Some are suspicious about the interviewer's true intentions, or may be concerned about disciplinary action. Secondly, quite often the driver cannot recall seeing the signal and/or the aspect of the signal that he/she passed at danger. This is not unreasonable as the driving task is so practiced and one at which drivers are so skilled that it is largely automatic. Interviewers have reported that it is often unclear whether the driver cannot remember specific details or if they have the information and do not want to reveal it.

Interviewers also found that drivers had a tendency for using the latest buzzwords to apportion blame or responsibility for their SPAD incident. For example, they might claim they were "suffering from fatigue" because they know fatigue had become a key management concern. Furthermore, it was felt that drivers could be influenced by colleagues who may encourage them to think of alternative reasons for the SPAD, or suggest what to say, so distorting their true account of what happened.

In addition to the problems reported by interviewers, research into the area of witness testimony indicates that typically witness reports are incomplete, partially constructed and unreliable. The investigation process could therefore benefit from a more structured approach to the collection of information from the interview.

Researching good practice

Evidence of the sort of difficulties faced by interviewers, as described above, was identified through two key activities:

- A review of transcripts from recorded interviews and/or the notes managers provided in support of their interview with the driver, and
- A workshop with driver managers who described what their current investigation process was, what the frustrations were, and what works well currently.

Having gained a good understanding of the investigation process and the areas for improvement, a literature review was undertaken which focused on 2 main areas:

- Whether good practice could be obtained from other sectors such as the police force where investigation is an integral and important activity.
- Whether the literature on false testimony would reveal useful information about appropriate interviewing techniques that could apply to the railway.

The outputs of the literature review were:

a) A recommendation that interviewers needed to understand more about memory and some of its limitations.
b) A specific technique called the Enhanced Cognitive Interview (ECI) that could be applied in the rail context to improve the quality, quantity and accuracy of the information obtained in an interview.

Why should interviewers need to know about memory?

Interviewers need to obtain information from those involved in an incident and are therefore primarily interested in memories about what has happened. Knowing how the memory works, how fragile it is and how they as interviewers can influence what interviewees tell them can impact upon the effectiveness of the investigation process.

Psychologists generally refer to there being 3 main processes underlying memory:

- *Encoding* concerns the process through which information is entered into memory.
- *Storing* is the process by which encoded information is then transformed into a representation to be retained.
- *Retrieval* refers to the processes by which we attempt to retrieve information from memory.

Memory can fail at any one or more of these stages.

Witness testimonies are often found to be flawed and inconsistent. Sometimes, accurate accounts of recalled information are deliberately masked for personal protective reasons, but even when people are keen to provide as much accurate information as possible, it is still rare to obtain fully accurate and consistent accounts of events because of natural limitations of, and influences on, people's methods of event observation, capture, storage and recall.

A specific example of a factor that can affect the *encoding* of a memory is attention. Our cognitive capacity is limited to dealing with approximately seven bits, or chunks, of information at a time so we must divide our attention and select areas to be taken note of. What we select will be influenced by many factors such as knowledge, expectancy, past experience, training, motivations and values, for example. This explains why you can have 5 people involved in the same event but get 5 somewhat different accounts of what happened (Fruzzetti et al. 1992). In order to encode information it must be attended to. So, the unavailability of information is not always due to retrieval or storage failure. Some of the information will simply not have been attended to. As a consequence, no interview technique, however sophisticated, will be able to retrieve all information about an event as it is impossible to recover information which was never stored in memory in the first place.

There are other factors that can affect our *retrieval* of information about an event too. Perhaps most important is the fact that memory for a complex event is largely constructive. What is stored in memory is not ready to be played back like a video-recording. Instead, it is thought that we store information in 'snapshots' – in various parts of our brains – not always under the most objectively logical filing system. So when it comes to retrieving information, we do not always know where to look, when we get there the information is often only partial representation of what went on, and we make sense of incomplete information by calling on our experiences and expectations of what we think ought to have happened in a given context or event.

The Enhanced Cognitive Interview

The ECI technique is based on an understanding of how humans encode, store and retrieve memories. The aim of the ECI technique is to seek out all stored information, thereby retrieving the maximum information possible. This interaction is displayed in Figure 38.1, one oval representing the stored information and the other the interview strategy. The shaded area in which they successfully overlap represents retrieved information. The purpose of the ECI is to increase this overlap, maximising retrieval.

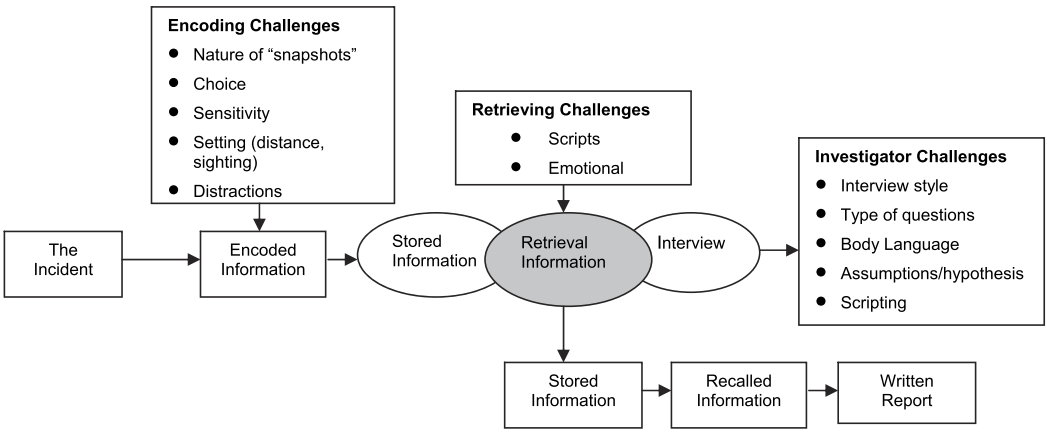


Figure 38.1 Incident memory and interview process (adapted from Milne and Bull, 2003)

The ECI was developed from the Cognitive Interview. The Cognitive Interview (CI) was a police interview procedure devised by Edward Geiselman and Ron Fisher, two American cognitive psychologists. The CI aimed primarily to increase both the quantity and the quality of information elicited from eyewitnesses. It comprised 4 basic techniques:

Reinstate Context Reinstating the context means that interviewees are asked to reconstruct in their minds the context of the incident, both physical (environmental) and personal (for example, how they felt at the time). Research indicates that returning to the context in this way aids remembering because elements of the context act as a cue for retrieving memories (see Malpass, 1996 for a recent review). This helps us to explain why we have a surge of memories about our past life when we visit places we've previously lived or worked. For example, the purpose of "Crimewatch" reconstructions is to reinstate the context of an event in order to jog people's memories of what happened.

It is not just the physical environment that can act as a memory jogger – our internal state can too. For example, a person who is feeling in a positive mood when experiencing an event may be better placed to remember the event in that state (i.e. a positive mood) (Eich, Macauley and Ryan, 1994).

Its effectiveness as a technique to aid memory retrieval is also supported by the fact that there is some evidence to suggest that context reinstatement is a technique that witnesses spontaneously use to remember events (Memon, Wark, Holley, Bull and Koehnken, 1997, cited in Milne and Bull, 2003).

Report Everything It is important that the interviewees are told to report everything they remember without editing even if they don't think the details are important or they can't remember them completely. Studies suggest that unless this is made explicit, interviewees have a tendency to edit information according to what they think the investigator views as important, omitting potentially useful details (Fischer, McCauley and Geiselman, 1992). An event is represented in memory at several different levels of precision ranging from very general to the very detailed. Unless asked to do so, interviewees tend to report from the general level. Telling the interviewee to report everything should therefore encourage him or her to report from more detailed levels of the event representation and this should result in more correct information being reported. In addition, many

people believe their confidence is a reliable indicator of the accuracy of their recalled information and, as a consequence, only report that which they are confident about. However, studies suggest that there is no relationship between the interviewee's confidence in information and its accuracy (Kebbell and Wagstaff, 1997).

Recall Events in a Variety of Orders This involves asking the interviewee to recall the event in a variety of different orders, for example in reverse, working backwards from the time of the incident. Memory is constructive and when people are asked to remember an event their memories are influenced by prior knowledge, expectations and the use of scripts (i.e., what typically happens in a SPAD). When an event is freely recalled most people report the event in real time (i.e., the order in which it took place). When recalling in this way people use their script knowledge of such events to help them recall a particular event. This results in the recall of information that is in line with the script. The danger is that if what happened in reality is different from the script, this does not get picked up. The instruction to change the order of recall invites the interviewee to examine the actual memory record, which in turn can result in the reporting of additional information that is incidental to the script. Research has shown that witnesses who were instructed to recall a crime in forward order and in reverse order remembered more total correct information than witnesses who recalled twice in forward order (Geiselman, 1987, cited in Milne and Bull 2003).

Change the Perspective People have a tendency to report events from their own perspective (Fisher and Geiselman, 1992) and it is recommended that interviewees are asked to recall an event from the perspective of another person who was present. The research that gave rise to this instruction involved a study in which participants were asked to read a story about two boys who went to the house of one of the boys when truanting from school. The story contained a lot of detail about the house such as the rising damp problem and the location of the burglar alarm. Half the participants were asked to read the story as if they were a house buyer, the other half as if they were a burglar. After some time had passed, the participants were told to write down as much as they could remember about the story. When they could remember no more they were asked to shift perspectives (i.e., the house buyers became the burglars and vice versa). It was found that the participants who thought they could remember no more about the story remembered more details when they shifted perspectives. Indeed, when subjects were asked why they hadn't recalled the additional information when using the first perspective they talked about having their memory "jogged" and information "popping" into their mind. It must be noted that care must be taken with this technique as interviewees may misinterpret the instruction of adopting someone else's perspective as an invitation to guess or fabricate an answer. It is imperative to explicitly tell the interviewees that they must only report details about events that they have actually witnessed themselves.

The structure of the interview

While researchers found the CI very effective at increasing the amount of information recalled by witnesses (George, 1992, cited in Milne and Bull, 2003), it did have some shortcomings. For example, it did not provide much guidance concerning the overall structure of the interview. Structure is important as it will influence the order of questioning; inappropriate sequencing of questions can hinder memory retrieval (Fischer et al. 1987). The Enhanced Cognitive Interview (ECI) was therefore developed, in which Fischer and Geiselman (1992) incorporated several new principles from memory research and the social psychology of communication. The ECI

is organised into seven interview phases. Each phase of the interview procedure makes a contribution towards ensuring the success of the interview as a whole (Geiselman and Fischer, 1997).

Phase 1:Greet and personalise the interview – establish rapport

Why?

The opening phase of an interview will determine how well the interview proceeds. An interviewee may feel nervous and anxious about being interviewed and this will hinder the information retrieval process. Therefore, the first phase involves establishing a rapport and helping the interviewee feel more comfortable and should allow them time to work through their feelings. The aim is to make the interviewee feel at ease, confident and secure.

How?

- Introduce yourselves by name.
- Make the interviewee feel at ease, confident and secure. This can be done by showing that you are concerned about their comfort and that the interview space is quiet and without distraction.
- Break the ice by asking some neutral questions not related to the event under investigation and which can be answered positively thus creating a positive mood, such as discussing the past weekend.
- Talking about yourself can help to personalise the interview and this openness can serve as a model to demonstrate what is required of the interviewee.
- Be empathetic and show you understand the situation from the interviewee's perspective.
- Show you are interested and you can listen.

Phase 2: Explain the aims of the interview

Why?

It is important that the interviewee knows what is expected of them during the interview as for most interviewees an investigation of this nature will be alien to them. People typically fear the unknown and by describing the interview process this fear can be reduced. It is during this stage that they should be told to report everything.

How?

- Explain what is expected of the interviewee.
- Transfer control – tell the interviewee that they are in control of the interview as they are the ones who hold all the necessary information. The interviewee will expect the interviewer to control the interview. However, the role of an interviewer in the ECI is as facilitator, to help the interviewee to remember.
- Explain that memory retrieval takes focused energy and concentration and that it is not an easy task.
- Report everything – it is important that interviewees are told to report everything they

remember without editing, even if they don't think the details are important or they can't remember it completely. For example:

•

You were involved in a SPAD this morning, tell me everything you can, even the things you think aren't important and even if you can't remember something completely. Everything which comes to mind tell me in your own time and pace.

Phase 3: Initiating a Free Report

Why?

This is where the instruction to mentally reinstate the context occurs. Returning to the context in which the information was encoded helps the recall process. In addition, asking the interviewee to explain in their own words what happened, at their own pace, in their own words, and without any distractions, helps them to feel in charge of the interview and allows them to recall more accurately what happened.

How?

- Encourage the interviewee to recall the event at their own pace, and in their own words and without any distractions.
- Encourage the interviewee to begin recall by thinking back to the place or context where the incident occurred to help them form an image or impression of the environmental aspects of the original scene, (e.g. the location of objects in a room). This is known as context reinstatement.
- Ask them to comment on their emotional reactions and feelings (surprise, anger, etc) at the time of the event, and to describe sounds, smells and physical conditions (hot, humid, etc) that were present. For example:

Take yourself back to the incident; Get the picture in your mind. Where were you? What did you see? How did you feel? What did you hear? When you are ready I want to hear about the whole incident, at your own pace, from the beginning. I want to hear all the details even if they seem unimportant or you're not completely sure. Take your time.

- The reason for including consideration of as many senses as possible is because people differ in their preferred sensory mode for information processing – some are mainly visual and some talk themselves through things.
- Initiate an uninterrupted free report. If the interviewee can imagine themselves back at the scene of the incident they are then encouraged to report everything they remember about the event, regardless of logical sequence or apparent importance. The invitation might be,

Tell me everything you can, even the little things you think are not important. Remember I was not there so just tell me everything in your own time.

- The interviewer should not interrupt or fill in blanks or pauses, even if they think of questions or need clarification, as this could influence recall or make the interviewee feel pressured. Rather, the interviewer should just maintain rapport and save any questions for clarification/ follow-up for later in the process.

Phase 4: Questioning

Why?

This is an opportunity for clarification. For example, *"Earlier you stated that even though you had booked on in time, it was later than normal, can you tell me more about why this was so?"* The interviewer should explain this phase to the interviewee so that they know they are going to be asked questions about the information they have just given, but only because this is proven to provide more information, and not because the interviewee's testimony is perceived as faulty.

How?

- Repeat the "report everything" instruction and explain it's okay to answer "I don't know" to the questions.

I am now going to ask you some questions based on what you have already told me. It is fine to say "I don't know" to any questions you do not know the answer to. Tell me everything you can remember in response to each question.

- Start with an open-ended question – along the lines of *"tell me more about.."*. *" you said x/y/z/,…did you notice anything else about ….at that time"*.. etc. If closed questions are needed to complete the picture they should be used with extreme caution to avoid introducing bias from the interviewer's perspectives/hunches, etc.
- During this questioning phase, it can be helpful to encourage the witness to go back to the scene again in their minds and try to recall images/sounds/feelings. It is also helpful if the investigator then focuses their questions on one topic at a time so that the witness can use their imagery to help them answer the questions that flow on from one to the next. If the interviewee has to jump back and forth, image to image, valuable information can be lost. Indeed, research has found an associated decrement in recall of around 20% (Fisher and Quigley, 1988, cited in Milne and Bull, 2003).
- Other questions/probes include:
- Confidence probe: e.g. *"How sure are you that you cancelled the AWS?"*
- Recall probe: e.g.*"How do you remember that you cancelled the AWS three times on your journey before the incident?"*
- General probes: e.g. *"How did you arrive at that answer? Was that easy or hard to answer? I noticed that you hesitated – tell me what you are thinking?"*

Phase 5: Varied and extensive retrieval

Why?

The more attempts the interviewee makes to remember a particular episode, the more information will be recalled. However, simply asking the interviewee to repeat things is unlikely to lead to new information. This is where changing perspectives and reversing the order of recall come in. However, the interviewer must explain clearly to the interviewee that the questioning technique will continue but only to provide more information, not because the interviewee's testimony is not believed or is perceived as faulty.

How?

Techniques include:

- Changing the temporal order – recalling the incident backwards. *"What is the last thing you remember? What happened before that?"*
- Changing the perspective – encourage the interviewee to view the incident from another person's view point e.g. an instructor in the cab or an onlooker.

Phase 6: Summary

Why?

The interviewer should summarise the information the interviewee has provided, and encourage them to add or question anything they think does not sound correct. This gives the interviewee the chance to clarify and correct information, and that the interviewer has recorded everything as intended. The interviewee should be instructed that it is acceptable to add new information at this stage.

How?

- Summary statements are a useful technique for demonstrating that you have heard and understood what has been said.
- Try paraphrasing or rephrasing information given by the other person. This is useful because it states the essence of the content in the listener's words. It can also give a slightly different perspective on the material, and lets the person know you are following what has been said, and that you understand.

Phase 7: Close

At the end of the interview, the interviewer should attempt to leave the interviewee in a positive frame of mind. The interviewer should:

- Gradually return to the neutral topics discussed during the rapport building phase.
- Thank the interviewee for their time, effort and co-operation.
- Make sure the interviewee is aware that they can contact the interviewer to talk through any concerns or any new information that might be remembered. It is natural for people to think about the event long after the interview and this may elicit further valuable information (Beckerian and Goodrich, 1997).
- If there is any demographic information that needs to be gathered, such as age or length of service, it should be done at this stage of the interview. Such information tends to create short answer questions, which may have a harmful effect on the rest of the interview and can inhibit rapport building.

Adopting the ECI technique within the rail industry

Overall, the ECI would seem to provide a structured and objective approach to conducting post-incident interviews as it is designed to address the challenges of memory faced by those being asked to remember an event. However, there are two important issues that need to be addressed before this technique can be fully integrated into the SPAD investigation process, or indeed railway investigation processes generally.

First is the issue of whether the CI/ECI is actually effective in the field or not. Since 1984, a large number of research studies examining the effectiveness of the CI/ECI have been conducted. The vast majority show that the CI/ECI elicits more correct information. Some studies have been laboratory based and others such as Fischer, Geiselman and Amador (1989) have examined the effectiveness of the ECI with real witnesses of crime. In their study, audio recordings of witness interviews were obtained from experienced detectives, some trained in the ECI and some not. Overall 47% more information was reported in the ECI group.

With a few exceptions the research has focussed on witnesses or perpetrators of crimes rather than accidents. Consequently, it is important to test the technique in the railway investigation context too. Unfortunately, at present there is a significant stumbling block to make evaluation studies of the nature described above difficult. Despite evidence that shows the value of tape recording interviews[2] there is great reluctance from managers as well as front line staff. Without such recordings it is very difficult to analyse interviewer and interviewee behaviour.

Second is the issue of training. Interviewers do not learn how best to interview through intuition and through observation of colleagues. It is a complex skill that requires appropriate training. There is a lot to be learned from the police on this issue. Following a review in 1992 by the Association of Chief Police Officers and the Home Office into investigative interviewing practices, the PEACE[3] training approach was introduced. This was the first time the police had a standardised interviewing training programmes across the country. Evaluations of the effectiveness of this programme highlight the sorts of issues that the rail industry would need to consider before embarking on a training programme to support the implementation of the ECI:

- Who needs to be trained and does it all need to be to the same level? Could a tiered approach to training be provided such that managers and those involved in investigation panels are given a level of training that enables them to understand the approach and support those undertaking the interviews?
- Do investigation processes need to be changed to support the implementation of the ECI? While trained interviewers might be keen to apply what they have learned they may be prevented from doing so by the processes that are mandated. For example, there are already concerns raised about some of the standard forms[4] that have to be completed and how they can end up guiding the interview rather than the interviewer.
- How is competence as an interviewer established and retained, particularly for those who may not carry out interviews on a regular basis?

2 Kohnken, Thurer and Zoberbier (1994) found that even if an interview is documented directly after the interview, only two-thirds of the information disseminated by the witness will be present in the written report.

3 The acronym PEACE identifies the steps of a model which provides an interview structure: Planning and preparation, Engage and explain, Account, Closure and Evaluation. PEACE training programmes incorporate the ECI.

4 GO/RT3552B form, which has to be completed by the train operating company as part of the mandated investigation process is a very long document (46 pages).

Actions to Date

Human Engineering, together with Cranfield University, has provided training in use of the ECI to several TOCs as part of a more general human factors training course for driver managers. The response to date has been positive – course attendees have appreciated guidance on how to structure their post-incident interviews and different techniques for helping drivers to overcome any problems in recalling the details of an incident. However, extensive practice is required to perfect the technique, and those responsible for interviewing staff in an investigation are likely to benefit from more detailed training in a more realistic context, e.g. by employing actors to carry out role plays of various scenarios. By answering the questions above a more tailored approach to ECI training could be adopted.

Conclusion

Significant improvements can be made to the quality of a SPAD investigation by increasing the skills and knowledge of those responsible for conducting investigative interviews. Evidence has demonstrated that providing interviewers with a structure to guide them through the interviewing process enhances both the effectiveness and accuracy of data collection. Based on the work done here, it is recommended that the ECI is adopted as such a technique and that interviewers, depending on their role in investigations, undergo specialist training to enable them to apply the technique effectively in the field. The optimum approach to ECI training could be guided by the further research into the Police's PEACE training programme. Consideration must also be given to how the interview fits into the investigation process as a whole, in order that the ECI becomes an embedded part of the process.

References

Beckerian, D.A., and Goodrich, S.A.1(997) Recovered memories of child sexual abuse. In G Berrios and J.R. Hedges (eds) *Memory Disorders in Psychiatric Practice.* Cambridge MA. Cambridge University Press.

Eich, E., Macauley, D., and Ryan, L. (1994) Mood dependent memory for events of the personal past. *Journal of Experimental Psychology: General*, 123 201-215

Fischer, R.P., Geiselman, R.E. and Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witness of crime. *Journal of Applied Psychology*, 74, 722 – 727.

Fischer, R.P., Geiselman, R.E. and Raymond, D.S. (1987). Critical analysis of police interviewing techniques. *Journal of Police Science and Administration*, 15 117 – 185.

Fischer, R.P., Geiselman, R.E, (1992). *Memory enhancing techniques for investigative interviewing: the cognitive interview*. Springfield Il: Charles Thomas.

Fischer, R.P., McCauley, M.R. and Geiselman, R.E. (1992). Improving eyewitness testimony with the cognitive interview. In D. Ross, J.D. Read and M. Toglia (eds) *Adult eyewitness testimony: Current trends and developments*. New York. Cambridge University Press.

Fruzzetti, A.E., Toland, K., Teller, S.A. and Loftus, E.F. (1992). Memory and eyewitness testimony. In M. Gruneberg and P. Morris (eds) *Aspects of Memory, Vol 1: The Practical Aspects*. New York: Routeledge

Geiselman, R.E. and Fischer, R.P. (1997). Ten years of cognitive interviewing. In D.G. Payne and F.G. Conrad (eds) *A synthesis of basic and applied approaches to human memory*. New Jersey: Lawrence Erlbaum

Human Engineering Limited (2004). The Driver Standards Manager (DSM) Post-SPAD Process - Human Factors Requirements and Best Practice. Report for Network Rail, reference HEL/NR/03847a/RT1.

Kebbell, M.R. and Wagstaff, G.F. (1997). An investigation into the influence of hypnosis on the confidence and accuracy of eyewitness recall. Contemporary Hynosis, 14, 157-166

Malpass, R. (1996). Enhancing eyewitness testimony. In S.L. Sporer, R.S. Malpass and G. Kohnken (eds) *Psychological issues in eyewitness identification.* Mahwah, NF: Erlbaum

Milne, R., and Bull, R. (2003). *Investigative Interviewing: Psychology and Practice*. West Sussex, England: John Wiley and Sons, Ltd.

Chapter 39

Changing Safety Critical Communications Behaviour

Emma Lowe and Peter Nock

A brief history

Communications between frontline staff have a critical role in ensuring the safe operation of the railways, and in the late 1990's it was highlighted as a key problem for the industry. Investigations of serious accidents pointed to poor communication as a contributory factor as did research looking at communication in railway maintenance: evidence that change was required. In 1997, the Communications Working Group (CWG), was set up and undertook work to analyse the full extent of the problem and produce a framework for competence in safety-critical communications.

In 1999 changes were made to the industry's Rule Book so that use of safety critical communication protocols became mandatory. The protocols included:

- instructions about how start a communication (i.e., by giving your name, function/duty and location)
- standard phrases such as "correction" to mean you have made a mistake and will now correct the word or phrase just spoken and "state again from…" if you need the other party to repeat all or part of their message
- use of the phonetic alphabet to identify letters of the alphabet and to spell words or names that are difficult to say or understand
- lead responsibility whereby one defined person takes responsibility for the outcome of the conversation and ensures that a clear understanding has been reached.

(Details of all the communications protocols can be found in Rule Book Section G1, Section 11 available at www.rgsonline.co.uk).

These protocols were developed based on feedback from frontline railway employees and from good practice from other industries, particularly air traffic control.

Despite these rule changes analysis from incident investigations a year later revealed that the anticipated improvements were not forthcoming. From a psychological perspective this was hardly surprising. The rules had been implemented without any training or briefing. This had an impact on individual's motivation to use them: while the rules in themselves were straightforward, their reason for being developed and their importance was never made clear. Therefore people could not see why they needed to change their behaviour and the undisciplined approach to communications persisted.

This was compounded by an industry which was split into many different companies thus creating a situation in which common culture and mutual loyalty between different groups of staff was lacking. The impact of this is inconsistency and a sense of mis-trust in the application of the communication protocols between the different groups of front line staff (i.e. signallers, drivers and track workers).

A summary of the initiatives

It became apparent that to achieve the improvements in communications whereby disciplined communications is the norm throughout the industry and anything else is unacceptable a multi-level approach was required to engender a massive attitudinal shift across such a diverse workforce. Such change is not easy and presented a huge and long-term challenge. A raft of initiatives were developed from 1999. At the time the basis for these initiatives was common sense and good practice. They can be categorised as initiatives concerning People, Process and Resources.

People

The People focused initiatives were designed to ensure that we employ people with the right abilities and the right attitudes and that they then receive the appropriate training in communications and that they then apply it consistently in their working environment. This led to initiatives in selection, training, briefing and on-going assessment.

a) Competence

These initiatives were concerned with ensuring that individuals had the appropriate aptitude and skills and knowledge for undertaking safety critical communications. Critical to the work was the development of a competence model for communications (see Figure 39.1). This has underpinned the work on training, development and assessment.

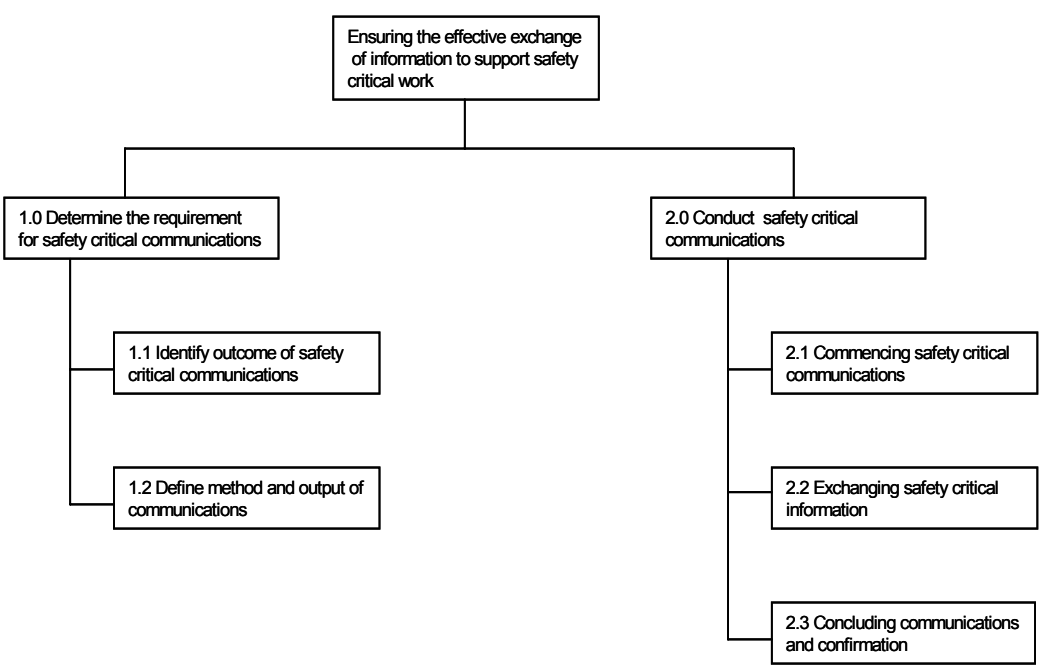


Figure 39.1 Overview of the Safety Critical Communications Competence Model (from Competence Assurance Solutions)

To date changes have been made to initial training programmes for signallers, track workers and track engineers. The aims of the changes were two-fold:

- to introduce a much more practical element into the training so trainees have an opportunity to use and practice the protocols rather than just be told what they are
- to provide information as to why the protocols are critical to safe working

It was hoped that by making these changes trainees would have the motivation to apply these rules and would be sufficiently skilled so that there would be little conscious effort in using them. Of course addressing the skills and knowledge of new staff is only one part of the jigsaw. A more challenging element was existing staff where the issue was not just knowledge and skills but an attitudinal one: "I've been on the railway 20 years and never had to use the protocols, why should I start now".

The 1-day cross-industry safety critical communications forum was designed to tackle this. The forum is not a training event but a large group workshop targeted at drivers, signallers, track workers and other front line staff. It provides individuals with the opportunity to work alongside their industry colleagues, considering the personal impact of poor communication, sharing stories, listening to real examples of good and bad communication and the implications of each plus getting the opportunity to practise some key skills using special telephone training equipment. These forums continue to be rolled out and have targeted thousands of front-line staff. The feedback also continues to be excellent. Comments include:

> Good to mix with the different grades of people and understand the roles, problems and pressures they face.
> Enjoyable throughout and not being made to feel like being taught to suck eggs …..
> A good chance to see my job from another perspective and to talk to the operators face to face.
> Would help to get rail dinosaurs to change their habits!

b) Assessment

Having provided arrangements to ensure employees have the appropriate skills, knowledge and motivation to apply the communications protocols the next challenge was to ensure that employees applied them in the workplace in a consistent manner. This led to initiatives around assessment and monitoring. Assessment and monitoring provide opportunities to check that people have the right knowledge and skills, and if not, improve their competence. However, the two activities also influence behaviour. Without them, standards tend to slip. When people know that their communications are assessed and monitored, they are more likely to follow the protocols. The two main initiatives in this area have involved training assessors in how to assess safety critical communications and improving monitoring arrangements.

Assessor training

A review of the quality of the assessments taking place and feedback from the assessors suggested that assessing safety critical communications was something they found very difficult to do. To address this a 1-day programme was put together for experienced assessors which involved examining good assessment practice and how to apply it to the issue of safety critical communications. An interactive CD combining video, voice recording and interactive exercises was used as part of the programme. Overall, assessors were provided with opportunities to carry out practical exercises to develop their assessment skills and were provided with a resource that could be used back in the workplace for coaching and development purposes.

This concept was later developed with the production of an interactive CD on assessing safety critical communications for the entire industry. As well as reviewing real conversations the CD emphasised the importance of communications and the role assessment can play in developing and maintaining communications competence in order to remind assessors about the critical role they have in managing this issue.

Communications monitoring

Improvements to the monitoring arrangements were first made in 2001 with the introduction of a standard which set out the requirements for the frequency of monitoring and provided a proforma to standardise the way in which communications were monitored and scored. This, together with improvements in the monitoring technology, has led to an increase in the amount of monitoring taking place. The arrangements have recently been reviewed in order to address a number of limitations.

For example, the old monitoring arrangements required conversations to be either A (good), B (need for improvement) or C (poor). Many managers thought that this was too rigid. They felt that they could not rate a communication as good unless it was technically perfect. So, even if there were only small errors, many managers rated the conversation as B. They also felt that they could not give a C, unless the conversation was very poor, and no effort was made to follow the protocols. As a result, the information gained from assessing communications wasn't as meaningful and useful as it could have been. So, a new scoring system has been introduced enabling reviewers to distinguish more accurately between different levels of performance and giving a more realistic view of the standard of communications in a team.

The new assessment is also more fine-grained than the old one in the sense that instead of having to score the communication overall, five communication categories are now scored individually: commences communication, structures communication, speaks professionally, demonstrates leadership and concludes communication. This means that a reviewer can see immediately where the individual performs well, and where they need development and makes it much easier to give feedback and form development plans.

These changes should mean that assessing and monitoring communications becomes more accurate, meaningful, useful and usable.

Processes

The Process initiatives have focused on the systems, processes and arrangements that need to be in place to support the appropriate management of safety critical communications. These have included the publication of a new industry standard detailing the management arrangements that companies should have in place for addressing the risk associated with safety critical communications including recruitment, training, assessment and monitoring.

As described above, Network Rail introduced arrangements for monitoring of communications which also included requirements for joint monitoring between companies. As the owners of the voice recording equipment, Network Rail control access to this data. Joint monitoring sessions provide other companies with access to the voice recordings and facilitates a consistent approach across the industry.

Most recently further changes to the Rule Book have been made to clarify some of the protocols. In particular clarification has been provided about when to use the standard phrases (i.e. not all of them need to be used if you are not using a 1-way radio system) and when it is appropriate to use numbers singly. While these changes are not major it is hoped that it will bring a substantial

improvement in application of the rules as the clarification has improved their credibility and usability.

Resources

The Resource focused initiatives are about ensuring the right equipment and tools are available to support effective communications. Two such initiatives have included the fitment of voice recorders to ensure the majority of safety critical communications taking place are recorded and the development of scripted checklists to support frontline staff in the field with their communications.

Voice recorders A major programme of fitment and upgrade of voice recording equipment into signalling and control locations will be completed by 2006. Changes in technology now allow remote real time monitoring and easier interrogation of recorded conversations which has resulted in an increase in the amount and quality of the monitoring taking place.

Scripted checklists Following concerns about the number of SPAD incidents caused by handsignallers a recommendation was made that a review be undertaken to evaluate the practicality of introducing "Livret Forumlaire" as operated by Eurostar and other continental operations. This work concluded that scripted conversations probably added most value at the training stage as examples of what a typical communications exchange should look/sound like for a particular activity. Where scripted conversations were used by existing railway staff they seemed to confuse rather than structure the conversation. However, the work did identify that a checklist which provides prompts about the order in which certain activities must take place would also help to structure the communications. This work is now being explored in more detail as part of an RSSB research project.

Applying organisational change theory

Rather than conducting an extensive search of the theory and supporting research on change programmes designed to improve safety critical communications, our focus has evolved largely from good practice elsewhere within and outside the industry. We have therefore applied a theoretical justification retrospectively – "this seemed to have worked: I wonder if the literature supports it". It is perhaps more useful to use the theory to understand why, despite all the initiatives that have been undertaken, significant improvements in the standard of communications have not been realised.

The research on changing people's behaviour at work and changing organisations is vast. Unfortunately it tends to focus on describing and depicting rather than providing proven "how to do change" models that are flexible enough to address different scales of change (i.e. whole scale organisational change and specific, micro-level behavioural change) and change in different contexts. However, it is this type of model that would be most useful in understanding and evaluating the safety critical communications change programme.

One model that does try to achieve this is the Burke-Litwin (1992) model of organisational performance and change (see Figure 39.2). This attempts to provide a causal framework that encompasses both *what* organisational dimensions are key to successful change and *how* these dimensions interact to achieve change. Burke and Litwin depict change as a function of both transactional (first-order) factors and transformational (second-order) factors. Transformational change occurs as a response to the external environment and directly affects organisational mission

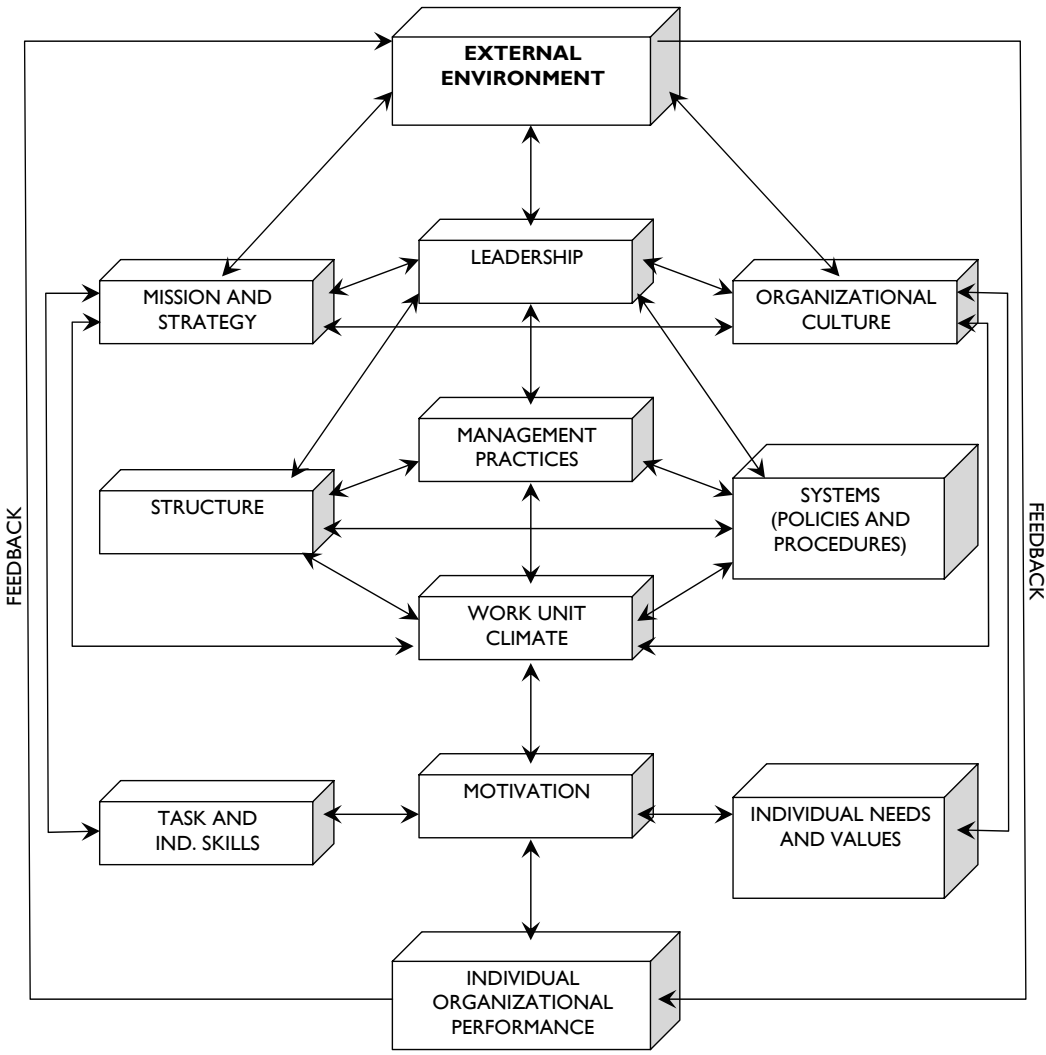


Figure 39.2 Burke-Litwin Model of Organisational Performance and Change (as cited in French and Bell, 1999)

and strategy, the organisation's leadership and culture. It is what produces the value shifts and fundamentally changes the organisation. Transactional change occurs when organisational structure, management practices and systems are affected. These transformational and transactional factors together affect motivation which in turn affects behaviour and performance.

It is important to note how the model also distinguishes between organisational climate and organisational culture. Climate is defined as people's perceptions and attitudes about the organisation and more specifically their work unit in terms of how it is managed and how they work together on a day to day basis with colleagues. Climate change then is a function of managerial and organisational practices which are relatively easy to change. Culture, on the other hand is defined as deep seated assumptions, values and beliefs that are enduring and difficult to change.

Burke and Litwin argue that transactional change could result in changes to the climate but that transformational change is required in order to achieve cultural change.

Applying this model to the safety critical communications change programme it becomes apparent that most of the initiatives to date have been targeted at changing the climate: they are transactional in nature. While few formal evaluation processes have been undertaken it would not be unreasonable to claim that individually and at a micro level these initiatives have worked. Feedback from the forums is positive. New entrants undergoing initial training programmes can apply the communications protocols effectively and there has been a considerable improvement in the standard of communications assessment and monitoring.

Moreover, the model suggests that the reason there has not been a significant shift in communications behaviour and culture is because there has been little in the way of transformational change. It could be argued that various attempts at transformational change have been made over the years. The primary purpose of the cross-industry Communications Working Group, for example, was to provide leadership. The inclusion of specific objectives within the industry's and subsequently Network Rail's Safety and Environment Plans could be argued as providing targets and a mission. Specifically, within Network Rail, an attempt at providing a safety critical communications strategy was made in early 2004 with the submission and approval of a strategy to the Strategic Safety Group and Tactical Safety Groups within Network Rail. This resulted in the Network Rail Safety Critical Communications Implementation Group (NRSCCIG) being formed to oversee the implementation of the strategy.

At this point then it is important to note that Burke and Litwin believed that organisational change stems more from external environmental impact (i.e., changes in the regulations, technological breakthroughs, and changes to the economic environment) than from any other factors. While not all organisational change research would agree with this premise there are two factors that indicate that as far as the safety critical communications change programme is concerned this is certainly the case.

Firstly, it wasn't until two significant, high profile accidents occurred that the company's Safety Critical Communications strategy was again endorsed by the senior management team. Only then were recommendations that it should be given the status of a company action plan with functional targets, the appropriate level of resource and a dedicated central resource to coordinate the implementation of the associated action plan taken forward.

Secondly, it is only recently that the industry is being able to describe in any objective sense the full extent of the safety critical communications "problem". For a long time the figure that 90% of all incidents and accidents had communications as a contributory factor was quoted. However, this figure was based on very broad-brush research and so was not as meaningful as it could be; ultimately it has been so over-used that it lost its impact. This prompted more recent detailed research by the Rail Safety and Standards Board (RSSB) to establish the causal role played by communications in incidents on the railways and costs associated with such communications errors. While this research has yet to be formally published initial reports suggest that for the first time the industry will be able to report with some confidence the number of incidents caused, at least in part, by miscommunication, the nature of the miscommunications and their causes.

This not only provides a level of detail that will enable initiatives to be appropriately targeted and a benchmark against which to measure initiative effectiveness, it also provides information about the economic and safety losses associated with poor communications. For example, knowing that a signal passed at danger (SPAD) caused by miscommunication costs £22,000 and that we have approximately 40 SPADs a year due to miscommunication is far more likely to engage senior management and release resources to address the issue than quoting 90% of all accidents and incidents involve miscommunications.

Obviously, in addition to providing insights as to why initiatives have not produced the desired culture change, the model can also be used to highlight areas to consider for future action. If, as the model suggests, transformational change won't be initiated until there is an external influence then we need to be very careful in our communication of that fact to prevent the (not surprisingly) cynical response of "things only change around here when people get killed or it costs money". Focussing on providing leadership and a strategy will clearly help to bring the transformational change being aspired to. One way of demonstrating leadership that is currently being explored is through a campaign of zero tolerance.

Zero tolerance campaigns have typically been used in other sectors and a useful industry comparison is the oil industry. A massive zero-tolerance campaign to accidents on rigs in the North Sea was run after the Piper-Alpha disaster (external factor prompting change). The management backed the campaign to the hilt and the company leadership held strong safety values that were well communicated and on the surface it seemed to work with reduced accident rates. However, when in-depth research was carried out on the rigs themselves it transpired that as a "zero-accident rate" was linked to the bonus for the entire team on the rig, unsurprisingly, accidents happened and were simply not reported.

However, there are a number of examples from the railway where such an approach has helped to bring about fundamental change in working culture in the past. For example, the railway now has a drugs and alcohol culture where it is accepted that coming to work under the influence is not an option. In 1992, it became an offence to carry out safety critical work while under the influence of drugs and alcohol. British Rail responded by introducing a zero tolerance policy to drugs and alcohol. This applied to all employees, including office-based staff. Alcohol was banned at all BR premises. Over 120,000 people were briefed at the workplace by trained teams and a number of significant high profile cases emerged where harsh sanctions were used against individuals for transgressions – the stories were passed around the industry very rapidly and helped to create the new culture that was being sought by management. Today, unannounced random testing helps to keep the standards high. People testing positive are the exception, and negative results the norm. It was a massive culture change, and one of the most successful ever in the rail industry.

In relation to safety critical communications a zero tolerance approach would mean not accepting poor, undisciplined communications and managers taking the same level of action as they would for other operating violations. The arrangements for monitoring, assessment and competence all help to achieve this but clearly stating and then demonstrating that poor communications will not be tolerated is another step.

However, considerable thought will need to be given to this goal in terms of how it is communicated without eliciting the same response of cover-up or denial that occurred in the oil industry and further research will be a useful step.

Overall, this model of organisational change has highlighted that the initiatives to date have had a positive affect at a local level only and that the reason why there has not been the significant shift in culture hoped for is because until the change in external environment occurred, as a result of two accidents and economic data becoming available, the senior management team had not been properly engaged in the strategic and leadership initiatives. This is an important lesson to learn for future programmes: while to some extent external influences are outside an organisation's control, change programmes must identify at an early stage what it can do to influence such events whether, as in this case, it is about providing reliable loss data or whether it involves exploring new technologies or lobbying for regulation change.

Conclusion

Working in a constantly changing industry on an issue that crosses a number of different work groups to encourage behaviour change at an individual level is a massive challenge. This chapter has described some of the initiatives that have been undertaken to change safety critical communications behaviour and which have started to show positive changes, albeit at a local level. Applying a model of organisational change, such as the Burke-Litwin model, has helped to highlight why such initiatives have had a limited effect and not produced the shift in culture that would bring a significant and permanent change in communications practice. It also highlights how any change programme designed to produce such a fundamental shift either in behaviour or performance must involve senior management commitment, leadership and strategy and that this is unlikely without some precipitating external event.

References

Burke, W.W. and Litwin, G.H. (1992). A Causal Model of Organisational Performance and Change. *Journal of Management*, Vol 18, No. 3, 523 – 545

French, W.L. and Bell, C.H. (1999). *Organisation Development: Behavioural Science Interventions for Organisation Improvement*. 6th ed. Prentice Hall.

This page intentionally left blank

Chapter 40

Safety Climate in Australian Railways

A. Ian Glendon and Bronwyn Evans

Introduction

Organizational safety cognition measures – mainly perceptions of and attitudes towards safety, commonly take the form of safety climate self-completion surveys, of which there are many examples (e.g., Cheyne *et al.*, 1998; Cooper, & Phillips, 1994; Cox & Cox, 1991; Cox *et al.*, 1998; Coyle *et al.*, 1995; Davies *et al.*, 2001; Dedobbeleer, & Béland, 1991; Donald & Canter, 1994; Flin *et al.*, 2000; Glendon & Litherland, 2001; Janssens *et al.*, 1995; Niskanen, 1994; Silva *et al.*, 2004; Varonen, & Mattila, 2000; Williamson *et al.*, 1997; Zohar, 1980). For reviews, see Cooper and Phillips (2004), Glendon *et al.* (2006), Guldenmund (2000), and Seo *et al.* (2004). These measures typically find between two and seven separate factors or safety climate scales. For example, Griffin and Neal (2000) identified five first-order factors, which in turn loaded onto a common higher-order factor, relating to perceptions of safety climate. Their first-order factors were: 'Management values', 'Safety communication', 'Safety practices', 'Safety training', and 'Safety equipment'.

However, no consistent factor structure has yet emerged for safety climate, which Seo *et al.* (2004) attributed to not specifying the influence of two critical safety dimensions – management commitment and supervisor support. Seo *et al.* (2004) demonstrated the importance of developing psychometrically robust safety climate scales. Of over 30 empirical safety climate studies published since 1980, most have developed their own measure. This opportunistic approach has resulted in no systematic study, for example by sector, which could provide an opportunity for customised valid safety climate measures that could be used, *inter alia*, to compare occupational groups within and between rail sector organizations, to benchmark different rail sector organizations – both nationally and internationally, and for rail sector organizations to evaluate safety interventions by comparing their safety climate between time periods.

There is dispute over defining safety climate, and how it links with other important factors to influence safety performance. Glendon and Litherland (2001) proposed that safety climate could be conceptualized as operating on three levels:

- *Operational* – accessing factors impacting most directly upon work performance and dealing exclusively with perceptions (e.g., Glendon & Stanton, 2000; Wilson, 1998).
- *Intermediate* – comprising perception-oriented measures but with some attitudinal items, reflecting generic factors, such as 'management commitment' and 'safety system' (e.g., Clarke, 2000; Flin *et al.*, 2000; Williamson *et al.*, 1997).
- *Highest* – using purely attitudinal measures (e.g., Donald & Canter, 1993; Niskanen, 1994), which could tap into some aspects of safety culture.

At the highest level described by Glendon and Litherland (2001), safety climate reflects to some extent the underlying culture of the organization with respect to safety; Cox and Flin (1998) considered safety climate to be a manifestation of safety culture expressed through workers' attitudes

and behavior, a description also given by Cheyne *et al.* (2003). 'Safety culture' is generally taken to be more embracing than 'safety climate', although the two terms have similar meanings. Whilst culture implies a notion of residing within an organization, climate has more passive connotations, reflecting attitudes and perceptions of organization members to both internal (e.g., management actions) and external (e.g., economic) influences (Glendon, 2005). The Ladbroke Grove rail accident inquiry concluded that: "climate is the observable, tangible part of culture. Culture is the understanding of people's fundamental values with respect to say, risk and safety" (Cullen, 2001, p. 2). Thus, safety climate can be defined as reflecting shared attitudes and perceptions of organizational members towards internal and external influences on safety. It may be regarded as an, 'indicator of safety culture within the workforce as a whole' (Mearns *et al.*, 2003), but only provides a 'snapshot' picture of the longer-term, more enduring safety culture of an organization. Cheyne *et al.* (2003) regarded safety attitudes as a component of safety climate, in turn was a manifestation of safety culture.

Some researchers have sought to determine whether different groups of workers within an organization or sector report different attitudes or perceptions of safety, or whether management and workers express different attitudes to safety. Cox *et al.* (1998) found differences between safety attitudes of workers, supervisors and managers in the UK manufacturing sector, for example that permanent workers had more positive attitudes on some issues than did other groups. Similar findings have also been reported in the UK rail industry (Clarke, 1999) and for US haulage firms (Arboleda *et al.*, 2003). Different sub-cultures or sub-climates are liable to exist at different levels within an organization (Gonzales-Roma *et al.*, 1999; Harvey *et al.*, 1999; Trice & Beyer, 1993). This can result in differences in safety attitudes at different levels (e.g., managers, supervisors, workers) within an organization, and between different groups of workers (Alexander *et al.*, 1995; Cox *et al.*, 1998). Gillen *et al.* (2002) found different perceptions of safety climate between unionized and non-unionized workers. Cheyne *et al.* (2003) found that while managers, supervisors and workers shared the same safety climate factor structure, their perceptions of the factors and their inter-relationships were quite different. Sampling different employee groups from the Australian rail sector, Glendon and Evans (2005) and McInerney (2005) found significant differences between employee groups, including managers and supervisors, on a safety climate measure designed for the rail sector.

Although studies have obtained varied results, there is growing evidence that safety climate reflects differences in safety performance, either directly or through some other channel, such as organizational climate – for a review, see Glendon *et al.* (2006). Kopelman *et al.* (1990) proposed that the influence of climate on behavior is mediated by cognitive and affective states, through particular pathways – for example, climate perceptions impact on work motivation, which in turn affects job performance, or climate perceptions impact on job satisfaction, which in turn affects psychological well-being and withdrawal. Irrespective of any behavioral link, a safety climate survey provides a valuable tool for identifying trends in an organization's safety performance (Cox & Cheyne, 2000; Coyle *et al.*, 1995; Seo *et al.*, 2004). Recent work has focused on group-level safety climate (Zohar, 2000, 2002), as opposed to organizational safety climate, where climate is defined in terms of how supervisors prioritize safety issues.

Method

The NSW rail organization was surveyed as part of the Special Commission of Inquiry into the Waterfall rail accident, in which seven people were killed (McInerney, 2005). The questionnaire included 34 questions on various aspects of safety, which respondents answered on a 5-point scale ranging from '1 Strongly disagree' to '5 Strongly agree'. Main occupational groups surveyed were:

Train drivers, Train guards, Signalling staff, Maintenance staff (rolling stock & track), Station staff, Customer service staff, Management and supervisory staff, and New employees (having less than 12 months service). A Commission staff member visited a number of locations to ask groups of employees to complete the questionnaire. This ensured both a very good response rate (459 employees completed safety climate questionnaires; only one of those asked declined), an adequate sample size for statistical analysis, and reasonable representation across key occupational groups. New employees comprised 11.5% of the sample. Mean length of employment within the NSW rail industry of the other respondents was 15.4 years.

In the second survey, questionnaires were mailed to the entire rail membership of the Queensland branch of the Rail Tram and Bus Union (RTBU), which represents about 6000 (95%) rail workers in Queensland. The survey yielded 514 usable responses (9% response rate). Sixty-four percent of respondents dealt directly with train operations (e.g., Train drivers, Track maintenance staff, Rolling stock maintenance staff, Train guards, Train controllers, Signalling staff) while another 22 percent were engaged in management and other support roles (e.g., Station staff, Management/supervisory, Administration staff, Truck drivers, Customer service officers, Procurement officers, and other onboard service officers). Mean time worked in the rail sector by respondents was 21.4 years (SD 9.88 years).

The survey was modified to reflect NSW study findings. Some original items were rephrased and, in response to comments to an open-ended question, further items on communication, equipment, maintenance, and shifts and rosters were included. The final survey contained questions on five topic areas hypothesised to predict safety climate at an organizational level (management commitment to safety, organizational communications, equipment and maintenance, shifts and rosters, safety training). Both samples exceeded the minimum number of cases required to conduct the proposed analysis (Hair *et al.*, 1995). Table 40.1 shows the occupational breakdown of respondents in the two surveys.

Table 40.1 Respondents' occupational group

Occupational group	NSW	QLD
Train drivers	56	91
Train guards	69	36
Signalling staff	48	18
Maintenance (Rolling stock/Track) staff	50	51/84
Station staff/Customer service staff	72	46/13
Management and supervisory staff	69	46
New employees	63	-
Others	32	129
Totals	459	514

Findings

NSW sample

Factor analyses reduced the questionnaire items to two main factors to represent safety climate. Factor 1, 'Management and Staff Safety' had 14 items and explained 54.74% of the variance (α .95). Factor 2, 'Safety Training and Rules' had 10 items and explained 5.75% of the variance

(α .93). More detail is given in McInerney (2005). Table 40.2 shows mean scores for the seven occupational groups on the two factors.

Table 40.2 Mean scores for seven occupational groups on two safety climate factors (NSW sample, N=459)

	Factor	
Occupational group	1	2
Train drivers	2.28[a]	2.82[a]
Train guards	2.66[b]	2.98[a]
Signalling staff	2.96[c]	3.11[a,c]
Maintenance staff	2.88[b,c]	2.88[a]
Station staff/Customer service staff	3.42[d]	3.52[b]
Management and supervisory staff	3.60[d,e]	3.27[b,c]
New employees (< 12 months)	3.75[e]	3.94[d]
Overall	3.11	3.24

Table 40.2 shows that mean scores for all respondents (excluding the 'Others' group) were 3.11 for the 'Management and staff safety' factor, and 3.24 for the 'Safety training and rules' factor. This means that overall, respondents perceived both safety climate factors to be just above the mid-point of the 5-point scale, where a score of '3' indicated 'Neutral'. Scores on both factors also show more than a 20% difference between the highest (New employees in both cases) and lowest (Train drivers in both cases) group scores, suggesting substantial differences between occupational groups' perceptions of the organization's safety climate. MANOVA showed significant differences between groups on both factors, indicated by different superscripts in the columns of Table 40.2.

In respect of their perceptions of Factor 1 'Management and staff safety', there are three separate clusters. Train drivers are in a 'cluster' of their own – agreeing with no other group in their perceptions of this safety climate factor. While Train guards and Signalling staff differ significantly in their perceptions of this factor, both agree with the perceptions of Maintenance staff. However, these groups' perceptions differ significantly from those of the other three groups. While there is no agreement between Station/ Customer service staff and New employees, the Management and supervisory group is in broad agreement with the perceptions of both these groups.

The picture in respect of Factor 2 'Safety training and rules' is one of greater agreement between Train drivers, Train guards, Maintenance staff, and Signalling staff, all of whom have similar perceptions of this factor. Signalling staff and the Management and supervisory group also share common perceptions. Management and supervisory respondents and Station and Customer service staff also share perceptions on this safety climate factor. However, the New employees group is completely isolated in terms of their perceptions of this safety climate factor.

QLD sample

Factor analysis yielded five factors with eigenvalues greater than 1, accounting for 61.6 percent of explained variance. These were '1 Communication and safety information' (11 items, 35.51% of variance, α .90), '2 Rosters and shifts' (6 items, 12.00%, α .91), '3 Signalling equipment' (2

items, 6.13%, α .80), '4 Equipment and maintenance' (5 items, 4.26%, α .80), and '5 Management commitment to safety' (4 items, 3.72%, α .87).

A MANOVA was conducted to identify any differences between occupational group means. Although 514 cases were available for analysis, because MANOVA is sensitive to differences in cell sizes, numbers of respondents in each occupational group were reduced by randomly deleting cases until cell sizes were approximately equal. The reduced sample means did not differ significantly from the larger sample means. Significant main effects were found for all factors apart from '3 Signalling equipment', which was dropped from subsequent analyses. Table 40.3 shows means for the various occupational groups on the four remaining safety climate factors.

Table 40.3 Mean scores for seven occupational groups on four safety climate factors (QLD sample, N=321)

	Factor			
Occupational group	1	2	4	5
Train drivers (N=53)	2.85[a]	2.24[a]	1.84[a]	2.96[a]
Train guards (N=36)	2.81[a]	2.32[b]	2.49[b]	2.54[b]
Station staff (N=46)	2.64[b]	2.49[b]	2.01[c]	2.97[a]
Track maintenance staff (N=53)	3.04[c]	3.26[c]	2.18[c]	3.32[c]
Rolling stock maintenance (N=51)	3.05[c]	3.17[c]	2.28[b]	3.20[c]
Management and supervision (N=46)	3.31[d]	3.37[c]	2.34[b]	3.20[d]
Administration staff (N=36)	3.32[d]	3.26[c]	2.97[d]	3.08[e]

Summarizing significant differences between occupational groups (means in Table 40.3):

- On '1 Communication and safety information', Station staff mean score was significantly lower than those of Managers/supervisors and Administration staff.
- On '2 Rosters and shifts' Train guards, Train drivers' and Station staff mean scores were significantly lower than those of Track maintenance staff, Rolling stock maintenance staff, Managers/supervisors and Administration staff.
- On '4 Equipment maintenance' Administration staff scores were significantly higher than those of all other groups. Train drivers' mean scores were significantly lower than those of all other groups. Station staff and Track maintenance staff were significantly lower than Train guards, Rolling stock maintenance staff and Management/supervision.
- On 'Management commitment to safety' Train guards' mean scores were significantly lower than those of Track maintenance staff, Rolling stock maintenance staff and Managers/supervisors. Train drivers', Station staff and Administration staff mean scores were significantly lower than that of Managers/supervisors.

A non-significant ANOVA result to determine whether differences existed between occupational groups on the combined safety climate scores indicated no differences between occupational groups on the summed safety climate scales.

Discussion

The finding from the first study of different occupational groups recording different scores on the safety climate factors within a rail sector organization was replicated in the second (State-wide) study. Occupational groups closer to daily operations tended to have lower scores than did groups further removed from daily operations. Train drivers' and Station staff scores differed significantly from those of Administration staff and Management/supervision on four of the five safety climate factors. The importance of measuring group differences on the separate scales was highlighted by the non-significant result across the combined scales.

The first factor, 'Communication and safety information', comprised mostly communication items with some training items. The scale seems to represent a need for good communication channels and for safety information to be communicated throughout the organization. The second factor was based on rostering and the impact of shiftwork on fatigue. This factor emerged as a strong and discrete factor, possibly due to the fact that only some operational groups work shifts, producing polarized responses. Although this factor may only apply to some workplace groups, its importance was evident from qualitative data gathered in the first study. The third factor, 'Signalling equipment maintenance', was also remarkably discrete, comprising only two items 'Signalling equipment is never left in use with safety critical faults', and 'Signalling equipment is maintained to a safe standard'. This rail-sector specific factor could not be part of a generic safety climate measure. As it did not discriminate between occupational groups in this study and, as factors with less than three items are not robust, this factor is problematic. Future development of a scale for the rail sector may benefit from additional items on this topic, or it might be better explored in other ways (e.g., via focus groups). The 'Management commitment to safety' factor emerges from most safety climate studies. 'Equipment maintenance', although not common to other studies, is relevant to a range of settings.

Conclusions

In the NSW sample, the overall perception of the sample was barely above the mid-points of both safety climate scales. Perceptions of Train drivers, Train guards, Signalling staff and Maintenance staff were all below both scale mid-points. Train drivers perceived safety climate to be significantly worse than did all other occupational groups sampled. Maintenance staff, Train guards and Signalling staff generally perceived safety climate to be significantly worse than did Station/Customer service staff, Management and supervisory staff, and New employees.

The second study yielded four stable factors that could be considered generic, that is, usable within different transport industries and other sectors. The 'Signalling equipment maintenance' factor is too specific to be useful outside the rail sector. Future studies could balance generic factors that produce useful comparisons across organizations or industries, with sector-specific safety measures. The four generic factors extracted in this analysis were: 'Communications and safety information', 'Rosters and shifts', 'Equipment maintenance' and 'Management commitment to safety'. Generic factors are required for comparing safety climate measures both within and between sectors. Inventories developed for other sectors could add sector specific scales to generic safety climate scales.

The two studies represent steps in the development of a rail safety climate measure. Both studies identified consistent differences between occupational groups. In the second study Train drivers', Train guards' and Station staff mean scores were significantly different from those of Managers/supervisors and Administration staff on three of the four factors discriminating between occupational groups ('Communications and safety information', 'Rosters and shifts', 'Management

commitment to safety'). Differences on 'Communication and safety information' and 'Management commitment to safety' may be due to the nature of groups different work environments. Train drivers, Train guards and Station staff perform their roles at a physical distance, sometimes remotely, from management supervision, and staff in these roles often work alone or with one or two other people. This may foster a sense of isolation from management and from the organization's communication mechanisms. Track maintenance staff and Rolling stock maintenance staff tend to work under closer management supervision and in larger teams. As a result, they may feel more informed about management's commitment to safety and current information in the organization. Management contact influencing safety climate was demonstrated by Clarke (1999) and Glendon and Litherland (2001).

Differences on the 'Rosters and shifts' factor is most likely to be due to managers, supervisors and administration staff for the most part not being required to undertake shiftwork. Track maintenance staff and Rolling stock maintenance staff work limited shifts, hence their intermediate scores on this factor. Train drivers reported lower levels of acceptable 'Equipment maintenance' than did all other groups. This factor, among other things, asked whether there was sufficient money and staff for maintenance and whether maintenance was carried out in a timely manner.

Researchers are increasingly using sophisticated statistical modelling techniques to unravel some of the complex relationships and interactions between safety climate components (e.g., Cheyne *et al.*, 1998, 2003; Griffin, and Neal, 2000; Neal *et al.*, 2000; Seo *et al.*, 2004; Siu *et al.*, 2004; Tomás *et al.*, 1999). Future research is likely to involve more detailed models of the operation of safety climate, its impact on safety-related behaviors, and its relationship with other facets of the work environment. The next stage of this research involves further development of the safety climate inventory for repeat use with a larger sample within the same NSW rail organization as in the first study described in this chapter. This is being undertaken as part of a larger safety culture project, which also includes qualitative measures.

References

Alexander, M., Cox, S., and Cheyne, A. (1995). *The concept of safety culture within a UK offshore organisation*. Paper presented at the Understanding Risk Perception Conference, Robert Gordon University, Aberdeen, February.

Arboleda, A., Morrow, P. C., Crum, M. R., and Shelley, M. C. (2003). Management practices as antecedents of safety culture within the trucking industry: similarities and differences by hierarchical level. *Journal of Safety Research, 34*, 189-197.

Cheyne, A. J. T., Cox, S. J., Oliver, A., and Tomás, J. M. (1998). Modelling safety climate in the prediction of levels of safety activity. *Work and Stress, 12*, 255-271.

Cheyne, A. J. T., Tomás, J. M., Cox, S. J., and Oliver, A. (2003). Perceptions of safety climate at different employment levels. *Work and Stress, 17*, 21-37.

Clarke, S. G. (1999). Perceptions of organizational safety: Implications for the development of safety culture. *Journal of Organizational Behaviour, 20*, 185-198.

Clarke, S. G. (2000). Safety culture: Under-specified and overrated? *International Journal of Management Reviews, 2*, 65-90.

Cooper, M. D., and Phillips R. A. (1994). Validation of a safety climate measure. Proceedings of the British Psychological Society: Annual Occupational Psychology Conference, Birmingham, UK, January.

Cooper, M. D., and Phillips, R. A. (2004). Exploratory analysis of the safety climate and safety behaviour relationship. *Journal of Safety Research, 35*, 497-512.

Cox, S. J., and Cheyne, A. J. T. (2000). Assessing safety culture in offshore environments. *Safety Science, 34*, 111-129.

Cox, S. J., and Cox, T. (1991). The structure of employee attitudes to safety: A European example. *Work and Stress, 5*, 93-106.

Cox, S. J., and Flin, R. (1998). Safety culture: Philosopher's stone or man of straw? *Work and Stress*, *12*, 189-201.

Cox, S. J., Tomás, J. M., Cheyne, A. J. T., and Oliver, A. (1998). Safety culture: The prediction of commitment to safety in the manufacturing industry. *British Journal of Management, 9*, 3-7.

Coyle, I. R., Sleeman, S. D., and Adams, N. (1995). Safety climate. *Journal of Safety Research, 26*, 247-254.

Cullen, W. D. (2001). *The Ladbroke Grove rail inquiry Part 2: Report*. London: HMSO.

Davies, F., Spencer, R., and Dooley, K. (2001). *Summary guide to safety climate tools*. Sudbury, England: HSE Books.

Dedobbeleer, N., and Béland, F. (1991). A safety climate measure for construction sites. *Journal of Safety Research, 22*, 97-103.

Donald, I., and Canter, D. (1993). Psychological factors and the accident plateau. *Health and Safety Information Bulletin, 215*, 5-12.

Donald, I., and Canter, D. (1994). Employee attitudes and safety in the chemical industry. *Journal of Loss Prevention in the Process Industries, 7*, 203-208.

Flin, R., Mearns, K., O'Connor, P., and Bryden, R. (2000). Measuring safety climate: Identifying the common features. *Safety Science, 34*, 177-192.

Gillen, M., Baltz, D., Gassel, M., Kirsch, L., and Vaccaro, D. (2002). Perceived safety climate, job demands, and coworker support among union and non-union injured construction workers. *Journal of Safety Research, 33*, 33-51.

Glendon, A. I. (2005). Safety culture. In W. Karwoski (Ed.), *International encyclopedia of ergonomics and human factors* (2nd ed.). London: Taylor and Francis. (in press).

Glendon, A. I., Clarke, S. G., and McKenna, E. (2006). *Human safety and risk management* (2nd ed.). Boca Raton, FL: CRC Press.

Glendon, A. I., and Evans, B. (2005). Safety climate in the railway industry. Paper presented at the 5th UQ Symposium on Organizational Psychology. Brisbane: University of Queensland, June.

Glendon, A. I., and Litherland, D. K. (2001). Safety climate factors, group differences and safety behavior in road construction. *Safety Science, 39*, 157-188.

Glendon, A. I., and Stanton, N. A. (2000). Perspectives on safety culture. *Safety Science, 34(1-3)*, 193-213.

Gonzales-Roma, V., Peiro, J. M., Lloret, S., and Zornoza, A. (1999). The validity of collective climates. *Journal of Occupational and Organizational Psychology, 72*, 25-40.

Griffin, M. A., and Neal, A. (2000). Perceptions of safety at work: A framework for linking safety climate to safety performance, knowledge, and motivation. *Journal of Occupational Health Psychology, 5*, 347-358.

Guldenmund, F. W. (2000). The nature of safety culture: A review of theory and research. *Safety Science, 34*, 215-257.

Hair, F. H. Jr., Anderson, R. E., Tatham, R. L., and Black, W. C. (1995). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.

Harvey, J., Bolam, H., and Gregory, D. (1999). How many safety cultures are there? *The Safety and Health Practitioner*, 17 December, 8-12.

Janssens, M., Brett, J. M., and Smith, F. J. (1995). Confirmatory cross-cultural research: Testing the viability of a corporation-wide safety policy. *Academy of Management Journal, 38*, 364-382.

Kopelman, R. E., Brief, A. P., and Guzzo, R. A. (1990). The role of climate and culture in productivity. In B.Schneider (Ed.), *Organizational climate and culture* (pp. 282-318). San Francisco: Jossey-Bass.

McInerney, P. A. (2005). *Final report of the Special Commission of Inquiry into the Waterfall rail accident.* Sydney: Government of New South Wales.

Mearns, K., Whitaker, S. M., and Flin, R. (2003). Safety climate, safety management practice and safety performance in offshore environments. *Safety Science, 41*, 641-680.

Neal, A., Griffin, M. A., and Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behaviour. *Safety Science, 34*, 99-109.

Niskanen, T. (1994). Assessing the safety environment in the work organization of road maintenance jobs. *Accident Analysis and Prevention, 26*, 27-39.

Seo, D.-C., Torabi, M. R., Blair, E. H., and Ellis, N. T. (2004). A cross-validation of safety climate scale using confirmatory factor analytic approach. *Journal of Safety Research, 35*, 427-445.

Silva, S., Lima, M. L., and Baptista, C. (2004). OSCI: An organizational and safety climate inventory. *Safety Science, 42*, 205-220.

Siu, O.-l., Phillips, D. R., and Leung, T.-w. (2004). Safety climate and safety performance among construction workers in Hong Kong: The role of psychological strains as mediators. *Accident Analysis and Prevention, 36*, 359-366.

Tomás, J. M., Melia, J. L., and Oliver, A. (1999). A cross-validation of a structural equation model of accidents: Organizational and psychological variables as predictors of work safety. *Work and Stress, 13*, 49-58.

Trice, H. M., and Beyer, J. M. (1993). *The culture of work organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Varonen, U., and Mattila, M. (2000). The safety climate and its relationship to safety practices, safety of the work environment and occupational accidents in eight wood-processing companies. *Accident Analysis and Prevention, 32*, 761-769.

Williamson, A. M., Feyer, A.-M., Cairns, D., and Biancotti, D. (1997). The development of a measure of safety climate: The role of safety perceptions and attitudes. *Safety Science, 25*, 15-27.

Wilson, J. D. (1998). *The development and validation of a safety climate survey for the Western Australian health sector*. Unpublished Master of Psychology dissertation, School of Psychology, Curtin University, Western Australia.

Zohar, D. (1980). Safety climate in industrial organisations: Theoretical and applied implications. *Journal of Applied Psychology, 65*, 96-102.

Zohar, D. (2000). A group-level model of safety climate: Testing the effect of group climate on microaccidents in manufacturing jobs. *Journal of Applied Psychology, 85*, 587-596.

Zohar, D. (2002). Modifying supervisory practices to improve sub-unit safety: A leadership-based intervention model. *Journal of Applied Psychology, 87*, 156-163.

This page intentionally left blank

Chapter 41

The Waterfall Train Accident: The Critical Role of Human Factors

Andrew S. McIntosh and Graham Edkins

Introduction

The Waterfall train accident occurred on the morning of January 31, 2003 approximately 2 kilometres south of Waterfall station on the southern border of Sydney. The driver and six passengers were fatally injured in the accident. This chapter reports on the critical role of human factors in the cause of the accident. The accident involved an outer suburban model Tangara (G7). The Tangara is a double-decker unit and was in a four car configuration. The accident came three years after the December 1999 Glenbrook train accident in the Blue Mountains close to Sydney. In both cases a Special Commission of Inquiry (SCOI) was established headed by Justice McInerney.

There was no obvious cause of the accident. Immediately after the accident there was a great deal of public speculation regarding the factors that may have led to the tragic loss of seven lives. This included whether ballast had been deliberately placed on the tracks to derail the train, whether the tracks had buckled due to the extreme temperatures the previous summer day, or if the train had 'surged' up to high speed and was 'out of control'.

It was evident that the Tangara had derailed at high speed, rolled, collided with a stanchion, ran up an embankment and finally returned to the track area. The collisions resulted in substantial damage to the rolling stock and rail infrastructure. The inquiry determined that the Tangara had rolled over at a speed of 117 km/h on a left hand bend after the driver became incapacitated from a cardiac event. The guard had not applied the emergency brake, and while the deadman system was electro-mechanically operable, it had not functioned. Mechanical train stops were present between Waterfall station and the crash site, but G7 had encountered no red signals so the only possibilities for braking the train were the deadman system and the guard.

Overview of inquiry and investigations

Inquiries were conducted by the Ministry of Transport and by the SCOI. The SCOI was the primary public inquiry and involved very detailed investigations that were conducted in two main phases: causation and safety management systems (McInerey, 2005). The initial investigations were divided into: infrastructure, rolling stock, human factors and management/administration. A lead investigator, Mr. Bob Lauby, was appointed by the SCOI for the first phase of the inquiry.

As the inquiry progressed public hearings were held at which investigators, as well as witnesses, rail personnel and other parties, gave evidence and were cross-examined. Further, as the infrastructure and rolling stock based investigations developed, it became evident that infrastructure defects had not caused the accident and that the train's power and braking systems were functional at the time of the crash. The gradual elimination of purely engineering factors and evidence that the driver had suffered a cardiac event placed more emphasis on the need to direct resources into determining the

human factors associated with the accident. It is generally agreed that *human factors* is an umbrella term for the study of people's performance in their work and non-work environments. From the perspective of accident investigation, human factors is about understanding how people interact with systems, use equipment, interpret policies, work with procedures and manuals, and operate within their working environment. It is also about their relationship with people.

The authors were appointed as consultants to the Commission to undertake investigations in these human factors issues.

Human factors investigations

Human factors investigations were undertaken in five main areas:

- deadman system function,
- driver and guard work and personal history,
- fatigue and shift work,
- safety culture and train crew communications, and
- medical standards.

The authors were the primary investigators in the first four areas, and contributed significantly to the latter.

Deadman system

Deadman system investigations took a number of forms: design history, measurement of operating forces and functionality on G7 and representative Tangaras, anthropometric analyses, and system assessments with volunteers.

Anthropometric data were used to assess the weight forces that could be applied to the deadman pedal by the driver population and the deceased driver. Pedal forces were measured with a Mecmesin system. Tests were conducted in an out of service cab on a non-revenue Tangara using 10 participants. Further tests were conducted using a driver training simulator with a sample of 41 participants. The population was chosen to include mainly drivers in the 90 to 120 kg range. Anthropometric measurements of stature, body mass, leg length, and pedal force were made. EMG was used to measure if the driver was applying any muscular effort. Descriptive statistics, regression and chi-squared analyses were conducted on the deadman and pedal force data.

The Tangara's deadman pedal functioned as a lever that could be operated by one or both feet (Figure 41.1). It was housed in an enclosed footwell under the console. Activation and deactivation pedal forces were measured at the distal end of the pedal with reference to the driver. The forces were measured perpendicular to the pedal surface. When the investigation commenced there were no relevant force measurements available. The range of force through which the G7 system was active was 50 to 199 N; if the force was less than 50 or greater than 199 N the system deactivated. This was typical for Tangara pedals. With the force measured in the middle of the pedal the range was from 96 to 370 N. The deceased driver's body mass was 116 kg. Estimation of the weight forces of one leg/foot/shoe and both were in the range 70 to 150 N. Based on anthropometry it was determined that it was possible for a driver of this body mass to rest one or both feet on the pedal without deactivating the system.

Ten 'drivers' participated in the test onboard a non-revenue service Tangara. A series of scenarios were assessed, with foot position, driving posture (including slumped over console) and body mass as the main independent variables. 'Drivers' were asked to operate the pedal with and

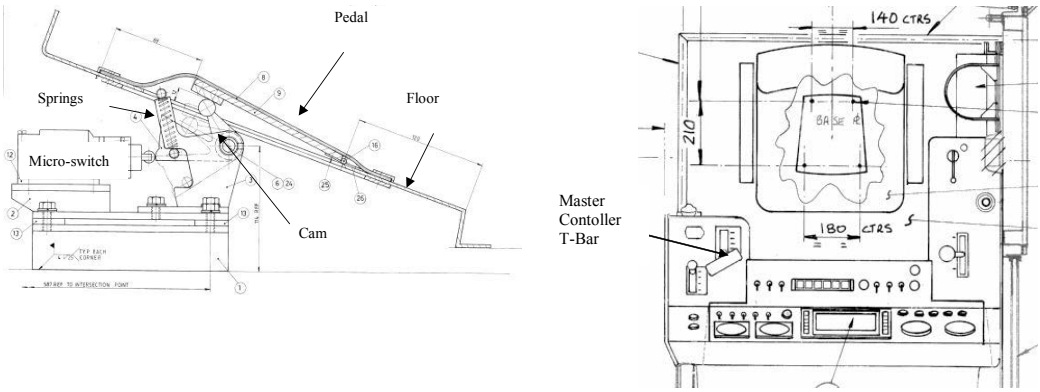


Figure 41.1 Schematic diagram of the deadman system: The pedal showing the floor, pedal, and cam/spring/micro-switch unit (left) and master controller T-Bar (right) shown in the deactivated position

without muscular effort. The mean body mass was 107 kg (SD=14). The footwell enclosure was found to restrict the movement of the legs and therefore, train motion did not appear to dislodge the legs from the pedal. The 'dynamic' tests indicated that all drivers above 105 kg could maintain activation of the deadman pedal with no muscular effort. Drivers also indicated that they found comfortable 'effortless' foot placements with which to operate the pedal. These tests were repeated in a static Tangara cabin simulator with the addition of the independent variable seat position and the measurement of the EMG signal from the lateral head of gastrocnemius. The latter confirmed that lack of muscle activation. Tests on 41 drivers (BM mean = 104 kg, SD=12) confirmed the previous observations and indicated that driver's of body mass as low as 90 kg could maintain deadman activation without effort using both feet. As anticipated foot position on the pedal influenced these results. The tests reinforced strongly the opinion that due to his body weight the deceased driver could have held the deadman's pedal in an activated position. The tests also revealed a range of problems with the pedal design, including the ability for a large proportion of the driver population to operate the pedal without effort and modes of inadvertent circumvention. An analysis of the SRA driver population (BM mean = 90 kg, SD = 16.9) indicated that approximately 18% of the drivers were at a high risk, and 26% at a medium risk, of being able to maintain the deadman activated without effort or while incapacitated (Figure 41.2).

Driver and guard work and personal history

A number of interviews with relatives, work colleagues and friends were conducted to develop a detailed work and personal background history of the deceased train driver. In addition, various documents were analysed including 31 police witness statements, investigator interviews with 8 employees working in the vicinity on the day of the accident, as well as a review of the driver's disciplinary and safety incident records. Observations of train driver performance on the route taken by G7 also assisted in developing a profile of the work load and tasks the deceased driver may have been typically occupied with.

These interviews and document reviews revealed a psychological profile for the driver that was representative of an individual that was a happily married and devoted family man to his two

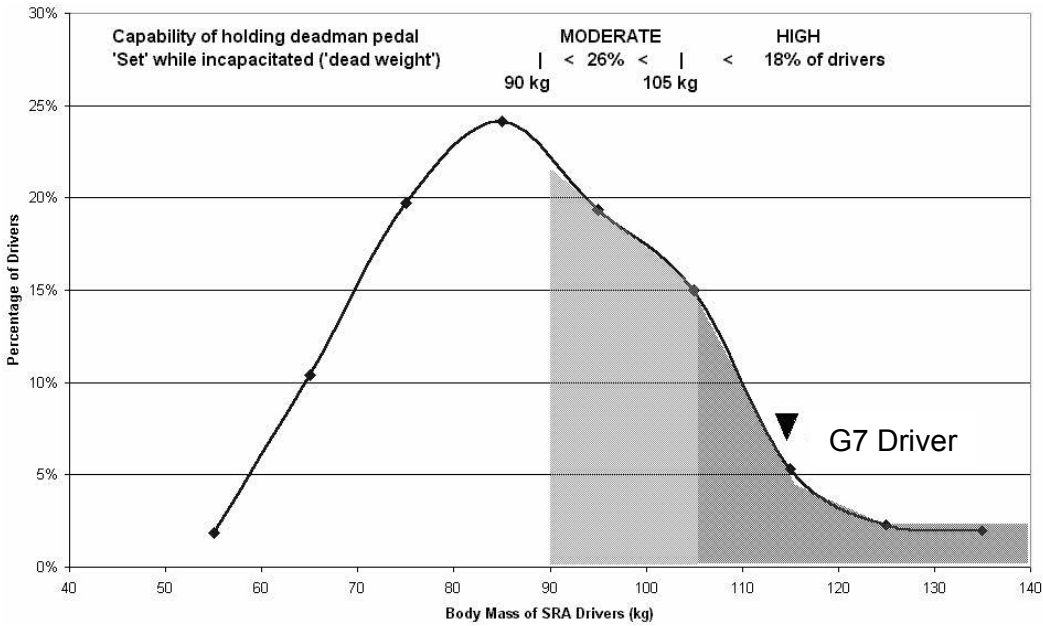


Figure 41.2 Estimate of proportion of SRA Tangara drivers able to hold the deadman's pedal in the activated position while incapacitated

children and grandchildren. There was no evidence that the driver was suffering from a level of stress that may have manifested in the form of distraction or inattention to his driving duties. The driver's activities in the days leading up to the accident revealed a man actively planning for the future, eg. booking a family holiday, and planning for an overseas trip. Friends and family indicated that the driver was not a moody person, rarely got depressed and did not become overly worried about things. He was looking forward to returning to work after a significant period of recreational leave and according to his wife enjoyed his job and rarely complained about management. There was no past history of conflicts or disagreements with work colleagues, and close friends described him as reliable, dependable and honest, well liked by his work mates.

The driver had been employed by the rail organisation for 37 years. He began his career as a trainee engineer and was endorsed as a Class 5 driver operating passenger trains in 1981. According to work colleagues the driver had an excellent knowledge of the Waterfall line from his many years of experience operating in the area.

Employment records indicated a number of safety and disciplinary incidents during his driving career. Between 1967 and 2002 he was involved in a total of seven safety incidents, of which three were considered Major Offences involving Signals Passed at Danger (SPAD). In a career spanning 37 years, the number of SPAD's are not considered unusual or indicative of a "cowboy" driver. A total of 24 disciplinary events were recorded between 1965 and 2002. These were mainly absences or lateness and the majority had been in the 1960s. The sheer number of incidents may also suggest poor management on the part of his train crew manager or problems he experienced with unrealistic shift patterns.

Medical records indicated that the driver consulted medical advice an average of 5 times annually between 1978 and 2002. This number of consultations is typical for most people in the general population. While most of the consultations were of a minor nature, his medical records indicated a problem with cholesterol. In 1994 his cholesterol level was 6.7 (any score greater than

5 for a male is considered high) and continued to rise to 8.6 in August 1999, prompting prescription of medication. In March 2002 he had not taken his prescribed medication for 12 months and it was restarted again combined with advice to reduce his alcohol intake and modify his diet.

His medicals between 1991 and 2001 indicate significant weight gain from 99kg to 116kg in 2001. His Body Mass Index (BMI) taken in 2001 was 34.3, which is considered obese. His weight taken together with his inactivity and lack or regular exercise regime, indicated that he was a significant candidate for coronary heart disease. For example, using figures from his last medical in February 2000, his age, Cholesterol and Systolic Blood Pressure was entered into the Coronary Heart Disease Risk Factor Prediction Chart developed by the American Heart Association. This chart provides a risk output ranging from a low of 1 to a high of 32. The result for the driver for all risk factors is 20. For any score greater than 15 it is recommended that further investigation and monitoring of the patient be conducted. The autopsy examination conducted on the deceased driver confirmed evidence of significant coronary artery disease.

During the course of the investigation the guard, age 39 years, was unable to be interviewed due to ongoing treatment for anxiety and depression. A complete 72-hour history of the guard's activities prior to the accident was unable to be determined due to his unavailability. His employment records indicated that he commenced work in the rail industry in 1982 and became a Guard Class 1 in July 1988. There were no records suggesting a pattern of disciplinary incidents. No details were contained in his personal files regarding safety incidents in which he had been involved in his capacity as a Guard. His personal medical records indicated a history of anxiety and depression focused on an inability to recall events. His medical history, presented an individual with poor coping skills and regular bouts of anxiety and depression. He sustained a number of physical injuries as a result of the Waterfall accident including a wedge compression crush fracture to the T11 vertebra and numerous lacerations and bruises. In summary, the guard's medical history presents an individual characterised by a self-defeating personality, poor coping skills and a pattern of regular bouts of anxiety and depression.

Analysis of the fatigue impairment using FAID™

The influence of fatigue on the driver's performance was examined using a fatigue analysis tool known as FAID (Fatigue Audit InterDyne™) developed by the Centre for Sleep Research at the University of South Australia (Fletcher and Dawson, 1997 and 2001). The Waterfall accident occurred on the driver's second consecutive workday after returning from a period of 5 weeks leave. The drivers' fatigue level on the day of the accident was calculated as a FAID™ score of 37.9 and on the previous day 14.7. These scores are well within the operational tolerance of 80 for a high-risk task as that of operating a train solo. However, FAID™ does not take into account an individuals non-work activities and quality of sleep. As previously indicated in the driver's 72-hour and 14-day history it appears that the relaxing nature of his holidays and family activities in the previous weeks leading up to the accident was an unlikely contributor to cumulative fatigue. According to his wife, the driver on the two nights prior to the accident slept restfully. Furthermore, he was well practised in preparing himself for shiftwork and had a pre-established routine. Based on this information, it appears that driver's quality of sleep in the previous 72 hours was good and because it was only his second shift back after holidays, work related fatigue was an unlikely contributor to the accident.

While the background history on the driver indicated that he was not sleep deprived, research on shift workers indicates that workers returning to shift work after a long period of absence are more likely to make routine errors in the first few days back as a period of adjustment occurs. Rostering records indicated that on his first day back at work, the driver reported for duty over 2.5 hours late. It is quite possible that the driver was still trying to adjust to returning to shiftwork in

his first few days back. This adjustment period may have had detrimental effects on his individual performance such as lack of concentration, difficulty in maintaining vigilance and slower reaction time.

The accident occurred on the guard's 5th consecutive workday in a block of nine early morning shifts. The guard's work roster, including actual hours worked, was analysed using the FAID™ fatigue modelling system. The guards fatigue level on the day of the accident was calculated as 90.8, well above the operational maximum of 80 and equating to a level of impairment equivalent to an alcohol intoxication level of 0.05 BAC, or greater. In the two days prior to the accident his FAID™ scores were 64.49 and 77.92 respectively. It is clear that his three early morning shifts (0323-1220) in a row were having a cumulative effect on his fatigue level.

Had the accident not occurred, and the guard had continued to work his planned roster of an additional four shifts, the FAID score would have peaked at 99 on day three and then dropped to 68 for the last day. A score of 99 is equivalent to a level of fatigue approximating 250% of the level of fatigue of the average working week. There is a high probability that the guard's performance may have been influenced by the effects of work related fatigue.

Discussions with drivers and guards revealed that the typical workload of Guards in the stations leading up to Waterfall from Central is relatively high, due to a greater frequency of stops and higher passenger loads in the urban areas. However, between Waterfall and Helensburgh the workload decreased and guards typically report that this is the first time that they are able to relax. Given the consecutive nature of early morning shifts worked by the guard, it is possible that he experienced a micro sleep and therefore may have been slow to respond to the unusual speed of the train as it approached the 60-kmh speed board.

Safety culture and train crew communications

In addition to the interviews described above, structured interviews were held with a small number of guards and drivers at two main stations. The objectives of the interviews were to identify: procedures and practice related to emergency braking, patterns of guard-driver communication and team work, and deadman system function. All guards and drivers participated voluntarily. The following is a summary of the findings:

- Ergonomics of the driver console were not considered to be particularly good. Both the hand control deadman lever and foot pedal were considered to be generally awkward and put a strain on the body of the users (arms, wrist, legs, ankles). Because of the strain of maintaining the required fixed postures some drivers resorted to jamming a foot between the deadman pedal and footwell heater.
- There were known ways of circumventing the deadman controls, such as with the foot pedal as noted above, jamming the driver's red flag between the pedal and console and possibly the use of rubber bands to hold the deadman's twist grip on the master control in place.
- The communications system in the driver's and guard's cabins was considered to be very poor. A major flaw was that there was no auditory signal to alert the driver to the guard initiating contact. Hence if a driver was asleep – he would not be woken up – or notice the guard's communication attempt.
- Communication between the driver and guard generally was minimal, other than for routine matters. Significantly there was a major cultural/hierarchal barrier to a guard 'pulling the tail' (emergency brake cock) on a driver. Although it occurred, it would for a guard to question a driver regarding the train's speed.

The cultural or hierarchal barrier, otherwise known as an "authority gradient", was explored in more detail. Evidence given in train crew statements throughout the inquiry, as well as anecdotal evidence and informal discussions, indicated that there was some cultural animosity between drivers and guards, which may have discouraged guards from speaking up about safety issues, or more importantly taking intervening action, such as applying the emergency brake. Specifically, junior guards revealed a reluctance to challenge drivers on various issues. This cultural animosity is not surprising given that the two members of the "team" operated largely in isolation at opposite ends of the train and their communication was limited to bell signals or occasional voice intercom. Drivers and guards often completed a journey without any face to face contact or voice communication and sometimes may not know who was at the other end of the train, particularly if there was a guard change mid-journey. Indeed there was no requirement for contact if personnel changes occurred during a shift.

The guard on the Waterfall train had an opportunity to take action to slow the train before it entered the curve and derailed. While the guards slow recognition of the evolving risk and associated fatigue level contributed to his inaction, at a more systemic level, guards were simply not equipped or trained in critical decision making skills or in undertaking well practised emergency action as a result of clearly defined risk events. In addition, at a cultural level there was a fundamental communication deficiency that discouraged guards from challenging a driver on his actions.

Medical standards

A review of the medical standards was undertaken as they related to the cause of the accident. This point is dealt with in detail in another contribution in this book. The driver of the Waterfall train had a history of cholesterol problems and obesity, which was not identified through the periodic medical examination regime implemented at the time. Had more rigorous standards been employed, and better rail industry familiarisation been provided to authorised medical practitioners, it is highly probable that the driver's risk factors would have been identified and his driving certificate withdrawn until his health was managed to acceptable levels.

Discussion

The investigations into the causes of the Waterfall train accident revealed the critical role of human factors. These factors were not errors of commission, but rather an error of omission by the guard in failing to brake the train, and more importantly a failure to utilise a system design approach for train operation, in particular train safety, with adequate consideration for human factors. The latter resulted in latent conditions that were exposed by the incapacitation of the G7's driver and inaction of the guard. The guard's error has its antecedents in system flaws, procedures, communication, training, selection and culture.

Human factors investigations were able to provide valuable input into the inquiry and assist in the identification of the primary causes of the accident.

The Waterfall accident was not the first time in Australia that a major train crash had occurred due to driver incapacitation and deadman failure; the Southern Aurora crashed into a freight train in Violet Town in 1969 and the more recent 2001 accident in Footscray are two examples (ATSB, 2001).

Acknowledgements

The authors were engaged as consultants to the Special Commission of Inquiry into the Waterfall Rail Accident. The chapter reflects the author's opinions not those of the Inquiry. The authors acknowledge the exemplary work undertaken by many involved in the Inquiry and the specific support provided by Justice McInerney, his legal team, and Mr. Bob Lauby.

References

ATSB, (2001). Australian Transport Safety Bureau, Rail Investigation Report, Footscray, Victoria, 5th June 2001. Collision between suburban electric passenger train 6369 and the empty express electric train 6371, November 2001, ISBN 1 877071 01 3

Fletcher, A and Dawson, D. (1997). A predictive model of work-related fatigue based on hours of work. Journal of Occupational Health and Safety – Australia and New Zealand, 13(5): 471-485.

Fletcher, A. and Dawson, D. (2001). Field-based validations of a work-related fatigue model based on hours of work. Transportation Research, Part F4: 75-88.

McInerney, P.A., (2005). Special Commission of Inquiry into the Waterfall Rail Accident Final Report Volumes 1 and 2, January 2005.

Chapter 42

The Role of Communication Errors in Railway Incident Causation

Paul Shanahan, Dik Gregory, Mike Shannon and Huw Gibson

Introduction

This study focuses on the verbal communications of railway personnel in safety-critical roles. These include primarily signallers, drivers, despatch staff at stations, and key workers on the track, notably the jobs of Controller of Site Safety (COSS), Engineering Supervisor (ES) and Person In Charge of Possession (PICOP).

The objective of this study was to achieve greater understanding of the extent to which *verbal* communication errors and failures involving these roles contribute to the occurrence of incidents and accidents on the railways, what kinds of errors and failures are most commonly involved, why do they occur, and what could be done to minimise the risk of their occurrence. The communication errors of interest here are those that are directly concerned in causing incidents, in other words errors that occur in the actual work situation. The incidents included within the scope of this study are those that resulted in a Formal Investigation. The classes of incident of major concern were SPADs and Possession Irregularities (more generally track work), although some consideration was also given to other types of incident.

The UK railway industry has long been aware of these dangers (Murphy, 2001; Amey Vectra, 2003; Gibson, 2002; Jones and Hickey, 2004). The Rule Book gives extensive coverage of the procedures for safety critical communication. Measures are in place to record, monitor, assess and improve safety critical communication performance. While there is evidence from communications monitoring and from incident investigations that the standards of communication are often not as high as they should be, this assessment is often based on failures to adhere to the laid-down rules and procedures.

It is less clear how much of a problem communications actually are in the sense of how much they contribute to safety-related incidents and accidents. In particular, if it is the norm for staff to make communication errors and to violate communications rules and procedures, then it is to be expected that the majority of incident investigations will discover such errors and violations. But this does not indicate that these communications lapses have in any way contributed to causing these incidents. Indeed since such incidents are comparatively infrequent and communications errors and violations appear to be endemic, for most of the time poor communications performance does not cause incidents or accidents.

This study sought to improve understanding of how much communications problems contribute to causing incidents and under what circumstances poor communications – which at most times do not appear to lead to undesired outcomes – sometimes do have negative consequences. This improved understanding can help ensure that management attention and effort, training activities, rule and procedure writing are all directed to the most essential aspects of communication performance.

The study method

An initial sample of 391 reports was taken at random from the RSSB library of Formal Investigation reports. A preliminary filtering of these reports was undertaken based on a review of the investigation recommendations to see if these indicated the probable causal role of communication problems. Following this review 123 reports were selected for more detailed analysis. This analysis process drew on a number of sources (those already mentioned and Shorrock, 2002; Reason, 1997). The analysis process is shown schematically in Figure 42.1. As a result of the detailed analysis a further 49 investigations were found not to involve communication as a causal factor. In most cases this was due to the fact that, although the investigation had uncovered communication irregularities, these irregularities had not contributed to the causes of the incident. The analysis of the reports was carried out by three members of the team. All reports were reviewed by at least two members of the team, and any differences of interpretation were explored until a common view was reached.

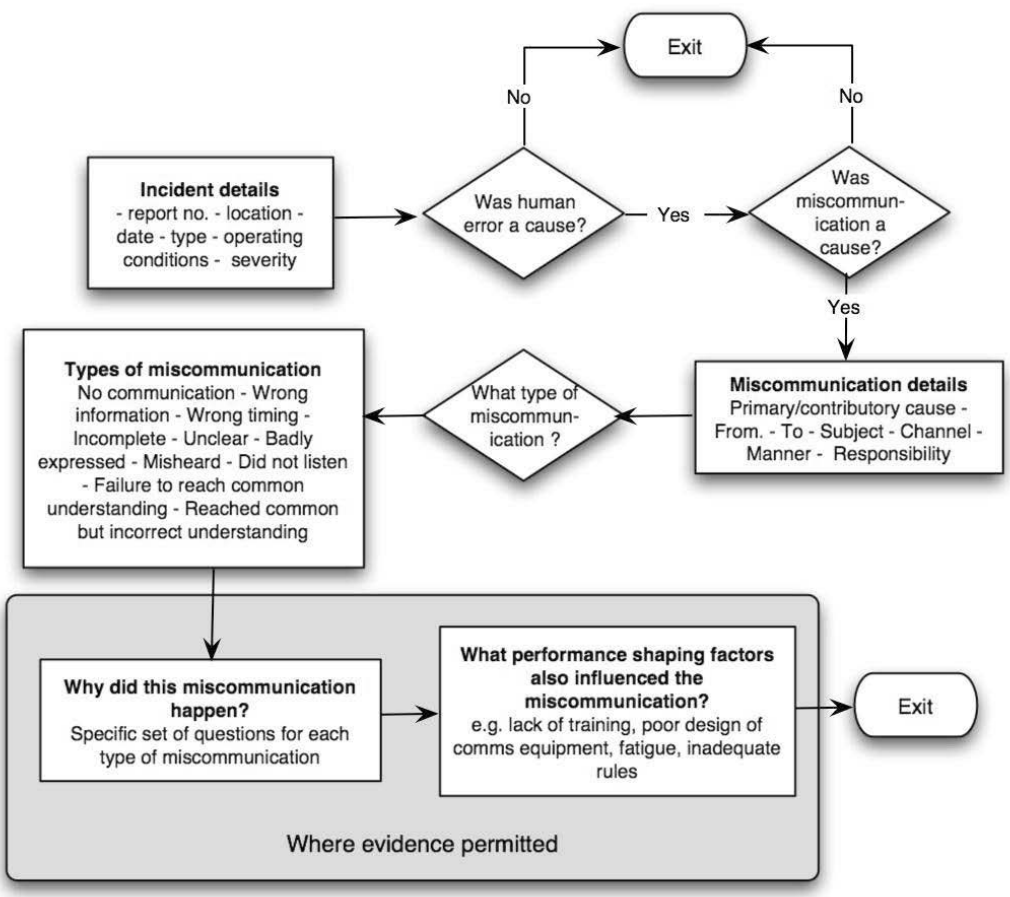


Figure 42.1 The process for analysing investigation reports

Results

The key result from this study is the proportion of incidents of various types caused – at least in part – by communication error or miscommunication. Table 42.1 shows for each type of incident the percentage in which communication problems played a major causal role, played a minor role, or did not play a causal role at all. Following the selection and initial analysis process, a total of 73 examples of miscommunications that contributed to the causing of incidents were identified for detailed analysis. The main points arising from the analysis are summarised here.

Operating conditions at time incident occurred: Two thirds of the miscommunications occurred during normal operating conditions and one quarter occurred during degraded conditions. Only 8% occurred in abnormal conditions and none in emergency conditions.

Table 42.1 Percentage of incidents in which communication problems played a causal role

Type of incident (no.)	Major role	Minor role	No role
Track work (72)	37.5	16.6	45.8
Crossing (30)	13.3	10.0	76.7
SPAD (193)	11.4	24.4	64.2
Derailment (56)	5.4	3.6	91.1
Miscellaneous (40)	5.0	10.0	85.0
TOTAL	14.8	17.6	67.8

Seriousness: Two thirds of the incidents resulted in no damage or injuries. 29% involved minor damage or injury, and only 4 of the 73 cases involved serious damage or injury. The great majority of incidents incurred limited disruption, although this is not usually specified in the incident reports.

Role of human error: 88% of miscommunications involved human error as the immediate cause, with the remainder involving human error as the underlying cause.

Role of communication error: In 60% of cases, miscommunication was the main cause of the incident. In 29% it played a direct role along with other factors. In 11% of cases communication contributed in a minor way.

Subject of communication: For nearly half the cases, the main subject of the miscommunication was a train movement. The second most common subject was the track state or activity on the track (15%) and the third most common was taking or giving up possessions (11%).

Who initiated the communication (or should have done): Three quarters of the communications were initiated by three roles – COSS (30%), Driver (25%) and Signalman (20%).

Who was the communication addressed to: 87% of the communications were addressed to the same three roles, although in a different order – Driver (37%), Signalman (32%) and COSS (18%).

The most common pairs involved in communications: Signalman/Driver (22 cases), COSS/Track Worker/Engineer (10 cases), and COSS/Signalman (8 cases).

Communication channel: Most commonly communications were face-to-face (27%); by mobile phone (18%); SPT (18%); and CSR (8%).

Who responsible for error: The person who initiated the communication (or should have done) – 58%; both parties – 21%; and the person addressed (or should have been) – 12%.

How did the error arise: The greatest number of errors were due to a failure to communicate at all (26%), then communicating wrong information (21%), and thirdly communicating incomplete information (14%). The following paragraphs elaborate the reasons for these miscommunications.

Why communication not initiated: Of the 19 cases of a communication not being initiated when it should have been, 4 were the results of decisions not to communicate, 3 were the result of not realising a communication should have been initiated, 3 were the result of being distracted by other tasks, and 2 were communications mistakenly addressed to the wrong person.

Why wrong information communicated: Of the 15 cases of wrong information being communicated, 6 were due to the person misreading the situation, 5 to the person confusing two items of information, and 3 to the source used for the information itself being wrong.

Why information incomplete: Of the 10 examples of a person communicating incomplete information, 7 were due to forgetting, 2 to a lack of knowledge of what to communicate, and one to being distracted during the communication.

Slips, lapses, mistakes and violations: In total it was possible to make some assessment of the underlying psychological mechanism involved in a miscommunication for 49 of the cases. These were classified according to the scheme of Reason (1997) as follows: Slips: 24 (49%); Lapses: 8 (16%); Knowledge-based mistakes: 10 (20%); Rule-based mistakes: 0; Violations: 4 (8%); Factor outside person's control: 3 (6%).

Discussion

The evidence from this study indicates that approximately one third of incidents on the railways are caused to some degree by miscommunication. Of those incidents in which miscommunication did play a part, the miscommunication was the main cause of the incident 60% of the time. The sample of incidents examined in this study was overwhelmingly drawn from SPADs and possession irregularities with a smaller representation of derailments, level crossing incidents, incidents at stations and a mixture of various other types of incident. Our analysis of all investigation reports received by RSSB in the first six months of 2004 suggests that the sample used here is reasonably representative of incidents in general.

These results indicate that miscommunication is one of the major causes of incidents, and should be considered one of the top priorities within the area of Human Factors in both the operational and research arenas. More generally, miscommunication should also be accorded a similar priority to other causes of rail incidents.

What kinds of incident are caused by miscommunications?

Well over half of incidents involving track work are caused by miscommunication. This compares with one third of SPADs and one quarter of level crossing incidents. Derailments are rarely caused by communication errors. If only incidents in which miscommunication was the main cause are considered, then the difference between incident types becomes even more marked. For track work, the rate of occurrence of incidents mainly caused by miscommunication is over three times that of SPADs and nearly seven times that of derailments.

The vulnerability of track work to miscommunication reflects several factors, especially the sheer number of participants involved. Whereas communications to do with SPADs typically only involve two parties, driver and signaller, those to do with track work can involve a variety of workers. To the numbers involved can be added the additional factors of the difficulty of working at night, in bad weather and in noisy working environments.

What kinds of miscommunication are involved?

There are two especially striking aspects of the findings presented here. First is the great variety of miscommunications found in the investigation reports. Second, in spite of the range of miscommunication types that did cause incidents, not one incident was found where a failure of safety critical communications procedure (use of phonetic alphabet, stating identity, location, etc) could be judged to have been the cause, direct or indirect.[1]

This is not to say, that there was no evidence of failures to adhere to the safety critical procedures. It is almost routine for investigations to note the poor standard of communication performance in this regard and to recommend that further training is required. However, there was no evidence that any of the incidents were caused by these frequent infringements.

Returning to the miscommunications that did cause incidents, then the three main types of problem were failure to communicate at all and communicating incorrect or incomplete information. These findings can be compared with results from the aviation sector for Pilot/Controller communication problems [7]. These authors found that "absent" communications were by a large margin the most common problem. Absent is defined as "failure to originate or transmit a required or appropriate message". This corresponds closely with the most common category in the present study, "failure to communicate at all". It appears that simply not communicating when one should is a widespread and serious problem.

Another view of kinds of miscommunication is provided by the breakdown into slips, lapses, mistakes and violations. Slips are the most common type of miscommunication by far. Slips can have serious consequences, for example, when a signaller confuses two similar train head codes so that s/he is talking to one train thinking s/he is talking to a different train. But perhaps more significant is the high number of knowledge-based mistakes. In comparison to slips, these are much more difficult to detect and correct.

Why do miscommunications happen?

The analysis revealed two main clusters of features of the communication errors. The strongest cluster concerns people making wrong assumptions about others' intentions, situations or

1 The high incidence of slips noted above largely concerned slips in content. It is arguable that more reliable use of readback would have identified some of these slips. Since only a few of the investigation reports contained transcripts of the communications, it is not possible to draw any definite conclusions on this point.

knowledge. This indicates a tendency to "get other people wrong". It also ties in with one of the main reasons why people communicated wrong information, namely, a tendency to misread the situation, in other words "to get the situation wrong".

In recent years much attention has been paid to team or shared mental models and the role these models play in helping people to work together on tightly coupled tasks. One of the key functions of communications is to enable co-workers to build up and maintain shared models of the situation (including each other). There is a danger of a vicious circle developing here. Poor communications mean that different people's mental models get out of alignment. When their models are out of alignment they cannot judge accurately when and what to communicate. This in turn means that their individual mental models become even more divergent.

The second cluster concerns people's adherence to the rules. As we have noted earlier, these are not the rules about communications procedure, but rather about whether the general operational rules apply (for example, when a train is authorised to pass a signal at danger), whether the person can do "better" by not following the rules, and so on. Although the evidence from this study does not really give enough information, one could speculate that this difficulty with the rules is actually connected with the previous cluster to do with failures in reading the world and others. Going beyond the issue of shared mental models, the deeper issue here may be to do with the processes of "organisational sense-making" (Weick, 1995) – how workers in a situation individually and collectively interpret and make sense of the world in which they are operating and co-operating. This may be a fruitful theme for future research.

In a great many of the cases of miscommunication, these were not isolated failures or errors, but occurred as part of a complex sequence of mishaps involving technical or planning failures, human errors that were not miscommunications, as well as miscommunications of different types made by different participants.

What can be done to improve communications on the railways?

It is almost impossible to draw a neat boundary separating communication from the other areas of work activity. To deal with miscommunication effectively demands a model of communication that recognises both the breadth of the area but also the many interconnected layers at work (Reilly et al; Pickering and Garrod, 2004; Gibson et al, 2006) – see Figure 42.2.

For two or more people to communicate successfully each must have a capability which matches the other at a number of functional layers. Each upper layer is dependent on the layers

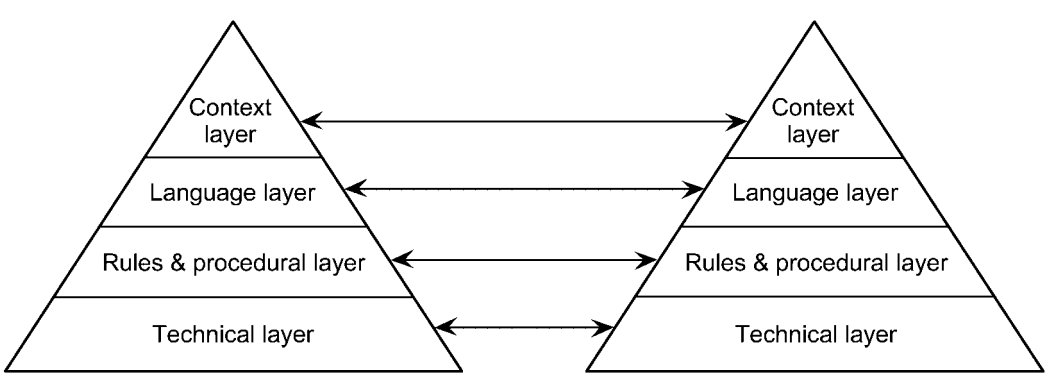


Figure 42.2 A layered model of the communication domain

below. The bottom layer (assuming communication is via some technology-based channel) is the *technical* layer. This assumes that each person has access to the equipment needed to communicate with the other. Implicit within this layer is the technical know-how that the person needs in order to operate the communications equipment. Factors affecting this layer include the training and experience the person has had in using the equipment, the ergonomics of accessing and interacting with the equipment, and environmental factors affecting use of the communications equipment (ambient noise, rain, etc).

Assuming that the technical layer is working satisfactorily, then the *rules and procedural* layer comes into consideration. The rules and procedures govern the to whom, what, when and how of the communications. The clarity, relevance and applicability of the rules and procedures are critical factors in effective communication.

Above the rules and procedural layer comes the *language* layer. This layer is concerned with how words and phrases are used together to create meaning. A communication may follow the rules totally correctly but be meaningless. A well known example is the sentence "Colourless green ideas sleep furiously". This sentence conforms to the rules of English grammar but most people would have trouble in attaching any meaning to it. The uppermost layer in the model is the *context* layer. It is essentially an extension to the language layer and is concerned with the way context can imply extra meaning to words or phrases, and so add to their value or content. In the railways considerable effort and resources are devoted to the first two layers (technical know-how, rules and procedures) that are largely skill-based. Much less attention is paid to building up the top two layers (language and context) that are largely knowledge-based. This knowledge which is essential to being able to read and exploit the context of communications is primarily acquired through informal, on-the-job experience.

This may have worked when most workers spent their entire working lives on the railways and spent years accumulating knowledge. But this is no longer the rule. Signallers are increasingly recruited "off-the-street" and are often placed in demanding workplaces such as large power boxes without having served extended "apprenticeships" in simpler mechanical boxes. Track workers may have very little experience of the railways and have a high turnover. For many workers English is not their first language and the existence of a large body of shared cultural knowledge can no longer be taken for granted. All this suggests that it would be worthwhile investigating further the kinds of career paths, training progression, and on-the-job experiences which are most likely to further the acquisition of the background knowledge that underpins *meaning*.

Lastly, we mention one specific area where training could make a major contribution. This is in what might be called "meta-communication skills". Much present training focuses – understandably - on the "mechanics" of communication. More training could be given in helping staff monitor and assess both their own and others communication performance. For example Fischer and Orasanu (2000) have done some preliminary work on "error challenging strategies", the ways in which airline pilots and co-pilots call attention to problems and get action on them from other crew members.

This is connected with the area known as critical thinking (Fisher, 2001), successfully tried in a number of areas, notably military command decision making (Cohen et al, 1998). The aim here is to become aware of one's own mental processes, assessing the quality of information and knowledge one is using, identifying and checking assumptions and accuracy, validating any logical inferences one is making, and so on. By challenging both one's own and others communications, quality can be improved and errors detected and rectified in a more timely manner.

Conclusions and recommendations

This report has indicated that about one third of all incidents on the railways are caused – at least in part – by miscommunications. Track work is particularly vulnerable to miscommunication, with well over half all incidents involving a miscommunication component. An analysis of the nature of the miscommunications involved reveals that failures to communicate at all and communication of incorrect information are the most common types of error. Simple procedural errors, whilst common, did not cause any of the incidents. The roles most commonly involved are signaller, driver and COSS.

Nearly one half of the errors were simple slips (e.g. confusing headcodes or misdialling) but a significant proportion involved knowledge-based errors. In a number of cases staff did not follow the rules for a variety of reasons: not knowing the rules well enough, judging that the rules did not apply, or simply ignoring the rules. The analysis made clear that miscommunications cannot be looked at in isolation. While some errors are "pure" communication problems, in many other cases the miscommunications are manifestations of other working problems.

It is recommended that further work is carried out to identify how career paths, training progression and on-the-job experiences could be better exploited to promote higher levels of communication performance. It is also recommended that the feasibility be explored of designing training to assist in developing and adopting strategies for detecting, challenging and getting action from others on errors in communications or actions.

Acknowledgements

We would like to acknowledge the contributions of the Safety Critical Communication Focus Group, John Peters (RSSB), Philippa Murphy (Network Rail) and Huw Bridges (London Lines).

References

Amey Vectra Ltd, *Communication risk analysis: Existing data review*, Report for RSSB, May 2003.

Cohen, M.S. et al. Critical thinking skills in tactical decision making. In Cannon-Bowers J.A. and Salas, E. (Eds) *Making decisions under stress: Implications for Individual and Team Training*. American Psychological Association, 1998.

Fischer, U. and Orasanu, J. *Error-challenging strategies: Their role in preventing and correcting errors.* In Proceedings of the International Ergonomics Association, 14th Triennial Congress and Human Factors and Ergonomics Society, 44th Annual Meeting in San Diego, California, August 2000.

Fisher, A. *Critical Thinking: An Introduction*. Cambridge University Press, 2001.

Gibson, W.H. *Core-Data Project: Analysis of transcripts of communication for the protection of the line*. Railway Safety Report, Industrial Ergonomics Group, University of Birmingham, 2002.

Gibson, W.H. Megaw, E.D., Young, M and Lowe, E. (2006). *A Taxonomy of Human Communication Errors and Application to Railway Track Maintenance*, in Cognition, Technology and Work (In Press)

Grayson, R.L. and Billings, C.E. *Information transfer between air traffic control and aircraft: Communication problems in Flight Operations*. In *Information Transfer Problems in the Aviation System*. NASA Technical Paper 1875, September 1981.

Jones, S. and Hickey, S. *T014- Improving Driver/Signaller Safety Critical Communications*, AEA Technology, report for RSSB, May 2004.

Murphy, P. *The role of communications in accidents and incidents during rail possessions*. In Harris, D. (Ed) *Engineering psychology and cognitive ergonomics, Volume 5, Aerospace and Transportation Systems*. Ashgate, 2001.

Pickering, M.J. and Garrod, S. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, 2004.

Reason, J. *Managing the Risks of Organizational Accidents*, Ashgate, 1997.

Reilly, R., Ferrari, G, and Prodanof, I. *Framework for a model of dialogue*. http://www.acl.ldc.upenn.edu/C/C88/C88-2114.pdf

Shorrock, S.T. *Error classification for safety management: Finding the right approach*, Proceedings of Investigation and Reporting Accidents, 2002.

Weick, K. *Sensemaking in organisations*. Sage, 1995.

This page intentionally left blank

PART XII
Human Error and Human Reliability

This page intentionally left blank

Chapter 43

Assessing the Potential for Human Error in Level 4 Rail Vehicle Maintenance

Steve Mason

Background

The importance of maintenance has long been recognised, but until recently remarkably little attention has been directed to: designing equipment and machines that can be maintained safely and speedily; providing procedures that are easy to follow; and addressing factors that can detract from strict compliance with these procedures.

'Level 4 maintenance' involves the regular inspection of vehicles and replacement of consumable parts and is performed during the relatively short time periods when the vehicles are not in normal use, for example, in the period between the last run of the evening and the first run the following day. Few accidents have been linked with failures in the level 4 maintenance operations, however the potential exists, especially if maintainers repeat common critical errors.

Apart from the potential safety issues, maintenance costs can be prohibitive if this work cannot be undertaken within the short available time and vehicles are delayed from returning to service. Cost can also be prohibitive if a routine maintenance deficiency results in the vehicle subsequently breaking down. A number of surveys have been conducted that, in part, have been looking for aspects where the level 4 maintenance can be improved. If poor maintainability features are accepted as inevitable, and if those human factors issues that detract from reliable performance are overlooked, then the overall costs of ownership and safety issues could arise.

HSEC have had an involvement at many rail maintenance depots for a number of clients as part of reviews on: depot performance, error potential, maintenance instructions, safety culture and incident investigations. This work has been conducted at city and rural depots and those maintaining their own fleets and those maintaining third party vehicles. Findings from individual depots remain strictly confidential. The purpose here is to distil the essence from these different investigations so as to present an overview of the main problems that can occur at depots. Although all of these reviews have aimed at helping depots address any problems or potential concerns, this chapter is strictly limited to raising an awareness of the issues affecting level 4 maintenance. It is hoped that this can show where management processes such as audits can be better focussed to consider the wider potential for errors.

Sample findings

There were clearly large differences between depots for a range of issues that could affect human performance. Many depots showed very positive results. This chapter presents some of the more critical problems. Individual problems were generally present at a minority of rail maintenance depots.

Original vehicle and depot design features

Original vehicle design features were primarily associated with problems of accessibility to inspect, use tools and replace some components. There were concerns also with the layout and complexity of some vehicles; designs that allow the wrong parts to be fitted; and the lack of consistency of some features between the different vehicle types being maintained. Problems can arise if maintenance procedures assume a level of access to vehicles that is not possible with their facilities. For example, some inspection pits have raised sections only to one side of the inspection pits. Others have both rails fully raised.

Supply chain

Errors made in the supply chain could become critical if a wrong part was issued and then fitted without question. Most depots recognise this. Errors can be made by those supplying/labelling parts being shipped into the stores and by fitters requesting the wrong parts. There are usually procedures to minimise this risk however more could perhaps be done to improve the quality assurance for the suppliers of the most safety-critical components. 'Goods receiving' often relies on experienced staff noticing any differences in parts. A system can be used whereby at least one part from each incoming batch is unpacked and checked against existing stock. Although the system essentially relies on the competence of the suppliers, when new components have to be sourced staff are aware of the potential for quality problems and much more attention is given to those parts.

Major difficulties were identified with obtaining some spare parts. This potential problem could be related to instances where there are no 'parts books' and where the exact parts numbers cannot be sourced from drawings of sub-assemblies. Fitters ask for a part by name or by part number. At some depots, stores personnel are trained to ensure parts numbers are obtained (perhaps from the Team Leaders) where there is any doubt. They are trained not to rely on the fitter's description of the part. There is still a potential problem and the current process is accepted as not being totally bullet proof. Nevertheless, consideration of means to further reduce the risks of supply error may be worthwhile.

There are also difficulties in finding suppliers of some obsolete parts. This has led to problems such as a component not being available and a modified alternative being made and fitted. Although those involved in making a replacement had the best intentions, the modified part was incorrect for service and a mechanical failure occurred when the modified part was in service.

A different issue is when a kit of parts to be fitted is not supplied. Some dampers are delivered to the stores with new split pins already fitted but others are not – even though they are both supplied by the same company! Two dampers were seen being fitted to a vehicle and neither was supplied with new split pins. In both cases the fitter had to return to the stores to obtain these parts. While this inconsistency in supply could be addressed if the fitter noticed the shortfall at the stores counter, the inconvenience and time involved could create a motive for the old pin to be reused. A greater use of supplying parts in a fitting pack (where all the necessary parts for a job were packaged) could be beneficial.

Finally there appear to be some problems of lack of guarantee of spares availability for the life of a vehicle. Examples have arisen whereby the stores have discovered that there were literally no further units in the world! If stores are not informed they cannot make alternative arrangements to seek continuity of supply prior to such parts being urgently needed. Vehicles are then robbed of parts in order to keep other vehicles running. This is both inefficient and needlessly time consuming in an environment when time is often very limited. A better integration of design and long-term supply capability is indicated.

Modifications

Modification programmes are common. Unfortunately, depots may not be given supplies to repair any faulty modified units until very late into the programme. If modified parts fail then trains become unusable if spares have not been allocated in advance of the full fleet being modified. Modified units are then 'robbed' to service other trains. It would appear that modification contracts do not always require the supplies of replacement parts and the necessary supporting information.

Recent modifications had been made for aesthetic benefits of the passengers. These can adversely affect maintainability. One example modification included replacing old seating with 'high back' modern seats, including those seats adjacent to the saloon end bulkheads. There is a maintenance access panel for electrical equipment in this bulkhead that had to be shortened to clear the top of the new higher seats. As a consequence, access to the bottom part of this panel has been significantly impaired and would require several hours of additional work to remove the seating to allow full access if this was necessary. When the depot only has a vehicle for a few hours this additional work could easily result in the vehicle missing its scheduled return to service. While access to replace units low down in this panel may be infrequent, it is understood that access is needed for certain fault finding work.

Although modifications are fairly common, some maintenance staff and stores personnel were found to be unaware of some modifications and consequently new parts had not been stocked. A lack of integration of the maintenance function, staff training, supply chain awareness, and updated procedures all seemed to combine to reduce the potential benefits from refurbishment/modification programs.

Training issues

Actions were seen, or reported, that suggest there are basic problems with the competence of some fitters. Observations showed several instances where EMU vehicles were being chocked on the bogies with the pick up shoes and on driven axles – both of which are against best practice. It is quite possible that the majority/all staff concerned may be genuinely confused as some 'class specific guidelines' may have been misunderstood to be applicable to all vehicles. Another example was of 'Yaw Dampers' being found fitted upside down such that they would not work.

Problems could be more frequent when modifications require fitters to possess new skills. For example, modifications included fitting CCTV in the passenger areas. It was understood that although these refurbishments had nearly been completed on the full fleet, nobody seemed to have been trained on the maintenance procedures for the CCTV system.

Paper-based maintenance procedures

Ambiguity in some procedures is an area of concern. For example, one procedure appeared to specify that only a simple visual inspection was required to check for leaks on dampers, but some staff thought that the check required the dampers to be physically compressed sideways to check for wear. This manual check was said to occasionally find suspect dampers. This reinforced the benefit of performing a physical check, despite the procedures only requiring a 'hands-off' visual check.

Similar problems occur over of how oils and greases are referred to. At one depot, there were 6 types of grease and 4 types of oil (for engine, fan, gearbox, etc) in general use. Some trade names are very similar. A lubricant may be referred to by its catalogue number – eg '62 – 1234' but this bears no logical relationship to the trade names such as: 'Esso Beacon', 'Shell Alvina', etc. Errors

could result in bearings collapsing with subsequent safety critical consequences. Some depots use different grease guns/nipples for different types of grease to make it physically impossible to apply the wrong grease.

Ambiguity also exists at one depot where a vehicle is said to have passed the horn check if it makes anything resembling a horn sound – irrespective of its loudness and effectiveness. Shop floor personnel do have the discretion to change a horn if they don't like it, but there is no set pass/fail procedure.

Another concern related to how certain components could be visually checked "for completeness, signs of fracture, loose components, or any other deterioration" when they were barely visible, heavily contaminated and cannot be steam cleaned due to their proximity to other key components. Low inspection pit light level also compromised visual inspection.

Procedures often fail to provide sufficient pass/fail criteria. Many rely on the qualitative or subjective judgement of the maintainer as to whether something is just acceptable or just failed. While there will always be some areas where such judgement is needed, it often seems that these pass/fail criteria omissions are primarily to makes the procedures easier to write. Where the desired level of reliability cannot reasonably be achieved using inexperienced qualitative judgements, then procedures need to adopt quantitative measures to more objectively specify pass/fail criteria in the procedures.

Even with new procedures, problems continue to occur that are likely to cause confusion or errors, especially to the newer members of the maintenance workforce who may expect and rely on these documents being fully accurate and easy to understand. This suggests that procedures are not adequately being checked for accuracy and practicality before they are signed off.

Computerised maintenance procedures

There seems to be a move towards computer based maintenance instructions that utilise HTML and hyperlinked documents. While these offer potential benefits for enhanced functionality, they can actually cause problems at those depots where computer terminal access is limited, or where paper-based print outs remain the preferred method of working.

Perhaps a greater emphasis could be given to provision of fault diagnosis information and guidance, especially on new or complex engineering systems. Vehicles appear to be getting more complex and several instances were found where even experienced staff were having difficulties in diagnosing some faults. More consideration therefore appears to be needed for technical authors to produce more and better diagnostic supporting information. Extending this to providing some basic driver diagnostic information may also facilitate maintenance engineers if it enables drivers to provide better feedback on complex problems.

IT systems

There is evidence that the lack of compatibility of the various software packages and databases employed to attribute delays to specific systems (to enable prioritisation of maintenance, presumably) is causing an excessive workload for supervisors and managers. This is inefficient, and reduces the effectiveness of management.

At one depot it was necessary for managers to access three different computer systems. This process took about 10 minutes for each incident to be processed. This occupied a considerable portion of the responsible managers' time (reportedly 20%), and reduced his capacity to deal with other issues. The delays are largely due to the (lack of) speed and the complexity of different databases that need to be consulted to check various defect reports.

A major problem is that the dates of some problems are usually not given. Faults can be received one week after being reported by another route. As there is no date given it cannot be assumed to have been rectified and must be investigated, however, many faults are found to have already been resolved upon investigation, thereby wasting maintenance personnel's time.

At another depot there was a problem regarding a lack of commonality in the terms of reference employed by some IT systems, as some systems refer to Coach Numbers whereas the depot refer to Unit numbers.

A number of difficulties using a new data logging system were identified when speaking to those who input raw data. One related to the pull down menus for fault causes/symptoms. These were clearly not mutually exclusive to the point that a range of single entries could be entered for certain faults. In practice the first reasonable entry is often used. Problems will arise when people subsequently interrogate the system for the frequency of each cause/symptom, as any fault could have essentially a near-random cause logged. The potential value of a subsequent trend analysis would be severely restricted. These problems are recognised and the depot has an ongoing programme to refine the system to ensure that it fully meets their needs.

Communications

A common complaint at some depots was that communication tended solely to be top-down. This was not, however, observed in practice. Observations at some meetings confirmed a very positive relationship between the Team Leaders and staff and communications were clearly two way. Despite this, the number of comments regarding poor and one-way communication suggests that there is probably some substance to this, but this does not seem related to an organisational problem, more likely the result of actions and style of individual managers.

Communications at some depots may over-rely on verbal communication alone, whereas at other depots there was a second level of documentation to support this. One specific concern was an apparent over-reliance on the notice boards to convey key items of safety information (eg correct use of scotches). Comments were made that these could be covered up by other notices, taken down, or simply not seen and that people need to be made aware (and coached) of new information and sign to the effect that they are aware of the changes. The essence of people's views is that items only communicated on notice boards make subsequent clarification very difficult. Any lack of opportunity for 'bottom-up' communication would be more critical if there was an over-reliance on notice boards.

Staff at some depots receive 'Red Letters' on anything safety critical and sign they have received and understood them. Most thought this was a generally good method of communication but commented that they had little easy access to ask questions or clarify certain points.

Although regular briefings were generally considered to be a good thing, some fitters stated that they could not always remember everything that was briefed to them. An additional written resource could be added to back up the verbal briefing. Language could also be a possible issue. At one depot, many were said to regularly speak in French and others had strong regional accents. This could introduce some mistakes, for example, by affecting the reliability of radio communications when shunting and verbally requesting supplies/parts from stores.

Communications within the organisations

Some communications problems seemed to exist between some depots and their headquarters. An example was found where a tool was supposed to automatically fill batteries and turn off at the correct level but this did not work well in practice. A new topping up tool was developed by HQ

and supplied to the depot to automatically top up the batteries but this did not fit. Although this tool is used to fill the cells, dipsticks are now necessary to check the levels. This suggests a lack of communication between HQ and depot staff to ensure that a design is appropriate.

Blame culture and incident reporting

There was evidence of blame cultures at a minority of depots that manifested themselves differently. At one depot this negative behaviour was found with the cause of delays and the extent of delays for vehicles that fail in service being attributed to key individuals. Shop floor personnel stated that whenever there is a mechanical fault there is always the question from management as to who was the last to work on that unit with the action usually being to allocate all the blame on that person – irrespective of whether there were more important causal factors higher up in the organisation or whether the fault could legitimately have occurred in the period between last inspection and failure. The assumption is also that that person 'signed that it was defect free' when this is not always possible or sensible. Generally, personnel can only sign that the inspected item was free of visible defects on those parts that were visible. For example, only about 90% of the bolts holding brake discs on to the wheels are visible at any one time. Additionally, other checks can be prone to interpretation errors – such as gauging brake pads condition, or wheel measuring. There were particular implications on blame whenever a train stops due to wear on the traction motor brushes. Some comments suggest that the blame culture is personality led rather than endemic throughout the depot.

At another depot the blame culture related to the practice of discouraging incident and near miss reporting. An interviewee stated that when safety related incidents do occur, shop floor staff are reluctant to report them, as it is perceived that management's prime objective is to establish blame, rather than find a solution. This appears to have resulted in near miss incident reporting drastically reducing.

Safety culture and violations

Our observations and the results of the informal discussions with a number of maintainers at some depots suggest that a large proportion of supervisors and managers are effectively condoning inappropriate actions by their lack of corrective action. It could not be determined whether this was a result of their own values and safety commitment, or a result of any pressure on them from a more senior level. Safety culture would appear to be at the core of many of the problems found. For example, non-compliance was found with: tags not being displayed by those working on vehicles; barrier procedures not being followed as vehicles move in and out of the shed; COSHH issues with many unmarked containers being used; and some mandatory PPE not being worn when working near/under vehicles. It is highly probable that those involved were fully aware of the appropriate procedures but have elected (for some reason) not to follow them.

The number of safety concerns or near misses reported can be a useful indicator of safety commitment. At one depot, only one concern was raised in the last two years or so (and none in the last year). It seems clear that there had been some events involving train movement when people were nearby. This poor reporting either suggests an excellent safety result or, more likely, motives for not reporting or complacency. Mention was made of people coming under peer pressure and management pressure not to report near misses. However, conflicting indications of management pressure were found. Questionnaires indicated no problem. Many of those spoken to did not offer this as an issue and some specifically said it did not occur. However, others indicated that management imposed direct pressure with threats of removing overtime etc. for those who did

not comply. This is clearly an issue that could be critical and may warrant more consideration to determine whether complaints are simply people 'getting back' at managers or whether these instances are actually occurring with some specific managers but not others.

One person volunteered to complete a questionnaire. The researcher gave the standard instruction to fill it in as honestly as they can. The reply was to the effect that "if I do that I'll get sacked". While this is often a joke, a number of similar comments were made by others, which at least suggests that such a perception may be believed to be true.

Checks at a depot confirmed that jobs were being undertaken during the survey where a number of fasteners had torque settings specified. None of the toolboxes seen had torque wrenches visible. None of the jobs observed involved torque wrenches. The stores records showed that only one torque wrench had been issued in the 24 hour period prior to the survey, and this was for a job not seen. On this basis it would appear that a 'systemic under-use' of torque wrenches had probably developed.

Some maintainers are required to perform shunting duties in and around some depots. Shunting incidents had occurred and some were repeat occurrences where drivers were not fully clearing points, perhaps to save time. Management seemed to have stressed that any delays were their problems and that drivers shouldn't think that they are expected to rush. Nevertheless there are a number of possible motives if these incidents were the result of intentional rule violations. These could include: pride in getting everything done in the shift; peer pressure; real or perceived pressure from some team leaders/shift managers to rush to save time; and poor perceptions of risk – although these may change following each incident.

Conflicting indications were obtained at some depots with some comments that staff are pressured to sign off vehicles they are not happy with and then management denying any subsequent involvement should an incident arise. Others at the same depots volunteered that the opposite applied with supervisors/managers signing things off themselves if it is their decision to release a vehicle.

Overall, very different views were found with those staff interviewed. Some showed management to be very supportive but others seemed to suggest a blame culture and a bullying culture. It is feasible at some depots that some employees may consider there to be a bullying culture if they were under-performing and the focus of management attention in an attempt to improve their performance. This would be consistent with a view expressed by a number of staff that management generally left them alone once they had become confident in their abilities. A number of other comments do, however, suggest quite clear problems exist at some depots that are probably the result of major personality conflicts.

Conclusions

These issues have arisen from a range of human factors surveys for different clients over a period of almost five years restricted here to an overview of the key areas that may be increasing the potential for errors at depots performing level 4 maintenance. The author has been involved with human factors aspects of maintenance for over 20 years in a wide range of industries, and rail maintenance depots present similar problems to those found in other industry sectors.

Some problems were identified that were the result of the vehicle design. The standard of maintainability is rarely specified. Although this may seem an obvious omission it is, at least in part, defensible. Defining and policing a specified level of maintainability can be difficult. Maintainability indices are available that could help improve the design interface and reduce such problems. Some retrofits are also introducing new maintainability difficulties and not taking the opportunity to resolve existing problems.

Depot design, facilities and physical capabilities may not be fully compatible with the requirements implied in the procedures. Examples were found where the specified procedures cannot be followed because of certain design features at some depots. Depot training capability, competence management, and competence assessment procedures may not be compatible with the skills/knowledge demanded by the maintenance requirements of the vehicles.

It is clear that the standard of technical writing differs for different vehicle maintenance instructions. This suggests that higher standards are being achieved by at least some organisations. Technical authors continue to produce procedures that could be confusing or otherwise increase the potential for error and resulting safety problems. One potential factor could be that a number of disciplines need to be involved in producing technical procedures that are both accurate (from an engineering point of view) and will be reliably interpreted (from a human factors point of view). It is possible that some organisations writing such procedures may not have sufficient human factors experience to draw upon. It is possible that insufficient time resource is being generally allocated to writing and checking new technical procedures. A non-technical review of procedures could specifically look at issues relating to usability, consistency of style, layout, etc. It may also be desirable for all technical procedures to be rigorously audited on the first application before it is finally issued to ensure that all the requirements are fully practical in the given workstations and with the given access and facilities.

At some depots, procedures seem to be more relied upon to supplement generic training. They are therefore more likely to generate errors when used by inexperienced employees if they are ambiguous, written to different styles, etc. Computerised management information systems (MIS) can introduce a foreseeable risk of error for those using them. Incompatibilities exist between systems and there can be poor training in their use.

As problems have been identified as a result of a variety of questionnaire, interview and observational surveys for a number of clients, it is difficult to objectively identify any factors that differentiate the generally better and less good depots. There are, however, two areas where there may be such associations. These warrant further study. Recruitment and retention of shop floor personnel seems to have a considerable impact on a wide range of issues. This is especially noticeable in the training function, which has to train new starters, and also assess the competency of existing staff. Depots with a low staff turnover tended to result in staff working better together and being focussed on meeting the Company's goals. Their culture tended to be more open and honest with fitters having no problems in reporting any concerns they have over vehicle safety. Staff took pride in their abilities and appeared to welcome audits to show off how good they are, but still genuinely welcoming any constructive feedback. Training appeared to be more effective with better systems in place to ensure on-going competence, reviews of training needs, and trainers regularly conducting audits. Competence testing also appears to be more effective.

Depots with higher financial penalties if vehicles are returned late into service, and especially those with limited control of the times that vehicles return to the depot, seemed to show higher levels of anxiety and stress amongst line managers/supervisors. Many management staff seemed to absorb this aspect of the business stress, others seemed to have passed it down to the workforce.

Chapter 44

Generic Human Reliability Assessment for Railways: Results

Ned Hickling, Liz Gaskell and Theresa Clarke

Introduction

This chapter describes work undertaken by Vectra Group on behalf of Network Rail. The principal objectives of the work were to establish whether generic human reliability risk assessment models could be produced which would be applicable to signalling operations and points maintenance, activities respectively. Additionally, the work was undertaken to provide exemplars of human reliability risk models and assessment techniques for Network Rail. As well as fulfilling the main methodological objectives it was considered very desirable that the work should provide some practical results. The work was undertaken between March 2003 and October 2004.

General description of the work

The process undertaken for the work closely follows that which has been used for undertaking human reliability and risk assessments in a wide range of other industries. That is: perform task analysis; undertake qualitative human error identification; formulate appropriate risk assessment models; populate risk assessment models with quantified human reliability assessments (HRAs).

This almost classical approach was replicated in this work, not only because it is a well-established framework, but also because it seemed entirely appropriate to the issues being addressed by the work.

Task analysis methods

Task analysis was undertaken by producing hierarchical task analysis trees, within which each node was documented within a row of a tabular decompositional analysis. Three different analysts undertook: a signalling task analysis, a mechanical points maintenance task analysis and a management of mechanical points maintenance task analysis. The analysts ranged in experience from 5 to 30 years of undertaking task analysis. All three analysts have professional qualifications in ergonomics or applied psychology.

Each analysis was undertaken using a stopping rule based upon the analyst's personal confidence. This was simply that the decomposition should stop at a point where the analyst became confident that any human errors of concern would have been identified and captured, or at the point where the first interaction with an interface object arose, whichever was more appropriate.

Whilst this rule is the subject of criteria unique to the analyst, it is appropriate because the complexity of human error means that different analysts will reach the point of confidence at different levels of task (or problem) decomposition. This is a factor, in part, of the analyst's

experience and pedantry, but more importantly of the nature of the task being analysed. For the signalling task analysis this stopping rule was modified to stop the tree decomposition at the level immediately above that which would invoke interactions with specific signalling technologies. This is described and discussed further below.

In each case, the objective was to aim for an analysis which would be generic, that is, it could be reused within different risk assessments. For this to be possible, the signalling task analysis considered all signalling operations that would be required within a shift and included the activities that would be necessary at the commencement and completion of a shift. Risk assessment is principally concerned with human interactions that either induce an unsafe state within a system or seek to mitigate an emerging unsafe state. Therefore, the signalling task analysis also included the consideration of signalling tasks demanded in abnormal situations and degraded operations. This was done to ensure that the tasks involved in responding incidents and that could make that incident worse, would be considered. Whilst the analysis covered all scenarios, it focused particularly on collision in the course of recovering a stranded train.

In contrast, the mechanical maintenance of points does not relate to risk in the same way. Here, human failures in mechanical maintenance will constitute initiating events that could subsequently cause a train to be derailed or a collision to happen. In risk assessment terms, initiating events are actions that introduce the potential for risks to be realised.

The model for the strategic management of points maintenance was different again. In this case, an extremely large number of potential tasks could be undertaken, that could lead to risk. The tasks could be dependent upon particular administrative systems in place, personal preferences and beliefs about task priorities, and different tasks or priorities in the face of contingencies or emerging situations.

Therefore, this last task analysis was undertaken at a very high level such that the principal goals of maintenance management were considered. Whilst hierarchical, it is probably better to view the maintenance management task analysis to be a goals analysis that is hierarchical, because the management tasks it contains are only decomposed to a functional level of description. For example: “Plan maintenance procedures to ensure that requirements are met with available resources”.

In all three HTA tree analyses, the analysts were encouraged to use a standard syntax that contains: an action verb, an object or state that is acted upon, and the statement of the sub-goal to be achieved. This approach was taken to ensure that a consistent level of good understanding by independent readers could be obtained for each analysis. The signalling hierarchical task analysis tree was developed using conventional HTA technique, however, the points task analysis and that the maintenance management were developed using the PAGODA technique (Hickling, 1992). This technique provides for the more systematic development of task analytical plans and provides methods that help to ensure greater coherence and completeness in the resulting analyses.

Qualitative error identification methods

Qualitative error identification was undertaken upon a portion of the signalling task analysis, and for the points mechanical maintenance task analysis. This was undertaken using human error guide words (Whalley, 1994). In addition, TRACEr Lite (Shorrock, 2002) was applied by a less experienced error analyst and the results were compared with the use of the guide word approach. Such an analysis could not be usefully undertaken for the analysis of maintenance management goals because, by definition, at such a high level any failure to meet the goal could result in system failure. Ideally, senior maintenance managers would have been involved in the definition of the

most likely human failures but industry restructuring and consequent workloads made them unavailable.

Risk model development

The qualitative error identification for the signalling task analysis very quickly revealed that it would not be possible to generate a generic risk model for signalling because of specific differences in track layout, timetable, and geographical surroundings. However, a fault tree was produced for the mechanical maintenance of points. In contrast, the higher level of goals stated in the maintenance management system task analysis meant that the model itself could be used directly for assessing risk. This is because, in any administrative system, coincident technical failures are not required for risks to be realised, and so a combinatorial model in the shape of a fault or event tree is unnecessary for assessing risk.

Fault trees were chosen in preference to event trees because they better accommodate different permutations and combinations of human error or failure events that could lead to systems failure with a given functional impact. For signalling, this was taken to be train collisions whilst the points maintenance train derailment was considered.

Human reliability assessment techniques

One objective of the work was to examine how well the existing human reliability assessment methods used in other industries would work in the railway sector. Within the UK, the HEART (Human Error Assessment and Reduction Technique) method (Williams, 1988) is the most widely used technique and is, probably, the most adaptable. Therefore, this technique was judiciously applied in this work. However, for points maintenance it does not provide useful data and the work of Dougherty and Collins (1996) on maintenance tasks was used in preference.

The methodological findings of the work fall into two broad categories. There are those concerned with the application and the applicability of the chosen methods, and there are those concerned with human factors findings for the system under examination.

Methodological findings

There are methodological findings about task analysis, qualitative human error identification, risk modelling and the applicability of HRA methods. These are now discussed separately.

Task analysis findings

It was found possible to develop a generic task analysis model for signalling. The structure of the model reflects the fact that train control in signalling tasks is the same at a functional level of description. However, differences in the interface technologies ranging from lever frames to a soft desk, and the corresponding increase in the number of trains controlled, means that the finer detail of signalling tasks cannot be described in a generic analysis. Therefore, for the purposes of producing a generic task analysis model, the stopping rule was applied where the level of hierarchical decomposition stopped at the level immediately above that at which specific technological or interface features would be invoked.

The signalling model was developed further assuming the application of N X. technology as an in feed to the later human reliability assessment process. For risk assessment purposes, a task analysis needs to represent a specific unsafe scenario. To provide an analysis for such a scenario, the task model was designed to be generic so that branches of the HTA tree could be selectively chosen for inclusion. Therefore, parts of the model needed to be selected and configured to represent a specific scenario. However, we later found that, whilst the task analysis model could be generic the subsequent qualitative human error identification became specific both to a scenario and to a geographical location. Therefore, we concluded that it is not possible to furnish a generic qualitative error identification and, by extension, a subsequent qualitative human reliability assessment of signalling.

The analyst undertaking the signalling task analysis has many years experience of analyses within the process industries. We were unsurprised to find that the ergonomics and human factors issues identified in the analysis were closely analogous to those involved in the centralised control of industrial processes such as chemical plants etc. This view persists through the subsequent qualitative error identification stages and qualitative human reliability assessment stage.

In contrast, we found that the analysis of mechanical points maintenance could be applied irrespective of specific location within which the maintenance was being undertaken. Therefore, we conclude that one generic task analysis for each type of points machine in use on the network will be sufficient to furnish subsequent qualitative and quantitative human error identification and probabilistic assessments.

As we expected, the hierarchical approach to signalling task analysis was appropriate. Higher-level HTA boxes were appropriate places to consider cognitive elements for tasks such as recognition and acquisition of goals, as well as those higher-level activities whereby signallers monitor their own signalling performance for error. However, we considered that the hierarchical analysis of points maintenance was actually unnecessary. This arose for two reasons. Firstly, the maintainer does not, largely, have to retain higher-level goals. The overall goal of the maintenance task is to undertake maintenance steps within the correct prescribed sequence and the sequence that is permitted by the physical configuration of the technology. Therefore, it follows that the principal goal is not to omit steps from the maintenance procedure. On that basis, there is very little value in a hierarchical task analysis. A simple linear description of the maintenance steps that are undertaken would be sufficient. If the written procedures that describe the steps are well engineered, then the analysis can follow directly from the procedural steps. However, whilst we found that the procedures in use were quite modern, some necessary detail was 'skated over' and, therefore, a task analytical stage was still found to be necessary to amplify vague or omitted detail.

The second reason for considering an HTA to be superfluous is the amount of effort required to produce the HTA documentation. We found that the resulting tree extended over some 16 pages when depicted in the smallest legible size. This makes documentation unwieldy in its representation of the necessary tasks. Therefore, we consider that for many maintenance tasks where higher-level goals cannot be defined, then a simple linear task description with an associated tabular analysis for each step is sufficient.

The task or goals analysis for the management of mechanical points maintenance was, by definition, a generic analysis because of its high level of description. At the outset, it was designed to cover all mechanical points maintenance management functions. However, we consider that the model might, with some further verification, be suitable as a generic analysis for other mechanical maintenance management undertaken on the network.

Qualitative error identification findings

We undertook most of the qualitative error identification using the human error guidewords that are part of the human HAZOP process developed by Whalley, (1994). Our brief comparison of TRACEr Lite (Shorrock, 2002) and normal qualitative error identification produced some interesting insights.

For an inexperienced analyst we found that TRACEr Lite was a valuable adjunct to error identification. This was true notwithstanding the fact that the version used was designed for application to air traffic control. We believe it would be possible to produce a generic version of TRACEr Lite that would be applicable over many different domains.

TRACEr Lite is a very comprehensive exposition of the many factors that can induce human error. Perhaps unsurprisingly, the two more experienced analysts found that it was long winded in use and, for them, rendered no additional insights. Rather, the insights that the analysts already had were held in mind and "parked" at an appropriate point in the TRACEr Lite analysis. Moreover, it made cross-comparisons of different tasks with similarities more difficult because the sheer volume of information for each assessment 'fogged' the later recall of the findings for a given task during the analysis. We felt the tool was easier to use by printing out the successive spreadsheet pages provided in the computer-based presentation. The less experienced analyst found that this provided faster access to allow consideration of the questions relevant to the matter in hand. Nevertheless, we conclude that TRACEr Lite is a valuable adjunct for qualitative human error identification and we would strongly recommend its use by inexperienced analysts and its adaptation to be more generic in form.

All three analysts had only limited experience of formal human error identification within the railway sector. Nevertheless, there was no difficulty identifying human errors based on the mechanical points maintenance and signalling task analyses. However, the much higher level description of task goals contained within the generic mechanical maintenance management analysis would require the expertise and knowledge of people involved in such a management task in order to identify the relevant errors and error inducing mechanisms that could arise.

Human factors modelling in risk assessment findings

We found that fault tree modelling was entirely appropriate for the modelling of human errors within a given signalling scenario. The chosen scenario was that of running a locomotive in the wrong direction to rescue a stranded train. However, as already stated in the qualitative error identification discussion above, the fault tree would need to vary according to the specifics of the layout, signalling arrangements and the position of the stranded train. This is the case because different layouts and arrangements provide different opportunities or affordances for human error and the application of protection that mitigates or controls them.

The generic maintenance management model was not subjected to a quantified risk assessment because constraints meant it was not possible to identify the human errors of concern in maintenance management.

We also found that the HEART augmented by Dougherty and Collins (1996) was entirely applicable to all the human reliability assessment needing to be undertaken to populate the fault tree models with quantified human errors. We have applied this approach with equal effectiveness in other sectors. We have noted the tendency in this sector, as in other sectors, to produce sector specific human reliability assessment models. Based upon our own broad experience in several sectors we would question the usefulness of sector specific developments. It is our view that the difficulties in the application of the quantification methods tend to stem more from lack of experience

in their application rather than from inapplicable assay to any given sector. Moreover, the character of human error tends to arise from psychological factors that remain constant irrespective of the sector. While some sectors believe that their personnel are better trained or more careful and others, in practice, the coarse estimates of human performance that are required (factors of three to ten) within human reliability assessment, means that such assessments are insensitive to these effects.

Systems findings

We now turn to the consideration of the system's findings that arose from the study. It is important to emphasise that these findings are not comprehensive; they arise simply as a side effect of meeting the principal aims of the project which were to produce generic human reliability assessment models.

In the signalling risk assessment, the principal matters of interest were the driven speed of the rescue locomotive and the knowledge or belief of its driver concerning the position of the stranded train. In signaller's terms, their main contribution to the risk would be in the actual description of the train's position and the landmarks that the driver of the recovery locomotive would expect to see. Additional but more obscure errors were concerned with the prospect of a locomotive approaching the stranded train from a direction opposite to that expected by the driver of the stranded train, due to a change in planning. In this case appropriate train protection in the form of detonators would not have been laid, therefore making collision more likely. As might be reasonably expected, the scenario specific risk assessment we carried out, suggested that the majority of risk arises from the conduct of the rescuing locomotive rather than that of the signaller. However, an appropriate risk assessment must integrate the actions and errors of both the signaller and the drivers of the two involved trains if the risk is to be properly assessed.

The success criterion defined for points maintenance was the avoidance of a train derailment due to a defect in maintenance. Our analysis showed that at least 35 human errors could result in train derailment. However, it is important to note that this analysis was not complete and these 35 were derived by considering approximately one third of the total surface area of the points maintenance HTA. In terms of the risk model, all of these errors would be placed as single errors leading to the consequence of derailment under a Boolean logic OR gate. Two human reliability conclusions of considerable practical significance can be drawn from this fact.

The first, and most significant conclusion, is that the large number of human errors in points maintenance that can lead to train derailment means that, by definition, the reliability of points technology is very vulnerable to human error. In practice, whilst there have been incidents resulting from failures in points maintenance, they are in fact very rare. The logical structure of the risk assessment model when taken together with the large number of events under the single OR gate, must mean that the levels of human reliability attained in practice are very high indeed. If this were not the case in practice, then we would expect to see a far higher incidence of train derailment arising from failures in the mechanical maintenance of points.

The second conclusion that can be drawn from this is that any changes in the task environment have a potential to systematically affect the reliability of points maintenance. In risk assessment terms this is called human dependency – where many nominally independent errors are affected by one common factor. One very good example of such an effect was identified during the task analysis stage of the points maintenance work. The recent introduction of RIMINII regulations means that a greater proportion of points maintenance work is now undertaken at night, so that maintenance personnel are not working in proximity to moving trains. However, a significant proportion of mechanical points maintenance tasks demand visual inspection or observation of their execution, to be successful. The task analyst noted that many maintenance personnel

were supplying their own head-mounted lamps for night-time working. However, these were not necessarily compatible with the provided safety helmets and, therefore, there was a choice of wearing a helmet or wearing a lamp. We could imagine a situation where an assiduous supervisor or manager concerned with safety would require workers to wear safety helmets, unwittingly, at the expense of their visual ability to perform the necessary tasks. Therefore, in principle, such an intervention could have caused a step change in the human contribution to failures in points maintenance. This illustrates the crucial point that any changes in the organisation, management or execution of mechanical maintenance of points should always be done with considerable care, and with all of the human factors and their potential impacts upon human error carefully analysed.

Conclusions

This work generated a considerable number of methodological and practical conclusions. Here, we summarise only those that are relevant to the content of this chapter. We conclude that task analytical methods to be used as a prelude to qualitative or quantitative human reliability assessment should be chosen with care. Hierarchical task analysis was found to be entirely appropriate for signalling tasks but was undoubtedly overkill for points maintenance. Considerable time and resource could be saved by using a much simpler sequential task description for mechanical points maintenance. This insight is not confined simply to analysis economy, but is reflected in the very limited meaning that could be ascribed to the higher level tasks and goals that we generated to "join together" lower-level tasks in the HTA tree.

We conclude that a generic task analysis for signals has been produced. However, this needs specific development at the detailed level for the particular signalling technology in use, and the track and traffic configurations that apply. In addition, the generic signalling task model must be made scenario specific according to geographical and track layout constraints, as well as those of the particular risk scenario being considered by selecting sub trees.

We consider that a single generic task analysis, qualitative human error identification and human reliability model would be appropriate for each type of points machine in use upon the Network Rail system. We conclude that methods for qualitative human error identification and the quantitative prediction of error probability are as applicable to railway tasks as they are to tasks in other industrial sectors.

Finally, the combination of task analytical, qualitative error identification and quantitative modelling methods provides insight on the relative importance of different human errors to potential risks. This assertion is clearly illustrated and verified by the fact that practical and important recommendations were generated within the project when the primary objectives were methodological.

Acknowledgements

We would like to acknowledge the contribution of those signallers and points maintainers who made this work both interesting and worthwhile. Without their cooperation and support, as in most human factors work, we would not be able to attain power project objectives. We must also thank Dr Andrew Shepherd for his work on the signalling task analysis. Without him there might well have been no generic signalling task analysis.

References

Dougherty, E.M. and Collins, E.P. (1996). Assessing the Reliability of Skilled Performance. Journal of Reliability Engineering and System Safety, 51.

Hickling, E.M. et al (1992). The Consistent and Structured Definition of Process Plant HCI-based Tasks Using the PAGODA Method. Contemporary Ergonomics 1992, Taylor & Francis, London.

Shorrock, S.T. and Kirwan, B. (2002). The Development And Application Of A Human Error Identification Tool For Air Traffic Control. Applied Ergonomics, 33, p 319-336.

Whalley, S.P. (1988). Minimizing the Cause of Human error. In Liberton, G.P. (ed). Reliability Technology Symposium. Elsevier, London.

Williams, J.C. (1988). A Data-based Method for Assessing and Reducing Human Error to Improve Operational Performance. Proceedings of 4th IEEE Conference on Human Factors in Power Plants, Monterey, California, June 1988.

Chapter 45

The Development and Application of a Rail Human Reliability Assessment Tool

Jenny Gilroy and Elaine Grimes

Introduction

Accidents on the railways often contain a strong element of human error as a causation factor. It is estimated that approximately 80 – 90 % of all major accidents have been caused in part by human error (RSSB; 2001). Human error also contributes to business risk, such as the reliability of the rail network. The ability to predict probable errors in new or changing rail systems, and to analyse errors that have already occurred, provides the opportunity to reduce the potential for error occurrence through error mitigation and reduction strategies. Therefore Human Error Identification (HEI) and Human Error Quantification (HEQ) techniques, collectively known as Human Reliability Assessment, are important risk assessment and risk reduction tools in the rail industry.

HRA tools are used in the rail industry when required, such as the Technique for the Retrospective and predictive Analysis of Cognitive Errors (TRACEr; Shorrock & Kirwan, 2002), an HEI tool; and the Human Error Assessment and Reduction Technique (HEART; Williams, 2003), an HEQ tool. However the available tools are not well suited to application to railway specific operations, as they have not been developed with this intention.

RSSB commissioned the development of a rail-specific HRA tool that comprised an HEI tool and an HEQ tool, which can be used in combination or separately. The tools were developed cooperatively by two different consultancies[1]. During the initial stages of the work, it was decided that the scope of the tools would be limited to train driver error, thus the tools that have been developed are specifically related to driver behaviour / driver tasks only.

The rail-specific HRA technique for driving tasks

Overview of the technique

The rail HEI tool is based on TRACEr, developed specifically for the rail industry and driver error. As with TRACEr therefore, the rail-specific HEI tool identifies errors with a focus on the underlying psychological mechanisms (including where in the person's thought patterns the error occurred, and the underlying causes of this error).

The rail HEQ tool has taken the elements of HEART, including some of its quantitative data, and again has developed and applied these concepts and data to driving behaviours. Within the HEQ tool, there are mapping devices to enable efficient use of the HEQ tool with output of the rail-specific HEI tool. However, the tool can also be used separately using input from other sources if required.

1 The Institute for Occupational Ergonomics developed the HEI tool and Atkins developed the HEQ tool.

HEI tool

TRACEr was designed to be applied to air traffic control scenarios, and it includes 'predictive' and 'retrospective' versions of the tool, and 'full' and 'lite' versions of each. The predictive tool is used to identify potential errors and the psychological mechanisms that will lead to them, and will be used to predict what errors may occur that could cause an incident; whereas the retrospective tool is used to pinpoint what errors have occurred in an incident, and to identify the psychological roots of those errors, in order to fully investigate an incident that has already occurred. The full version has a wider range of error classifications and a more in-depth analysis of the psychological mechanisms involved in the error than the lite version; however both encompass the same range of error categories. Each of the four versions of TRACEr were developed to be rail-specific; the rail-specific retrospective HEI tool stands alone and the predictive HEI tool and HEQ tool combine (if required) to act as a complete rail-specific HRA technique.

The structure of the predictive HEI tool and its key concepts are as follows:

- *External Error Modes (EEMs):* These are error types, which centre mainly on omission and commission errors that relate to actions and communications (e.g. 'driver forgets to…' or 'driver's action too late' or 'information not transmitted'). A list is provided in the tool as a structured method to allow the user to consider all error types.
- *Domains:* Each EEM is classified into an error domain ('Perception'; 'Decision'; 'Memory'; or 'Action') or is classified as a Violation. Within the HEI procedure, these domains then steer the user to the correct group of Internal Error Modes.
- *Internal Error Modes (IEMs):* These provide specific information on what cognitive functions failed, and how they have failed. Each error domain contains separate and specific IEMs, for example, in the 'action' domain, the driver may have made a 'selection error' or there may have been an 'action omission.' Within the violation domain, there are 'violation types' instead of IEMs. These include 'exceptional violation' and 'optimising violation' for example. The user must interrogate the list of IEMs within the necessary domains to select the IEMs that describe the faulty thought processes that have caused the error.
- *Psychological Error Mechanisms (PEMs):* These are the inferred psychological reasons for the occurrence of each IEM, providing a finer level of detail about the IEM which can aid in the subsequent identification of error reduction strategies. For example, within the 'action' domain, two PEMs are 'manual variability' and 'habit intrusion'.
- *Performance Factors (PFs):* These are conditions that influence the occurrence of an error. A comprehensive set of PFs are presented, which are sub-divided into nine categories including training, procedures and in-cab environment. The assessor can also indicate whether the effect of the PF on the likelihood of occurrence of the error is negative (increasing its likelihood) or positive (decreasing its likelihood or increasing the likelihood that it won't occur.
- *Error Detection and Recovery:* A guide is provided to aid the assessor in determining the likelihood level of successful recovery from the error. The level of recovery success likelihood (RSL) is categorised broadly into low, moderate or high.

Whilst the predictive tool is aimed at the assessment of potential human errors that are related to incident risk, the retrospective HEI tool has been developed to aid in incident investigation. This considers domains, IEMS, violation types, PEMs and PFs in the same manner as the predictive HEI tool however there are some differences. Instead of EEMs the assessor considers 'task errors' which allows the systematic identification of the specific part of the activity that the error occurred in. Additionally, the assessor classifies the subject/topic of the error with 'error information

keywords.' Finally, the 'error detection and recovery' step is different from the predictive tool because it captures how the driver became aware of the error and how he/she gained feedback of the error (if this occurred).

The main aim of the work for RSSB was to develop a HRA technique for risk assessment, thus the major focus of the work was on the predictive rather than the retrospective HEI tool.

The HEQ tool

The error quantification technique HEART can be applied within the rail industry, however the language and definitions used within the tool mean that it is more suited to nuclear and process industries. In the development of the rail-specific HEQ tool, the structure of HEART was retained, requiring error classification and assessment of influencing factors to calculate a human error probability, as described below:

- Identification of the *Generic Error Type (GET):* The rail HEQ defines errors by their *type* rather than the *task* that the error occurs during, as carried out in HEART. Thus an error will be defined as one of the following seven GETs: detection failure, perception failure, decision failure, memory failure, interpretation failure, response error or omission.
- Choose appropriate *Generic Error Probability (GEP):* Each GET has a nominal GEP, and Upper and Lower Bound GEPs which are based on the specific situation under consideration. For most there is a differentiation between the type of train service, e.g. driver of a freight train versus an express train, etc.
- Select relevant *Performance Factors (PFs):* An extensive list of potential PFs that can affect and reduce the reliability of human performance is included in the tool for assessors to consider. Based on HEART's Error Producing Conditions (EPCs), the list is different from those included in the HEI tool, but mapping from HEI PFs to HEQ PFs is possible. For any PFs that are selected, the relevant multiplier value must be used.
- Assess the *proportion of effect:* For each PF the assessor indicates the strength of influence by specifying the proportion of effect (between 0.1 and 1), and a detailed tool is provided to help guide the assessor.
- Input the details / data from the HEQ process into the HEQ calculation, to find the likelihood of the occurrence of the error. The calculation is the same as in the HEART calculation.

At this stage the HEQ tool is not designed to quantify violations (a facility that was not within HEART either), although this is a possible avenue for development in the future.

The entire HEI and HEQ process using the rail-specific tools is portrayed in Figure 45.1.

Original development of the tools

Development of the HEI tool

The original work to develop a rail-specific HEI tool for driving tasks involved adaptation of TRACEr to enable the classification and prediction of train driving errors. Initially this involved a review of broad information on driving tasks, including task analyses produced by RSSB (2003). An in-depth review of driving tasks was then carried out in different rolling stock types, followed by the subsequent development of a high-level generic task analysis that encompassed tasks for all rolling stock. A validation assessment was carried out on this task analysis using subject matter experts (SMEs) and cab rides.

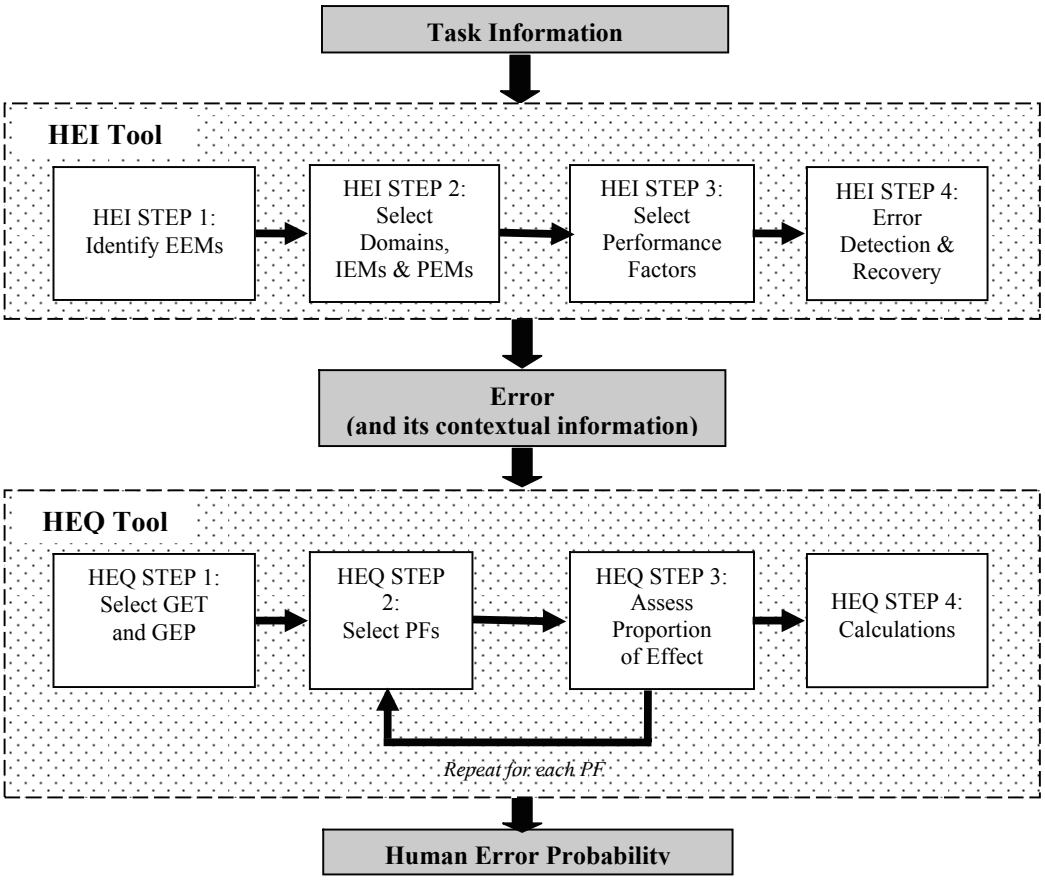


Figure 45.1 Representation of integrated use of HEI and HEQ tools

It was then necessary to identify the types of errors that drivers could make whilst attempting to achieve task goals. Error examples were generated through the application of all versions of TRACEr to the identified driving tasks. Additionally, using incident investigation reports, RSSB SPAD reports and RSSB Risk Profile Bulletins (RPBs), error factors were classified using TRACEr to ensure that all errors could be classified and that there were no gaps in the capability of the taxonomies for driver-related errors. No gaps were found within TRACEr taxonomies.

From this initial work, the two rail-specific HEI tools could be developed. For the retrospective HEI tool, the task analysis was used to derive 'task errors' for driving tasks. 'Information keywords' were produced through interrogation of the task analyses and by using the TRACEr IEMs as prompts. For the predictive HEI tool, and partially for the retrospective HEI tool, the following aspects of TRACEr were developed:

- Existing TRACEr PFs were assessed and all non-relevant PFs were removed. Additionally, literature searches, reviews of existing HEI / HEQ tools and incident reports were undertaken to identify further PFs, and task data were used to include driving factors. Train driving SMEs aided in the refinement of the PF set.

- The original EEMs were used for the predictive HEI tool, as they are context-free.
- Although IEMs and PEMs were non-industry specific from their original development, a further literature review was undertaken. This identified no additional error types for inclusion. Additionally, error types that were originally identified but rejected as irrelevant to air traffic control were also reviewed. None of these were included. The only changes made were adaptations in terminology.
- The error detection and recovery classifications in the detection and recovery taxonomy from TRACEr retrospective were originally based on non-domain specific literature and research, thus the originals were retained for the retrospective rail tool. However this taxonomy also includes context-specific aspects which were re-constructed with the help of train driving SMEs. The error detection and recovery taxonomy for TRACEr predictive has no context-specific elements, thus the original was used in the predictive HEI tool.

The Lite version of the HEI tool was then developed, which mainly involved grouping and simplifying IEMs and PEMs. The knowledge elicitation technique of card sorting was used to achieve this. Additionally the terminology used was simplified to allow more accessible use by non-human factors assessors.

User manuals were developed for each version of the tool to provide background information and clear guidance on their use.

At this stage, validation of the tool was necessary. All four versions were assessed by SMEs, an intended user group discussion and user trial, and a human factors appraisal. Usability, face validity, construct validity and content validity were all assessed. Additionally, an inter-reliability test of the 'full' and 'lite' retrospective HEI tools was carried out in user trials. The validation phase enabled some further development but it is recognised that a more comprehensive formal validation exercise is required.

Development of the HEQ tool

The work to develop a HEQ tool involved adapting HEART to be relevant and specific to train driver tasks / errors, including developing the structure of the tool, populating it with new data and producing appropriate supporting materials.

Initially a review of available approaches for data generation of probability estimates was conducted, (including the rail industry's Safety Management Information System, COREDATA and RSSB's Safety Risk Model). The preferred approach was selected as Absolute Probability Judgement (APJ), and a pilot APJ workshop was conducted with train drivers to confirm feasibility and refine the approach to meet project requirements.

Additionally, a similar piece of work to develop a refined HEART-based tool called NARA for the Nuclear Industry Management Committee and British Energy was sought out in order to establish useful links and areas for mutual development.

The initial development work carried out involved replacing the Generic Task Types in HEART with a list of GETs from which assessors would choose the error type rather than task type. A number of different options for the categorisation of errors were reviewed and considered: different types of error form (e.g. omission versus commission; slips, lapses and mistakes); task behaviour types (i.e. skill / rule / knowledge based errors); and categorisation by information processing types / stages which make a distinction between information input, processing and output. The cognitive domains within the HEI tool are founded in this latter approach, and it was also considered to be the most applicable for the HEQ tool thus the GETs were developed from

the same approach. However the final list of GETs reflects more finely distinguished information processing stages.

To generate the probability values (GEPs) for each GET, train drivers of different rolling stock / train service types were used as expert judges in three separate APJ workshops: express; commuter; and freight trains. The results therefore provided three separate sets of data according to each service type. The mid-value of the three judgements was used as the nominal GEP, and each other value as the upper and lower bounds for use according to the service type being assessed. However, the GEPs for the GET 'decision failure' were derived from real SPAD data due to a discrepancy between estimates generated in the APJ and real data. Other than for this GET, it was not possible to validate estimates against empirical data due to a lack of appropriate data.

In order to develop the list of PFs for the rail-specific HEQ tool, the error producing conditions (EPCs) from HEART were revised to rail-specific PFs with detailed definitions and examples. Additionally the EPCs from NARA were also reviewed and appropriate ones were included. The multiplier values of the EPCs from HEART were used, as it was not in the scope of the work to review this data, however where multipliers in NARA differed from those of HEART, the reasons for the change were considered and with guidance from its developers, the most appropriate multiplier for rail scenarios was chosen. Finally, guidance was developed for selecting the appropriate assessed proportions of effect, providing specific guidance for anchor points between 0.1 and 0.9 for every PF.

The user manual was developed for the HEQ tool, and feedback was gained from the rail industry on the tool at RSSB's 2004 Risk Management Forum. Finally, the rail-specific HEI (predictive) and HEQ tools were combined with mapping tools developed for users to effectively use the tools together.

Application of the tools to Selective Door Operations (SDO)

The formal user testing of the tools is currently taking place, and will be discussed later. However the first full application of the rail-specific HRA technique took place in a piece of work commissioned by HSE. Human errors associated with the use of manual and semi-automatic Selective Door Operation (SDO) systems on trains (where the major risk of this highly human-dependent task is that the wrong passenger doors will be opened on a short platform) were identified and quantified for a sample of different SDO systems. The aim is for the work to act as a reference point for TOCs and other rail organisations that are assessing the risks associated with an existing SDO system, or that are comparing different SDO systems and methods of operation. They will be able to either use the data to ensure that all potential errors identified in their own work have been assessed / considered, or in circumstances when the SDO system is sufficiently similar, data from the report can be directly applied. However, it is recommended that users assess this carefully and consider the PFs further to ensure that the PFs reflect the context they are being applied to.

The application was conducted on the following SDO systems:

- Class 170, Southern Trains (Bombardier system): a manual system controlled by the guard on one service and the driver on a different service.
- Class 377, Southern Trains (Bombardier system): a semi-automatic GPS-based system controlled by the driver.
- Class 180, First Great Western (Alstom system): a manual system controlled by the driver with some initial activities controlled by the guard.
- Class 221, Virgin Voyager (Bombardier system): a semi-automatic odometry-based system controlled by the driver, but not yet in use on any service.

Information was gathered about each system from the TOCs, HMRI Inspectors and the companies who designed and built the systems - how the system works, the equipment, procedures developed to use them and history of use. Task analyses were developed for use of the SDO systems in normal and abnormal working scenarios. Then the rail-specific HEI (full) and HEQ tools were applied to the most appropriate level of the task analyses for complete but reasonably non-repetitive output.

The intention of the work was to select a range of systems that would enable all human errors (as far as is possible) to be identified, for any system that could be used. Once completed, other sources of identifying SDO-related errors were interrogated, including SMIS, Network Rail's National Log and the CIRAS national database of confidential rail reports. Of all errors included within these sources of actual reported incidents, no additional errors were identified, providing evidence that the output of the work was as complete as possible, and that it was applicable to a wide variety of SDO systems and scenarios. The activity of applying the tools also enabled sufficient user testing to develop the tools for use as SDO-specific tools, which was the second phase of this HSE study.

Development of tools for SDO-specific use

The application of the rail-specific tools to many errors provided generic feedback about the use of the tools, including issues regarding the usability of the tools and their instructions, the transparency and the relevance of the output, etc. It also provided specific feedback about the use of the tools for this situation; i.e. assessing the use of an SDO system. HMRI have found inconsistent use of HRA methods in relation to SDO systems, due to the variety of tools available and the lack of guidance available to TOCs who need to carry out HRA for the purpose of assessing the risk of SDO. They want a tool for the rail industry to aid the accurate and consistent assessment of human errors and violations.

The experience of the application of the tools indicated that they are indeed applicable to SDO use, and led to an evaluation of aspects of the tools requiring development, within a structured approach. All issues raised in the evaluation were applied to the developments other than the data, the content of the PFs and the structure of the tool. Changes were not excessive and included:

- The instructions / guidance were altered to be SDO-specific and more usable, explain the relevance of activities and output, and provide guidance throughout use of the tool as well as the initial instructions. Additionally, guidance on the level of task analysis to be assessed was included.
- All references to drivers were changed to 'drivers / guards' and all examples were changed to be SDO-specific.
- The entire 'full' version of HEI was removed from the tool, so only the 'lite' version can be used which provides a sufficient level of detail.
- GEP data was changed so that there is no 'nominal' GEP; rather, the assessor uses the data for the specific type of service – express or commuter.
- Some changes were made to the mapping tools, including mapping of HEI output to the HEQ tool.
- The overall presentation and usability were improved.

HMRI are not planning to publish this work yet, due to the development work that is still being conducted on the original rail-specific tools; the tools will be updated at a later date if necessary.

The report and results of the application of the tools to SDO activities will be available via the HSE website however.

Development of the HRA tool

To ensure confidence in the tool, the reliability, validity and usability of the rail-specific HRA tool were to be assessed. A user trial is currently being conducted for RSSB. A Community of Interest Network (COIN) has been established to participate in the trial. The COIN is comprised of human factors experts from a variety of organisations, including consultancies, universities, and rail industry stakeholders, along with non-human factors experts who would be required to use a HRA tool in their work. These include risk assessment professionals, and operational experts within the Train Operating Companies (TOCs).

Users will analyse a range of operational scenarios using the rail HRA tool. The scenarios developed are driver specific, and two examples of each of the Generic Error Types (GETs) have been included, to assess the ease by which each of the GETs could be identified. The scenarios have been driven by the availability of 'real-world data'; it was felt important to compare the HEP values generated by using the tool with existing data to assess whether the values presented in the HEQ tool were accurate. The scenarios were generated through discussion with risk analysts, operations experts, and review of appropriate documents, to provide sufficient information for users to use the HRA tool, but without guiding them in their choices of the GETs or PFs. A task analysis was provided with each scenario, sourced from existing task analyses or generated for the scenario. An example of a scenario was 'driver misjudges environmental conditions', as follows:

> A driver is approaching a red signal aspect. S/he decides that s/he should apply the brakes in order to come to a stand and selects a braking point and rate. However, the driver misjudges the prevailing environmental conditions, fails to brake in time and experiences a SPAD.

As well as tool validity, the HEPs calculated by the users for the scenarios compared with real-world data, the inter-rater reliability, i.e. the extent to which there is consistency between users in how they analyse the scenario, and the HEP generated will be assessed. Therefore, the analysis and selection of GETs and PFs, and the HEP calculated will be compared between users.

A questionnaire on the use of the tool has been developed and will be distributed to the user trial group. This seeks feedback on all aspects of the tool, but particularly the HEQ tool. The HEI tool will be assessed relative to its usability when used in conjunction with the HEQ tool, in both full and 'LITE' forms.

Analysis of the outputs of the user trial and usability questionnaire will be used to identify areas for further development. The analysis will investigate whether there was a link between experience in using the tool and accuracy in selecting and calculating the HEP. The trial will also seek to assess whether the HEPs generated for each scenario are consistent with the real world validation data. It is envisaged that the rail HRA tool will be used in a variety of ways, including assessment of new and existing systems and procedures, and prediction of human error for the UK Rail Safety Risk Model.

Summary

The rail-specific HEI and HEQ tool for driver errors developed from TRACEr and HEART, has been and successfully applied to the use of a variety of SDO systems. It has also been developed into an SDO-specific version. The driver error tools are undergoing user testing and subsequent developments which will improve reliability for confident wide-spread use.

Acknowledgements

The following individuals participated in the development and/or application of the tool:

Colin Dennis, Huw Gibson and Katherine I'Anson from RSSB; Liz Cullen; Ben McCaulder and Lucy Swinson from Atkins Rail; Stuart Smith from Institute for Occupational Ergonomics; Debbie Lucas and Peter Randall from HMRI; Laura Sutton, Laura Sutton Associates. The relevant section of this work (application of the tools to SDO activities and development of SDO-specific versions of the tools) was funded by the Health and Safety Executive (HSE). Its contents, including any opinions and/or conclusions expressed, are those of the authors alone and do not necessarily reflect HSE policy.

References

RSSB (2001) "SMIS Phase 4: Briefing notes." RSSB, London.

RSSB (2003). Hierarchical task analysis of the UK train driving task, 22nd January 2003. Prepared by Human Engineering for RSSB.

RSSB (2003). Tabular task analysis of the UK train driving task, 22nd January 2003. Prepared by Human Engineering for RSSB.

RSSB (2003). Task analysis of the ERTMS train driving task, 22nd January 2003. Prepared by Human Engineering for RSSB.

Shorrock, S.T. and Kirwan, B. (2002). Development and application of a human error identification tool for air traffic control, Applied Ergonomics, 33(4).

Williams, J.C. (1993). "A user manual for the HEART Human Reliability Assessment method". C2547-1.001, DNV Technica.

This page intentionally left blank

Chapter 46

Improving the Predictive Power of an Existing Human Reliability Assessment Method

J.L. Bell, J.C. Williams, C. Daniels, and N.D. Warren

Introduction

Like many rail systems around the World, the UK Rail network is reliant on human behaviour and the attainment and maintenance of consistently high task reliability. Despite increased use of train protection systems and driver assistance devices on trains, drivers are fundamental to safe train operations, as are signallers, despatchers, electrical control and maintenance personnel.

Many of the tasks require high levels of situation awareness, effective recognition, excellent recall, and consistently reliable detection. Although many improvements have been made, and continue to be made, over a very long operating history, the system remains vulnerable to human error. Acknowledging that human unreliability might invalidate safety assumptions, The Railways (Safety Case) Regulations (2000) require all duty holders to assess safety and, where necessary and appropriate, show that potential hazards are managed and controlled, so that any risk will be as low as reasonably practicable (ALARP).

In order to demonstrate that risks will be ALARP, Quantified Risk Assessments (QRA) are sometimes performed in order to determine what the major components of risk might be, how susceptible to human failures the systems might be, what foreseeable accident sequences might arise and, where appropriate, what the impact of human error might be, so that suitable safeguards can evaluated and incorporated into relevant operating and management systems.

In recent years, Human Reliability Assessment (HRA) has become an integral feature of the UK railway industry's approach to QRA, safety management and investment evaluation (Maidment, 1998). As well as the main system employing these methods, the London Underground Jubilee Line Extension and the Channel Tunnel Project Management Teams have, at the design stage, both used HRA to provide assurance that levels of risk on these new railways would be acceptable. Design and operational features have been evaluated to provide insight into beneficial practices and cost-effectiveness so as to ensure that sound, risk-based, principles are employed.

Human reliability assessment

In order to assess the predicted impact of human-centred system and work design features on rail operations, the industry has, to date, tended to use actuarial data, laboratory and field research data or human reliability assessment methods to inform and underpin its judgements. Typical of the actuarial data are Signals Passed at Danger (SPAD) reporting and analysis - e.g. Lock and Li (2003) - and Confidential Incident Reporting Analysis System (CIRAS Core Facility, 2004) reports. Typical of the laboratory studies would be the research conducted to explore ways of making signals more conspicuous by changing the nature of the aspect or the backboard (DERA, 2000), or by studying the eye movements of drivers (Groeger *et al*., 2003).

Typical of the sorts of HRA that are used in support of safety arguments and investment decision-making are those that have been undertaken by Railtrack (A D Little, 1997) and the Health and Safety Executive (Cowell, 1998). The Railtrack study investigated, amongst other things, the appropriateness, or otherwise, of installing Driver Reminder Appliances (DRA) and activating them via the Automatic Warning System (AWS), whilst the HSE study investigated the risks associated with fatigue in railway safety critical tasks such as pre-planned isolation, detection of, and reaction to a gas alarm.

In both cases where a HRA method was used, quantified judgements were made using HEART (the Human Error Assessment and Reduction Technique) (Williams, 1985 and 1988). Assessments of the predicted difference between, for example, the additional reliability judged to be associated with active (AWS) setting of DRAs and the risks associated with fatigue that might affect pre-planned isolation were stated to be of the order of 50%. The DRA assessment for Driver Only Operation predicted that the nominal task failure probability difference between passive and active setting would be (0.00241 active) and (0.00352 passive), a factor of 1.46. For effective response to a gas alarm (which was judged to contain the task components of visual search, information processing, perceptual-motor and memory) and including the predicted nominal likelihood of task failure associated with fatigue, it was judged that the difference between those experiencing fatigue and those not would be a factor of 1.64. The "Expert" Electrical Control Operator nominal probability of task failure was 0.23 when not fatigued, as compared to 0.378 predicted nominal probability when fatigued.

It will be apparent in the HRA examples above that predictions were obtained which showed human reliability differences in relation to the likelihood of task failure in the ratio of about 1:1.5 for both the potential system design impact and work design change. As rail risk assessments show that assumptions about human performance can have profound effects on the way in which risks would be managed, having confident estimates of human reliability and of the influences on human reliability is clearly important in relation to managing risk. So, any improvements to the predictive power of existing human reliability techniques should prove beneficial and be welcomed by the industry. This chapter discusses some of the improvements that can be made to the predictive power of the HEART methodology.

Like most UK industries, it is predicted that the rail industry will be employing a statistically older workforce in future. Current projections suggest that, even allowing for any plausible influx of immigrant labour, the workforce will become noticeably older. Therefore, for a safety-critical industry such as rail, it will be important to assess and understand whether age would be likely to have any impact on safety and reliability. With a predicted older workforce, and current interest in issues such as the effects of age on dynamic visual acuity, it is clearly important that the rail industry has as complete an understanding of the impact of age as possible.

At present, few HRA techniques assess age-predicted impacts on human reliability and, although HEART does, this particular Error Producing Condition multiplier was first proposed at a time when there were relatively few robust age/reliability studies available for consolidation into a single maximum EPC limit value and so an estimate was offered for consideration.

Method

Since the HEART method was first proposed, the human factors literature has grown enormously, with EPCs that were previously largely-unresearched now becoming amenable to incorporation into the model. Those that, to start with, had limited support are now becoming much more easily analysed and assessed. Accordingly, in order to check the veracity of the original age EPC and to gather information relevant to aging, a wide-ranging literature search was conducted.

This included a detailed search of a range of specialist journals, including *Psychology and Aging, The Journal of Gerentology* and direct contact with the principal researchers of the day. The literature search flagged up some very important developments which had occurred since the mid '80s.

In total, some 82 relevant papers have been identified to date and of these, some 15 were found to be useable in terms of the data that could be extracted for further processing. Data were regarded as being useable if the results of the various experiments could be plotted as probabilities of failure, if the experimental sample sizes were substantial (above 10 participants per study), and if age was a factor in the experimental design.

The range of experimental studies considered was very great and it is noticeable that, over the last 20 years, the information on aging and the impact of aging on human performance that was needed to "complete" the HEART model in this respect has now largely become available, to an extent that was unthinkable in the mid '80s.

Each experiment identified as being potentially useable was analysed for evidence of the effect of age in relation to recall, recognition and information-processing. Task failure data were converted to probabilities of failure and these were plotted on a logarithmic-scaled spreadsheet for initial visual comparison. The wide range of studies used quite different measures which in some cases produced an error probability. Other results were presented as a percentage of information lost or the percentage below the maximal score. For the purposes of this study, all results were converted to a common measure of human performance between 0 (best) and 1 (worst), which is interpreted as an error probability.

Statistical methodology

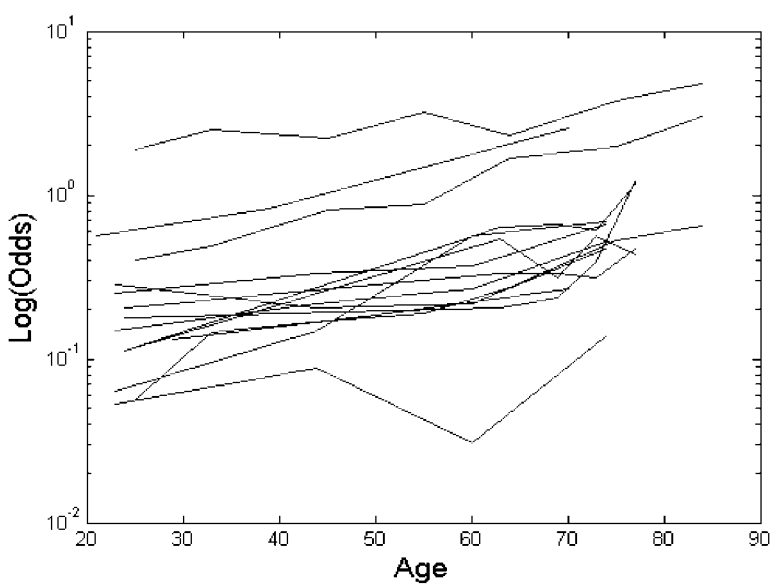


Figure 46.1 Probability of failure data taken from each relevant study

In order to provide a common basis for analysis across a range of studies, to permit meaningful comparison, data from the experiments of Ardila *et al* (2000), Atchley and Kramer (2000), Baddeley

et al (1994), Detweiler *et al* (1996), Gamboz *et al* (2000), Kirasic *et al* (1996), Kvavilashvili *et al* (2005), Maylor and Lavie (1998), Ostrosky-Solis *et al* (1998), Parkin and Walters (1991), Roge *et al* (2003), Salthouse *et al* (1996), Salthouse (2004), Sliwinski *et al* (2003) and Whiting *et al* (1997) were calculated as "odds", where:

$$\left(\text{odds} = \frac{p}{1-p}\right), \text{ and p is the probability of an error occurring.}$$

These "odds" were then plotted as natural logarithms (Figure 46.1).

It will be apparent via visual inspection alone that there is a consistent change in odds that is associated with age and that this change is relatively uniform across a wide range of ages. This means that reliable predictions can be made about the impact of age on human reliability.

This analytical approach is consistent with the widespread use of logistic regression to model the probability of an event occurring where dichotomous observations are available. For our analysis, direct observations of the occurrence of errors are unavailable and logistic regression inappropriate. Instead, our analysis uses a weighted linear regression model to examine the effect of age on the log odds. The weight for each data point is determined by the number of participants in the underlying experiment with the variance of each data point's error term being inversely proportional to n (because each data point is the mean of n participants). Using weighted regression ensures that, despite widespread variation in the size of studies, appropriate conclusions are drawn. The model for the log odds can be written as:

$$\log(\text{odds}) = ③ + \text{task effect} + \text{age effect x age}$$

Here ③ represents some nominal base line value for the log odds over all studies, the task effect describes the adjustment for a specific task and the age effect the incremental effect for every year of age. The age effect determined by this analysis represents an integrated estimate of the effect of age based upon the entire set of experimental studies. A 10 year increase in age increases the odds by a factor of

$$\text{Exp}(10\text{x age effect})$$

This approach deals effectively with data which have large values corresponding to probabilities close to one. This is a major advantage for it allows the model to be extended to multiple error producing factors without the danger of their combining to give impossible results. It is worth noting that the probability of an error can be easily recovered from the odds using:

$$p = \frac{\text{odds}}{1 - \text{odds}}$$

For rare events, where the probability p is small, the probability and the odds are almost numerically equivalent.

Work is currently underway to develop a mixed effects model to investigate variation in the effect of age with different generic task groups. This analysis will also model the dependencies between 'data points' taken from the same study and investigate the impact of some additional Error Producing Factors.

Summary of the effects of age on human reliability

The coefficient for age in the current weighted regression model is 0.015 (95% confidence interval 0.011 – 0.019, p-value <0.0001). This value implies a relative increase in the odds of an error occurring of 1.16 or 16% (95% confidence interval 12% - 21%) for every 10 year increase in age. For small error probabilities (of the order 10^{-2}) this corresponds to a 16% increase in the error probability itself.

Some experimental results were kept to one side, so as to determine whether the impacts predicted by the data treatment would actually predict the outcome in the unanalysed experiments (Glisky *et al*, 2001; Fernandes & Moscovitch, 2003; Kausler & Puckett, 1981; McIntyre & Criak, 1987; Madden & Langley, 2003; Mantyla & Backman, 1992; Maylor *et al* 1999; Park *et al*, 1989; Rankin & Collins, 1985; Shinar *et al* 2005; Singer *et al*, 2003; Taylor *et al* 2005). Evidence from this withheld information suggests that some fairly accurate predictions can be made, based solely on the strength and direction of the rest of the data set. These are shown in Figure 46.2.

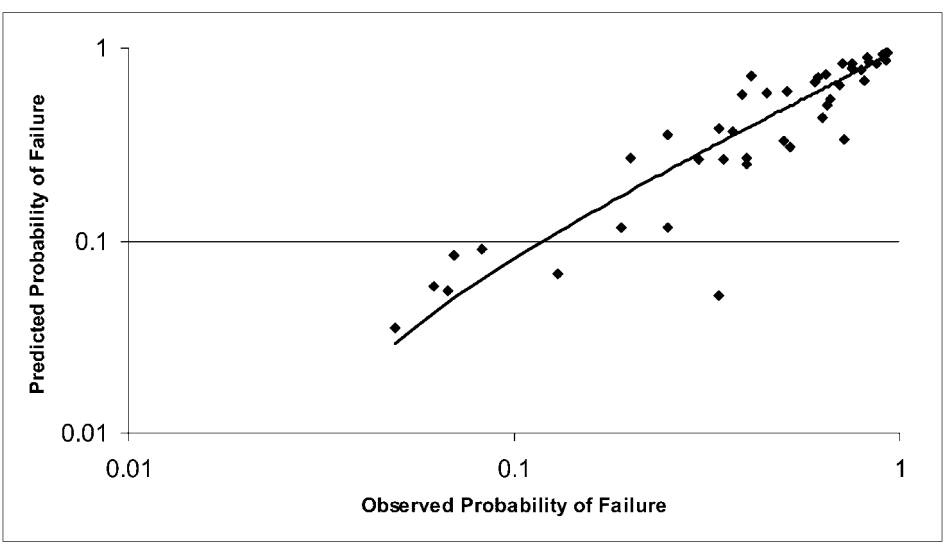


Figure 46.2 Correlation between observed and predicted experimental results

The correlation coefficient is 0.916, p<.01 which suggests that accurate predictions of the impact of age on human reliability can be made based solely on an EPC of 1.2 for every 10 years age, over the range 25 to 85* years (*non demented). Further work to refine this EPC is currently underway, but it is not expected to depart radically from initial indications.

Discussion

What is immediately apparent from this research is that the Age EPC would benefit from revision. A proposed value of 1.2 for every 10 years age over the range 25 to 85* (* undemented) would seem appropriate until further evidence can be brought to bear on this particular unreliability multiplier. The second most obvious finding is that, potentially, age could be a much more important human reliability influence than has, so far, been thought to be the case. Thirdly, the impact on human

reliability that is predicted is comparable to the amounts of reliability change that have so far been of interest to the rail industry, with respect to systems design and work design impact. It will be apparent that, for a 27 year age difference (for example, drivers aged 40 versus those who might be aged 67) the group predicted reduction in reliability for an aging workforce might well exceed that which would be of interest to safety assessors in other contexts.

Not all the industrially-relevant EPCs in HEART have been identified, as yet, and some of those that have been identified are not supported by as much evidence as one would like to have in order to have confidence in an assessment in order to take appropriate action. In addition, over the last 20 years, the human factors literature has grown enormously, with EPCs that were previously largely-unresearched now becoming amenable to incorporation into the model and with those that, to start with, had limited support now becoming much more easily analysed and assessed. For example, the Illness EPC was originally assessed to have a value of 1.2 and more recent research (e.g. Smith 1993 and 1995) provides support for this assessment.

The nett result of all the above is that there are many improvements that can now be made that were simply infeasible in the mid '80s. These include enhancements to content and applicability, as well as to usability by non-specialists. One area where further development is clearly desirable would be in the further identification and quantification of EPCs and Violation Producing Conditions (VPCs). As many rail tasks such as driving, signalling and electrical switching require highly reliable retention, recall and recognition of information and compliance with procedures and rules, it will be apparent that there is considerable scope to improve the completeness of some of the modelling of task reliability in the rail sector.

Although the principal feature of this chapter is the quantification of the effects of age on human reliability, there are a great many other potential EPCs that would benefit from data collation, processing and quantification – for instance, impaired information access, irrelevant information, inadequate information, depleted information, competing information, distraction, loss of information, amount of information to be retained, amount of information to be processed, information processing delay, impaired memory, incorrect risk perception stimulus uncertainty, choices per decision and fatigue all have the potential to be incorporated in the HEART methodology.

The principal finding from the current study is that within the 25-85* years age range (*non-demented), younger operatives appear to be considerably and consistently more reliable than their elder counterparts. The finding that, for every 10 years over this age range, human unreliability increases by about 20%, has important implications for the rail industry, particularly against the background of an aging workforce. Whilst it is clearly the case that experience can "counteract" the decrement in performance that is associated with age, this can only be relied upon to a limited extent. As Morrow *et al* (1994) have shown, expertise can reduce age differences in reading back air traffic control communication spoken messages but at similar levels of experience younger people will be more reliable.

Conclusions

The purpose of this particular study has been to extend the general-purpose nature of the HEART method by capturing data for those EPCs that are currently under-substantiated and, where feasible and appropriate, to extend the number of EPCs available to the practitioner. Identification and quantification of additional EPCs and Generic Task Types is the subject of ongoing research, which will be reported in due course. This particular study suggests that, within the 25-85* years age group (*non-demented), younger operatives do seem to be considerably and consistently more reliable than their elder counterparts, by a predictable amount. This may have important ramifications for safety-critical industries which are predicted to have an aging workforce, such as rail.

References

Ardila, A., Ostrosky-Solis, F., Rosselli, M., and Gomex, C. (2000). Age-Related Cognitive Decline during Normal Aging – The Complex Effect of Education. *Archives of Clinical Neuropsychology, 15, 495-513*.

Atchley, P. and Kramer, A. F. (2000). Age-related changes in the control of attendtion in depth. *Psychology and Aging, 15,* 78-87.

Baddeley, A. D., Emslie, E. and Nimmo-Smith, I. (1994). *Doors and people: a test of visual and verbal recall and recognition.* Bury St Edmunds : Thames Valley Test Company.

CIRAS Core Facility (2004). *CIRAS System Report – January 2003 to December 2003*, Ref. 5024928/National Report 04.

Cowell (1998). *Assessing the risks associated with fatigue in railway safety-critical tasks: Guidance for Railway Managers*. Health and Safety Executive Contract Report No. AM5122/431/03, April 1998.

DERA (2000). *The Visual and Mental Acuity of Drivers*. DERA CHS DERA/CHS/PPD/CR000323

Detweiler, M.C., Hess, S.M. and Ellis, R.D. (1996). The Effects of Display Layout on Keeping Track of Visual-Spatial information. In W.A. Rogers, A.D. Fisk, and N. Walker (Eds.), *Aging and Skilled Performance: Advances in theory and applications,* pp.157–184. Mahwah, NJ: Lawrence Erlbaum Associates.

Fernandes, M. A. and Moscovitch M. (2003). Interference Effects From Divided Attention During Retrieval in Younger and Older Adults. *Psychology and Aging,* 18, No 2. Pages 219-230.

Gamboz, N., Russo, R., and Fox, E. (2000). Target selection difficulty, negative priming and aging. *Psychology and Aging,* 15*,* 542-550.

Glisky, E.L., Rubin, S.R., Davidson, P. S. R (2001). Source Memory in Older Adults: An Encoding or Retrieval Problem? *Journal of Experimental Psychology,* 27, 5, 1131-1146

Groeger, J.A., Bradshaw, M.F., Everatt, J., Merat, N. and Field, D. (2003). *Pilot study of train drivers' eye-movements*. University of Surrey Technical report for Railway Safety, London.

Kausler, D. H. and Puckett, J. M. (1981). Adult age differences in memory for sex of voice. *Journal of Gerontology,* 36*,* 44-50.

Kirasic, K. C., Allen, G. L., Dobson, S. H. and Binder, K. S. (1996). Aging, Cognitive Resources and Declarative Learning. *Psychology and Aging,* 11*,* 658-670.

Kvavilashvili, L., Kornbrot, D. E., Mash, V., Cockburn, J., and Milne, A. (2005). Effects of age on prospective and retrospective memory. *Manuscript in preparation.*

Little, A.D. (1997). *Assessment of Human reliability in the use of Driver Reminder Appliances*. Report for Railtrack plc.

Lock, D. and Li, G. (2003). *An Analysis of SPAD Occurrences in the Southern Zone in Comparison with UK National Data*. Human Engineering Limited, Report HEL/RSSB/03963/RT2 for Rail Safety & Standards Board.

Madden, D. J and Langley, L. K. (2003) Age-Related Changes in Selective Attention and Perceptual Load During Visual Search. *Psychology and Aging*, 18, 1, 54-67.

Maidment, D. (1998). Responding to Public Criticism of Safety Management Systems – is the Response Always Effective and Appropriate? In Hale, A., Wilpert, B. and Freitag, M. (Eds.) *After the Event – From Accident to Organisational Learning*. Oxford: Pergamon.

Mantyla, T. and Backman, L. (1992). Aging and memory for expected and unexpected objects in real-world settings. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 6, 1298-1309

Maylor, E. A. and Lavie, N. (1998). The influence of perceptual load on age differences in selective attention. *Psychology and Aging,* 13*,* 563-573.

Maylor, E.A., Vousden, J. I., Brown, G. D.A. (1999) Adult Age Differences in Short-Term mewmory for Serial Order: Data and a Model. *Psychology and Aging, 14,* 4, 572-949

McIntyre, J. S., Criak, F.I.M. (1987). Age Differences in memory for item and source information. *Canadian Journal of Psychology 41*, 175-192.

Morrow, D., Leirer, Altieri, P. and Fitzsimmons, C. (1994). When Expertise Reduces Age Differences in Performance. *Psychology and Aging*, *9*, 1, 134-148.

Ostrosky-Solis, F., Jaime, R. M., and Ardila, A. (1998). Memory Abilities during Normal Aging. *International Journal of Neuroscience, 93,* 151-162.

Park, D.C. , Smith, A.D., Dudlerm W.N. and Lafronza, V.N. (1989) Effects of age and a divided attention task presented during encoding and retrieval on memory. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 15, 1185-1191.

Parkin, A. J. and Walters, B. M. (1991). Aging, short-term memory and frontal dysfunction. *Psychobiology, 19,* 175-179.

Rankin, J. L. and Collins, M. (1985) Adult Age Differences in Memory Elaboration. *Journal of Gerontology, 40*, 451-458

Salthouse, T. A., Hambrick, D. Z., Lukas, K. E., and Dell, T. C. (1996). Determinants of adult age differences on synthetic work performance. *Journal of Experimental Psychology: Applied, 2,* 305-329.

Salthouse, T. A. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science, 13,* 140-144.

Shinar, D., Tractinsky, N. and Compton, R. (2005). Effects of practice, age and tasks demands on interference from a phone while driving. *Accident Analysis and Prevention*, *37*, 315-326.

Singer, T., Lindenbergerm U and Baltes, P. B (2003) Plasticity of Memory for New Learning in Very Old Age: A story of major loss? *Psychology and Aging, 18*, 2, 306-317.

Sliwinski, M. J., Hofer, S. M., Hall, C., Buschke, H., and Lipton, R. B. (2003). Modeling memory decling in older adults: The importance of preclinical dementia, attrition, and chronological age. *Psychology and Aging, 18,* 658-671.

Smith, A. (1993). Viral illnesses and performance. In Attention: Selection, Awareness and Control. A Tribute to Donald Broadbent. Edited by A.D. Baddeley and L. Weiskrantz. Oxford University Press.

Smith, A. (1995). Colds, flu and performance. Occupational Health Review. November/December 1995, 29-32.

Taylor, J. L. O'Hara, R., Mumenthaler, M.S., Rosen, A.C. and Yesavage, J.A. (2005). Cognitive Ability, Expertise, and Age Differences in Following Air Traffic Control Instructions. *Psychology and Aging*, *10*, 1, 117-133.

Whiting, W. L. and Smith, A. D. (1997). Differential age-related processing limitations in recall and recognition tasks. *Psychology and Aging, 12,* 216-224.

Williams, J.C. (1985). *HEART - A Proposed Method for Assessing and Reducing Human Error*. In Proceedings of a Symposium on the Achievement of Reliability in Operating Plant, Safety and Reliability Society, 16th September 1995, Southport.

Williams, J.C. (1988). A data-based method for assessing and reducing human error to improve operational performance. In Hagen, E.W. (Ed). *Proceedings of IEEE Conference on Human Factors in Power Plants*, pp. 436-450, Monterey, California, June 5-9.

PART XIII
SPADs – Signals Passed at Danger

This page intentionally left blank

Chapter 47

The Impact of Returning from Rest Days on SPAD Incidents

W. Huw Gibson, John Shelton and Ann Mills

Introduction

This chapter is concerned with signals passed at danger (SPADs) which are caused by some form of driver error (termed Category A SPADs in the UK) and are one of the key railway risks focused on by the UK railways (RSSB, 2005). There has been concern about an increase in the risk of SPADs when returning to work following rest days.

The association between rest days and increased risks of SPADs has been identified as an issue through analysis of SPAD data by Wharf (1993), Gibson (1999) and Li (2003). However, Crick et al. (2004) through review of SPAD data, collection of field data and qualitative data collection from drivers did not find strong support for this effect. The studies by Crick et al. (2004) and Li (2003) have issues associated with the normalising data used. For this study, a new source of normalising data have become available from the collection of driver roster data using driver work times collected during a diary study (QinetiQ, 2005). These data provided an important opportunity to look in further detail at the impacts of returning from rest breaks on SPAD incidents and in particular the impacts of rest breaks of different durations.

With regard to the wider research literature related to the impacts of rest days on performance and risk, reviews have been undertaken by Folkard et al. (2004) and Crick et al. (2004). The key findings tend to support the findings from Akerstedt et al. (2000) who states that "one day of recovery is never sufficient, two days usually is, whereas 3-4 days are necessary after periods of severely distributed circadian rhythmicity" (p. 251). This conclusion is drawn primarily from the application of a subjective sleepiness rating scale for personnel in a number of industries involved in a number of different shift patterns. It is therefore primarily related to sleepiness rather than direct measures of risk.

The finding that two or three day rest breaks, when compared with rest breaks of a single day, improve performance (in particular subjective ratings of alertness) is supported by Totterdell et al. (1995). Also, on the first day returning to work after a rest break, benefits were identified across a number of subjective measures such as 'alertness, ' calmness', 'social satisfaction' and 'workload' for breaks of two rather than one day's duration. Data from train drivers are reported in Akerstedt et al. (2000). For train drivers, recovery from sleepiness was complete on the first day off. This was for drivers undertaking irregular work hour patterns involving a range of shift patterns including earlies and lates. This finding appears to contradict the overall conclusions of the chapter, although it may be due to the extremely varied nature of the driver's shift patterns masking any strong effects.

Folkard et al. (2004) also identify that difficulties can arise after several rest days. However, this is only clearly supported by a single reference from Lewis and Swaim (1986) and based on subjective reports on increased problems in shift handover following eight consecutive rest days

when compared to four consecutive rest days. Another interesting finding presented by Akerstedt et al. (2000) is that several days of extended rest may yield unusually low levels of sleepiness. While the chapter notes that the functional significance of this is yet to be demonstrated, it does highlight that longer rest periods lead to an unusual state which may be a factor in accident causation. Perhaps we are too relaxed on return from extended holidays?

Method

Processing SPAD data

The SPAD data used included a description of the activities the driver was undertaking on the days prior to a SPAD. These data enable an analysis to be undertaken to identify the frequency with which a SPAD occurs following a work day or after days not worked. The sequence of non-worked days preceding the SPAD can be determined from the reporting forms, which record driver activities for the 12 days prior to the SPAD.

Table 47.1 present the current categories used to record work/rest activities. The key factor of interest in the analysis was defined to be the simple distinction between days worked and days not worked and their sequence of occurrence. The actual categories presented in Table 1, Column 1 were therefore replaced in the database with three revised categories, working, not working or not identified.

Table 47.1 SPAD categories and the revised classification

SPAD Category	Revised Category Name
A – Annual leave	Not working
B – Booked off	Not working
D –Worked day of rest	Working
N – Worked Sunday	Working
O – Other	Not identified
R – Rest day off	Not working
S – Sick	Not working
V – Worked overtime	Working
W – Worked ordinary time	Working
X – Special Leave	Not working
Z – Worked – Emergency call out	Working
Not identified	Not identified

The total number of SPAD records supplied from 1998 to 2004 were 2938. It should be noted that a full sample of data for 2004 were not available at the time of the analysis. Records were r ejected if:

- The day on which the SPAD occurred had been classified as a day not working, which is an error.

- The 12 day sequence included one or more days where it was not identified whether a person was working or not working.

The total number of SPADs remaining after this was undertaken was 2473. Of these, 600 occurred after one or more days not worked and 1873 were preceded by one or more days worked. The frequency and percentage of SPADs occurring by breaks and break duration are presented in Table 47.2.

Table 47.2 SPADs and preceding days work/not worked

Break duration preceding SPAD	Frequency	Percentage
1 day	333	13.5%
2 days	111	4.5%
3 days	40	1.6%
4 days	20	0.8%
5 days	27	1.1%
6 days	8	0.3%
7 days	5	0.2%
8 days	11	0.4%
9 days	11	0.4%
10 days	5	0.2%
11 days	2	0.1%
>=12 days	27	1.1%
No break prior to SPAD	1873	75.7%
Total	2473	100.0%

Normalising data

The normalising data were collected from drivers as part of the RSSB project Fatigue and Shiftwork (McGuffog, 2005). They were collected as part of a diary study, where respondents kept a record of their shift patterns. There were 105 drivers who returned diaries by the cut-off date for inclusion in the analysis. The diaries were completed by drivers from the following train operating companies (percentage of respondents in sample from that train operating company presented in brackets): Arriva Trains Northern (22.9%); Eurostar (6.7%); London Lines (38.1%); Midland Mainline (15.2%); ScotRail (17.1%). The data come from a variety of Train Operating Companies and they therefore provide a useful cross-section of shift patterns. In total 4140 days were recorded by respondents. Table 3 summarises the frequency data after they have been processed to match the SPAD table above.

Results

The analysis has identified that there is an increased SPAD rate following a rest day when compared with SPADs following a day worked. Chi-square analysis of the SPAD and roster frequency

data identified this difference to be significant (17.2, p<0.00). The data are presented in Table 47.1. The SPAD data have been normalised by dividing the SPAD frequencies by the sampled roster frequencies. The normalised SPAD rates should not be considered as absolute measures in themselves, as they reflect the arbitrary sample size for the rostered data. However, the data are important in relative terms, larger values reflect a higher SPAD rate. The key finding is a 33% increase for SPADs occurring following a rest day.

Table 47.3(a) Days working and not working

Break Duration Preceding a Day Worked	Frequency	Percentage
1	200	7.6%
2	111	4.2%
3	68	2.6%
4	51	1.9%
5	27	1.0%
6	11	0.4%
7	7	0.3%
8	17	0.6%
9	2	0.1%
>=10	19	0.7%
No break prior to a day worked.	2122	80.5%
Total	2635	100.0%

Further analysis was undertaken to investigate how the duration of a break affects the SPAD rate. The data have been chunked due to cell count requirements for chi-square analysis. However, the categories selected serve a useful purpose because they replicate the key daily frequencies identified in the literature and associated with different levels of sleepiness and assumed risk. These are breaks of one day only, 2 day breaks, longer breaks (2 to 7 days) and breaks of 'several days' (> 7 days). The data are presented in Table 4. As with Table 47.1, the normalised SPAD rates should not be considered as absolute measures in themselves, as they reflect the arbitrary sample size for the rostered data. The data are also summarised in Figure 47.1.

Chi-square analysis shows statistically significant differences in the SPAD rate based on the duration of a break (45.6, p<0.001). The highest SPAD rate was when having only a single days break. This presents the rest break duration providing the greatest risk. Breaks of greater than 7 days also lead to higher SPAD risks. The benefits of a break of 3 to 7 days can be seen by the fact that breaks of this duration provide a lower SPAD rate than occurs for SPADs immediately following a day worked.

As data are available for a number of years, it was possible to carry out an analysis to identify if there were differences by year in the frequency of SPADs for the different numbers of rest days. Chi-square analysis determined that there were not significant differences in the frequencies by year (19.52, p = 0.19), suggesting that the findings are consistent over a number of years. This finding requires an assumption that roster patterns have not changed significantly between 1998 and 2004.

Table 47.3(b) SPAD rates following a rest day and following a day worked

Data Type	Following a Rest Day	Following a Day Worked
SPAD Frequency	600	1873
Sampled Roster Frequency	513	2122
Normalised SPAD rate (SPAD Data/ Sampled Roster Frequency)	1.17	0.88

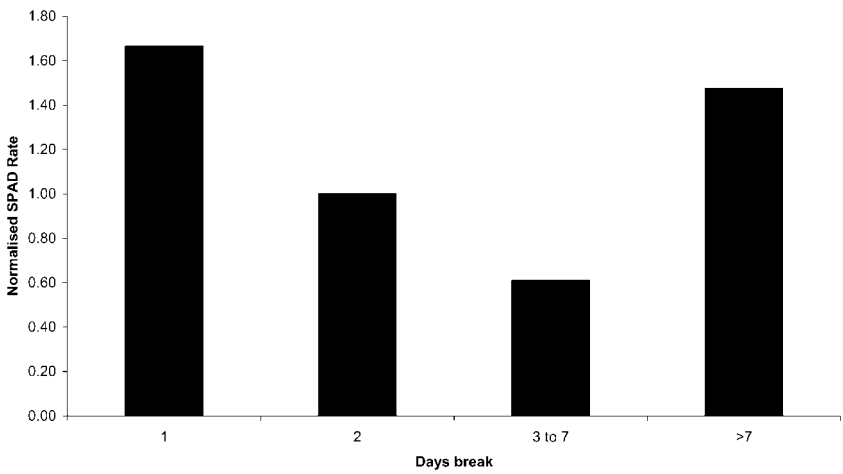


Figure 47.1 SPAD rates for different rest break durations

Conclusions

The key conclusions from the analysis are:

- Breaks do lead to an increased risk of SPADs.
- This increase particularly relates to breaks of a single days duration. This finding is supported by McGuffog et al. (2005) who also analysed SPAD data to investigate this factor.
- There is also evidence of an increased likelihood of SPADs following a long break of seven or more days.
- A decrease in SPAD risk for breaks between 3 and 7 days duration is also identified.

While previous studies have reviewed the relationship between fatigue and accident/injuries (Folkard and Akerstedt, 2004), the data presented in this chapter provide an indication for the impacts of rest days on accident risk. In the context of the wider non-rail research literature this provides a new avenue of support for previous studies into the impacts of rest days which have relied on subjective ratings of factors such as tiredness and alertness (Totterdell et al., 1995; Akerstedt et al., 2000; and Jansen et al., 2003). In the context of the rail industry, it helps define this generally accepted issue in further detail, in particular highlighting the risk from breaks of only a single day's

duration. These data could be used to support fatigue models (e.g. Akerstedt et al., 2004; Roach et al., 2004; and HSE, 1999) as well as used as a consideration in shift planning.

This review has focused on one context factor associated with SPADs. In reality it is well understood that SPADs are mutli-causal. The data would therefore benefit from further examination in the wider context of SPAD causal factors. Secondly, it is widely recognised that there is a general interaction between the requirement for rest days and the types of shift working undertaken up to that rest day. For example, the impact of extended night shifts on the need for recovery is widely reported (e.g. Folkard et al., 2003) and this factor is taken into account in, for example, the UK Health and Safety Executive Fatigue Index (HSE, 1999). Consideration should be given to looking more closely at this interaction, though as noted by Akersted (2000) train driver roster patterns can be varied and therefore be difficult to analyse in this context.

The analysis is also based on the assumption that the sampled roster data reflect the UK driver population roster data. Before any policy decision were made on the basis of these data, consideration should be given to the collection of further normalising data. This could be taken from roster information or a larger diary study. In particular this is required because the diary study sample size was relatively small and from a limited time period.

Acknowledgement

This work was funded by the Rail Safety and Standards Board and the input from the RSSB Human Factors team is gratefully acknowledged.

References

Akerstedt, T., Folkard, S. and Portin, C. (2004) 'Predictions from the Three-Process Model of Alertness' in *Aviation, Space, and Environmental Medicine* 75(3), pp. A75 – A83.

Akerstedt, T., Kecklund, G., Gillberg, M., Et al.. (2000) 'Sleepiness and Days of Recovery' in T*ransportation Research Part F: Traffic Psychology and Behaviour* 3(4), pp. 251 – 261.

Crick, J. et al. (2004) Driver Error Data Collection QinetiQ Report QINETIQ/KI/CHS/TR041855/3.0. Rail Safety and Standards Board: London.

Folkard, S. and Akerstedt, T. (2004) 'Trends in the Risk of Accidents and Injuries and Their Implications for Models of Fatigue and Performance' in *Aviation, Space, and Environmental Medicine* 75(3), pp. A161 – A167.

Folkard, S., Sutton, L. and Yates, A. (2003) 'Impact of Shiftwork and Fatigue on Safety' in (2004) *RSSB Human Factors Research Catalogue* CD Rom Rail Safety and Standards Board: London.

Gibson, W.H. (1999) 'Human Factors Review of Category A SPADs' Number 004 in (2004) *RSSB Human Factors Research Catalogue* CD Rom Rail Safety and Standards Board: London.

Health and Safety Executive (1999) *Validation and Assessment of a Method for Assessing the Risks Arising from Mental Fatigue* Contract Research Report 254/199 H.M. Stationery Office, London.

Jansen, N.W.H., Kant, I., Van Amelsvoort, L.G.P.M., Et al.. (2003) 'Need for Recovery from Work: Evaluating Short-Term Effects of Working Hours, Patterns and Schedules' in *Ergonomics* 46(7), pp. 664 – 680.

Lewis, P.M. and Swaim, D.J. (1986) 'Evaluation of a 12-Hour/Day Shift Schedule' in A Cradle for Human Factors. Proceedings of the Human Factors Society 30th Annual Meeting, Dayton, Ohio. The Human Factors Society, Santa Monica, California, Volume 2, pp. 885-889.

Li, G., Lock, D. and Hamilton, W.H. (2003) 'Analysis of the May/Summer Peak in SPAD Occurrences' Number 537 in (2004) RSSB Human Factors Research Catalogue CD Rom Rail Safety and Standards Board: London.

McGuffog, A, Spencer, M., Turner, C. and Stone, B. (2005) Working Patterns of Train Drivers: Implications for Fatigue and Safety QinetiQ T059. Rail Safety and Standards Board: London.

Rail Safety and Standards Board (2005) The Railway Strategic Safety Plan Rail Safety and Standards Board: London.

Roach, G.D., Fletcher, A. and Dawson, D.A. (2004) 'Model to Predict Work-Related Fatigue Based on Hours of Work' in Aviation, Space, and Environmental Medicine 75(3), pp. A61 - A69.

Totterdell, P., Spelten, E., Smith, L., Barton, J. and Folkard, S. (1995) 'Recovery from Work Shifts: How Long Does It Take?' in Journal of Applied Psychology 80(1), pp. 43 – 57.

Wharf, H.L. (1993) 'Working Time Patterns of Train Drivers and Safety' Number 005 in (2004) RSSB Human Factors Research Catalogue CD Rom Rail Safety and Standards Board: London.

This page intentionally left blank

Chapter 48

Driver Assessment of the Effectiveness of SPAD Mitigation Measures

Tidi Wisawayodhin, Suzanne Heape, Guangyan Li, Claire Turner,
Ann Mills and John Peters

Introduction

Signal sighting is controlled by Railway Group Standards (principally GE/RT8037). In practice, the activity still relies heavily on the professional judgement of the signal sighting committee (SSC) members, as standards cannot cater for the contextual uniqueness of particular signal arrangements.

Numerous mitigation measures have been used to reduce the risk of signals being passed at danger (SPADs). However, the effectiveness of these mitigation measures has never been formally recorded or reviewed. The Human Factors Signal Sighting Framework was developed to assist signal sighting practitioners in arriving at appropriate mitigation measures and to provide some evidence for their effectiveness. Originally developed for all signal types and for retrospective risk assessments, for simplicity of use, the framework has since been re-scoped to cover only signal sighting work carried out during the design and pre-commissioning stage of new signals.

The framework represents a comprehensive source of information regarding solutions to signal sighting arrangements and mitigation measures currently available. The framework allows all signal sighting practitioners to approach signal sighting in a consistent manner. In the years since the development of the framework, many studies have been carried out concerning signal reading and the reduction of SPAD risk. A validation of the framework content was, therefore, necessary to ensure its continued applicability.

Aim

The aim of the study was to validate the effectiveness of the mitigation measures contained within the Human Factors Signal Sighting Framework. This was achieved through the elicitation of train drivers' opinions on the effectiveness of the mitigation measures.

Method

The mitigation measures were validated through the following four stages:

- Stage One: Literature review to ensure continued applicability of the mitigation measures contained within the framework.
- Stage Two: Application of the framework to multi-SPAD signals to arrive at a list of suggested mitigation measures.

- Stage Three: Obtain drivers' ratings of the effectiveness of particular mitigation measures at the same multi-SPAD signals used in Stage Two.
- Stage Four: Comparison of the results from Stages Two and Three.

Stage one: Literature review

A review of relevant research and studies published from January 2003 through to January 2004 was carried out. The work reviewed covered the following topics:

- Signal and sign co-location.
- The effect of train speed on signal and sign reading tasks.
- The effect of the extended use of AWS magnets.
- Signal background design.
- The design and use of banner repeaters.
- The design and use of line identifiers.
- The use of light emitting diodes (LED) in the rail industry.

The new information obtained could provide further validation of the existing mitigation measures, identify additional conditions of use, suggest new mitigation measures, or provide evidence for the rejection of existing mitigation measures.

Signal selection

In order to conduct Stage Two and Three, multi-SPAD signals were selected based on the presence of performance shaping factors (PSFs) identified during the original development of the Human Factors Signal Sighting Framework. PSFs are features present at the signal location that could have an effect on the signal reading task. Multi-SPAD signals with different patterns of PSFs were selected such that all possible PSFs, and thereby all testable mitigation measures, could be evaluated. A total of ten multi-SPAD signals were selected for the driver trials.

Stage two: Application of the framework to the signals selected

The framework was applied to the selected multi-SPAD signals. This generated a list of appropriate mitigation measures in order of priority of implementation, depending on the pattern of PSFs present at the signal location.

Identify testable mitigation measures

In order for the effectiveness of a mitigation measure to be assessed, the PSF that it addresses must be visible on the video footage. All mitigation measures for which this was the case are presented in Table 48.1.

Stage three: Quantitative and qualitative drivers' opinions

Trials were incorporated into TOC safety briefing days to obtain drivers' opinions on the effectiveness of the mitigation measures. Each trial lasted between 1.5 to 2 hours and was held

either before or after the briefing. A total of five workshops were run with 28 drivers from various different TOCs.

Table 48.1 List of testable mitigation measures

Testable Mitigation Measures
Increase the size of the signal backplate by 50%
Increase the height of the signal so it is parallel with the other signals
Alter the height of the signal so it has greater visibility
Install line ID plates
Install spread lenses
Paint signal post white against a dark background or matt black against a light background
Install a co-acting signal
Remove all signal signage from the signal backplate
Increase siding signal height
Where there are multiple parallel signals, vertically stagger adjacent signals systematically
Change signal configuration and type to ensure all parallel signals match e.g. both are 4-aspect signals
Where stanchions obstruct the signal, remove them if possible, or move them further to the left away from the track and signal
Paint the area immediately behind the signal backplate white
Where an extra line is visible, for example one leading to a siding or yard, provide a blank backboard
Relocate the signage that interferes with the driver's signal reading task
Remove sources of obstruction as far as possible
Remove all retro-reflective signs that are near the signal aspects
Move signal from left to right side of the track (may be useful where there is a sharp left hand curve or reverse curve)
On a right hand curve, move the signal further to the left
Extend the gantry to include the siding aspect
Beyond a bridge locate the signals in a straight line and as low as possible
Move post signal to ensure that all signals appear parallel
Move the signal further away from the tunnel exit
Locate the signal at ground level inside the tunnel exit
Provide small, dim lights or paint small bricks with retro-reflective paint at regular intervals along the inside wall of the tunnel
Re-locate the signal closer to the junction points ahead
Where the signal after the junction is visible from the junction signal, shield or relocate it
Install a banner repeater to warn drivers of a signal ahead
Install countdown markers
Provide regular briefings and training
Reduce the line speed

In order to introduce the study to the participants and ensure that they understood what was required, each participant was given a briefing note to read on arrival. A presentation outlining the background, the mitigation measures to be tested, the rating process, and a worked example were also given.

Procedure

The participants were presented with video footage of the selected multi-SPAD signals against which the mitigation measures were to be evaluated. Each video clip commenced 2 signals on approach to the target signal and was presented at a normal approach speed. The footage was shown from start to finish to enable the target signal to be identified by the participants and to give them the opportunity to familiarise themselves with the signal setting. The footage was presented for a second time and paused at the point when the AWS horn sounds. In those cases where the signal was not visible from the AWS marker, the footage was paused at the point when the signal first became visible. The participants were then asked to evaluate the effectiveness of the proposed mitigation measures using a rating card.

Picture cards

Picture cards were developed and presented to drivers when clarification was sought on the meaning of the mitigation measures proposed. The majority of the picture cards showed a before and after implementation picture for the mitigation measure (Figure 48.1) and some gave examples of the mitigation in existence (Figure 48.2).

Rating cards

The drivers assessed the mitigation measures against the target signal by rating its effect on a seven-point scale (Figure 48.3) after viewing relevant signal footage. The scale allowed the participant to indicate whether they think the mitigation will make the sighting of the signal worse or better. Comments could also be made in the space provided on the card.

Figure 48.1 Examples of before (a) and after (b) picture cards illustrating 'backplate enlargement' mitigation measure

Figure 48.2 Examples of the implementation of banner repeaters

ID No. 2

Backplate Enlargement

Increase the size of the signal backplate to increase the signal to background contrast.

1	2	3	4	5	6	7
Drastically worsen sighting	Noticeably worsen sighting	Slightly worsen sighting	No change	Slightly improve sighting	Noticeably improve sighting	Significantly improve sighting

Comments:

Figure 48.3 An example of a mitigation measure rating card

Following the completion of the mitigation measure evaluation, a group discussion was led by the researcher where topics such as the methodology used, the target signals presented and the mitigation measures proposed were discussed. Any comments made by the participants were recorded.

Data analysis

It was hypothesised that:

- Experienced drivers (2+ years of fully qualified mainline train driving experience) would have a similar opinion regarding the effectiveness of a mitigation measure when applied to a particular signal.
- Their judgments of the effectiveness of the same mitigation measure against different signals would be comparable, such that for signals displaying similar attributes, mitigation effectiveness will be scored at a similar level.

It was judged that inexperienced drivers may not be able to provide reliable opinions on the effectiveness of the mitigation measures. Their data were, therefore, excluded from the analysis. Of the 28 drivers who participated in the study, two had less than 2 years experience. Their scores were removed from the data pool leaving 52[1] sets of data for analysis. The non-parametric test, Spearman's Rank Correlation Coefficient, was conducted to test the two assumptions stated above.

Stage four: Comparison of the results from stages two and three.

A comparison was made between the mitigation measures tested. Only those that received an average score of 5 or more from the drivers were considered to be effective mitigation measures. The scores also provided material for the prioritisation of the effectiveness of the tested mitigation measures.

It was expected that the mitigation measures proposed by the framework would match those favoured by the drivers and that the priority order given to the mitigations by the framework would reflect the priority order proposed by the drivers.

Results

Correlation of drivers' opinions

The correlation analysis of the rating scores from different drivers for each mitigation measure has shown that there is a significantly high level of agreement amongst experienced drivers as to which mitigation measures are effective at a particular signal, at 0.01 confidence level ($r = 0.62$, $p<0.01$). The analysis of signal specific scores has also found that experienced drivers' opinions on the effectiveness of the same mitigation measure against different signals were comparable at 0.01 confidence level ($r = 0.52$, $p<0.01$). This suggests that the data obtained are meaningful and can be used reliably to validate the effectiveness of the tested mitigation measures.

Effectiveness of mitigation measures

A comparison was made, for each signal, between the mitigation measures voted as effective by the drivers and those suggested by the framework. Ten multi-SPAD signals were used to test the effectiveness of the mitigation measures (see Table 48.2).

At seven multi-SPAD signals, more than half (50%) of the mitigation measures suggested by the framework were viewed as effective by the drivers. Out of these, all of the mitigation measures suggested at two signal locations by the application of the framework were matched with those that drivers viewed as effective (100%). At three multi-SPAD signals, the match between the effectiveness rating of the mitigation measures and those suggested by the framework was less than half, but all were higher than 40%. There were a number of mitigation measures (up to 4) that were viewed as effective by the drivers at improving signal visibility and readability at various signals that were not suggested by the framework. These were added into the framework. The ratings obtained from the drivers were also used to prioritise the tested mitigation measures contained within the framework.

1 Each driver rated 2 signals (28 x 2 = 56). Data for 2 drivers were removed, therefore 4 sets of data did not qualify for the analysis, leaving 56 – 4 = 52 data sets for analysis.

Table 48.2 Comparison of the mitigation measures rated effective by the drivers and those suggested by the application of the framework

Signal no.	No. of testable effective mitigation measures identified by framework	No. of effective mitigation measures rated by drivers	No. of mitigation measures rated effective by drivers and suggested by framework	% suggested by the framework and viewed as effective by the drivers.
K443	20	12	10	50
K467	13	11	8	61.5
L1213	8	12	8	100
L460	12	6	5	41.7
MP393	13	7	6	46.2
NK401	17	8	7	41.2
N129	20	13	13	100
N224	18	13	10	55.6
T20	14	11	7	50
WS21	12	10	9	75

Discussion

It was anticipated that the mitigation measures identified in Stage Three would be similar to those generated through the use of the framework in Stage Two. The resulting 50 – 100 % agreement between drivers and the framework on effective mitigation measures at 7 out of 10 multi-SPAD signals tested suggests that the majority of the time, the SSC members and drivers are in agreement. This indicates that the mitigation measures suggested in the framework reflect drivers' opinions, as well as those of SSC members.

The discrepancies between SSC members' and drivers' opinions on the effectiveness of the mitigation measures have highlighted the importance and value of expert knowledge within signal sighting committees and in the application of the framework. The specific locations of the PSFs, how they affect the signal visibility and readability, and the expertise of the SSC members, are all required to select suitable mitigation measures.

In general, the results indicate that the participating drivers view the mitigation measures suggested by the framework to be effective. From the comparison, the results indicate that the drivers could capably identify the problems presented at these signals. This suggests that their involvement at an early stage of the design process could provide valuable input.

Conclusion

The aim of the project was achieved through the validation and prioritisation of the effectiveness of the mitigation measures contained within the Human Factors Signal Sighting Framework by train drivers.

The use of video footage of signals, rather than still images, ensured that the drivers assessed the effectiveness of the mitigation measures in a more realistic setting. Therefore, the results are more meaningful and more directly related to the real world.

A number of modifications were made to the Signal Sighting Framework as a result of the driver trials. A total of sixteen mitigation measures, which were not applicable to the new scope of use or not substantially supported by evidence, were removed from the framework. Twenty-one mitigation measures and conditions of use were modified as a result of the literature review and discussion with the drivers.

The ratings allowed mitigation measures to be ranked and prioritised based on quantitative measures. This was used to re-prioritise the mitigation measures within the framework.

The mitigation measures have been validated and drivers' knowledge and expertise have been incorporated into the framework making it a potentially powerful tool for improving signal positioning at signal design stage.

Chapter 49

Predicting the Causes of SPAD Incidents

Linda Wright, Mark Dabekaussen and Tjerk van der Schaaf

Introduction

The common cause hypothesis or CCH (that the patterns of failure causes of accidents and near misses are the same) as first proposed by Heinrich (1931) has been little tested over the years (see Hale 2000 and Wright 2002). Despite the lack of testing and evidence for the model the theory appears logical and is illustrated in Van der Schaaf's (1992) accident causation model (figure 1 below). Although van der Schaaf's accident causation model provides a common sense view that near misses and actual accidents have the same root causes, the need remains to establish whether or not near misses are indeed precursors to later actual accidents: this 'common cause' assumption is the foundation of near miss reporting systems which are found in more and more domains these days. If accidents and near misses have widely different causal pathways then the collection and analysis of near misses will not help decrease the number or severity of accidents. This is an area of great importance for all sectors and may have implications for the usefulness of near miss reporting (van der Schaaf & Wright, 2004).

Below we describe a test of the CCH using data gathered by the Dutch railways. The data comprised recent incidents (from 2001 to 2004) where a train passed a signal at danger without permission (commonly abbreviated to SPAD – Signal Passed at Danger). These incidents (both their causes and consequences) were categorised according to a SPAD-specific version of the Bow Tie model (Ale, 2002) which consists of a combination of a Fault Tree and an Event Tree.

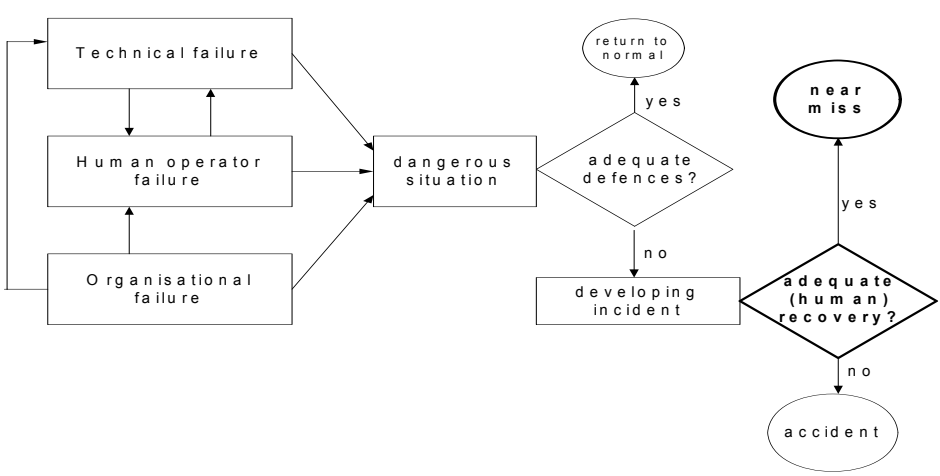


Figure 49.1 The PRISMA incident causation model (van der Schaaf, 1992)

The SPAD Bow Tie Model

The Dutch railway sector has developed the principles of the Bow Tie model specifically for SPAD incidents. This SPAD Bow Tie model is used to analyse and identify the causes of SPADs, and due to its specificity is not able to be generalised to any other event type. The model includes both causes and consequences, and context factors are also included in the final database (e.g. age and experience of the driver). The model includes 10 main causal factors and 10 possible consequences. It is possible to have both multiple causes and multiple consequences. An example of the SPAD Bow Tie model is shown in Figure 49.2 with the main causal factors only.

Each of the 10 main causal factors as shown above comprises a number of sub-categories. The consequences do not break down into further categories. Figure 49.3 shows the sub-categories of causal factors for the main category of 'braking operation'. In total there are 63 sub-categories of causes which comprise the SPAD Bow Tie model. The sub-categories are not mutually exclusive although each category can be used only once.

As the model can only be used to analyse SPAD incidents and no others it means that in effect the causes of SPADs cannot be compared to the causes of other accident or incident types. While the model is not wholly scientifically underpinned, it does include similar causal factors as used by other countries (such as the UK) to analyse and classify SPADs (see Dabekaussen 2005).

SPADs are classified by staff from the Dutch Railway Inspectorate (IVW), who are experts in the use of the SPAD Bow Tie model. The classification is based on data comprising statements by those involved and investigation reports. An inter-rater reliability trial was performed to ensure that the categorisation of the data was reliable. Three independent IVW-raters were involved in

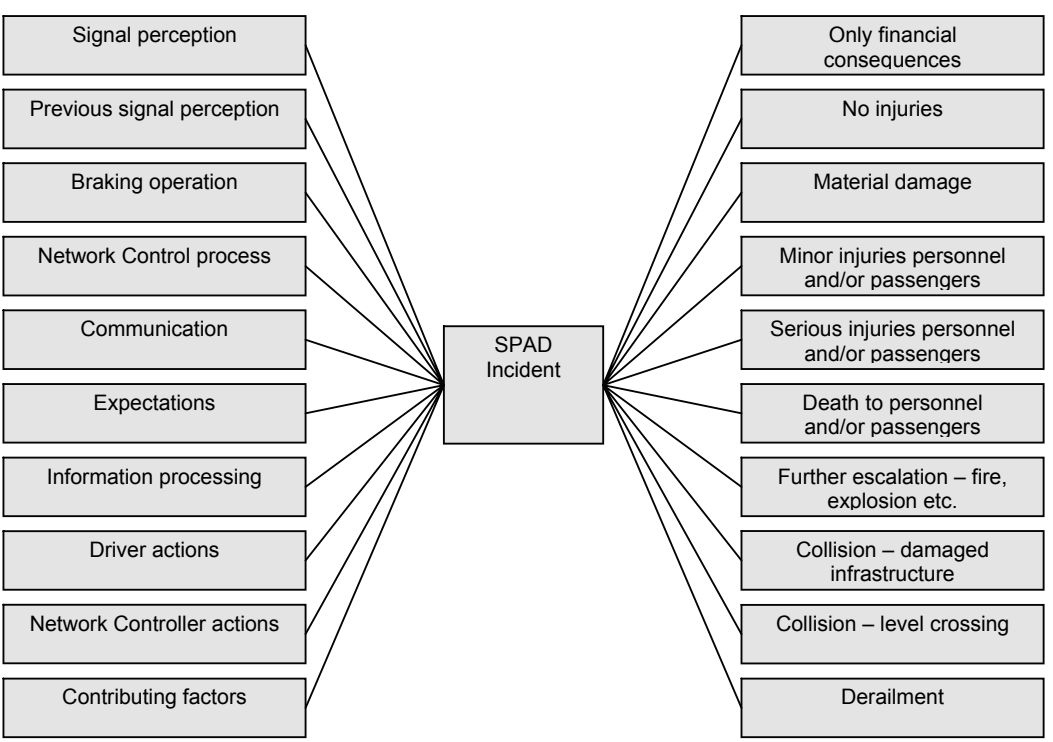


Figure 49.2 The main components of the SPAD Bow Tie Model (IVW 2005)

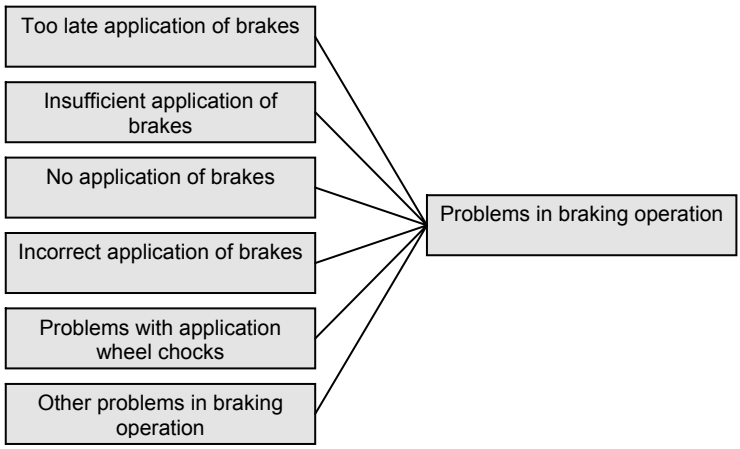


Figure 49.3 The sub-categories contained in the main category 'braking operations'

the trial using a randomly selected sample of 20 incidents. The result was a Kappa of 0.49, which is moderate agreement according to Landis and Koch (1977). After discussion of the results and the reasons for certain categories being chosen a second inter-rater reliability trial was performed using a further 20 incidents. The result was a kappa of 0.69 which is substantial.

Criteria for testing the CCH

In order to properly test the CCH certain criteria need to be fulfilled (Wright, 2002). It is necessary to have a range of consequences which can be compared for their distribution of causal factors. While the most ideal test historically would be to have similar consequence categories as originally proposed by Heinrich (1931), it is more important that discrete levels of consequence can be compared, such as incidents resulting in a consequence versus no consequence, or no consequence versus injuries.

This has not been performed in most of the previous tests of the CCH (see Wright & van der Schaaf, 2004 for a review).

The unit of analysis for performing the test is not the incidents themselves: the incidents are simply the data source for the identification and subsequent classification of their causal factors. However, using the *absolute* number of causal factors assigned to the number of incidents is likely to be misleading due to the fact that serious incidents are more likely to be more deeply investigated than minor incidents with no adverse effects. This tendency commonly leads to the situation in which serious incidents have many more causes assigned to them than inconsequential near-misses. Therefore any statistical differences between absolute frequencies could be attributed to either a real difference in causation or a difference in investigation methods. In order to avoid this problem the *relative* proportions of causal factors are compared.

Method

The data used for testing the CCH were collected and categorised by the Dutch Railway Inspectorate IVW. Data were available on 648 incidents which had occurred in the years from 2001 to 2004. Specific information such as statements from drivers and other involved staff, is required to be

submitted to the Inspectorate. The analysts at the Inspectorate use this information to code the SPADs according to the SPAD Bow Tie model.

Data were included in the sample if they fulfilled the following criteria: at least one causal factor of the 10 main categories in the SPAD Bow Tie model was present, and the actual consequence was clearly indicated. As the SPAD Bow Tie model also includes consequences the data were already categorised by the Inspectorate for the actual consequence following the occurrence of the SPAD.

The consequences in the SPAD Bow Tie model can range from no adverse effect to deaths and injuries as a result of a collision or derailment. Fortunately there have been no deaths and very few injuries on the Dutch Railways as a result of a SPAD. Therefore it was necessary to divide the incidents into discrete groups based on the consequences which had occurred. It was decided to test the hypothesis using only the two categories of 'consequence' and 'no consequence'. The 'consequence' group comprises damage to infrastructure, derailments and injuries to trackworkers and/ or their equipment. The 'no consequence' group includes SPADs which resulted in a delay but no damage to assets or injuries. As delays are almost inevitable after a SPAD occurs (due to the initial form filling required by the Network controller) it was decided to include these in the 'no consequence' group.

Results

Of the 1181 incidents from 2001 to 2004, 648 SPADs fulfilled the criteria and were included in the analysis. The table below shows the distribution of SPADs to the 'consequence' and 'no consequence' groups with the average number of causal factors assigned.

Table 49.1 Number of incidents and mean number of causal factors

Group	No. of incidents	Mean no. of causal factors per incident
Consequence	107	2.7
No consequence	541	2.9

Figure 49.4 below shows the relative distribution of the 10 main causal factors for each group of SPADs as proscribed by the SPAD Bow Tie model.

Data analysis was performed using Chi square for proportions (Fleiss 1981) to test whether significant differences existed between the two groups of 'safety consequence' versus 'no safety consequence'. The results are shown in Table 49.2.

From a total of 1873 causal codes, distributed over 10 causal factors a significant difference was found for only one causal factor: the behaviour of the 'network controller'.

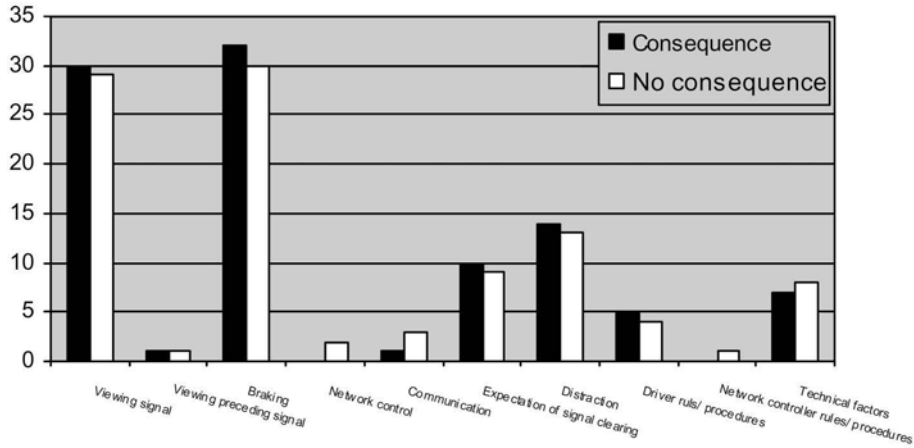


Figure 49.4 Percentage of causal factors per category

Table 49.2 Results of chi-square for proportions

Causal code	Chi square	Significance level (2 tailed)
Viewing signal	0.07	P = 0.786, n.s.
Viewing preceding signal	0.02	P = 0.893, n.s.
Braking	0.56	P = 0.454, n.s.
Network control	6.4	P = 0.011
Communications	2.79	P = 0.095, n.s.
Expectations	0.55	P = 0.460, n.s.
Distraction	0.2	P = 0.655, n.s.
Driver rules/ procedures	0.56	P = 0.453, n.s.
Network controller rules/ procedures	1.73	P = 0.189, n.s.
Technical factors	0.79	P = 0.372, n.s.

Limitations of the study

In terms of the testing of the CCH the data are limited in a number of regards. Firstly, for an optimal test of the CCH it would be preferable that the consequence groups conformed to the original 1931 hypothesis i.e. that sufficient numbers of incidents were available for the groups 'major injuries', 'minor injuries', and also for the 'property damage' category as later added by Bird (1976). As already stated, despite data being selected for four years, there were (fortunately) too few serious consequences as the result of a SPAD on the Dutch railway infrastructure. This necessitated the use of combined groups, which while not ideal still provide a valid test of the hypothesis. Secondly, the data used to test the hypothesis are all the same kind of event i.e. SPADs which limits the

generalisation of the results. A stronger test would be to compare a wider range of railway event types.

Discussion

Despite the limitations of the data, this focussed study, like that of Wright (2002), provides support for the common cause hypothesis. The study by Wright (op cit) compared near misses and accidents in the UK railway (including SPADs and other types of railway incidents) using the CIRAS causal taxonomy (Davies et al 2000). The results provided support for the CCH, with only three of 21 causal factors being significantly different between the three levels of 'injury', 'damage' and 'near miss'. In all three cases the proportions of causes were greater for the more serious incidents.

There are a number of differences between this study and the previous study. While the previous study compared many different types of railway incidents, this study was limited to SPADs. The results would have been stronger had a wider range of incidents been included. However despite this limitation, the results add support to previous findings.

The taxonomy used in this study is also different from that of Wright (op cit). Her study used the CIRAS taxonomy to categorise the failure factors, which has been applied successfully to many different types of railway incident (collisions with track workers, derailments, SPADs and near misses). As already stated the SPAD Bow Tie model has been specially developed to understand the causal factors that are involved in SPADs. The fact that two different studies using different taxonomies come to similar conclusions supports the CCH more strongly because the results are independent of the taxonomy used. In order to test this further, we plan to perform a test of the CCH using PRISMA-Rail (van der Schaaf & Wright 2005) on a wide variety of railway incidents collected by ProRail, the Dutch infrastructure manager.

Only one factor 'Network controller' was significantly different for the group 'no consequences' as compared to the group 'consequences'. This relatively minor factor (see figure 4) only occurred a total of 34 times, all of which resulted in no consequence. The causal factor 'network controller' relates to the task which the network controller is responsible for but does not include the rules and procedures or the communication between driver and network controller. The main category 'network controller' includes five sub-categories: 'signal re-called', 'signal re-called without communication', 'signal re-called with communication', 'signal re-called communication unknown', and 'other'. On closer examination it is clear that the organisation of this category results in more causal factors being assigned than are present. The data present in the 'signal re-called' category are in fact duplicated in the categories relating to the communication that occurred after the signal was re-called. Despite this duplication, re-analysis shows that this cause occurred 18 times and remained a significant difference. This problem of the classification model is under discussion with IVW and hopefully such duplications of the data will be removed.

Within the Dutch railway sector there have been relatively few major accidents over the past 10 years. This presents a statistical problem when it comes to providing not only sufficient numbers, but also the correct type of data to test the CCH. There are two possible solutions to this problem: one is to collect data from the European rail community in an attempt to gather sufficient incidents with serious consequences to compare to the large number of near misses which the railway companies themselves collect (although the large organisations which collect data such as the UIC [an international body which is specific to the rail sector] and the newly formed European Railway Agency (ERA) only collect data on serious accidents); and the other is to amalgamate data from the different severity levels over a range of incident types, both within the railway sector and from different industries. Both these solutions raise questions regarding how the data are gathered, what the investigation methods are and which taxonomy should be used to classify the incidents.

In terms of SPAD prevention, this study suggests that effort should be concentrated on the causes which contribute to SPADs most frequently, or can be most easily eliminated. As there are no major causal differences between major and minor SPADs, then the elimination of the most frequently occurring causes should lead to a reduction in the serious SPADs as well as the less serious SPADs.

Conclusions

The results of this study provide further support for the common cause hypothesis, at least in the railway domain, and despite the limitations of the data add further evidence to previous studies .For a total of 1183 causal factors, subdivided over ten main category causes, a significant difference was only found for one relatively infrequent cause 'Network controller'. However, this causal factor was identified only when there was no adverse consequence as a result of the SPAD, making this result difficult to interpret. Although this study adds more evidence to the applicability of the CCH, we plan to further test the hypothesis on more data using the PRISMA taxonomy.

References

Ale, B. J. M. (2002) Risk assessment practices in the Netherlands. *Safety Science* Vol 40, issue 1 – 4, p. 105 – 126.

Dabekaussen, M. (2005) On track for reducing SPADs: an analysis of the root causes of SPADs using PRISMA-Rail. M.Sc. Thesis, Eindhoven University of Technology.

Davies, J. B., Wright, L., Courtney, E., and Reid, H. (2000) Confidential incident reporting on the UK railways: The CIRAS system. *Cognition, Technology and Work* Vol:12, Number 3 pp 117 – 125, 2000, Springer-Verlag London Limited

Fleiss, J. L. (1981) Statistical methods for rates and proportions. New York, John Wiley & Sons

Hale, A.(2000). Conditions of occurrence of major and minor accidents. *2me Séance du seminaire Le risque de defaillance et son controle par les individus et les organisations.* 6-7 novembre. Gis sur Yvette.

Landis, R.L. and Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics 33 Issue 1,pg 159 – 174, 1977*

Van der Schaaf, T. W. and Wright, L.B. (2005) Systems for near miss reporting and analysis.In: *Evaluation of Human Work,* Eds. John R. Wilson and Nigel Corlett pp 877 – 900. CRC-Taylor and Francis Group, Boca Raton.

Wright, L. B. (2001*)* The analysis of railway accidents and incidents: a comparison of their causal patterns. PhD Thesis, University of Strathclyde

Wright, L. B. and Schaaf, T. W. van der (2004) Accident versus near miss causation: a critical review of the literature, en empirical test in the UK railway domain, and their implications for other sectors. *Journal of Hazardous Materials,* vol. 111, pg 105 - 110

This page intentionally left blank

PART XIV
Human Factors Integration and Standards

This page intentionally left blank

Chapter 50

Integrating Ergonomics into Engineering and Engineering into Ergonomics

Mike Carey

Introduction

Network Rail has taken the step of establishing ergonomics as a core engineering discipline in its organisation. This has created significant opportunities for the application of ergonomics within engineering processes in Network Rail. Henceforth, ergonomics has a role at the core of engineering design and system maintenance as well as covering the full system lifecycle including operations and maintenance. At the same time, being part of engineering for a major rail infrastructure company places responsibilities upon the way ergonomics is practised. The author works within the engineering team for the major improvement works on the West Coast Route Modernisation Project. This experience has led to ideas on what is required to maximise the integration of ergonomics practice with engineering processes across the company.

Role of ergonomics in system design

Ergonomics (or Human Factors) has long sought to be accepted as a core element of the system design process. It is contested that without due application of the knowledge and methods of ergonomics, opportunities can be lost to maximise the performance of the human element of a complex system. However, in general engineering design has generally continued to focus on the hard, equipment elements of systems (the specific engineering deliverables) and the opportunity to date for fully integrated involvement of ergonomics has been limited.

Various approaches have been tried to integrate ergonomics into system design including:

- The application of systems thinking to describe man-machine systems and the particular training and staffing considerations (e.g. Singleton, 1974);
- Generation of guidance material for use by engineering designers (Carey, 1994);
- Extending considerations of human reliability in the high hazard industries as a platform to provide wider system level design for human performance (Ball, 1991);

The current predominant strategy being applied to drive forward the adoption of ergonomics methods is the promotion of Human Factors Integration (HFI), which consists of a set of policies, processes and defined deliverables developed for use in the design of military systems (Booher, 2003). HFI is all embracing of human roles, tasks and contributions to system performance.

Progressively, the HFI doctrine is being adapted for use in other non-military domains, such as air traffic control, process control and rail (Bourne and Carey, 2001). However, the experiences of the author suggest that simply specifying this approach is not sufficient to ensure the appropriate

involvement of ergonomics in design. Indeed, the full HFI approach, which is largely designed for novel and bespoke system design, may be over engineered for more conventional non-military engineering projects and may be a barrier to acceptance. More importantly, imposing such approaches without high level organisational support and enforcement is prone to failure.

The position of ergonomics within Network Rail

In 2002, Railtrack (later Network Rail) decided to establish an Ergonomics team. The decision was made to place this team directly within engineering, with the Professional Head of Ergonomics reporting directly to the Chief Engineer. Hence, ergonomics is now placed alongside the other core engineering disciplines that have collective responsibility for engineering assets and performance in Network Rail.

The Ergonomics team has grown from its initial inception to meet current and anticipated demands. The team currently has 2 full-time and 9 part-time technical staff plus 1 support post, which includes staff from a variety of backgrounds, including ex-operations staff that are undergoing cross-training in ergonomics.

The integration of ergonomics into engineering

The engineering function in Network Rail is responsible for defining how the physical assets of the company are managed, maintained and renewed. It is also responsible for maintaining engineering competence and assuring the Board that the railway engineering system is delivering the required level of performance. Unlike other engineering disciplines, ergonomics does not own any particular asset group and hence does not have direct responsibility for asset renewal or maintenance. However, there are many aspects of the systems and processes managed by the other engineering

Value of placing ergonomics within engineering

The decision to incorporate ergonomics into engineering is considered to be of major importance in ensuring the integration of ergonomics into projects. Within an organisation that is committed to being 'engineering led', the position afforded to ergonomics provides authority to ensure that suitable consideration is given to human requirements in design. This authority is supported at a senior level in the organisation. The responsibility of the ergonomics team in essence only then remains to clearly define what needs to be done, when and to what quality.

It should be noted that with ergonomics, there are other alternative parts of the organisation where it could have been located. During the development of a strategy for human factors within London Underground (Bourne and Carey, 2001), it became clear that key stakeholders existed within not only the Engineering Directorate, but also within Operations (Trains and Stations), Human Resources (including Occupational Health) and Safety. However, the difficulties experienced during major underground projects had highlighted how difficult it can be to influence engineering decisions from outside of engineering. For example, challenges on engineering design from a safety perspective tend to be limited to those aspects which can be shown to directly influence safety risk.

Example of ergonomics integration within West Coast

The West Coast programme is responsible for the renewal and upgrade of the West Coast route to enable the introduction of 125mph tilting trains, to increase capacity and to replace or remove life-

expired infrastructure. West Coast Engineering (WCE) was created to provide an engineering lead for specification and acceptance for the teams delivering the works. The organisational structure for WCE mirrors that of the Network Rail engineering executive team in part, with an ergonomics post alongside that of signalling, telecommunications, track, structures, electrification and plant.

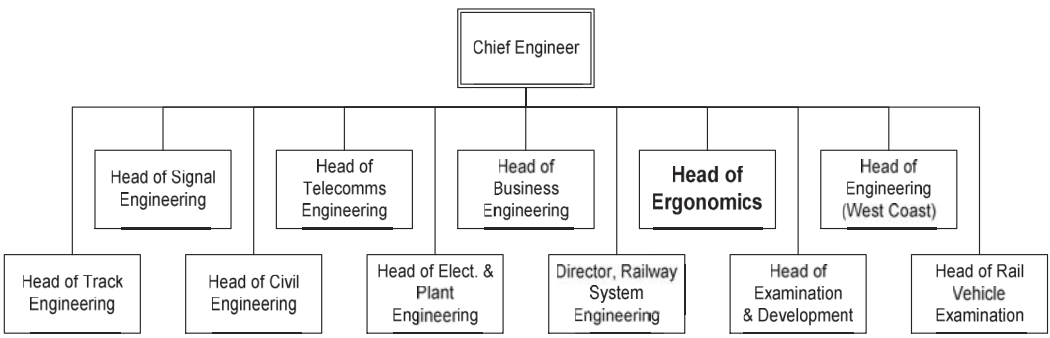


Figure 50.1 Network Rail engineering executive team

Working alongside the other asset engineers, areas have been identified where ergonomics has been given the lead in specification and acceptance. This includes:

- Specific signalling and speed signage issues;
- Signalling centre strategy, design and approval;
- User interface design for new and upgraded signalling control systems;
- Simulator provision and building facilities for new signalling centres;
- Input to the development of a new Automatic Route Setting product;
- Control system display design for a new Overhead Line power distribution system
- System and interface design of new remote condition monitoring systems;
- Installation design for new hot axle bearing detection systems.

Ergonomics has contributed to technical specifications issued to projects. The Senior Ergonomist has also been consulted by other members of the Engineering team on aspects of the strategy for re-signalling projects on the West Coast. Ergonomics also has a significant role in the approval of work on West Coast. This includes being part of the West Coast Project Review Group (WC-PRG). This body is a safety review body for project works. Routine submissions to WC-PRG include:

- *Project Definition Documents* – Highlighting the scope of the project and within it declaring any specific identified human factors issues;
- *Signalling Control Centre Integration Reports* – Demonstrating that projects have managed the technical and operational staged changes to control centres;
- *Operational Readiness Reports* – Presenting evidence that all required aspects of engineering and operational change have been completed prior to commissioning.

With Ergonomics represented on WC-PRG, submission of these documents has provided the opportunity to confirm that projects are giving appropriate consideration to ergonomics issues. The

need for specific ergonomics work streams is also identified. The approach that is then taken has the following main components:

- *Ergonomics Remit* – A remit or specification is written that identifies for the project the works that need to be undertaken. Assistance may also be provided in selecting a suitable specialist contractor to carry out the works;
- *Ergonomics Integration Plan* – A plan of works (or proposal) is required from the project for review and approval by the Ergonomics team;
- *Technical Deliverables* – Depending upon the nature of the change, various technical deliverables may be required. This can include a workload assessment, operational concept, design evaluation report, human reliability assessment, equipment design report, specification for working environment and training needs analysis. These will generally reviewed by the Ergonomics team;
- *Operations Liaison Group* – Liaison with local operations may take a number of forms depending upon the extent of the change. The minimum requirement is to ensure that the local Signalling Manager is consulted, informed and approves changes planned to take place. For the more extensive projects, an Operations Liaison Group is established at an Area Operations level.
- *Summary Report* – If there is a substantive amount of work required for a particular change, then the project is required to develop a single report that summarises all of the works in advance of commissioning, listing any outstanding works remaining to be completed or deficiencies that will be carried into service. This report is submitted for formal approval by the WCE Senior Ergonomist.
- *In Service Review* – Post commissioning, a review is carried out of the installed system to identify snagging issues and deficiencies. This may include a workload assessment and environmental measurements.

Part of the engineering change process is to ensure that all completion works, snagging and handovers are carried out before agreeing to the formal closedown of a project. Ergonomics has a role within this process, particularly for the signalling control centres. An issues database is maintained within WCE Ergonomics and is being used to track and drive completion. This has also involved ensuring that the buildings, services and equipment have been handed over into effective maintenance.

There is little doubt that the position of ergonomics within the WCE team has assisted greatly in ensuring that ergonomics is considered in design. The authority of WCE has been important in achieving some significant changes to the benefit of long term operations. It also has enabled the ergonomics team to ensure that the initial in-service issues with systems are suitably addressed.

Integration into other areas of engineering

Ergonomics now has direct involvement in a number of the major system engineering developments being undertaken by Network Rail. These include:

- The GSM-R train to shore radio system;
- The UK implementation of the ERTMS signalling system;
- The next generation of signalling centres;
- Integrated Control Centres that integrate Network Rail and Train Operating Company operational control staff;

- Initiatives to automate forms for signallers;
- Signalling scheme design tools based on driver workload;
- Evaluation of cab and information systems provided with new on-track plant;

It is the extent of these commitments that has contributed to the growth of the Ergonomics team. The ethos of Network Rail is to bring engineering in-house and for many of these initiatives, the ergonomics work is being led by the in-house team.

Interaction with other areas of Network Rail

As already discussed, ergonomics is also relevant to other areas of a Rail company outside of Engineering, such as Operations, Human Resources, Operational Safety and Occupational Health. The Ergonomics team provides services and are involved in a number of initiatives that cross over into other areas of the organisation such as:

- Improving the management of Signals Passed at Danger and other operational risks
- Initiatives to reduce risks from communication errors;
- Stress management;
- Workload assessment (on request from Operational managers);
- Issues arising with Display Screen Equipment;
- Major incident investigation.

To be effective, ergonomics has to take a whole system perspective and consider organisational influences on human performance and well being. The placement of ergonomics into Engineering has not prevented this broad approach being applied.

The integration of engineering practice into ergonomics

Placing ergonomics within engineering brings responsibilities and challenges as well as opportunities. There is a need to ensure that the focus applied by the Ergonomics team supports the business priorities of the engineering function within Network Rail. The Chief Engineer has defined 'good engineering' as follows:

> Good engineering is decision-making, design and construction that meets business needs and addresses the risks facing the organisation as cost-effectively as possible.

Ergonomics is looked upon to be part of the engineering decision-making process that takes into account the whole-life costs and benefits of any initiatives or changes. To do this it needs to understand the business priorities of the company and the particular risks the company faces. The risks relate to performance as well as safety.

Policies

Within Network Rail policies are required to be published by each engineering group, which indicates the technology preferences and development plans within each area. The Network Rail Ergonomics policy document covers the following wide-ranging areas:

- Control Facilities;
- Human Machine Interaction;

- Job and Task Design;
- Work Organisation;
- Integration with overall railway system;
- Communications;
- Stations and Related Facilities;
- Interaction with the Public.

Inevitably this includes coverage of aspects of assets and operations that are owned by other groups within and external to engineering. Ensuring that the policy is met requires ongoing collaboration and development initiatives.

The policy defined by Ergonomics needs to targeted effort in those areas that can be shown to pose the most significant occupational or major accident risk or which could contribute most to improvements in performance or costs. This requires an understanding of the extent and impact of deficiencies in facilities, equipment and practices. This understanding is being achieved through extensive surveys of workplaces and progressive analysis of various forms of incident data.

Standards

Within the UK rail engineering environment, standards provide vital controls over the detail of engineering change. Each project is required to comply with all relevant standards and where they cannot be compliant, seek derogation.

Current standards published by Ergonomics establish a framework for ergonomics in engineering projects and state requirements for control centres (Network Rail, 2004a and 2004b). Other standards are under active development including a standard on alarm management and on standardisation of control desk design. Ergonomics also is providing an input into the standards of other asset groups. For example, there is interaction between Signal Engineering and Ergonomics on the content of the standard governing VDU-Based Signalling Control Systems (Network Rail, 2003). It is inevitable and appropriate that a significant proportion of ergonomic requirements should form part of asset-based standards and not be part of an isolated set of ergonomic standards.

Continuous development is required to refine the standards owned by Network Rail Ergonomics. Standards provided by other areas of engineering are highly specific and use the derogation process to ensure that all deviations are approved and tracked. The current standards within Ergonomics rely more upon direct involvement with projects to manage compliance and deviations. Only with experience of project applications, will it be possible to consider developing requirements that are sufficient, scaleable and prescriptive.

Commercial framework

The commercial arrangements for renewals works have recently undergone a major overhaul within Network Rail. An in-house project management capability has been established that will form the 'hub' of all renewals projects. Most of the key design and construction resources will remain external to Network Rail but managed by the hub team via a structured set of standard contracts.

A key challenge is to ensure that ergonomics does not suffer due to the contractual relationships established between the various parties involved in infrastructure renewal. One strategy has involved the development of a single contract for the control systems and related furnishings in major control centres. This is intended to ensure that there is a single point of responsibility for the configuration control of the control centre as further areas of re-signalling are added. However, other

works such as the building, telecommunications and the signalling system interlocking renewal will be performed by other parties. Delivering effective ergonomics will require co-operation and integration of all parties across the contractual boundaries.

To achieve this, Network Rail Ergonomics has been directly consulted in the drafting of the contracts to determine how ergonomics work will be addressed. The option has been left open to perform the ergonomic studies through the hub Network Rail team or as part of the contract of the control systems supplier. Whichever route is chosen as advantageous on a particular contract, all parties are contractually bound to co-operate with the ergonomics studies and to take on board its findings. Network Rail Ergonomics remains directly involved in such decisions, with a responsibility to balance the technical risks against the commercial risks in all decisions that are made.

Competence management

All Network Rail engineering disciplines have specific requirements for the competence of designers and all parties involved in making changes to the infrastructure. Competence requirements are defined in various Network Rail and Railway Group standards, with specific focus on those acting as a design authority.

The importance of supplier competence is recognised by Network Rail Ergonomics. In order to be able to work effectively with a project team it is essential that contractors engaged in design work have a good grasp of rail industry technology, terminology and organisation. Fundamental misunderstandings or impractical recommendations can affect the confidence of the project in the design, cause delay and ultimately impact upon the credibility of ergonomics within engineering projects.

It also follows that the contractor needs to have suitable experience and capabilities for the particular design or analysis task. For example, different analysis and engineering design skills are required to specify and procure a control console that will remain in use for upwards of 30 years, to that required to determine the optimum number of consoles from an analysis of tasks and workload.

Network Rail Ergonomics advises projects on suitable suppliers for carrying out specific tasks. These are generally drawn from those organisations pre-qualified under a corporate framework agreement. Support is also provided during the procurement process, with attention paid to the skill and experience of the specific individuals that are being proposed.

With a relatively small pre-qualified supplier base and with the detailed supervision that takes place of all ergonomics design work, it is possible to maintain knowledge of individual and company capabilities. It is possible that this situation may change in the future and processes for qualification and accreditation may need to become more formalised in line with other engineering processes.

Efficiency in design

As previously discussed, a key theme in good engineering is ‘cost-effectiveness’. This applies not only to the products of design, but also to the design process itself. Network Rail Ergonomics has to seek to ensure that any design services that are rendered are appropriate and effective. Network Rail Ergonomics seeks to ensure that the methods proposed in integration plans are proportionate to the design issues being addressed. For example, for the development of a brand new form of control system or device, it might be expected that a suitable simulation or trial would be implemented as part of the design process. However, for the implementation of an established off-the-shelf signalling control system at a new site, providing a full simulation for user testing is

unlikely to be cost-effective. Similarly, where significant new procedures are being proposed, full task analysis, human reliability assessment and trials would be expected to provide the assurance that the procedures are safe. Yet, where established and tried procedures are being applied for first time at a location, it is not deemed to add value to carry out a fundamental and detailed re-evaluation of the procedures.

Whilst the development of new systems that radically change established operational processes can require a suite of analyses to ensure that all aspects of the human role are accounted for, smaller system modifications or deployments can require less sophisticated analysis. Task or scenario analyses, for example, are essential for understanding how roles and activities might change with a completely new type of system. Yet, for more routine renewals, that will apply well understood procedures and training processes, the value of detailed task description is questionable. The focus of analysis needs to be placed on understanding key task characteristics and parameters that affect the design, rather than detailing every task activity. Hence, the methods selected need to be carefully tailored to the risks and uncertainties that are posed by a particular design context.

Design approval

Network Rail operates a structured life-cycle for projects based on defined stages and stage gate reviews. This is defined within a Guide for Railway Investment Projects (GRIP) and consists of eight stages which control projects. Of particular importance is the need to have achieved engineering Approval in Principle (AIP) by the end of single option outline design (GRIP 4), prior to entering detailed design (GRIP 5). This is a critical point for a project as it often marks the point beyond which major funds will be committed externally for detailed design and construction.

Each engineering discipline has a defined process and deliverables necessary for achieving AIP. The need for these submissions is well understood by projects and provides a necessary control over abortive project engineering.

Ergonomics would also benefit from aligning its processes to the approvals framework used across engineering in Network Rail. Not only can ergonomics make use of the control provided by such approvals, but also projects would then receive co-ordinated input to enable them cost, scope and programme any works required. Currently, ergonomics standards are based on a project lifecycle, but do not define a specific requirement for an AIP submission. It is recognised that there is still work to be done to define the nature of submissions required for AIP that can equally apply, for example, to a major control centre design (which has to have completed extensive analysis of operations and layout by GRIP 4 in order to feed into structural and architectural designs) as well as to a set of minor changes to a signal box panel arising from a signalling renewal scheme (which will have defined the signalling scheme in some detail for GRIP 4, but can leave the details of panel changes to a contractor GRIP 5).

For most engineering disciplines, detailed design and construction is delegated to project engineers and no other formal approvals are issued during GRIP 5 and 6. However, within ergonomics, relatively detailed aspects of design can have a disproportionate effect upon human performance and health. The current ergonomics standard does require an Ergonomic Operational Readiness Report to be submitted in advance of commissioning that provides a milestone for approving the fully detailed design (referred to as a Summary Report on West Coast). Progressively, as detailed design elements are standardised (e.g. standardising desks, building systems and user interfaces), the detailed design information required at this stage may be reduced.

It is clear that there is further work to do to ensure that a clear and prescriptive approvals regime can be provided for ergonomics that aligns with the stage gate submissions of other engineering disciplines. This will apply most directly to renewal projects and enhancement projects that are applying conventional technology. System development projects will probably require a very different set of approval deliverables at progressive stages of design. These issues are to be explored in the next planned review of the current ergonomics design standard.

Conclusions

The embedding of ergonomics at the heart of engineering in Network Rail has provided significant opportunities to achieve practical and effective integration of ergonomics into engineering projects. There are further benefits to be achieved by closer alignment of ergonomic design processes with those of the other core engineering disciplines.

Network Rail has not found the need to adopt the full HFI set of processes applied in military projects, though many key management principles have been adapted for use. This is considered to be appropriate for the types of conventional engineering projects that currently dominate the focus of work in Network Rail. It is rarely the case that radically new technology is being developed or new tasks/job roles being created within such projects. Even where new systems are being developed (e.g. ERTMS, GSM-R), the scope for change is limited by interoperability requirements.

Yet, Network Rail has an aspiration to continue into the future the progressive improvements in performance that it has achieved over recent years. It is recognised that once the scope for improvements through better management have been exhausted, further improvement is going to require a transformation in the means by which the railway infrastructure is managed. This is likely to require technical innovation and as a result, means of modelling, quantifying, trading off and measuring the benefits of various options for human functions and tasks. The types of trade-off and appraisal tools being applied to the engineering of military systems are likely to be essential in such circumstances.

References

Carey, M.S. (1994) *Application of Human Factors Requirements to Offshore Facilities Design*. In: *Third European Seminar on Human Factors in Offshore Safety, Aberdeen, 27-28 September 1994*. Business Seminars International, London.

Ball, P.W. Ed. (1991) *The Guide to Reducing Human Error in Process Operations. Human Factors in Reliability Group*, SRDA-R3, The SRD Association, Warrington.

Booher, H.R. (2003) *Handbook of Human Systems Integration*. John Wiley & Sons Inc. New York.

Bourne, A. and Carey, M.S. (2001) Integrating Human Factors into the Development of Railway Systems. In: *People in Control: An International Conference on Human Interfaces in Control Rooms, Cockpits and Control Centres, June 2001*. IEE, London

Network Rail (2004a) I*ncorporating Ergonomics within Engineering Design Projects: Requirements*. Ref: RT/E/P/24020, Issue 2, April 2004.

Network Rail (2004b) *Control Room Design: Specification, Process and Guidance*. Ref: RT/E/S/24017, Issue 2, April 2004.

Network Rail (2003) VDU Based Signalling Control System. Ref: RT/E/S/10067, Issue 2, August 2003.

Singleton, W.T. (1974) *Man-Machine Systems*. Penguin, Harmondsworth.

This page intentionally left blank

Chapter 51

Focussed and Effective Human Factors Integration Programmes for Railways

Martin Dooley, Ned Hickling and Paul Stephens

Introduction

The comprehensive fulfilment of a human factors mission requires the execution of very many human factors activities. The necessary activities are, amongst others:

- The design of human systems interfaces;
- The formulation of operating instructions, procedures and rules;
- The selection and competence assurance of personnel;
- Organisational design;
- The measurement and maintenance of continuous improvement and safety culture.

In practice, the majority of railway projects do not begin with a greenfield (fresh) site. Therefore, effective and responsible human factors engineering requires a clear and justified identification of those human factors processes that are necessary as a result of system modification. The fundamental requirement for such a selective approach is to have a clear definition of the criteria that can be used to establish when a human factors process is required. For practical purposes, a working definition would be "The implementation of an appropriate human factors process is required where significant differences in roles undertaken, or where predicted levels of human performance, implementation of new or modified systems and interfaces give rise to system risk modifications."

Human factors processes that might potentially be invoked because of applying the above criterion are many. These could range from an analytical process such as task analysis or human reliability assessment through the provision of guidelines to a designer. Equally, human factors specialists may be directly involved within a system's design or assessment team. Processes might be directed at human performance in operation, systems design, systems maintenance or performance during decommissioning, or staged system modification and accident scenarios. It is a fundamental misconception to think that the discipline of human factors should only be applied to routine systems operation.

This chapter reflects upon experience of designing and implementing human factors integration plans (HFIPs) within a variety of railway applications over a number of years. These have ranged from the refurbishment and relatively minor modification of existing information systems to the engineering of an entirely new section of railway.

Enabling human factors involvement

It is self-evident that any railway application involves human beings. However, whilst this is true, this does not automatically lead to the explicit recognition of the need for human factors processes

delivered within an HFIP. Indeed, amongst engineering organisations, there sometimes appears to be a working assumption that the human factors issues associated with, for example, new information technology, are already well understood, or can be handled without further professional assistance, and that no further human factors input is necessary. Whilst this might be true for human systems interface design, the introduction of a new information system will most certainly change working practices, workloads and expectations of system user performance. Therefore, if for no other reason, systematic consideration of potential HFIP content should be undertaken. Within the UK there are compelling reasons why a HFIP should be prepared. Not least is the existence of the clients' standards including Network Rail Human Factors Standard (Network Rail, 2003) and the equivalent London Underground Standard (LUL, 2003).

Selective planning

Within the military sector, it is common practice to undertake a comprehensive HFIP for each new engineering or refurbishment project. Taking a strategic overview rapidly reveals why such an approach is almost always appropriate. Typically, military systems have a long lifetime which means that a major change in technology will have occurred. Moreover, major changes in the characteristics of military missions have occurred within the last 15 years. In addition, and of most importance, human performance within military systems is usually mission-critical. Sometimes exceptionally high levels of human performance are expected in a large number of stressful conditions and with degraded capability of the engineered systems. This is often reflected in rigorous personnel selection criteria and continual training and re-training programmes.

Unfortunately, there appears to have been some tendency amongst the UK human factors community to apply the military doctrine in a mechanistic, procedural manner to the very different situation of the railways. Here, the contribution of the human to success is very variable. Often, human effects on mission criticality can be avoided or minimised by the use of technology such as interlocking. However, interlocks only provide protection for trains and many signallers' safety-related tasks are to ensure the safety of the maintainers. These tasks rely almost exclusively on procedural constraints and, thus, human performance. Therefore, it is appropriate to produce an HFIP that is selective in defining the human factors processes deemed to be necessary.

This can be achieved this using a selective planning process, considering the principal human factors domains and HFIP management tasks then cross tabulating these with the defined engineering programme. Each cell of the resulting table is considered in turn and the human factors objectives to be met are identified. At this stage, the specific methods to be used generally remain undetermined but the objectives of a human factors activity are defined. Where particular methods offer themselves as the appropriate choice to fulfil objectives these are put forward for later endorsement. In addition, and of equal importance, the project team members or organisations who are considered to be best placed to deliver the necessary activities are also identified.

This approach has been deliberately chosen, explicitly recognising not only that others may be better placed than human factors specialists to deliver elements of the HFIP, but also that human factors specialists are generally contributing members of a much wider client team or an engineering alliance. This approach does not prejudge the solution to the delivery of an HFIP activity, but rather provides a basis within which project team or alliance dialogue can then occur to arrive at the most appropriate methods to achieve the solution. This is an absolutely essential prerequisite to integration and overall human factors mission success.

HFIP integration with engineering programmes

Our approach to HFIP design, as described above, automatically aligns human factors processes and objectives with stages of the engineering programme and results in project team engagement with the human factors mission. However, this by itself is entirely insufficient to ensure effective integration. Therefore, we work with the specific client to explore the project interfaces; particularly with respect to information transmission and access to prototype or working facilities. This ensures that project dependencies are explicitly recognised and addressed and that any requirements for access to resources (e.g. existing rolling stock or signal boxes) by the human factors specialists, is feasible within the required timescales. Identifying resource access is particularly important for access to staff, where it may be necessary to re-arrange rosters to ensure participation in trials or design assessment workshops. The process also fulfils a crucial requirement of ensuring that outputs from HFIP activities are received by appropriate target elements within the wider project team. Similarly, the project programme for the HFIP explicitly contains dependencies that show where engineering outputs (from analysis or design) are necessary contributors to the successful completion of the HFIP.

This process of integration is far from being a mechanistic project planning activity. Appropriate negotiations must occur so that responsibilities for activity delivery are properly defined and personnel responsible for delivery are suitably informed about the human factors issues that they need to address. It is our experience that the successful completion of the activity is far more heavily dependent upon negotiation diplomacy skills than it is upon technical project planning. In many instances, systems analysts and engineers are encountering a need for an HFIP for the first time. This process of active involvement forms the beginnings of stakeholder buy in and is discussed in further detail below.

Users and super-users

The ultimate client, i.e. the operating railway, is the repository of users and super users. Users are those who generally have current experience and skilled expertise in one area. Super users are those who have experience and expertise in multiple areas of knowledge and have the authority to make, and override, decisions. Both are essential to the execution of a successful HFIP, irrespective of the technical content of the engineering program or the HFIP itself. However, clear relationships need to be set up between the contracting alliance or team and the system end users or super users. For example, there can be potential contractual impediments if the views of users that are required to support human factors activities, can be taken, contractually, as representative of the client's requirements.

As in other industries, it is essential for rail systems to ensure that user desires or needs identified during human factors processes, for example, user requirements capture, rapid system prototyping or formal prototype testing, are suitably moderated before they impact upon the engineering design. This is not done to compromise or reduce the impact of the users upon system design, quite the contrary. However, it may be necessary for an ergonomist or applied psychologist to apply intelligent interpretation to the views that have been expressed. Superficially, this might appear to be a patronising or even godlike perspective. However, this is an erroneous view. Instead, this moderation of inputs is necessary to ensure that the views of one or two users involved in trials are not overly partial and ignore important other opinions. Indeed, one important role of the human factors specialists in any engineering project is to maintain continuity in understanding of users' needs and opinions. The human factors specialist may need to moderate opinions over a far wider spectrum than those involved in any one particular exercise or study. Another role of the human

factors specialist is to ensure that these user requirements are managed, as any other technical requirements would be. The professional human factors role may also be necessary because some users may persist in a "current world" view of systems operation, with its attendant limitations or assumptions, rather than address where the design intentions are actually changed and the future system that is being assessed.

Super users are those who have the contractual power to require implementation. They may also be those who are most likely to induce systems 'scope creep'. From a human factors perspective they, therefore, bring both benefits and limitations to the process. If they are supportive and endorse the human factors mission then they will ensure that human factors findings are incorporated into the design. However, the design team should not simply comply without question with the super users requirements and instructions. It is necessary for them to deliver a human engineered design, which is acceptable irrespective of whether or not the super user has identified a particular requirement. By the same token, the power of the super user may be misused to overturn the opinions of end-users undertaking structured human factors studies. Unfortunately, we have some experience of this happening. When this occurs it is necessary to weigh the super user opinion against, hopefully, more systematically derived views obtained through human factors process. It is particularly important that the super user acknowledges that their opinions may be distorted by the passage of time, because of changes in methods, procedures or general working practice. For example, there may be faster career progression, meaning that staff in a specific role today have less experience than the super-user had when in that role.

Perhaps one area where an improvement is still to be brought about is a clearer understanding, which is set into project procedures, of the relationships between users, super users and human factors personnel. Their respective roles should be formally factored into design reviews and assessment activities. This would then parallel the approach used in many other industries where high human performance levels are required, for example, in aviation, military, nuclear, chemical, banking and finance.

Stakeholder buy-in and commitment

The production of a selective HFIP and negotiating its content and project interfaces with engineering and user team members starts the process which ensures that stakeholders, understand the benefits of the process, their involvement within it and, ultimately, the requirement for delivery. It is fair to say that whilst the level of formal acceptance of the need for the human factors mission continues to improve, the attitude of individuals and their beliefs about the necessity of such processes varies considerably more. Irrespective of contractual boundaries and commercial interfaces, a successful HFIP requires a wholehearted commitment by those involved. It is too easy for members of the human factors community to forget that a process which is novel and to which potential users may feel ambivalence, may not obtain the level of buy-in that the human factors professional would desire and the HFIP process requires! And yet, wholehearted buy-in by non-discipline key stakeholders is a significant contributor to the successful completion of a human factors mission.

Within a project team, the minimum level of buy-in to human factors that should be considered tolerable is where there is clear acknowledgement that compliance with human factors standards is required to obtain systems acceptance. However, the aim should be for more developed and informed buy in. We have found that a key element in achieving this is to demonstrate a clear understanding and appreciation, to those involved, of the engineering and operational missions. In addition, it is necessary to 'give away' human factors insights and expertise so that other team members have ownership of particular human factors processes and issues, rather than the onus resting entirely with the human factors specialists. Of course, if the structured process which we have

described above is applied, then the necessary process of explanatory negotiation is certainly a way to win minds. However, this does not necessarily win hearts! It is not unusual to see project team members who are intellectually convinced of the need to undertake human factors but emotionally seem more reticent. Of course, this should not be too surprising. The ergonomist or human factors specialists can be seen as a challenge or an implied criticism to the role of the industrial designer or engineering designer. Of course this is not the intention, nor should it be the case in practice. However, it is important to recognise the system designers may well also have useful insights about human factors and the consequent solutions should be accepted where appropriate. Otherwise, a clear explanation should be given why the suggested solutions are not considered appropriate.

Another distinction between the mission of the human factors specialist and that of the designer is their time horizon. Throughout, the human factors specialist is concerned with ultimate operational performance. Whilst this is also true in principle for the systems designer, their horizons tend to be more closely defined by the need to solve immediate engineering problems. The skill of projecting the use of the finished design into the future and of anticipating future changing system use may not be one which comes to easily to all design engineers. There is clearly a role for the human factors specialist in assisting with this imaginative projection into future operational and use scenarios.

Delivery verification and validation

A fundamental point about any human factors process is that the passage of time clarifies human performance issues and allows a more accurate understanding of the ultimate levels of human performance to be realised. It is important that the iterative nature of this improvement in understanding and the need for continuous human factors involvement is well understood by the project engineers. Continuity of knowledge about design decisions is essential. For example, if the human factors specialist is withdrawn, an engineering design decision may be made which is believed to have no impact but which, either with closer scrutiny or with greater knowledge, is shown to have a significant impact. Obviously, the easiest way to judge the likely user impact lies with the human factors specialist.

Within the tradition of British systems engineering, it is customary to expect the finished system to perform satisfactorily upon completion. The human factors process introduces verification and validation stages into the design process whereby it can be shown whether or not the required levels of human performance will in practice be achieved. To achieve this demonstration, user testing and prototyping are required. We have found that this best succeeds when a representative range of users are involved rather than just the odd nominal user.

Conclusions

The key to a focused HFIP is the use of a systematic technique applied at the project planning stage. From the outset, it is necessary for the project, specifically systems engineers and project planners, to understand that human factors is not confined to the physical dimensions of desks or rooms or the design of user interfaces but has far wider implications. In itself, the HFIP the planning process helps to achieve that understanding if it is absent or limited.

The existence of a well-formulated HFIP is actually of little benefit unless members of the project team are signed on to the relevance and importance of the human factors mission. In addition, designers, end users and super users need to clearly understand that the human factors specialist is not there to compete with or undermine their opinions and roles. Instead, the human factors specialist is there to complement them and integrate with them within the context of the

systems design process and over arching project requirements management. Whilst this statement may seem obvious to experienced members of the human factors community, it is one that bears repeating at regular intervals within a systems design process. Finally, in addition to the existence of a plan and the intellectual recognition that the human factors mission is of relevance, it is equally important to ensure the design team members are thoroughly convinced on an emotional plane, that human factors:

- Is necessary,
- Is in proportion with other disciplines, and
- Adds something in the currency of the project.

In itself, the existence of a carefully structured HFIP plan helps to achieve this because, like the engineering activities, efforts are focused in a rational and systematic fashion. Those who would wish to apply HFIPs in a dogmatic fashion would do well to consider the extremely deleterious effects that this has. In particular, the execution of systematic activities which provide little or no palpable benefits to the project are seen as making work and time wasting. Real systems engineering has no time for such diversionary activity. Moreover, there are always sufficient surprises and complexities within the domain of human performance to keep the human factors professional gainfully occupied within the issues of systems design, application and consequent operation that really matter.

Ultimately, we believe that our success can be ascribed not only to the use of systematic planning techniques which eliminate inefficient and unnecessary activities, but also the appreciation, experience and acute awareness of the needs of the engineering project requirements to be fulfilled in a timely, effective and economical fashion.

References

LUL (2003) I*ntegration of human factors into systems development.* LUL Number: 2-01018-003 A1.

Network Rail (2003). *Company Specification: Incorporating Ergonomics Within Engineering Design* RT/E/P/24020 Issue: 1.

Chapter 52

Experience of Rail Human Factors in Hong Kong

Karen Priestley

Introduction to Hong Kong Railways

Since 1992 there has been considerable expansion of the rail network in Hong Kong as depicted in Table 52.1. A summary of the characteristics of the main systems as they are today is presented in Table 52.2.

Table 52.1 Major projects on MTRCL and KCRC since 1992

Mass Transit Railway Corporation	Kowloon-Canton Railway Corporation
• Airport Express Line • Tung Chung Line • Operational Control Centre • Tseung Kwan O Line • Disneyland Resort Line • Retrofit Projects including:	• West Rail • Control Centre • Tsim Sha Tsui Extension • Ma On Sha Line • Lok Ma Chau Spur Line • Retrofit Projects including:
Upgrade to Station Control Rooms Platform Screen Doors to existing railway	ATP on East Rail ATO on East Rail

Table 52.2 Summary of system characteristics for Hong Kong Railways

Criteria	KCRC East Rail	KCRC West Rail	MTRCL Urban Lines
Network Length	35Km	30.5Km	91.2Km
No. Stations	14	9	52
Operating hours	19.5	19.5	19
Current Peak headway	180s	180s	126s
System Capacity headway	50s in critical section	105s	105s
Weekday patronage	800,000	170,000	2,500,000
Train service delivery Sept 2004 [a]	99.95%	99.5%	99.9%
Train punctuality Sept 2004 [b]	99.75%	99.4%	99.6%

Note: train service delivery which measures actual train trips against train trips scheduled in a month train punctuality which measures punctual train trips against actual train trips in a month

Much of this work to upgrade the Mass Transit Railway's central control room to VDU technology was undertaken before or during the advent of standards such as ISO 13407 Human Centred Design and ISO 11064 Ergonomic Design of Control Rooms. However, requirements for consideration of human factors issues were built into the specifications of these projects, prior to the availability of such guidance.

Regulation

MTRCL and KCRC are required to maintain a proper and efficient service at all times in accordance with the Mass Transit Railway Ordinance (Cap. 556) and the Kowloon-Canton Railway Ordinance (Cap. 372) respectively. In addition, MTRCL must meet the targets defined in the Operating Agreement made with the Government. Failure to meet the requirements in these documents could result in fines being levied by the Government.

The Hong Kong Railway Inspectorate (HKRI) monitors the safety of the railway systems and ensures that the railway corporations have addressed all safety issues in the design, construction, operation and maintenance of the railway systems. The Transport Department (TD) monitors the reliability of the train services and impact on the overall transportation system of Hong Kong. For new railways, HKRI and the TD will conduct inspections and monitoring activities to ensure that the railway is safe before it can be opened for public service.

It is the responsibility of the railway corporations to demonstrate to the satisfaction of HKRI that their railways are safe for use. Although not explicitly mandated, both MTRCL and KCRC use a risk-based approach to address safety in the design of their railways. They make use of an equivalent process to safety cases and risk assessment in their design process and develop their own internal safety and design standards and safety management systems which are compatible with international standards. Triennial reviews of the safety management systems in line with the requirements of the ordinances are held to ensure that best practice is being followed and to identify potential for continuous improvement.

So far this represents a fairly similar model to that in the UK. The major differences being that MTRCL and KCRC are substantially vertically integrated organisations that encompass design, operation and maintenance of their own systems. This provides a significant advantage over the UK privatised and segmented arrangement with regard to having oversight and control of all the interfaces in the railway system within one organisation. In addition there are no external parties such as the EU or RSSB setting standards or requirements to be met.

Why should Hong Kong Railway companies consider human factors?

From the point of view of the human factors practitioners it may be that we see improvements to the effectiveness of integration of human factors into existing processes providing multiple benefits including identifying and addressing issues early in the design, avoiding costly retrofits, efficient operations and prevention of error.

However, given the modern design of the systems including the incorporation of safeguards in signalling and train design that have eliminated in most part the threat of SPADs and other major safety risks such as derailment, and the fact that in general the projects have been delivered on time and within budget, it might not be unreasonable for the rail companies to take the view that there is no need to do anything more than what they are doing already. One of the arguments to be made here is that although human factors input is included in many of the large projects, better value could be obtained from this input.

From personal experience it seems that there are four main drivers for improvement in human factors input to Hong Kong railways, namely:

- Over-confidence in automated systems
- Media and public scrutiny
- Continuous improvement
- New Technology

Value from human factors input

Despite advantages over the UK industry with respect to simpler organisational structures, the nature of contractual arrangements for large projects can still present barriers to effective human factors input. The issues are not confined to the rail industry nor to Hong Kong and appear fairly universal, similar issues are described by Courteney (1996) with regard to the aviation industry. She presents a table that indicates the trends and practices in the modern business culture that pull in the opposite direction – as 'inhibitors' – to those required for improvement of the human factors in the design process – 'facilitators'.

Table 52.3 Contractual and cultural influences on the success of human factors implementation (based on Courtney, 1996)

Facilitators	Inhibitors
Adherence to the spirit of the requirements	Minimum compliance to the 'letter of the law'
Staff in positions for long periods of time	Staff move on after two or three years
Staff concerned for long term outcome	Staff concerned for immediate outcome
Collective responsibility in the industry	Individual responsibility for contract task
Success measured by Operational feedback	Success measured by immediate cost/time
Delivery date dictated by time required	Demanding, fixed, delivery dates
Strong financial resources, money available	Financial resources tight, minimum costs
Costs attributed to the industry collectively	Costs attributed to separated business units
High staff levels of experience / expertise	Staff with minimum qualifications
Overall tasks completed by single company	Task divided between subcontractors
Preventative activities justified / resourced	Resources only for 'proven' need: 'fire fight'
'Judgements' independent of pressure	Commercial pressure on judgements
Select subcontractor by product quality	Select subcontractor by lowest bidder
Clear responsibility for issue	Issue shared among areas; (e.g. driver error – cab design, signalling, platform design)
Issue is simple to define, bound, resolve	Issue is complex, unbounded and difficult

While in Hong Kong the client staff turnover is low and typically there are good levels of resources available in the client organisation this is not always true for the contractors. Other inhibitors such as the selection of the lowest compliant bid, contractors seeking to meet minimum requirements,

issues crossing contract boundaries and demanding, fixed delivery dates have also been known to influence human factors input. The cost recovery of human factors activities both within a contract and across contracts may be deferred and indirect, the benefits of a smooth interface between human and system are long term and ultimately to the client, with little apparent short term benefit to the contractor.

From a human factors perspective there is another significant difference between the UK and Hong Kong. While HMRI, RSSB, Network Rail and LUL employ human factors personnel directly, this is not the case in the equivalent organisations in Hong Kong. Thus there is no-one with an overview of the system wide, system specific and interface human factors issues, capturing lessons from operations to feed into the requirements for new projects or to brief human factors consultants employed by the various parties on the information they need to know to conduct their studies. This leads to repetition at each project phase between contracts, and possible omission as each party tries to determine the context in which users are undertaking their tasks within the boundaries of their own contract.

As the rail companies in Hong Kong do not employ human factors staff directly they tend to utilise human factors consultants either to advise contract managers or stipulate that contractors use human factors consultant support. This results in one of the three arrangements represented in Figure 52.1.

The first arrangement (Contract A) can lead to confrontation between the human factors specialists assigned; this can result from different motivations of their clients, different understanding of the task requirements resulting from being briefed by different parties and lack of understanding of the context of the client organisation by one or both parties, especially if recruited from overseas. In one instance a human factors acquaintance attended a series of meetings arguing points such as whether yellow text on a red background for VDU display was advocated in ergonomics text books, when as she pointed out they couldn't actually read it clearly on the screen.

The second arrangement (Contract B) often comes about because the client stipulates requirements for human factors consultant support in the contract. However, this does not necessarily instil motivation in the hearts of the contractors. On my first day at the contractor's office to support the design of new rolling stock, I was told in no uncertain terms 'the senior management are not interested in human factors', a little more honest than some, but nevertheless extremely limiting. Banned from speaking directly to the client, it proved almost impossible to raise issues or concerns regarding human factors, especially those identified that might be interface issues with other contractors. The only activity required by the contractor was demonstration of

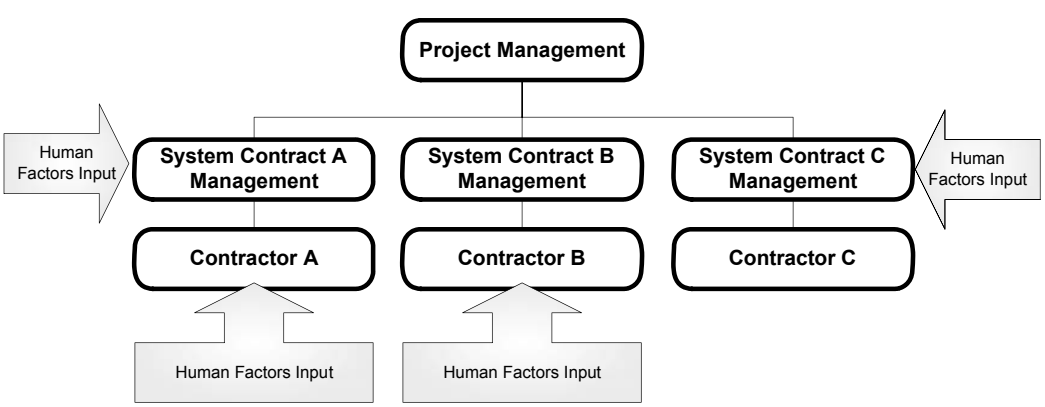


Figure 52.1 Representation of human factors input to projects

compliance with the specification. Although the specification contained many requirements with regard to ergonomics a proportion appeared to be copied from textbooks and irrelevant to the design of trains. For example, details about the design, dimensions and separation of the markings for a slider switch seemed inappropriate for the driver's cab, especially where no slider switches were specified.

The third arrangement (Contract C) can again tend towards the negative with expert review provided on design submissions. At this stage the comments often come too late and may be too vague, because of the distance from the design process, for the contractor to respond to. This was the case on a project to develop a graphical user interface for a decision support system for station controllers. Whilst the design at the milestone submission had captured the flowchart provided by the client on the steps required to respond to a given incident, the interface was virtually unusable because there had been no understanding of the user's task in the situation when the decision support system would be required. Rather than simply send back comments on the design, the project manager requested that a workshop with representatives from the end users, client design team, system assurance, operations management with myself as facilitator be conducted to brainstorm the design. By thinking about and analysing the task the operator needed to undertake in response to an emergency rather than the particular procedure to be followed, the team were able to develop a storyboard using Microsoft Excel to describe how they expected the interface to perform. The concept was developed in a single day.

In fact this collaborative, multi-disciplined approach is that which is advocated by human factors process standards, it achieved many benefits for both the contractor and client, including reduction in wasted time and resources for both sides, a clearer description of client expectations and a resulting product that was more able to meet client expectations and better communication between parties regarding the design.

Unfortunately, as the project progressed and user testing was undertaken on preliminary designs the reception to the feedback gained was a variation order against each change.

Overconfidence in automated systems

Improvements in technology and good operations procedures have brought the Hong Kong railways to a position where they are achieving between 99.5 and 99.95% reliability. One of the consequences of achieving such high levels of reliability can be the conditioning of people to expect total reliability. For example, a recent incident reported in the press involved a KCRC train that failed to stop at a station because the driver forgot that the automatic train operation system was not working. In Hong Kong, passengers do not plan their journeys with sufficient contingency for delay, expecting the system to be reliable. Thus a two minute delay becomes very stressful for many people, especially those working in organisations that penalise staff for being late.

Media and public scrutiny

The modern design and incorporation of safeguards in signalling and train design have eliminated in most part the threat of SPADs and other major safety risks. For those systems that are heavily congested, especially in the peak hours, such as MTRCL and KCRC East Rail, the consequences of a one or two minute delay on congestion can be severe. At around 6 minutes it can become necessary to close some heavily used stations to ensure the safety of those already inside. Any delays over 8 minutes must be reported to the Transport Department.

Thus while the system has been designed to 'fail safe' and avoid injury or harm to passengers and staff, any delay in itself has the potential to become a safety issue. The focus is therefore

on reliability and performance, not just to improve business performance but also to ensure safety.

In the last year the rail companies have been the focus of media attention with respect to the reliability and safety of their services. In September 2004, the Hong Kong media started to report on a 'catalogue of failures' at MTRCL. A number of incidents occurred that received wide coverage in the media; a typical example was a case where the brakes hung on generating smoke. The train was routed to an emergency platform and passengers were safely and efficiently detrained and collected by a following train. There were no injuries and some passengers thanked the staff for their handling of the incident. Yet the perception was widely propagated in the press that the MTRCL network was about to hit a problem with ageing infrastructure and that the reliability and safety of the system was in question. Indeed some of the local press were offering cash rewards for anyone phoning in with information of a delay on the network.

LR Rail were commissioned to undertake an independent review of MTRCL's asset management and performance. As part of this review it was identified that MTRCL did in fact have a good asset management system and that the number of incidents had been steadily declining, 2004 was actually the best year of the four years analysed with regard to system performance.

What the study did reveal was a number of operational human factors issues that have been highlighted by the high reliability of the technical systems and the increased service expectations of the travelling public, namely:

- the importance of preventing human error in response to abnormal events on the railways to minimise any associated delay and prevent them from becoming reportable events (anything over 8 minutes).
- The importance of the management of passenger behaviour in response to abnormal events on the railways (for example using the passenger alarm to ask when the train will be moving again) to minimise any associated delay and prevent them from becoming reportable events (anything over 8 minutes).
- The importance of the management of customer expectations – that 100% reliability is not achievable, 99.9% is world class but still results in some delays.
- The importance of enabling passengers to make informed decisions about their journeys in the event of a delay.
- Raising the bar with respect to safety performance – is it sufficient to 'fail safe' or are we now looking to 'fail safe without distress to customers', thus needing to define how distress might be caused e.g. smoke, loud noises, loss of lighting, uncomfortable rides in the event of component or system failures and how to quantify the value of 'distress' to the organisation for the purposes of cost benefit assessment.

Continuous improvement

There is a strong drive for continuous improvement within Hong Kong railways. Options to increase redundancy in infrastructure and equipment in order to achieve further gains in reliability are disproportionately costly in terms of the associated reliability gains and maintenance costs. As discussed above, minimising human errors with regard to the cause of or extension to duration of incidents therefore becomes a more viable option and requires consideration of less tangible operational interventions such as crew resource management training, fatigue and stress management in addition to design and system improvement.

But as the reliability of the train service becomes harder to improve upon within the physical and cost constraints of the organisation, other ways of viewing and measuring improvement may be required.

Just as in product design where goals have moved from making products usable to making them 'pleasurable', other goals such as avoidance of distress as mentioned above or indeed the achievement of a 'pleasurable' journey may be required in the rail industry. This would require the use of scientific methods to increase understanding of the parameters of these goals (e.g. both pleasurable and displeasurable, distressful and not distressful) and applying these parameters to the design and evaluation of changes, upgrades and the design of railway systems.

Thus for fatigued Hong Kong people who take any given opportunity for a nap, the introduction of noisy TV advertising to transportation systems is not necessarily an improvement to the service and has lead to complaints that it is disturbing the sleep of paying passengers. The balance between commercial needs for advertising and the passenger's 'pleasurable' experience of their journey may therefore require further consideration.

New technology

Although high level customer requirements specifications may require new systems to operate in the same manner as existing systems it can often be the case that new technology providing the same functionality does not necessarily provide the same operability which can in turn cause operability problems. The introduction of electric rather then pneumatic doors to new trains caused changes to the behaviour of the passengers and drivers. Passengers used to surging right up to the doors in order to save vital seconds getting on to the trains got a fright when the plug doors moved out towards them before sliding open as this was not the behaviour of the pneumatic doors they were used to. Prior to the introduction of platform screen doors they had to learn to take a step back until the doors were fully open. For the drivers it was necessary to relearn how to press the door open button. A simple press needed to become a press and hold action for electric doors and took a while to adjust to.

However, Hong Kong railways are not simply replicating existing systems with new technology, the MTRCL Airport Line introduced different methods of operation and requirements for baggage handling and security, KCRC's West Rail has in-cab signalling systems and the first driverless passenger systems have been introduced in a high profile situation taking customers to Disneyland. Consequently the nature of the human factors issues are changing with the introduction of the new technologies and need to be systematically considered throughout the design lifecycle of the existing and new systems.

Discussion

In the UK the need for the integration of human factors into the design, development and change management processes have been clearly highlighted by incidents such as Ladbrooke Grove. In response, we have seen the development of guidance such as the Yellow Book Application Note on Human Error and human factors standards within Network Rail and LUL and the employment of human factors personnel within operator, regulator and standards bodies.

The motivation for interest in human factors in Hong Kong is not as obvious as SPADs or derailments. Hong Kong has some of the safest most reliable railway systems in the world. The utilisation of ATO and ATP technology and other design measures within Hong Kong railways means that some of the most serious hazards such as collision have been effectively mitigated under normal conditions. However, due to the high patronage, especially in peak hours even short

delays can lead to congestion problems that may threaten passenger safety. Paradoxically the high reliability of the systems can lead to increasingly higher expectations of the travelling public and the scrutiny of the media and greater focus on staff and passenger response to abnormal operations or delays. Thus, currently, the main driver for interest in Human Factors would appear to be the generation of improvements to service and reliability. These issues are perhaps less easy to pinpoint and require greater sophistication and effort to identify, measure, mitigate and monitor.

In addition, the introduction of new technology and the quest for continuous improvement are typically raising the bar in terms of the consideration of human factors issues in the railway and in particular in the context of Hong Kong railways. To provide the contextual input for the understanding of the impact of new technology and the softer issues identified for continuous improvement requires a more detailed understanding of Hong Kong railway clients than can typically be gained on a contract by contract basis.

It is the belief of the author that the current approach to human factors is too fragmented and that the Hong Kong railway organisations could gain better benefit from the human factors input than that which they currently receive by addressing the problems observed through existing project arrangements including:

- Loosely defined user and organisational objectives for human factors issues and measures for acceptance.
- Creation of confrontational rather than co-operative working arrangement.
- Lack of human factors awareness for client and Government to be able to evaluate human factors submissions
- Lack of human factors continuity through the lifecycle of the project through recruitment of different consultants at each stage
- Isolation of human factors consultants from end users during some project stages
- Lack of whole system overview of human factors consultants

Problems may be addressed by consideration of the following:

- Employment of human factors personnel to support both operations and projects, or long term relationships for consultancy support to cross the operations / project boundary.
- Contracting human factors support to the design authority rather than the contract manager to allow input and evaluation of early concepts and whole system / interface view of the design development
- Development of human factors contracts to address tasks / roles rather than systems to allow facilitation and evaluation of systems across design boundaries.
- Direct reporting of human factors consultants to client or provision of human factors support to client from client staff pool.

Conclusion

Hong Kong railways are often at the forefront of best practice and as such recognised the need for human factors input to projects more than 10 years ago, before technical and process standards were available to provide any guidance.

To achieve gains in service, performance and reliability and to smooth the introduction of new technology, that now seem to be the emerging needs with respect to human factors, it is necessary to push the boundaries again. This requires the examination of how better to address human factors issues in operations in a systematic fashion in order to achieve the incremental improvements in performance required and to better facilitate the interface with projects and project contractors.

It is encouraging to note that both railway companies take the issue of human factors seriously and are currently examining the way in which they want to integrate human factors principles into their everyday activities.

References

Courteney, H.Y., 'Practising What We Preach', in the *Proceedings of the First International Conference on Engineering Psychology and Cognitive Ergonomics, 1996*. Available on the web: http://www.he-alert.org/displayArticle.aspx?articleID=HE00150.

This page intentionally left blank

PART XV
Impairments to Performance

This page intentionally left blank

Chapter 53

The Prevalence of Obstructive Sleep Apnoea Syndrome (OSAS) in Train Drivers

Melissa Hack, Elaine Grimes, Toyin Davies and John Shelton

Introduction

Sleepiness is a known risk factor in up to 30% of accidents and near misses in the transportation industry in the USA. Examples in industry where sleepiness has been identified as a factor leading to major accidents include the Exon Valdez accident in the marine industry, and the Chernobyl the nuclear explosion in the nuclear power industry.

Sleepiness describes the propensity of an individual to fall asleep. This can cover a range from very mild symptoms of tiredness which the individual can manage, to severe symptoms when the individual struggles to stay awake or "nods off" for periods of time. In areas of industry where an individual is employed in a safety critical role the awareness of healthy sleep and knowledge of the consequences of sleepiness are important. Many factors contribute to an individual suffering from sleepiness including poor sleep hygiene due to social situations, shift work requirements and the presence of sleep disorders.

Over eighty different sleep disorders are recognised in the International Classification of Sleep Disorders. Some are common disorders including Obstructive Sleep Apnoea Syndrome (OSAS). This medical condition occurs in 4% of adult males and 2% of adult women in the general population (Young et al., 1993). OSAS is a result of narrowing of the upper airway during sleep periods. In severe cases of OSAS the individual will have repetitive narrowing or complete closure of the back of the throat and breathing is much reduced or ceases completely. This is called an episode of Hypopnoea or Apnoea respectively. There is usually accompanying loud snoring which can be considered a "marker" for the possible presence of apnoeas. During the Hypopnoea or Apnoea the oxygen level falls resulting in hypoxia and there is an increased inspiratory effort to try to overcome the occlusion. At a critical point an arousal occurs in the brain allowing the upper airway obstruction to be cleared. As a result the back of the throat opens and breathing starts. This cycle can occur hundreds of times in a sleep period. Arousals and hypoxia result in fragmented sleep quality which leads to waketime symptoms of sleepiness (Gaustaut, et al., 1966).

Some individuals are aware they have symptoms and request a consultation for advice. Many are not aware of a problem and may be told by a partner or colleague that symptoms exist. Individuals have also been told to seek advice by their employers when work performance is reduced or problems have been detected in the work place.

Untreated individuals are more likely to have accidents at the work place and up to six times more likely to be involved in a road traffic accident (Teran Samtos, et al., 1999). Laboratory studies confirm that many individuals with untreated OSAS demonstrate impaired performance in a variety of tasks including driving simulators (Juniper et al., 2005). Occupational accidents have been found to be increased in individuals suffering from sleepiness and snoring which are symptoms suggestive of underlying OSAS (Lindberg et al., 2001). Other health associations with

OSAS include hypertension, myocardial infarction, strokes, and mood and personality changes (Peker et al., 2002).

OSAS can be diagnosed following an overnight sleep investigation involving monitoring of breathing, cardiovascular, and brain wave changes. A widespread working definition of OSAS in use is five or more apnoeas and hypopnoeas per hour of sleep associated with at least two major symptoms (American Academy, 1999). The symptoms and results from the sleep investigation will determine the best possible therapy option for each individual. The most effective therapy for moderate to severe OSAS is called nasal Continuous Positive Airway Pressure (CPAP). This therapy is used for the main sleep period and provides a positive pressure that holds the back of the throat open preventing Hypopnoeas and Apnoeas from occurring. Oxygen levels remain normal and there is no hypoxia. Consequently there are no large increases of inspiratory effort which reduces the arousals of the brain and enables sleep quality to be maintained.

Robust methodology has been used to study the effectiveness of CPAP and the results from many studies confirm significant improvements in symptoms of subjective and objective sleepiness, sleep quality, quality of life and performance including driving following CPAP use (Hack et al., 2000). A study from Canada used data from real road driving and accidents and showed that the increased risk of involvement in a road traffic accident for individuals with untreated OSAS returns to normal following CPAP therapy (George et al., 2001).

Heavy Goods Vehicle drivers and Public Service Vehicle drivers with OSAS in the United Kingdom are able to continue in their driving jobs provided that symptoms of OSAS are considered to be controlled adequately by a sleep specialist, and the Driver and Vehicle Licensing Authority supports their fitness to drive. Airline pilots with OSAS in the USA are able to continue in their job as pilot provided that symptoms of OSAS are controlled adequately and they demonstrate adequate performance regularly on laboratory tests.

There is little information on the prevalence of sleepiness or OSAS in the rail industry. There are no published data on the medical condition of OSAS in railway employees. However in 2003 a report of a collision between two trains in USA investigated whether OSAS was a factor in the rail accident[1]. The probable cause determined by the National Transportation and Safety Board was neglect of the crewmembers to act on wayside signal information because they were asleep. The engineer had been diagnosed with moderate to severe OSAS and had received CPAP but the device was never properly adjusted and was likely to have been ineffective in treating the condition. The health of the crewmembers was determined to have contributed to this accident.

In the UK individual train drivers occasionally present to National Health Service clinics for sleep disorders by chance and may be confronted with long waiting lists before assessment. The level of provision of investigation, management and follow up of individuals with OSAS in the National Health Service varies widely throughout the UK. Untreated individuals with OSAS demonstrate impaired performance in a wide variety of tasks and this may be relevant to rail industry. Difficulty in accessing advice and delay in receiving effective treatment would also be relevant to the railway industry.

Hypothesis

The purpose of this study is to understand the prevalence of OSAS among a sample of train drivers and it has been hypothesised that unrecognised and untreated OSAS is present in employees of the rail industry. Previously unrecognised OSAS can be identified and treated thereby enabling

1 http://www.ntsb.gov/publictn/2005/RAB0503.htm.

employees to return to or continue in work. This study aims to estimate the prevalence of sleepiness and OSAS in train drivers using a railway specific questionnaire methodology. The overall aim of this study is to develop an effective screening tool which can be used by TOCs to identify staff with OSAS. Following the development and collection of data using this questionnaire, further work is being completed to classify the respondents into subgroups considered at high, intermediate or low risk of OSAS. Individuals of each subgroup will be randomly chosen for diagnostic sleep studies in order to set the score levels for the screening tool. This study was granted full ethical approval by an independent ethics committee. The study was also reviewed and endorsed by the rail industry trade union.

Method

Questionnaire

A questionnaire was developed for use in this study that includes a variety of questions regarding symptoms of sleepiness, snoring, OSAS, and other sleep disorders. The Epworth Sleepiness Scale and the Karolinska Scale of sleepiness are validated and used internationally in general populations.

The Epworth Sleepiness Scale (ESS) is a self rating scale for subjective sleepiness and ranges from 0 which represents no sleepiness to 24 which represents severe sleepiness and difficulty in maintaining alertness in any situation. A score greater than 9 is considered to indicate abnormal sleepiness in a general population. The Karolinska Scale is a self rating scale for subjective sleepiness at the time of completing the questionnaire and ranges from 1 which represents fully alert to 9 which represents fighting to stay awake. These scales were included to estimate subjective sleepiness.

This questionnaire was submitted as a pilot study to 10 males in the local community. Feedback from participants resulted in alteration of the wording of two questions for clarity. Railway specific questions regarding shifts, job details and comment boxes were added and the questionnaire was submitted to 12 train drivers in one TOC. Feedback from this pilot was satisfactory.

All of the UK Railway Train and Freight Operating Companies (FOCs and TOCs) were invited to participate. The first three companies to offer assistance with the study were selected. Confidential lists of train drivers employed by each of the participating TOCs were provided to the study team, and information sheets, consent forms, questionnaires and stamped envelope were sent to each of the train drivers at the contact address. Completion of the questionnaires was on a voluntary basis, and questionnaires and consent forms were returned to the study team by mail.

Completed questionnaires were analysed for demographics, general health, symptoms of sleepiness, symptoms of snoring and symptoms of OSAS. Data not presented in this chapter were also collected on medication, sleep habits, insomnia, shift work and shift work disorders.

Responses to the questions were analysed to assess the likelihood that an individual was suffering from OSAS. Firstly, questions regarding one symptom were grouped together. One of the markers for possible OSAS is severe snoring. For an individual to be considered to have positive symptoms of severe snoring they would have to show evidence of snoring greater than five times a week, snoring heard through a closed door, waking themselves with snoring and their partner moving out of the bedroom. Responses of non snorers were negative to the above. Responses positive to some, but not all, of these questions placed an individual into an intermediate group. Other questions regarding symptoms were grouped for sleepiness, choking and gasping, and apnoeic episodes. For an individual to be considered to have a high risk of symptoms of OSAS they had to satisfy all criteria based on all main symptoms of this condition. Answers to the groups of

questions about self reported sleepiness, snoring, gasping and choking with apnoeic episodes were combined. Using strict criteria the possible cases of OSAS were estimated by strongly positive responses to all of the groups of questions. This method is likely only to identify severe cases with a clear indication of OSAS.

A ranking system of the data included in the questionnaire was used which included Body Mass Index, collar size, sleepiness and symptoms of OSAS. These divided the respondents into categories used for random selection of individuals for laboratory investigation. The respondents in the upper quartile were considered to be at high risk for OSAS, the lower quartile at low risk and the middle quartiles at intermediate risk.

Results

Questionnaire response

The results are presented combined for the three participating TOCs. 2351 train drivers were sent questionnaires. A total of 508 questionnaires were returned. Five questionnaires were incomplete and were not included. Three questionnaires were returned not completed by trainee drivers who felt unable to provide full information. One questionnaire was returned not completed with a covering letter to state that the individual did not think any driver would take part due to worry about job security. A total of 499 completed and usable questionnaires were returned which gave a response rate of 21.23%. The ratio of male to female respondents was 489:10. The mean age of respondents was 42.12 years, BMI 27.39, collar size 16.79 inches.

Sleepiness rating

The mean Karolinska Scale score was 1.31. The mean Epworth Sleepiness Scale score was 6.98, and 23.2% of respondents self rated on the Epworth Sleepiness Scale as having symptoms of waketime sleepiness above normal (ESS>9).

Snoring

86.2% of respondents self reported snoring.
34.87% reported frequent snoring of greater than one night per week.
19.30% reported snoring that could be heard through closed doors.
10.09% reported waking themselves from their own snoring at least one night per week.

Apnoeic episodes

16.67% reported waking with gasping or choking at least one night per week.

12.28% reported apnoeic episodes that they had noticed or had been witnessed on at least one night per week.

OSAS prevalence

Using the strict criterion of OSAS as defined by the presence of excessive sleepiness, severe snoring and frequent apnoeic episodes were present in 7.3% of the respondents. There was no significant difference between TOCs in the main responses.

General Health

The questionnaire contained a section asking for information on a variety of medical conditions including hypertension, asthma, chronic obstructive airways disease (chronic bronchitis), stroke, heart attack, diabetes, and other health problems. OSAS has a prevalence of 7.3% and was placed in the top five medical conditions in this study (Table 53.1).

Table 53.1 Prevalence of smoking and medical conditions in the study population

Condition	Percentage of responders %
Current smoking	23.25
Hypertension	10.53
Asthma	10.2
Chronic bronchitis	6.58
Diabetes	0.2
Headaches	0.2

Conclusion

This is the first study to investigate the prevalence of OSAS in the rail industry. In this study of voluntary participants the response rate was 21.23%. This is a similar response rate to that obtained in previous surveys within the industry. The prevalence of OSAS based on strict criteria in the questionnaire is 7.3%, which is greater than the quoted prevalence of 2% to 4% in the general population. The underlying factors for the increased prevalence are unclear and require further study. Symptoms of sleepiness are present in OSAS but may also be due to many other factors.

The average results from the sleepiness scales used in this questionnaire are within the normal values for the general population. However 23.2% of the train drivers self rated as having ESS greater than a normal value of 9 which indicates excessive sleepiness and this is a higher proportion than would be expected in the general population. This is important when considering the safety critical role of train drivers. Consideration should be given to providing education about healthy sleep.

The results of this questionnaire place OSAS in the top five of health problems in the railway industry based on this sample of train drivers, as identified using a screening questionnaire. The second part of the study will provide additional information as to how many of the drivers identified as having OSAS from the questionnaire were confirmed from the sleep study. Consideration needs to be given to strategies to raise awareness of sleep disorders such as OSAS with help available for individuals who have symptoms.

The railway specific sleep questionnaire is easy to administer and provides a practical tool for assessing the possible risk of OSAS in employees in the rail industry. It is possible that some train drivers may have been wary of participation and declined to return the questionnaire because of concern over losing their jobs. OSAS can be treated and with appropriate control of symptoms and specialist follow up it is considered appropriate for individuals to remain in safety critical jobs in other areas of the transportation industry. We plan to test the sensitivity of this questionnaire in

the next part of the study. This involves in depth investigation of subgroups of drivers identified as having high, medium, and low risk of OSAS, to identify where it is actually present, and set the scoring system for the screening tool so that it could be administered by a wider audience of occupational health practitioners in driver medical screening.

This study confirms that unrecognised symptoms of OSAS are present in individuals working in safety critical roles in the rail industry. This study indicates the need for further consideration of the relevance of unrecognised OSAS within the rail industry.

References

American Academy Sleep Medicine Task Force. Sleep disordered breathing in adults: recommendations for symptom definition and measurement techniques in clinical research. *Sleep* 1999; 22(5):667-89.

Gaustaut H et al. Polygraphic study of the episodic diurnal and nocturnal (hypnic and respiratory) manifestations of the Pickwickian syndrome. *Brain Research* 1966;2:1667-86.

George C et al. Reduction in motor vehicle collisions following treatment of sleep apnoea with nasal CPAP. *Thorax* 2001; 56:508-12.

Hack M et al. Randomised prospective parallel trial of therapeutic versus subtherapeutic nasal continuous positive airway pressure on simulated steering performance in patients with obstructive sleep apnoea. *Thorax* 2000; 55(3):224-31.

Juniper M et al. Steering simulation performance in patients with obstructive sleep apnoea and matched control subjects. *European Respiratory Journal* 2000; 15(3):590-5.

Lindberg E et al. Role of snoring and daytime sleepiness in occupational accidents. American *Journal of Respiratory and Critical Care Medicine* 2001; 164:2031-35.

Peker Y et al. Increased incidence of cardiovascular disease in middle-aged men with obstructive sleep apnoea: a 7-year follow up. *American Journal of Respiratory and Critical Care Medicine* 2002; 166(2):159-65.

Teran Santos J et al. The association between sleep apnoea and the risk of traffic accidents. *New England Journal of Medicine* 1999;340(11):847-51.

Young T et al. The occurrence of sleep disordered breathing amongst middle-aged adults. *New England Journal of Medicine* 1993; 328(17):1230-5.

Chapter 54

Fatigue Management on the Operational Railway: Developing a Strategic Approach

Fiona Kenvyn

Introduction

The demand for a 24-hour operable rail network poses a crucial issue for the rail industry with respect to maintaining a workforce capable of providing a safe and effective service around the clock. As owner of the rail infrastructure across England, Scotland and Wales, Network Rail is challenged with not only managing and maintaining the infrastructure itself, but also ensuring the dedicated workforce charged with undertaking that work are able to perform their roles in an effective and optimum way. As with any 24-hour industry, fatigue is an issue for rail operations, particularly given the safety-critical nature of much of the work involved and the reliance on shift rostering as a means of organising working hours among staff. It is thus essential that companies such as Network Rail are aware of the risk fatigue can pose to employees, and have management systems in place.

This chapter will look at the issue of fatigue as it impacts upon the rail industry, and Network Rail in particular. It will provide an insight into where the company is at present as far as providing a structure for managing fatigue among its employees, particularly those in safety critical roles, and will outline the company's developing fatigue management strategy, identifying the areas where improvement and enhancement of processes already in place are required. The chapter aims to add to the existing discussion on fatigue management as it relates to the rail industry, providing a case study that describes the reality of trying to integrate good practice into a complex and in many ways unique organisation.

Why manage fatigue?

Fatigue as a construct can be difficult to define given its multidimensional nature; however, it is generally accepted to refer to a reduction in a person's capacity and inclination to work (Pheasant, 1991). Put simply, it involves a loss of alertness that can eventually end in sleep. In the context of work, the key issue surrounds the impact that tiredness can have on a person's performance in their work tasks, on their health and well-being at work, and on the effects with respect to safety. Beyond the legal requirements for Network Rail it is an important issue for the company for several reasons.

Impact on human performance

Many of the operational tasks within the railway industry pose a significant risk of accident and injury when performance is compromised. If an individual's alertness and awareness are not functioning at their optimum level, the consequences can have wide-ranging implications for

the safety of others and for the integrity of the rail infrastructure. With a considerable body of research suggesting fatigue may have a significant impact on human performance, (e.g. Gold, Bock, Tosteson, Baum, Speizer and Czeisler, 1992; Folkard, Totterdell, Minors and Waterhouse, 1993; Williamson and Feyer, 1995; Matthews and Desmond, 2002), it is essential that fatigue management be an integral part of the company's health and safety processes.

Shiftwork effects

Though it has long been recognised as an issue for attention, significant influence has inevitably come from the "Hidden Report" into the Clapham Junction incident in December 1989 (Hidden, 1989), which made recommendations on the control of rostered working hours given the concern that fatigue may have played a role. The idea that shiftwork has been identified as having an impact on the development of fatigue symptoms among workers is not new. This may be largely due to the fact that shiftwork, particularly that involving night work on either a regular or rotating basis, requires workers to sleep at a time they would normally be engaging in activity, and to work when they may be least able to carry out most tasks (Scott, 1994). Shiftwork by its nature requires workers to continuously adjust their sleep-wake schedule to meet a roster of changing working hours, and this disruption to the body clock can impair both the quality and duration of sleep. For this reason, shiftworkers are perhaps the most vulnerable to the effects of fatigue.

Research has provided a number of recommendations for good practice shift scheduling that can assist to minimise the impact of shiftwork on the fatigue levels of workers. While it is recognised that there is as yet no one "best fit" shift schedule design (Folkard, 1992; Knauth, 1993; Kogi, 1995), there seems a general agreement that certain characteristics are of more benefit than others, including:

- Forward rotating shift rosters (Barton and Folkard, 1993).
- Rapid rotation of shifts (2 to 4 consecutive shifts), rather than slower rotation (5 to 7 shifts) (Totterdell and Folkard, 1990; Knauth and Schonfelder, 1990)
- Restriction of night shifts due to their having a greater risk of impaired performance (Folkard et al., 1993; Williamson and Feyer, 1995; Smith, Folkard, Tucker and Macdonald, 1998).

As of October 2005, there was approximately 11000 safety critical staff, including Signallers, Mobile Operations Managers (MOMs), Electrical Control Operators (ECOs), Controllers and Track Maintenance workers employed by Network Rail, and all working to a shift schedule. On its own, this is reason enough for having a fatigue management system that ensures there are sufficient measures in place to protect employees from developing fatigue at work, as well as processes for educating employees on their own responsibilities where fatigue management is concerned.

Impact on health and well-being

Another reason for managing fatigue at work is that it has been identified as having an adverse effect on both the health and well-being of workers, particularly where it results from shiftwork (Costa, 1996; Martens, Nijhuis, Van Boxtel and Knottnerus, 1999). It has also been suggested to impact a person's relationships with family and friends, which can consequently negatively influence both well-being and work performance (Monk and Folkard, 1992).

Economic implications

In addition to the safety implications, poor fatigue management can also generate an economic problem for organisations. It has been estimated that sleep-related accidents can cost companies

some £115 – 240 million per year in the UK (Folkard, 2000). Given the considerable research that has been undertaken into means to reduce the impact of fatigue in the workplace, it seems sensible to try to integrate recognised good practice, where applicable, into the day to day functioning of any organisation. Network Rail is no exception.

Why have a strategy for fatigue management?

A fatigue management strategy provides an organisation with a focused plan for addressing the issue of fatigue from all angles, from identifying and managing the sources of fatigue, to implementing proactive strategies for mitigating against its effects. Having a formal strategy also emphasises an organisation's commitment to addressing the issue, importantly recognising that the process of managing fatigue as a risk to safety and performance is a shared responsibility between the organisation and the employees who work within it. A formal strategy also provides the means for coordinating the many existing efforts that contribute to the management of fatigue under one clear banner.

A review across other industries, including the off-shore oil industry, the nuclear power industry, and other transportation industries, has identified several common factors that should be considered in any organisation's fatigue management strategy, specifically for operating a 24-hour timetable. A robust and effective strategy for managing fatigue in employees is one that addresses the following:

1. Responsible scheduling of working hours
   - Current legislation on working time recognises that the hours a person works, and how those hours are scheduled, can have significant implications for safety and performance, as well as the individual's own health and well-being.
   - There must be a measure in place providing limits on working time that can be audited against. This will ensure workers do not suffer excessive fatigue by virtue of the work schedules they are required to follow.
   - There must be a focus on producing fatigue-friendly rosters to minimise the risk of fatigue for employees where possible. These should reflect current rostering good practice.
   - There must be a form of risk assessment to identify shift schedules that may pose a greater risk of fatigue.

2. Monitoring of compliance with working time regulations
   - It is crucial that where a policy is in place outlining working time limits, that compliance with these limits can be evaluated.
   - There must be a process for reporting exceedances of working time regulations so that information on the extent of non-compliance may be obtained, and the appropriate risk management processes may be put in place.
   - There must be clear control measures for managing instances where the standard on working time cannot be complied with.
   - There must be a focus on identifying and managing the causes of non-compliance with working time regulations, where possible, so that mitigation measures can be put into place to manage these for the future.

3. Briefing and training of fatigue management strategies and procedures
   - It is essential that employees are aware of the procedures in place for managing fatigue at work, as well as the strategies they themselves may undertake to mitigate against it. There must be commitment to providing education on fatigue and its management to all staff, particularly those vulnerable to the effects of fatigue, including shiftworkers.

- There must also be a focus on education with respect to the legal responsibilities related to shift scheduling, as well as to good rostering practice. The responsibilities of both the organisation and of employees should be clearly established and briefed.
- Education should also focus on the procedure for managing staff who are not fit for work due to the effects of fatigue.

4. Integration with health assessment arrangements
   - Poor health and fitness can be contributing factors in exacerbating the risk of excessive fatigue at work. Health management systems must be sufficient to identify and assist employees at risk due to their physical well-being.
   - For staff undertaking safety critical roles in particular, a focus should be on attaining sufficient information on medical history, including a history of sleep disorders.
   - Health promotion with respect to the impact of diet, exercise, substance use and abuse and lifestyle on fatigue should also be included.

5. Assessment and monitoring of workplace conditions
   - Workplace conditions that are unsuitable or unsafe can also contribute to the development of fatigue, and so there should be a clear procedure for assessing both workplace environment and work roles to ensure appropriate Health and Safety and organisational standards are being sufficiently met.

Working hours and fatigue – The current picture

While the implications of fatigue for health and safety have long been recognised, Network Rail has not as yet been able to implement a formal fatigue management strategy, though there are certainly procedures in place aimed at minimising the effects of fatigue on employees. The reason for this is unclear, though the organisation has undergone considerable change in recent years with the demise of Railtrack Plc, several internal reorganisations and the considerable influx of staff with the taking in-house of maintenance contracts in 2004. It may be that articulating a strategy in the shape of a formal document is not the most appropriate way forward. Given the company is already tackling the issue via a number of measures, the best strategy may instead involve addressing fatigue management from a continuous improvement perspective, coordinating current efforts under one cohesive approach that is agreed upon across all company functions.

That the current arrangements are not coordinated by any one Function within the company is not in itself an issue, though historically much of the work on fatigue management has focused on employees within the Operations and Customer Services function: that is, Signallers, MOMs, ECOs and Controllers. With the number of safety critical staff working for Network Rail increasing dramatically over the last 18 months, particularly within the Maintenance Function, plans for extending the focus for fatigue management beyond Operations is still being fully realised.

To date, fatigue management within Network Rail has consisted of a focus on four key areas: Shift scheduling, monitoring of working time exceedances, education on managing fatigue, and workplace design and assessment.

Shift scheduling

Company procedure The working time of all safety critical staff employed by Network Rail is controlled by the Company Standard on working hours (RT/LS/P/003), itself guided by the

principles of the European Working Time Directive and the relevant Rail Safety Regulations on managing fatigue in the context of safety critical work. The Company Standard is currently under review, with a plan to ensure it extends its coverage to include safety critical employees outside of a purely Operations setting.

At present, the Standard makes the following assertions for staff in safety critical work roles, effectively derived from the "Hidden limits" outlined in the Appendix of the Railway Group Standard (GH/RT4004):

- No less than 12 hours to be taken between shifts
- No more than 13 days to be worked consecutively without a rest
- No shifts to be longer than 12 hours
- No more than 72 hours to be worked per calendar week

Roster design and implementation A set of National Rostering Principles, developed in conjunction with the relevant trades union body, govern the rostering arrangements primarily for signalling staff, though there are plans to revisit these to ensure they cover all safety critical staff within the organisation. Shifts are designed by rostering clerks with some consultation with the teams to which they apply, though there is no one consistent approach to rostering across the company; instead, shifts have tended to be arranged to best fit the needs of the team in question. While there has been some attempt to brief out information on fatigue-friendly rostering in accordance with research-based best practice, this process has perhaps not been as effective as it could be, with the result that there are many different rosters utilised, from 8 to 12 hour shift schedules, with advanced or delayed rotation, of varying start and finish times, and in blocks of varying number of consecutive shifts. Perhaps an exercise in demonstrating the flexibility of the working time regulations, all of these shift patterns, despite their variation, have met the limits set by the relevant legislation.

Automated rostering The first step towards attempting to gain some consistency over rostering across the company has come with the National Rostering Application for Network Rail (NRANR), with roll-out completion at the end of the 2005. Arising from the former Operations function (now Operations and Customer Service), the project had as its goal the development of an automated rostering system, for use by signallers and other operational shift working staff, with the following aims:

- To facilitate the distribution of shifts and rest days in accordance with current legislation and good practice;
- To maintain centralised data on working hours to provide better visibility of overtime and other exceedances, as well as of the costs associated with additional working time.

The system aims to utilise a number of generic rosters, as agreed through consultation with the relevant trades union representatives, as the basis for the programme, although the varied body of rosters available and in operation across the network made this a difficult process. Following the roll-out, the next step will involve setting up a communication between the NRANR system and the Human Resources System (HRS) to enable automatic calculation of hours worked for payroll purposes. This will also provide an opportunity for identifying inefficiencies in both rostering and in staff resource levels. The roll-out of the system to areas outside the O&CS Function will potentially be an issue for the future, though would in all likelihood not precede an evaluation of the system's success once the current roll-out is completed.

Fatigue Index In order to assist employers to carry out an effective risk assessment when reviewing staff working hours, the Health and Safety Executive have developed the Fatigue Index (FI), a tool designed to gauge the risk of fatigue posed by certain shift patterns (Rogers, Spencer and Stone, 1999). By taking into account the factors that can have an impact on fatigue levels when shifts are altered, the tool can provide an indication of the level of fatigue that is likely to result as a consequence of a particular shift schedule. Though it is not without its limitations – it is not suitable for long patterns of night shift and does not consider factors outside of the workplace that may be contributing to fatigue – it does provide a guideline for roster development and is able to highlight areas of concern. As a result of industry feedback, the FI is currently being revised and an enhanced edition should be available towards the end of 2005.

At present, the FI is utilised in two main areas within Network Rail:

- The Investigations team have applied the tool following critical incidents to review the working hours in the days preceding the event of the staff involved. The index, used in accompaniment with an interview, can highlight whether fatigue is likely to have been a contributing factor to the incident.
- The FI is also used to assist with roster development, where proposed rosters can be entered into the FI spreadsheet to ascertain whether the roster is adequately minimising fatigue risk. The Network Rail Ergonomics Group is currently involved in facilitating this process.

Monitoring of working hours

Traditionally, the focus for fatigue management has been on the control of working hours and on monitoring the number of instances where these working time limits have been exceeded. There is currently a process in place for line managers for reporting exceedances, though this procedure is under review in line with the re-examination of the Company Standard on working hours. The aim is to ensure it provides more useful information on the reason for the exceedances, enabling more active problem solving where an exceedance may have been avoidable.

Network Rail has received a warning in the past due to exceeding working time regulations, having an improvement notice issued by the HMRI in 2003 due to exceedances in the then Midlands Region. While this has been the only instance and the issue has since been closed out, it provides a reminder that Network Rail must be vigilant about exceedance monitoring given the ongoing compliance monitoring. Audits and reviews over the last two years have indicated that while there is a general attempt by line managers across the company to follow the reporting process, it is not fully adhered to, and so further briefing and education is required.

It is recognised that in certain cases an exceedance of working time regulations may not be avoidable; resource levels, and the resulting impact in cases where annual leave or sick leave are taken, are a primary cause. However, it is imperative that an exceedance be considered the exception rather than the rule when it comes to the management of working hours.

Education

There has been an attempt to educate operational staff in the past on the effects of fatigue on performance, the risks of working while fatigued, and on the steps that can be taken to manage it. Education has been provided through a number of means, including integration into the signaller training programme, a fatigue management briefing provided for the quarterly operational staff briefings, and a component on fatigue integrated into the Investigations training module where fatigue may have a role in contributing to human error.

Education on fatigue management is not currently part of the training programmes for other shiftworkers outside of signallers, nor is it provided for other safety-critical and non safety-critical staff within Network Rail. Good feedback has been received where fatigue management education has been provided, and a proposal has been made to extend fatigue education to other areas of the company. Plans are in hand to include information on managing fatigue on the company's intranet as part of a wider drive on employee well-being.

Workplace design

Network Rail has in place a Company Standard on integrating ergonomics into all projects where a human interface exists, and this includes a requirement for specialist input into issues of workplace design, such as in the case of signal box refurbishment, or workload, as may be relevant for a resignalling project.

In addition, fatigue has been monitored via the Baseline survey process and the Railway Ergonomics Questionnaire (REQUEST) as part of a general assessment of ergonomics and human factors in the rail environment. To date only operational staff have been assessed, though there is a plan to extend the REQUEST to other safety-critical staff. The most recent wave of assessments has identified that fatigue is an issue for the employees sampled, raising a number of key points including:

- 72% of workers reported having worked more than 72 hours in one week over the previous 12 month period
- A high level of variation in roster patterns exists across the Network Rail Territories
- A majority of respondents reported having at least 12 hours break between shifts
- Many respondents identified night working and problems with shift patterns as contributing to difficulties with sleep

A Strategy for the Future

With a period of considerable change now behind it, and having moved into the second year of the organisation's new phase, it is of benefit that Network Rail take stock of its position with respect to the management of fatigue and identify a clear strategy to take into the future. Reflecting on where Network Rail sits currently in comparison with where it could be is helpful for developing a clear strategy for the management of fatigue as the organisation moves forward. While there are many areas that demonstrate good fatigue management practice, the organisation needs to apply a continuous improvement approach to this area as it does to others as part of managing risk if its aim of providing excellence in all areas of the rail industry is to be fully realised. While there is progress, there is still much more to be done.

In light of both the company's current strengths, and the identified areas for improvement, it is proposed that a workable and effective fatigue management strategy for Network Rail should entail the following:

1. Scheduling of working hours
   - Review of Company Standard (RT/LS/P/003) on working hours to include all safety critical employees.
   - Integration of good practice into rostering procedures where practical to ensure rosters are fatigue-friendly. This should also include clear policy for rostering following return to work after leave or illness.
   - Utilisation of appropriate risk assessment (HSE Fatigue Index) to review rosters to ensure the risk of their contributing to fatigue in workers is minimised as best as possible.

- Renegotiation of National Rostering Principles to ensure they are consistent with good practice where appropriate.

2. Monitoring
   - Enhancement of the process for reporting non-compliance with working time regulations, including greater focus on causes.
   - Provision of clear procedures for managing situations that may arise in non-compliance
   - Enhancement of the monitoring process, including data collection and data reporting, with a greater focus on managing and mitigating against causes where possible.
   - Monitoring should also extend to other areas, including collection of data on instances where fatigue is identified as a contributing factor in investigations, better use of signaller error data, and ongoing review of employees' subjective experience of fatigue, such as with the Railway Ergonomics Questionnaire (REQUEST), for example.

3. Education and Training
   - Extension of education to other safety critical employees, as well as non-safety critical staff, on managing shiftwork and fatigue.
   - Education for roster clerks on good practice rostering for mitigating against effects of fatigue
   - Briefing for all line managers on reporting of non-compliance with working time regulations, including education on procedures for managing staff who are unfit for work due to the effects of fatigue, and on how to identify fatigue in employees.

4. Health care arrangements
   - Extension of Managing for Health procedures to include identification of staff at risk of fatigue due to issues of health, diet and lifestyle. Should include focus at medical review on history of sleep disorders.
   - Enhanced health promotion with respect to the impact of diet, exercise, health and lifestyle on fatigue.

5. Workplace assessment
   - Promotion of workplace assessment resources currently available within the organisation, including through Occupational Health and Safety and the Ergonomics Group.
   - Clarification of line manager responsibility to ensure staff are working under appropriate workload and environmental conditions so as to mitigate where possible against the effects of fatigue.

There are likely to be a number of challenges faced in the process of realising this strategy beyond theory, though none of these are insurmountable if fatigue management is given its due recognition as a significant health and safety issue for the organisation.

Successful implementation of the strategy will be dependant on the extent to which the following issues can be addressed:

- There is some resistance to the notion of "fatigue-friendly" rosters from employees who have negotiated rosters at a local level for a number of years, though who also have what may be considered non-standard rostering practices. A way of addressing this must be found given implementing prescriptive change may not be the most effective means. The old adage "if it ain't broke, don't fix it" seems to be present in many spheres which makes implementing change all the more difficult.

- There is considerable discussion as to the need for guidance on commuting time, and whether a maximum time should be set to minimise the effects of travelling time on fatigue at work. With a large number of commuting staff across the organisation, this is a very real issue that will need to be considered, and it may be that further research into this issue is required before next steps can be taken.
- There is a need for Function Directors, particularly for HR, Maintenance and Operations and Customer Service, to agree an acceptable vacancy gap not only with respect to the financial implications where overtime is concerned, but also in terms of the effect it has on fatigue for those employees required to work overtime to fill the vacancies. This is made more difficult with there being an uneven vacancy gap across the organisation, along with different staff turnover rates.
- There needs to be commitment across the organisation to taking a proactive stance to managing staff turnover, with appropriate interim measures taken where prior knowledge of job losses is held. This will in turn assist with the issue of managing vacancy gaps, generally the biggest cause of exceedances within the Company.
- There must be a long-term goal of creating a culture where fatigue management is accepted as part of the normal functioning of the organisation, and exceedances of working time regulation are the exception and not the rule.
- A challenge to good practice fatigue management is posed by the lack of clear consensus in research among a number of issues relating to shift scheduling (e.g. benefits of 8hr and 12hr rostering); it may be that further research must be undertaken within the rail industry, and particularly with operational and maintenance staff, to provide a better understanding of shift effects.

Conclusion

Fatigue and its management are substantial issues not just for the rail industry but for organisations worldwide. The impact of fatigue on performance and safety, as well as on the health and well-being of workers, provides an even greater argument for ensuring a robust strategy is in place for mitigating against its effects. For Network Rail, the demands for meeting the requirements of 24-hour operation means the job of ensuring 100% compliance with both working time regulations and recognised good practice for fatigue management can be a difficult one. There is certainly a positive move forward for enhancing current practice where possible, ensuring the management of fatigue is an issue that remains high on the organisational agenda, tying in with the company's goal to provide a safe and efficient railway for all who work and travel upon it.

References

Barton, J. and Folkard, S. (1993). Advancing versus delaying shift systems. *Ergonomics;* 36 (1-3): 59-64.

Costa, G. (1996). The impact of shift and night work health. *Applied Ergonomics;* 27: 9-16.

Folkard, S. (1992). Is there a 'best compromise' shift system? *Ergonomics*; 35: 1453-1463.

Folkard, S. (2000). Foreward. In *The Impact of shiftwork and fatigue on safety*. Railtrack Safety and Standards Directorate publication, November 2000.

Folkard, S., Totterdell, P., Minors, D. and Waterhouse, J. (1993). Dissecting circadian performance rhythms: implications for shiftwork. *Ergonomics*; 36 (1-3): 283-288.

Gold, D.R., Bock, N., Tosteson, T.D., Baum, T.M., Speizer, F.E and Czeisler, C.A. (1992). Rotating shift work, sleep, and accidents related to sleepiness in hospital nurses. *American Journal of Public Health*; 82: 1011-1014.

Hidden, A. (Q.C.) (1989) *Investigation into the Clapham Junction Railway Accident*. Presented to Parliament by the Secretary of State for Transport, November 1989. London: HMSO.

Knauth, P. (1993). The design of shift systems. *Ergonomics*; 36 (1-3): 15-28.

Knauth, P. and Schonfelder, E. (1990). Effects of a new shift system on the social life of shiftworkers. In G. Costa, G. Cedana, Kogi and A.Wedderburn (eds.) *Shiftwork: Health, Sleep and Performance.* Peter Lang: Frankfurt. 537 – 545.

Kogi, K. (1995). Increasing flexibility in shift work arrangements. Work and stress; 9: 211-218.

Martens, M.F.J., Nijhuis, F.J.N., Van Boxtel, M.P.J and Knottnerus, J.A. (1999). Flexible work schedules and mental and physical health: A Study of a working population with non-traditional working hours. *Journal of Organizational Behaviour,* 20: 35-46.

Matthews, G. and Desmond, P.A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology*; 55A (2): 659-686.

Monk, T. and Folkard, S. (1992). *Making shiftwork tolerable.* London: Taylor and Francis.

Pheasant, S. (1991). *Ergonomics, work and health*. Macmillan Press, England.

Rogers, A.S., Spencer, M.B. and Stone, B.M. (1999). *Validation and development of a method for assessing the risks arising from mental fatigue*. HSE Books: Norwich

Scott, A,J. (1994). Chronobiological considerations in shiftworker sleep and performance and sleep scheduling. *Human Performance*; 7 (3): 207-233.

Smith, L., Folkard, S., Tucker, P. and Macdonald, J. (1998). Work shift duration: A review comparing 8-hour and 12-hour shift systems. *Occupational and Environmental Medicine;* 55(4): 217 – 229.

Totterdell, P. and Folkard, S. (1990). The efforts of changing from weekly rotating to a rapidly rotating shift schedule. In G. Costa, G. Cesana, K. Kogi and A. Wedderburn (eds.) *Shiftwork: Health, Sleep and performance*. Peter Lang: Frankfurt. 642-650.

Williamson, A.M. and Freyer, A.M. (1995). Causes of accidents and the time of day. *Work and Stress*; 9 (2-3): 158-164

Chapter 55

Alcohol and Drugs as Risk Factors in Railway Traffic

O. Ervasti, L. Pitkämäki, H.K. Otterstad, M. Sternudd, and J. Worm-Petersen

Introduction

The abuse of alcohol, drugs and narcotics is certainly a serious threat to safety in the railway industry as to traffic in general. This is an area where medicine has a lot to contribute, both as provider of knowledge, but also because routine medical examinations of all employees involved with safety critical functions are considered as a major safety barrier.

We report here on a joint effort with contributions made by five Nordic colleagues lasting from January 2004 to April 2005. The facts are derived from several sources, with selection and documentation based on what has been readily available in the four Nodic countries.

Alcohol is a major risk factor in many ways, because of its widespread use, and the long exposure time before incapacitating alcoholism has developed. The increased risks are due to several modus operandi:

- *Acute intoxication* with a high incidence of accidents, which can be found in many persons without a chronic problem of abuse.
- *Heavy drinkers* suffering from impairment of sleep and lack of concentration and negative influence of hangovers. This is a chronic situation which may last for many years. It is important to identify this group because of the obvious safety risks lasting for years.
- *Chronic alcoholism* - with brain-, liver- and other organ damages which can usually be detected by proper medical screening.

When it comes to abuse of drugs and narcotics, the risk pattern will in principle have features in common with alcohol abuse. In many cases there exists an abuse of a combination of several substances. Such issues call for high level professional knowledge.

Problem description

The number of individuals and employees with problems of heavy drinking and use of drugs has increased in the last decades. This scenario has resulted in an acceleration of the prevalence of human risk factors in traffic.

Consumption of alcohol, especially wine, has increased a lot in the Nordic countries in the last 20 years. Since there is a close correlation between total consumption in the population and the proportion of heavy drinkers and alcohol abusers, the risk-profile has changed with an increase in the problem in the railway.

In Finland a recent autopsy study of alcohol as a causal factor in fatal accidents in the workplace found one in three of such cases among the 30-40 fatal accidents per year. The recommendation was that a test for alcohol should be mandatory in the investigation of all fatal accidents at work.

Due to a general increase in the average age for employees and the prescription of more potent drugs, the medical treatment in itself will also introduce more risks. We do not know to which extent use of illegal drugs and narcotics takes place among railway-employees, but it is fair to state that this scene has also changed for the worse.

The prevalence of people with alcohol-problems is presently around 15 percent of the working population.

The accepted formula has been that one percent of the workforce are alcoholics, while ten percent are "heavy drinkers". A study in Finland found that the prevalence of heavy drinkers among men in Primary Health Care patients was 20 percent, while the proportion was 17 percent for clients in Occupation Health Care, and 16 percent in the general population. For women the corresponding figures were nine, six and 13 percent. The method was a questionnaire (CAGE) – Aalto (1999).

Addictions and misuse of chemical substances like alcohol and drugs, shows increasing prevalence with age. This can be explained both by the increased time of exposure, the probability of being addicted, and the direct short term effects of these substances on a lot of problems caused by a modern lifestyle. A recent experience in Norway illustrates this. In September 2002 more than 20,000 litres of smuggler-alcohol for sale, containing 23% methanol, was brought in to Norway. This crime was soon detected and announced in all media, but during the initial period 33 persons with methanol-poisoning were treated in hospital. Their average age was 54 years, only three were under 40 years. There was a majority of men: 23 compared to 10 women. In 2005 the average age of the employees in the Norwegian Rail Agency is around 47 years.

Prevalence of heavy drinkers among Danish civil aviation personnel in 2003

A prevalence study was conducted in Denmark in 2003. In a population of 2000 civil aviation pilots and air traffic controllers, testing with biological markers found 100 cases of increased gamma GT. Out of this subgroup five persons also had increased CDT-values. The prevalence of "heavy drinkers" in this group could be calculated to two per thousand. The important fact is that a quantifiable risk exists and that this small risk group can be identified with proper techniques during improved routine examination.

It is a problem that the responses from employers and authorities are often weak when addiction and dependency are revealed. The consequence is that prevalent risks are not eliminated.

A recent study among Norwegian physicians in the years 1998-2000 revealed that 36 cases had been reported to the authorities. Most of them were men, average age 50 years and they had worked for more than ten years, mainly in Primary Health Care. Eight of them were managers. Some abused a combination of drugs and alcohol. This sample represents doctors who have lost professional control and are intoxicated in their work. Some had even stolen medicine. The employer had reacted in less than half of the obvious cases. When such cases are reported to the Ministry of Health most of the cases result in loss of authorisation (Hoddevik and Nygaard, 2004). The message here is however that local supervising bodies and employers are often reluctant to react and stop the abusers from continuing their work.

Prevalence of abuse of alcohol and drugs among railway employees

The statistics from the recent years from the four Nordic countries are lacking in coverage and specificity. The following numbers are the best available:

Denmark – After 1990 all major train-accidents have been investigated with blood tests for alcohol, without a single positive test.

Finland – Since the 1970s one random breath test per year on alcohol from every employee is mandatory. There are no national statistics on the prevalence, but based on the numbers of positive tests from one of the four regions, the incidence of alcohol use/abuse in service are around 30 cases per year. With 14.000 employees the incidence rate will be approximately two persons per 1000 employees per year. Many of them are expelled from duty without further warning. This implies that most of the cases are "new".

Norway – Yearly statistics shows that approximately 20 cases have been reported and the employees are either banned from safety functions or work under special supervision. There are around 4000 persons with safety functions in Norway and the general conclusion is that the problem is grossly underreported
.
Sweden – The Swedish Rail Authority have the following official record for the three year period 2001-2003 for the approximately 9000 safety classified personnel. Numbers of cases are:

alcohol abuse: 54, drug abuse: 2, abuse of prescribed drugs: 1, total: 57 .

The data indicate a prevalence rate of around two persons per thousand employees. The statistics, or lack of statistics, reflect both an uncertainty and an underreporting, and a lack of a proper institutional database.

Facts from alcohol- and drug testing in the transport industry in Stockholm

Table 55.1 Results of alcohol and drug testing in the transport industry in Stockholm 1998 – 2004

Categories	Types of testing				
	Recruits	Random testing	Accidents	Suspicion, control of treatment etc.	All
Number of tests	19543	9603	825	1862	31833
Number of positives	88	36	2	41	167
Postive tests per thousand	7.4	6.5	4.8	4.3	5.2

Source: AB Storstockholm Lokaltrafik. Letter dated January 17 2005

Since 1998 the local transport in Stockholm has been out-sourced. One condition for the operators (railway-, bus- and underground network companies) is that random testing of employees will be mandatory. Five per cent of all employees should be tested each year as a part of quality control of public transport.

The results for the seven year period from 1998 to 2004 includes 31822 tests, of which 167 or 5.2 per thousand were positive. The details are given in Table 55.1.

Different substances were detected in the tests, and in the latest two year period from 2002 to 2004 the distribution was:

cannabis : 27, opiates : 2, amphetamine : 6, cocaine : 3, alcohol : 7.

In a big city with a test program concentrating on young recruits, it is not surprising that cannabis is the most prevalent substance that was found. The results also affirm that testing on suspicion and in the control of treatment, is important. Random testing and testing in accident cases will also give valuable information. The conclusion seems to be that testing is a valuable instrument in the quality control of the human factor in transport, and that the prevalence of alcohol and drug use in Stockholm local traffic is low, but significant.

Human risk factors for drivers

A large study named IMMORTAL, financed and organised by the Transport RTD 5th Framework Programme of the EU, has identified 28 risk-factors relevant to accidents in road driving across seven countries. The list of these factors should be of great relevance for safety in railways. The use of alcohol has a prominent place as a leading hazard, as shown in Table 55.2.

Table 55.2 Different level of risks in road-traffic and driving (Average normal risk level is set to 1.0)

Types of risks	Level of risk
Alcohol level more than 1,5	65,00
Alcohol level between 1,0 and 1,49	25,00
Alcohol level between 0,5 and 0,99	13,20
Drivers with sleep disorders/ sleep apnoea / narcolepsy	3,71
Male drivers 75 years compared with male drivers 45- 54 years	3,10
Driving on ice/snow compared with dry surface	2,50
Chronic alcoholism	2,00
Misuse of legal drugs	1,96
Use of opiates	1,83
Neurological diseases	1,75
Use of cellular phone while driving	1,72
Mental diseases	1,72
Use of cannabis	1,70
Use of narcotics and medicine	1,58
Diabetes	1,56
Cardiovascular diseases	1,23
Reduced hearing	1,19
Reduced mobility	1,17
Reduced vision	1,09
Renal diseases	0,87

Source: IMMORTAL

The findings must be read with caution, especially in relation to the medical disorders, since those persons with the largest loss of function and reduced ability to drive are banned from driving and will therefore not be included in the study. But it should give us a clear message that driving under influence of alcohol, is not compatible with safety in rail-traffic.

What do the railway companies do to prevent and reduce the risk factors?

Describing the tasks and accepting the challenge

This has been outlined in an official Swedish document. The use of alcohol and drugs is equated from the risk point of view to those illnesses that are dealt with in this ordinance. Their influence (intoxication) during work is an immediate danger, comparable with the danger of an acute heart attack. Abuse outside working hours is a risk factor for future influence during working hours too, comparable with angina pectoris being a risk factor for a future heart attack. Operators therefore have every reason to involve themselves in problems of abuse. If a decision is taken to use random drug tests, they should be carried out to such an extent (at least 25 % per year) that they give a reasonable chance of detecting possible abusers.

Rules and regulations

The Railway Act (Lov om jernbaner) in Denmark has recently been revised. The revision also tightened and extended the articles regulating alcohol and drug consumption in relation to safety classified functions. A person who exerts safety-classified duties must not exert these or intend to do so if the person is under influence of alcohol or other compounds influencing consciousness. Alcohol concentration in blood of 0.2 parts per thousand (promille) or higher is incompatible with safety-classified functions. The right to exert safety-classified functions must be withdrawn for a minimum of 6 months or permanently when a person has violated these rules.

Zero tolerance

In principle the companies have adopted the policy of zero tolerance when employees are at work. An example of this comes from Finland. In VR-group (Finnish state railway) there is a zero tolerance rule. That means 0% alcohol in breath tests at work. The cases over 0.5% are always police-court-cases. Cases under 0.5% in previous times received a warning and they had to be away from the safety sensitive work for at least six months. Nowadays they lose their jobs.

In Norway there has since 1936 been a general law demanding that employees in safety work are not allowed to consume alcohol the last eight hours before the work starts.

In Denmark DSB has had a zero tolerance policy for more than a decade.

Identification of abuse. Role of medical practitioners and Occupational Health Services (OHS)

Regular medical routine examination has been the key method for screening of all types of illness that may impair safety standards. It consists of three parts: a) Interview, present and previous complaints and illnesses, b) Clinical examination and c) Biochemical testing. People with a problem with abuse will regularly try to conceal their problems because they fear the consequence of disclosure. This represents an additional challenge to medical practitioners who traditionally expect a great deal of co-operation from their clients and patients.

In order to detect an alcoholic or a drug abuser the whole range of instruments must be utilised, and the proof will often come from biological tests. The routine interview is not a good screening method especially in the early stages of abuse. An unpublished study from Sweden showed that the clinical judgement was correct in only seven percent of cases. Detection rates are much better after introducing specific questionnaires (such as CAGE, AUDIT) during examination.

Biological Markers (gamma-GT, ASAT, ALAT, CDT) are of central importance, but the problem is specificity. In Denmark, Finland and Norway gamma-GT will be suggested as the initial screening test. However, the specificity is only 75 percent, which means that the result must be qualified by further examination, for instance CDT.

In a Swedish Occupational Health survey only five percent reported their alcohol problem, while 10 percent of the workers had pathological CDT values. The conclusion was that in control and intervention programs to detect abuse, the whole range of tests must be utilised.

In which ways are medical personnel involved in the investigation of minor and major accidents?

Denmark:

The departmental order on health requirements states that safety classified persons who have been involved in a safety related accident or adverse situation must be withdrawn from safety classified duty and reported to the Railway National Agency. The health certificate is temporarily withdrawn, and a full medical examination must be performed before considering re-certification. On the site of accident only the police can order blood samples for alcohol test or medical examination.

Finland

In minor or major accidents the health examination is made if the employer wants it, otherwise not. The employer wants the health examination made if he suspects the employee has some health problems. OHS always ask about the alcohol consumption as part of routine health examinations.

After an accident, the police will always take a breath test for alcohol. The police decide if a blood test also should be taken.

Norway

Health personnel are only consulted when the technical investigations of the accident are not sufficient to make a relevant conclusion, or when there is suspicion of abuse. Police investigation is required when a blood or a breath test is demanded. Health examination will be in the form of a review of the medical files, and if necessary a new health examination.

Sweden

The current Swedish regulations on medical examinations and health status for personnel with duties of importance to the safety of rail traffic operation (BV-FS 2000:4 §5.3) say: "The following circumstances constitute impediments to the performance of an employee's duties: Direct involvement in an accident or a near miss in which the person involved may have been affected by the occurrence in such a way that his or her duties cannot be continued in a reliable way from the point of view of traffic safety". Blood-concentration of alcohol after accidents is today not tested as a routine, except in more serious accidents.

Case histories from Nordic countries

It is important to recognise that cases of abuse of alcohol or drugs will be systematically under-reported (the iceberg phenomenon) especially on the railways with their traditional heavy reliance on engineering solutions.

Case 1 (From Denmark)

A locomotive driver was killed after a SPAD-incident (Signals Passed At Danger) resulting in a collision with another train. There was no post mortem investigation of alcohol in the blood, but the autopsy revealed bodily pathologies associated with long-term alcohol abuse.

Case 2 (From Finland)

A points man, 40 years old, had been under surveillance by VR´s Occupational Health Service since 1999 due to heavy drinking. In December 2000, fatty liver was found in an ultrasound test. In 2001, the man was referred to an A-clinic for treatment, after which his drinking somewhat decreased, but did not stop. No irregularity was observed at work. In November 2003, he was found unconscious in a point hut during duty hours. He was taken to the local Central Hospital, where a head CT was performed. On CT, there was a clear finding suggesting brain atrophy. EEG was normal. In the neuropsychological test, cognitive performance was found to be at a good level. According to the neurologist's statement, the incident had been a brain-derived epileptic-type fit, provoked by alcohol use. A couple of days before the fit, the man had consumed a lot of alcohol, but had been found sober at work. He has now been permanently removed from rail safety work.

Case 3 (From Finland)

A track machine driver, 52 years old, had been for years under surveillance by VR´s Occupational Health Service due to elevated liver enzymes. The man repeatedly denied heavy alcohol use. In the summer of 1999, he endangered rail safety by driving a track machine on an open line without the dispatcher's permission. The breath test result was 0%. He was immediately removed from rail safety work and underwent a health examination. EEG and brain CT were normal. In the neuropsychological test, signs of organic damage to his functions were found, suggesting heavy alcohol use. Fitness for driving a vehicle had diminished. A treatment agreement was made, and the man was referred to an intoxicant clinic. He reported there that he drank 10 bottles of beer daily as well as strong spirits occasionally. After a month of treatment, he refused to visit the clinic any longer. He still drinks, but says that he now drinks less. In his own words, his weekly consumption of absolute alcohol is 240 grams. The man is now employed in track maintenance work. He cannot return to rail safety work unless he quits drinking and a sufficiently long surveillance time that his cognitive functions are restored.

Case 4 (From Norway)

In November 1998 a locomotive driver fell asleep at 0825 in the morning and the freight train he was driving went into the end butt at Bodø station, causing major damage to the infrastructure, calculated at more than 4 million Norwegian Kronor. The driver was unhurt. An extended blood test showed high concentration of Flunitrazepam in correspondence with intake of major doses of drugs before and during duty. The driver had earlier been taken out of service due to drug-abuse, but had been granted an exemption from the health requirements under surveillance from OHC.

Case 5 (From Sweden)

A freight train, consisting of the engine and nine wagons loaded with LPG (Liquid Propane Gas), passed a stop signal at Borlänge station. The train continued at about a speed of 70 kph, and passed two points, which was dimensioned for 40 kph, upon which the seven leading wagons derailed, and five of them overturned. However, no LPG escaped. The driver was under the influence of alcohol, with a blood concentration of 1.53 promille.

The driver was experienced, qualified since 1982, and he had been working for SJ for ten years. He left driving for a few years but started again in 1994 and he became an "hours-worker" for the BSM Company. He has gone through the further training each year without problems and also health examinations. At the last health examination in the year 2000 he had written in the health declaration that he once thought that he was drinking too much alcohol. The examining doctor did not take up that statement for further analysis at that time.

Conclusions

The problem of alcohol and drug abuse by workers with safety critical functions in the railways has increased in the Nordic countries in the last ten years, due to heavier consumption of alcohol in the general population, a more liberal attitude towards narcotics and heavier use of pharmaceuticals.

The role of the medical profession in the investigation of accidents and near-accidents in the railway in the Nordic countries does not include a standard and mandatory health investigation of the employees involved. The result is that important background factors related to human factors are overlooked.

Today, accident reporting has a heavy bias towards the technical causes, and underlying human and organizational factors will not usually be included. There is therefore a need to restructure the teams dealing which this type of investigation so that the reports become more comprehensive in covering all possible reasons for accidents. The competence of the medical authorities in this field is not properly utilised today. Furthermore, the methods applied by the medical authorities and OHS are often inadequate, relaying mainly on clinical judgement. It is therefore the obligation of the medical expertise in railways, to come forward with better tools and more sensitive methods, especially more systematic use of structured questionnaires and biological tests.

References

Aalto, M., Seppa, K., Kiianmaa, K., Sillanaukee, P., Aalto, M., Seppa, K., Kiianmaa, K. and Sillanaukee. (1999), *Drinking habits and prevalence of heavy drinking among primary health care outpatients and general population*. Alcohol Research Centre, National Public Health Institute, Helsinki, Finland. P.Addiction. 9 Sep;94(9):1371-9.

Hoddevik, G.H. and Nygaard, M. (2004), *Abuse of alcohol or drugs among Norwegian physicians. Norwegian Medical Journal*, 7/2004; 124: 955-957.

Chapter 56

Research on Fatigue and Mental Workload of Railway Drivers and Traffic Controllers

Anabela Simões, José Carvalhais, Pedro Ferreira, Júlia Correia and Miguel Lourenço

Introduction

The Public Transportation industry, particularly the light railway sector, has known a considerable growth in recent years. This trend is mostly motivated by an increasing quality demand on public transportation networks, and also by the fact that mobility has with no doubt, become a basic need for urban societies. When compared to a heavy subway service, the tramway network provides a profitable and adequate solution for the urban contexts, mostly because of considerable lower infrastructural requirements. In fact, this kind of service is based on reliability and comfort for small distances, whilst the traditional railway service aims mostly for high speed long distance travelling.

However, the fast growth in the sector and its pursuit for higher safety and quality standards have brought deep job changes for workers such as tramway drivers and traffic controllers and need for working conditions adequate to a higher complexity and increased cognitive job demands. Therefore, management concerns are drawn to issues related to system efficiency and safety and comfort of workers, in order to ensure a quality service, not only for customers, but for all those concerned with the commercial operation.

Driving tasks impose continuous information processing, regarding the driving environment and the vehicle status. On the other hand, traffic regulation tasks are performed in dynamic and complex environments, under which workers must be able to perceive and understand multiple characteristics and anticipate events. Driving and traffic regulation tasks are characterised by a continuous information processing and decision making.

With the purpose of providing guidelines for health and safety, research was carried out by the Ergonomics Department of the Technical University of Lisbon, answering to a request from a company operating a light railway network. Two different working contexts were studied more or less independently, tramway drivers and traffic controllers.

Objectives

In order to assess whether or not fatigue and mental workload factors could represent a potential hazard, a thorough ergonomics analysis was carried out. Based on this analysis, health and safety preventive and promoting measures were established. The research element of the project consisted of the identification of mental workload factors and the establishment of correlations with potential hazards and inefficiency elements. The intervention phase aimed at the definition and implementation of ergonomic improvements. To comply with these two different steps, the following objectives were defined for both working situations:

- ·Identification and assessment of risk factors and harmful elements, regarding workers and whole working system;
- Assessment of mental workload and identification of corresponding factors;
- Identification of training and information needs relevant for workers, regarding stress and fatigue management;
- Study of preventive and corrective measures for each working situation;
- Provision of recommendations as a best practice guide for equipment design and organisational issues, as well as on the job management of stress, fatigue and sleepiness symptoms;
- Design of a leaflet, for an increased awareness on health and safety promotion and stress and fatigue symptoms prevention.

Methodology

The same methodology was used for both working contexts, whenever necessary adjustments were made, particularly during the ergonomics analysis.

Questionnaire for subjective assessment of working conditions

The first analysis step consisted of the development and application of a questionnaire for the subjective assessment of working conditions, focusing mostly on the characterization of discomfort and fatigue symptoms. This questionnaire was built taking into account specific details of each context, using several kinds of questions, such as open answers, "yes or no" answer and scoring scales. Since the application was carried out as an interview (using the questionnaire as a guide), additional notes were taken regarding particular remarks made by workers, which provided important information on several work related issues. Basically, the questionnaire was built into 7 different sections: Identification of work site; Characterization of the working population; Characterization of the working station; Working schedules; Job Characterization; Workload assessment; Health related issues.

Regarding the drivers' population, 88.5% of them (N=54) were interviewed, whilst for traffic controllers, the sample was the total population (N=12). This questionnaire made it possible to identify and characterise the causes of stress and fatigue, and it also provided an overview of important organisational aspects relevant to subsequent analysis.

Ergonomics analysis

A job analysis was carried out, using video recordings for data collection, which allowed further detailed observations. For this procedure, as each one of the working contexts had different requirements, specific protocols for video sampling and analysis were defined for:

- Registration and categorisation of observable elements, such as eye movements, posture and other behavioural aspects;
- Registration of fatigue signs such as yawning;
- Characterisation of all significantly different working periods and environments (day and night, beginning and end of work shifts);
- Identification of particular differences in working methods between workers.

Regarding tramway drivers, the sampling of routes and drivers aimed to obtain an accurate representation of all different driving environments. As for traffic controllers, the considerable

variability of tasks performed required the collection of continuous video records throughout a whole working shift. For both working contexts, data have been collected in different working shifts in order to cover all the working schedules, including day and night shifts. A total of 6 drivers and 3 traffic control were recorded in three different shifts. All video records were collected using a set of 4 cameras connected to a multiplexer and a video recorder.

Since the variables under analysis weren't quantifiable through direct means or absolute values, risk factors were characterised by means of indirect and observational methods. For both groups of subjects a chart was prepared to support a self-confronting interview, based on relevant issues gathered during video observation. Basically, these interviews consisted of the interpretation of different observed behaviours, in order to understand the hidden aspects required by task performance. Subjects verbalized certain aspects of their behaviour and identified potential fatigue and stress causes, as well as factors contributing to a mental workload increase. The interviews aimed at:

- Getting workers to identify their different behaviours and establish a relation between them and the evolution of a fatigue condition;
- Establishing the particular events that justify, on a regular basis, those specific behaviours;
- Getting workers to verbalise the consequences they considered fatigue to have for their performance.

Workload subjective assessment From all the known subjective methods for mental workload assessment, the NASA-TLX, developed by NASA is probably the one most used for the driving work context. Pauzié and Pachiaudi (in Lenné, M. et al., 2004) recently developed the DALI method, adapting the NASA-TLX to automotive driving. This method was used in both contexts, and within traffic regulation an additional specific analysis was carried out for violation of safety procedures and regulations.

This DALI protocol integrates a set of 6 factors for mental workload assessment: Attentional effort; Visual demands; Hearing demands; Time pressure; Interference; and Stressful situation.

Each one of these factors is assessed on a scale of 1 to 6. Factors are then organized into 15 pairs, from which subjects are asked to select the more significant ones.

Analysis of safety procedures and violations Violations consist of deliberate deviations from the established rules, procedures or instructions. Within traffic regulation, safety rules and procedures are crucial for the whole operation. A failure to comply can jeopardise safety. The analysis carried out considered:

- Identification of organisational factors that, combined or separate, increase the potential for the occurrence of violations within traffic regulation;
- Provision of guidelines on priorities and courses of action for improvement.

The methodology used was developed and published by the HSE – Health and Safety Executive (Improving compliance with safety procedures – Reducing industrial violations, 1995). The theoretical basis of this tool considers 4 different categories of violations:

- Routine violations consist of common events that, although contrary to rules, have become usual;

- Circumstantial violations are based on contingency situations or events;
- Exception violations are related to very specific circumstances in which usually, something has gone wrong;
- Optimisation violations occur when there's an intention of improving a working situation.

Based on these categories, it is possible to establish intervention areas for the minimisation of the potential for the occurrence of violations. The first step should be the selection of procedures or rules to be submitted to analysis, after which interviews are conducted with those who are responsible by the design of these procedures (management) and to those are compelled to apply them (workers). It is possible to characterise not only the companies' safety trends and main concerns, but also to compare different perspectives. The selected procedures are then characterised through a set of 48 concise statements, which workers should score from 0 to 3 according to the degree to which they concur with each statement. The results of this classification are then input into a set of matrices, which will output the final scores and the guidelines for intervention, according to 13 possible categories of measures as shown in Table 56.1.

Table 56.1 Categories of measures for intervention

A.	Rules and procedures – objectives
B.	Rules and procedures – application
C.	Rules and procedures – training
D.	Hazards and risks – training
E.	Commitment to safety – workers
F.	Commitment to safety – management
G.	Supervising – control and detection
H.	Supervising – supervising style
I.	Modification in equipment design and productive area
J.	Job design
K.	Working conditions
L.	Logistics support
M.	Organisation

Results

The results are presented separately for tramway drivers and traffic controllers.

Tramway drivers

The driving task and subtasks Tramway driving involves a complex set of subtasks, continuously in interaction with the steering itself.

- Besides driving the vehicle in commercial operations, the driver has to perform all the necessary steps for the preparation of vehicle before going into operation and coming out, as well as vehicle manoeuvring outside commercial lanes (parking and shops).
- In case of breakdown or any malfunction, drivers must perform a set of technical procedures to support continuity of operations
- Drivers must be able to communicate at all times any changes in the vehicle or tracks.
- Drivers are also directly responsible for certain tasks of commercial operation such as ensuring the safety and comfort of passengers and keeping up with schedules.

Questionnaire for working conditions assessment The 88.5% of drivers' population corresponded to 54 drivers on duty at the time of the study. The population is young, having only one driver above 40 years old, while the majority is below 30. All individuals except for one were male.

– *Workstation characterisation* In general, workers considered the working conditions acceptable or good. However, some complaints were noted in relation to the windshield's high surface, which caused difficulties in regulating the cabin's temperature when outside temperatures were extreme and visibility difficulties. A number of complaints were also noted in regard to noise sources. The layout of some of the instruments was criticised in terms of hand and visual reach. Drivers considered that, due to their frequent use, the position of instruments such as the speed dial could be improved.

  Special attention was given to the "automatic surveillance system" which controls the state of awareness of the driver. Although it can only detect if the driver is reacting at all or is unconscious, it was recognised as an important security device. However, since it requires the pressing of a button every few seconds, drivers complained of fatigue, especially as they approach the end of the work shift.

– *Organisation of working time* Two different complaints were identified in regard to driving schedules and the organisation of work breaks. While some were displeased with the fact of sometimes, having lunch out of regular hours, others were more concerned with some of the schedules imposing long driving periods before or after the meal break. There were concerns with the distribution of driving periods during the day and of breaks within the shift.

– *Job characterisation* Every driver considered their job as very demanding in terms of awareness and attention, because of driving in urban environments, the unpredicted reactions of pedestrians and passengers, inadequacies of road signs and the sense that mistakes could have serious consequences or jeopardise their career.

  Nevertheless, the majority of drivers consider themselves well prepared for this job. It is clear that drivers are confident that their technical knowledge and training provides them the means to perform quality and safe driving. There is a high sense of commitment to the job and strong determination to perform well.

– *Stress and Fatigue symptoms* The need for high levels of awareness and attention are also seen as causes for stress. In addition, driving in monotonous environments was also referred to as a stressing element. Although the next project phase would allow a more accurate assessment of fatigue, the questionnaire provided indicators as to what consequences it could have over driver's performance. The effect most frequently mentioned was the increase of reaction time to all driving situations. Also when feeling more fatigued, drivers admitted a higher uncertainty regarding dynamic elements of the driving environment, such as road signs.

Video analysis and self-confronting interviews As driving tasks were systematic, the video analysis was supported by a checklist of events, which were to be identified and registered along the driving journey. This checklist referred to events like the frequency of glimpses at the speed dial, the number of movements on the thrust handle, and others. Approximately 10 hours of video records were collected, across 25 samples of 6 different drivers. During self-confronting interviews, 2 of these drivers were no longer present at the company and so, they were excluded from the procedure.

From these four drivers, the more experienced revealed a higher awareness of dangers, which led them to systematically lookout for evidence in the environment that could help them anticipate potential risks. Thus, it became clear through this process that more experienced drivers were generally more focused on their job, while inexperienced ones sometimes revealed hastiness on potential dangerous situations, such as crossroads and pedestrian crossings.

Another significant fact was the difficulty that inexperienced drivers revealed in identifying the moment in which fatigue was beginning to affect them, while experienced ones were able to clearly identify their first signs of weariness. Despite the fact that these drivers recognised fatigue effects after driving for a period of approximately two hours, it was also clear through video analysis that these effects were usually observed after 1 to 1.5 hours of driving time.

Experience is with no doubt a critical factor for drivers' performance, perhaps more then in other working contexts, as it provides the necessary knowledge of all driving environments of the network, and perfects skills for adequate reactions to unexpected events.

Subjective assessment of workload The application of the DALI protocol revealed high levels of mental workload in drivers. However, in order to correctly validate these outcomes, more systematic assessments with this method were necessary. Thus, due to unforeseen constraints, the DALI was limited to the four drivers that undertook the previous analysis procedure (Table 56.2). Driver 4 is one of the inexperienced ones and also one of older workers (36 years old).

Table 56.2 Mental workload indexes obtained with DALI

	Workload index
Driver 1	88
Driver 2	85
Driver 3	81
Driver 4	92

Traffic controllers

The traffic regulation task and subtasks In a dynamic and complex environment, traffic regulation of a tramway network involves performance of multiple subtasks. These subtasks often require simultaneous performance imposing divided attention demands on the traffic controller. Time management becomes crucial for performance of:

- Monitoring network performance and in case of need, carry out immediate corrective measures or modifications;
- Providing costumers of the transportation services with the quality standards demanded;
- Supervising the operation of all equipment directly related to the network;
- Instructing drivers and other workers on courses of action under all circumstances;
- Logging information regarding incidents or any events occurring in the network, and provide information of their status;
- Activating the proper emergency devices in case of accident or incident.

During network operation, there are two controllers at the CCP – Central Command Post (the traffic regulation office) and a third one on location. Thus, traffic regulation is based on a distribution of tasks, mainly between the two controllers at the CCP, as the third one operates as on site support for critical situations. This job organisation permits efficient problem solving for traffic regulation.

Questionnaire for working conditions assessment At the time of this study, only 12 controllers were on active duty, all males with ages ranging between 24 and 43 years old.

– *Characterisation of the workstation* Regarding technical and environmental aspects at the workstation, no problems were reported due to recent efforts of management and workers to improve the layout of workstations and equipment in the CCP. Several complaints were made related to noise levels. Due to the large number of controlling tasks performed at the CCP, frequent audio alarms required the controllers' attention. In addition, there was a repeated coming and going of people from outside the CCP. Although these noise levels were found to be below the legal requirements, they clearly represented an important disturbance factor and an increase of mental workload, stress and fatigue.

  Traffic regulation is also very depending on communication. In order to control the network operations, controllers must efficiently make decisions and promptly pass them through to those responsible for their implementation. This capability depends entirely on telecommunication devices available in the CCP. For this reason, controllers were very critical of these communication devices, especially the quality of sound of the radio system on board each vehicle, which sometimes made difficult to clearly pass through instructions to drivers.

– *Organisation of working time* The responsibilities of traffic controllers make it impossible for them to perform regular meal breaks and working schedules. It was established as normal and necessary to sometimes skip the lunch break in order to attend to a critical event or emergency situation. Although in this case they would have a break whenever possible and eat, this break was always made with certain hastiness. Under these circumstances, it was noticed that workers were deprived of a fundamental rest

period and that stress and fatigue indicators became more significant. On the other hand, the daily rotation of shifts had sometimes sudden variations of working hours. Although these schedules always complied with a minimum required by law, of 11 hours rest between two working shifts, workers considered that due to the job's high demands this was sometimes insufficient for an adequate recovery.

– *Job Characterisation* Like drivers, traffic controllers characterized their job as very demanding in terms of awareness and attention. The reasons found were related to a constant and simultaneous supervising of all vehicle manoeuvres and operations in the network, as well as decision making on necessary courses of action to those situations.

– *Stress and Fatigue symptoms* Regulation tasks were unanimously considered very stressing, especially when operating the network under a malfunctioning condition, because of systems breakdown or accidents on the railways. Because of their impact on regulating activities, communication difficulties and malfunctioning were considered one of the most stress – inducing elements. Workers stated that whenever communications were down, they were powerless to get through any instructions out to on-site co-operators. The need for constant attention and awareness were the main causes for high levels of stress and fatigue. The effects perceived by workers were a reduced ability of reasoning and consequently a delay in decision making.

Video analysis and self-confronting interviews

The checklist used for drivers wasn't suitable for the analysis of traffic controllers video records. Thus, a more flexible description of events was used, in order to grasp the high variability of task performance in this work context and to identify fatigue inducing elements and note evidence of performance decrease. For this purpose, the collection of simultaneous audio records was necessary, in order to understand the on going events and the corresponding reactions of controllers. This process included 22 hours of video records from two different working shifts, including the following variations on operation mode:

- During the night when vehicles were retrieved from circulation and parking management was carried out, as well as vehicle maintenance and cleaning;
- The dawn when circulation began with the replacement of vehicles in circulation;
- A period before noon and before afternoon rush hour, when circulation assumed normal standards.

A total of 8 controllers on duty were observed, to obtain a reasonable sample of different levels of experience. The self-confronting interviews were based on a set of hypotheses raised during video analysis. Thus, the interviews would either confirm or refute these hypotheses. They were:

- Under more demanding circumstances, are there less important tasks that can be shed?
- What is the importance of informal job strategies within the operation?
- What is the importance of cooperation between on duty controllers for the efficiency of the operation?

Traffic regulation relies on the definition and constant reassessment of priorities. Inevitably, when faced with more pressing matters the lower priority issues are postponed or might not even

perceived. Regarding stress and fatigue effects, controllers associated them more with bad humour and aggravating moods during tense situations, rather then the usual yawning and stretching, which were singly associated with undemanding and monotonous situations.

Experience was considered a fundamental component of traffic regulation, and as controllers are able to achieve an increased promptness and confidence when facing complex situations, they tend to value this component of their job. Nevertheless, training was considered as the fundamental ground that each one can then develop and perfect through experience.

Teamwork was also characterised as a crucial element for traffic regulation. In fact, all individuals stated that their performance greatly depended on the ability of both on duty controllers to cooperate.

Subjective assessment of workload

The DALI protocol was applied to 6 traffic controllers and although high mental workload levels were found validation was not possible (Table 56.3).

Table 56.3 Mental workload indexes obtained with DALI

	Workload index
Controller 1	100
Controller 2	99
Controller 3	93
Controller 4	75
Controller 5	80
Controller 6	77

Analysis of safety procedures and violations

Although designed for industrial purposes, the HSE methodology provided an adequate identification of inadequacies of procedures content and organisation. From an initial list of 40 operation and safety procedures, 20 were selected according to their significance. Three controllers were selected for interview based on different degrees of experience, and two interviewees represented the CCP management and top management of the operation. For the second step of this procedure, 10 controllers filled out the 48 statements list, representing approximately 75% of all controllers on active duty.

The interviews

The information collected from all 5 interviews was used to establish comparisons on three different levels:

- Comparison between perspectives of different traffic controllers;
- Comparison between perspectives of different levels of management;
- Comparison between perspectives of traffic controllers and management.

Traffic controllers more or less shared the same perspective on safety issues. Management was considered a participative element and generally open to debate, especially on issues concerning the CCP management. Nevertheless, controllers were unanimous in recognising a strict attitude of management in regard to errors, no matter how small they might seem. As for violations on safety or operation procedures, it was also unanimously assumed that they sometimes occurred, even if they were at all times supervised by the CCP management.

Regarding the different levels of management, the interviews stressed different perspectives on the application and content of procedures. Whilst top management revealed total confidence in the efficiency of procedures, the CCP management does not rely so much on them, as they are able to identify and follow up on everyday constraints that these procedures impose on traffic regulation. Given the high complexity of the operation, the development of documents and procedures has to go through an extensive process of teamwork discussion, and it depends on the approval of many different entities.

When comparing traffic controllers and management, it was first noticed that the CCP management has a perspective rather closer to controllers than top management. In fact, the participative attitude is reflected mostly by this level of management, and not so much by top level. On the other hand, any warning, sanction or punitive action comes from the top, and CCP management usually adopts a sympathetic position in regard to controllers.

The analysis of procedures

Based on the list of statements filled out by controllers, different priorities for intervention were obtained. The issue that out came as a first priority was the "Modification in equipment design and productive area". This concerns the adequacy of systems, technical requirements and limitations, for the individuals' aptitude and safety and comfort conditions. The definition of procedures should aim for a better compromise between job demands and the workers' aptitude, thus reducing the potential for the occurrence of violations. Through this analysis, it was made clear that the non compliance with procedures by traffic controllers was mostly due to the acquired notion that their violation simplifies the task to be performed. The reason for this attitude was found in the inadequacy of the following systems, which strongly contributed to the increased potential for violations:

- The radio communications that sometimes make it difficult to pass through the necessary information;
- The high frequency of audio alarms that constantly draw the controllers' attention from other tasks;
- The complexity of the computer software and devices that, despite their reasonable reliability are subject to flaws, which means controllers might question the information that these systems provide.

Discussion

The data collected during all analysis procedures provided solid grounds to support health and safety recommendations on several matters and for both groups of workers under study. The clear identification of fatigue symptoms and the establishment of their relationship with job demands was presented to the company's management as a basis for the need of improvements on shift scheduling, training content and organisation, and safety procedures.

For tramway drivers, the core of recommendations was the organisation of working time and rest periods. Also references were made to the layout of the driving post and specific devices on board the vehicle, as well as to the placement and organisation of railway signs. For traffic controllers, the most critical aspect concerned the high variability of tasks and implications for job demands. As for the systems used in traffic regulation, such as communication devices, recommendations were made to improve their adequacy for the high variability of working scenarios.

Knowing that major constraints related to both of these work contexts are inevitable, the goal was to provide means for their minimization and control, mostly through a more efficient management of working times and an improvement of working situations. Special attention was also given to the importance of prevention of health problems from fatigue and stress effects during a working day, and in the long run.

References

HFRG/HSE (1995) *Improving Compliance with Safety Procedure – Reducing Industrial Violations*, HSE Books, ISBN 0 7176 0970 7. HSE Books.

Lenné, M. et al. (2004) – *Review of recent research in applied experimental psychology: Implications for counter measure development in road safety*. In Monash University Accident research Centre- Report documentation.

This page intentionally left blank

Chapter 57

Fatigue and Shift Work in UK Train Drivers

Barbara Stone, Alison McGuffog, Mick Spencer, Claire Turner and Ann Mills

Introduction

Fatigue, resulting from schedules of work and rest, is an important issue in the rail industry. Unfavorable shift systems can have wide-ranging effects, on the personnel concerned, on the efficiency of the organizations for which they work and, most critically, on the safety of the rail system.

The requirement for this work primarily arose from recommendations resulting from the public enquiry following the serious rail accident at Southall in 1997. The public enquiry recommended that 'current rules governing drivers hours should be reviewed in light of current research into human behaviour', this recommendation in turn led to the United Kingdom (UK) rail regulator to apply pressure for a review of the limits in working hours of all safety critical workers. These limits are given in an appendix in a Railway Group Standard (1996).

A research program was undertaken by QinetiQ for the RSSB to understand the risks of current shift patterns in UK train drivers and develop strategies for risk reduction and control. It was hoped that this research would include all safety critical workers but unfortunately it was only able to focus on passenger train drivers. The aims of this research program were to understand the risks of current shift patterns in UK train drivers and develop strategies for risk reduction and control, specifically:

- to determine the risk associated with the current situation;
- to determine mitigation strategies that could be applied across the industry to effect reduced levels of fatigue-related accidents;
- to determine a method of monitoring the rail system to provide on-going assessment of fatigue related risk;
- to deliver useable and practicable guidance for designing shift patterns for drivers within the rail industry.

Methods

The study involved a range of approaches to provide an in depth understanding of the culture of shift working practices, methods to monitor fatigue, relationship of accidents and incidents to factors associated with shift work and analysis of data from train drivers

Initial interviews were carried out with roster clerks in order to understand the practices used in different train operating companies (TOCs) (McGuffog, 2003). The main component of the programme involved the collection of information from the drivers themselves, by means of two surveys. The first survey was a questionnaire relating principally to the drivers' shift patterns and attitudes to various aspects of shiftwork. It also included questions on the health and well-being of drivers. The second survey was in the form of a diary of drivers' sleep and duty, which

was completed over 28 consecutive duty periods (McGuffog et al, 2004). An investigation of accident risk based on the occurrence of Signals Passed at Danger (SPADs) was undertaken. The programme also involved the evaluation and identification of effective strategies for drivers to combat the effects of fatigue (Turner and Stone, 2004). The final aspect of the study involved an evaluation of the tools and techniques used for estimating risks associated with shift work (Stone, 2004).

Here we are presenting the results in the form of guidelines for the management and reduction of fatigue in train drivers.

Results

Guidelines for management of shift patterns

These guidelines have been formulated using information from the questionnaires, diaries, and the SPAD analysis . In addition, information has been drawn from the scientific literature and other knowledge bases on fatigue and accident risk. The main factors that increase fatigue and incident risk have been identified and, from these, guidelines have been developed that, if implemented, would reduce the risks associated with demanding shift patterns. It is appreciated that commercial imperatives and other factors may complicate the introduction of the suggested guidelines. For this reason, examples are given of the likely impact of their implementation on the operations of those TOCs whose drivers participated in the RSSB survey[1]. Because the risk is related to the actual hours worked, which may differ from those originally planned, it is important that the guidelines are applied not only to the rostered hours but also to the patterns worked. This would ensure that they included factors such as overtime, shift swapping and spare turns.

Although each issue is discussed in turn, it should be recognised that there is a considerable interaction between the individual factors and that it is only by considering the roster pattern as a whole that sympathetic rostering can be achieved.

Shift length

The duration of the shift is a key factor leading to fatigue. Although duty periods are governed by the Hidden Limits, the RSSB survey found some examples of duties lasting for 13 hours or more. Other research has shown that accident rates in shift workers are 25% higher on 12-hour compared with eight-hour shifts. It is therefore important to restrict the amount of overtime taken at the end of a shift. Moreover, the impact of long duty periods on fatigue are likely to be most severe on night and early morning shifts.

A reduction in shift length by limiting night and early shifts to 10 hours would mitigate fatigue. Based on the data from the RSSB survey, this would affect 5.1% of early shifts and 10.1% of night shifts. It is also suggested that very early shifts (e.g. those starting before 05:00) be further restricted to eight hours (Figures 57.1 and 57.2).

1 Guidance in the Safety Critical Work Regulations 1994 is based upon the 'Hidden Recommendation 18' from the Clapham Junction railway incident of 1988. The recommendations are: a maximum of 12 hours per duty; 12 hours rest between booking off from a duty to booking on for the next duty (this may be reduced to eight hours at the weekly shift changeover, in the case of staff working a shift pattern which rotates/alternates on a weekly basis; a maximum of 72 hours per week (Sunday to Saturday); and a maximum of 13 consecutive duties in any 14 day period.

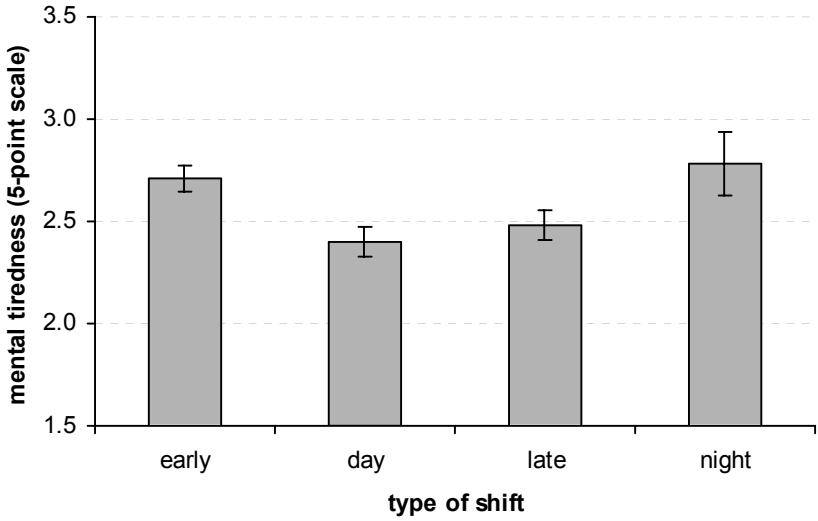


Figure 57.1 Mental tiredness related to shift type

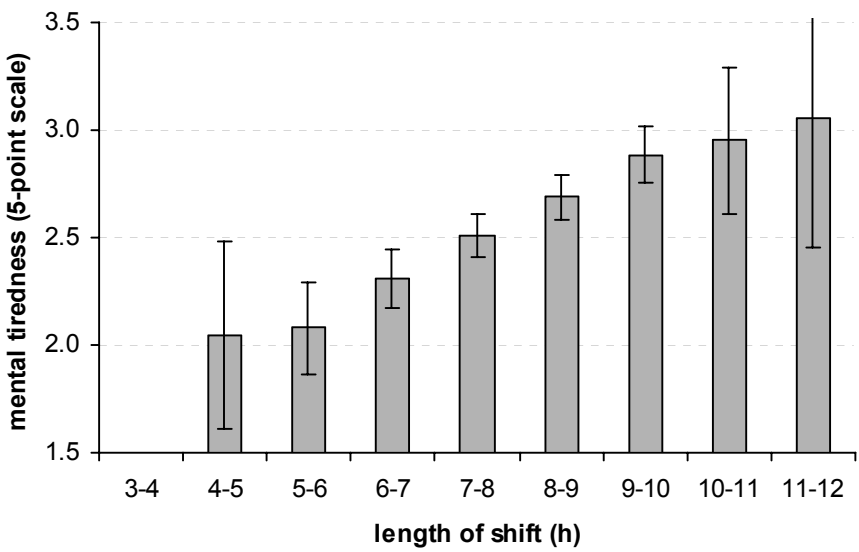


Figure 57.2 Mental tiredness related to shift length

Continuous driving time

The second factor identified by drivers as contributing most to fatigue was the duration of continuous driving. In particular, long periods of continuous duty without a break are particularly fatiguing, and this was supported by the analysis of SPAD data (Figure 57.3).

It is therefore important to ensure that rest breaks at appropriate times are taken in order to reduce risk. Ideally, breaks should be of at least 15 minutes duration and free from any work-related activities. However, a break of any duration is better than none at all. As with shift length, it

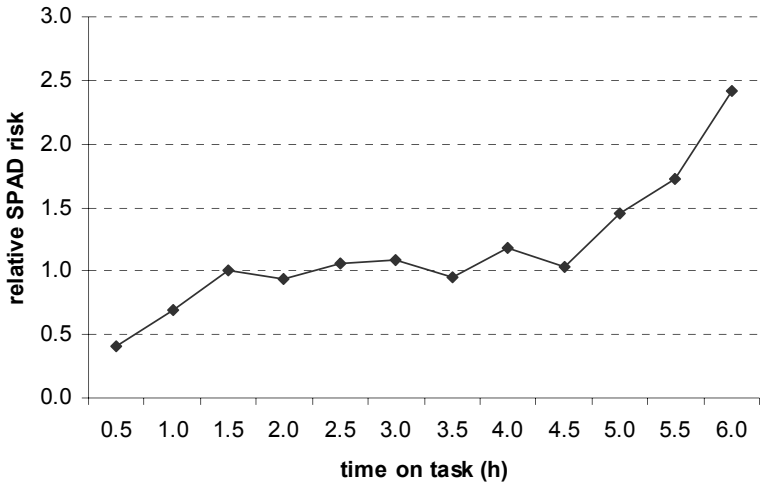


Figure 57.3 Relative SPAD risk related to time on task

is night shifts and early starts that are most affected by fatigue as the duration of continuous driving increases. Early starts were found to be particularly associated with long periods without a break. It is therefore proposed that continuous periods of driving time be limited to four hours.

Hours worked per week

Although the average number of hours worked each week was 39, this was exceeded on many occasions. Drivers reported an average maximum hours per week of 52.5 hours, but 25% of drivers had worked for 60 hours or more on at least one occasion in the last year. Long hours inevitably involve long individual shifts or short recovery times or both.

Limiting maximum hours over a rolling week to 55 or 60 would allow sufficient recovery time between shifts. Based on the data from the survey, a restriction to 55 hours would affect 5% of shifts, and only 1.8% of shifts would be affected by a restriction to 60 hours. This is comparable with restrictions on working hours in other industries. For example, there is a 55-hour rolling week limit for aircrew and 56-hour weekly limit for heavy goods vehicle drivers.

Consecutive shifts

There is a marked increase in SPAD risk when duties continue over more than six consecutive days. In particular, consecutive early or night shifts are associated with a greater incidence of cumulative sleep loss, which could further increase the SPAD risk (Figure 57.4).

It is therefore recommended that consecutive night shifts should be limited to three before a rest day, early shifts to five before a rest day, and other shifts, including mixed patterns, to seven before a rest day. A limitation of seven consecutive shifts would affect only 7.8% of the shifts surveyed.

Shift variability

Some variability in the start time of shifts is inevitable when running an efficient railway operation. Changing the shift start time from one day to the next is especially fatiguing in an

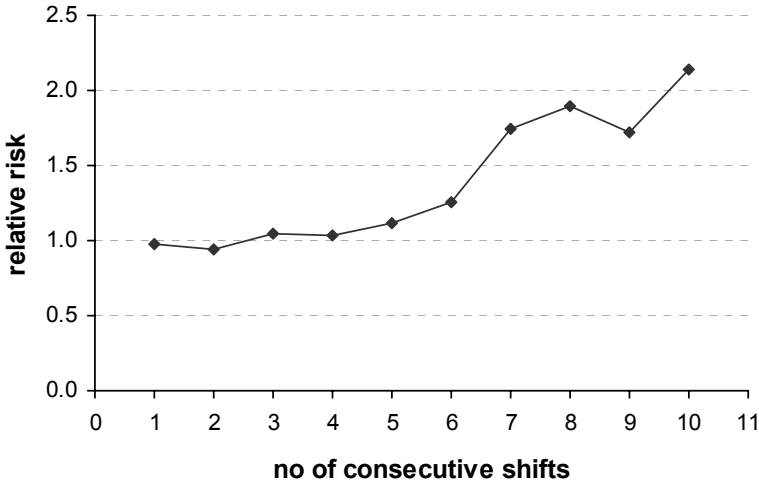


Figure 57.4 SPAD risk over consecutive days

advancing schedule when the duties are early starts. A rapid switch from a late finish or a night shift to an early shift is also particularly fatiguing. The use of spare turns also contributes to shift variability, with the four-hour window of start time leading to some loss of time for sleep between shifts.

Although shift swapping may improve work/life balance, it may lead to additional variability in start time. Careful monitoring is required to prevent the revised shift patterns being more fatiguing than the original rosters. Controlling the variability of shifts would reduce fatigue and a rapid change from a late finish or night shift to an early start should be avoided.

Rest between shifts

Drivers require enough time between shifts to obtain sufficient sleep. In the RSSB survey, 15% of drivers reported one or more rest periods of less than 12 hours in the previous 12 months. Commuting time can in some circumstances contribute to the loss of sleep required for recovery. It is particularly important to have sufficient time between consecutive night shifts to allow for a nap before the night shift, especially when sleep after the preceding shift has been insufficient. It is therefore recommended that there is adequate rest between shifts. A 12-hour minimum rest period is the current requirement and, if adhered to, this should be sufficient for most shift types. A rest period of 14 hours between consecutive night shifts would be desirable to allow sufficient recovery.

Rest days

The survey found that most drivers work for five or six days before a break of at least one day, although the maximum number of days worked consecutively was nine. One quarter of drivers worked on their scheduled rest day between two and three times each month. This loss of rest days increases risk associated with working on consecutive days. This is particularly problematic when returning to earlies after late or night shifts. It is therefore recommended that the change from nights to earlies incorporates at least two rest days (e.g. 54 hours minimum) and that the

change from lates to earlies is separated by at least one rest day (e.g. 30 hours minimum). In other situations, at least one rest day should be provided.

Fatigue countermeasures for drivers

Various coping strategies that may be useful in counteracting the fatigue of train drivers were evaluated by considering the scientific evidence for their effectiveness and practicality of implementation in the rail industry. The countermeasures evaluated were: napping, breaks, social interaction, work variety, shift swapping, caffeine, health monitoring, exercise, education and training, bright light, dietary strategies, gum-chewing, temperature, noise, and alertness monitoring devices. Bright light was considered impractical for train drivers, even though it has been used successfully in several work environments and could possibly be useful for signallers. Dietary strategies, gum-chewing, temperature, noise, and alertness monitoring devices were not considered further either because insufficient information was available on their efficacy, or because they were considered impractical for train drivers.

As a result of this review a Good Practice Guide for Drivers 'Coping with Shift Work and Fatigue' was developed with a view to supporting driver education programs.

Fatigue monitoring

As part of the program, an evaluation of tools and techniques currently available for estimating risks associated with irregular work-rest schedules and shift patterns was undertaken *(6)*. The following tools were reviewed:

- the HSE Fatigue Index Risk Assessment Tool;
- the Sleepwake Predictor;
- the Fatigue Audit Interdyne (FAID);
- the Two-Process Model of Sleep Regulation;
- the Sleep, Activity, Fatigue, and Task Effectiveness Model (SAFTE);
- the Circadian Alertness Simulator;
- the Interactive Neurobehavioural Model;
- the System for Aircrew Fatigue Evaluation (SAFE).

The review concluded that, as far as the available models are concerned, the HSE Fatigue Index is currently the best option for use in the assessment of the shift patterns of safety critical rail workers. The HSE Fatigue Index is freely available, it is simple to use, and can be applied by rail staff when adequately trained. It provides an overall assessment of the risks associated with fatigue in any shift pattern. Furthermore, it was designed specifically for the variable patterns in the railway industry by DERA (now QinetiQ) in support of Railway (Safety Critical Work) Regulations, 1994. The Fatigue Index is currently being revised by incorporating new data, including that from the current RSSB programme, and by gathering information from existing users about potential areas for improvement.

A tool such as the HSE Fatigue Index could be used by the TOCs to demonstrate the safety of their proposed rosters and to monitor the shift patterns that are worked in practice, including

overtime and spare turns[2], [3]. The Index could be integrated into the rostering software of the TOC. For example, it is currently incorporated into the IRMA and CrewPlan scheduling tools. TOCs would then be able to define their own schemes for the management of fatigue and submit these to the Regulatory Authority, with the Fatigue Index forming part of the on-going assessment and audit process. In addition, a fatigue reporting system was proposed so that staff could report issues of concern anonymously. It was suggested that the avoidance of excessive fatigue should be a responsibility shared jointly by the TOCs and their staff. The drivers would need to receive training to ensure that they would be able to evaluate their own fitness to drive, make sure they are fully rested before duty, and employ the appropriate strategies to mitigate fatigue.

It was proposed that, during the reporting of incidents and accidents, information on fatigue should be collected, irrespective of whether fatigue is judged at the time to be implicated. The data collected should include: the timing of duty and of any preceding breaks; the previous shift pattern; immediate sleep history; commuting time; a fatigue checklist including specific symptoms (for example, unintentional closing of eyelids, problems completing tasks, becoming easily distracted) as well as environmental factors.

Concluding remarks

The findings arising from this research programme were used to produce draft guidelines for reducing risk and improving the management and monitoring of shift patterns. An industry-wide workshop was then held in February 2005, at which the results of the research, together with the draft guidelines, were presented and discussed. Feedback was received from representatives of ATOC, Train Operating Companies (TOCs), Freight Operating Companies (FOCs), IMCs, the Associated Society of Locomotive Engineers and Firemen (ASLEF), and the Health and Safety Executive (HSE) concerning the practicalities of implementing these guidelines. The input from industry has been incorporated into this chapter.

The programme of work has shown that fatigue is an important issue in rail operations and specifically on train driver performance. It is hoped that further work will address the specific shiftwork issues that relate to track workers, freight drivers and signallers. As a result it is not proposed to change the current requirements that are given in Railway Group Standards as this governs all safety critical workers, however companies will be encouraged to consider their current rostering practices in light of the research findings.

The results from the evaluation of potential countermeasures have been used to develop a good practice guide for drivers to help them cope with shift work and fatigue. The leaflet will be made available to train operating companies to distribute to their drivers. RSSB has developed additional resources that can be used by companies during their briefing cycles.

The introduction of new Railway Safety legislation in the UK in the forthcoming year will include an approved code of practice on managing fatigue in safety critical work. It will require the controller of safety critical work to introduce a fatigue management system, it is hoped that

2 During an industry workshop to discuss these findings it emerged that some representatives considered that overtime and spare turns were effectively monitored at the present time, and therefore there was not really a need for a new system.

3 A report on rail rostering culture concluded that there appears to be no set process for rostering and differences exist between companies 0. One area of particular concern is monitoring procedures, with some companies placing the onus on rostering clerks and company management, and others placing it firmly with staff themselves. Responsibilities should be clearly defined to reduce the risk of limits being exceeded.

the revisions to the HSE Fatigue Index which are currently ongoing incorporating new data from this research programme will help organisations to ensure that safety critical workers do not carry out safety critical work where they are already fatigued, or have such work patterns that would be liable to cause them fatigue.

References

McGuffog A. *Rail rostering culture: Interview findings*. QinetiQ Centre for Human Sciences Report, QinetiQ/KI/CHS/CR032371; 2003

McGuffog A, Spencer M, Turner C, Stone B. *Working patterns of train drivers: implications for fatigue and safety*. QinetiQ Centre for Human Sciences Report, QinetiQ/KI/CHS/CR043098; 2004.

Railway Group Standard GH/RT4004: *Changes in Working Hours – safety critical work*, Issue One, August 1996.

Stone BM. *Evaluation of current tools and techniques used for estimating risks associated with shift patterns*. QinetiQ Centre for Human Sciences Report, QinetiQ/KI/CHS/ CR032327; 2004.

Turner C, Stone BM. *Review of coping strategies to mitigate fatigue of train drivers*. QinetiQ Centre for Human Sciences Report, QinetiQ/KI/CHS/ CR043243; 2004.

PART XVI
Competencies, Training and Procedures

This page intentionally left blank

Chapter 58

Educational Activities in Relation to Affective Commitment, Voluntary Turnover, Perceived Inner Motivation and Work Effort

Cathrine Brustad, Anna Gunningberg and Marie Svensli

Introduction

During spring 2005 research was carried out in Norwegian Rail (JBV) to measure possible correlations between courses and educational activities and affective commitment, voluntary turnover, perceived inner motivation and work effort. A questionnaire was sent to 306 employees in different parts of the organization. There were 219 answers, which gives a reply rate of about 72 percent.

The goal of this research was to examine possible connections between courses and educational activities and affective commitment, voluntary turnover, perceived inner motivation and work effort. Courses and educational activities were divided into strategic rooting, support from leader and work colleagues, and relevance for the work situation. This gave 12 hypotheses. Since the factor analyses revealed that the factor work effort had to be left out in the further analysis due to low validity, only the following 9 hypotheses were further analysed.

H1: Support from leader and colleagues when participating on courses and educational activities leads to:

a) lower voluntary turnover
b) stronger affective commitment
c) stronger inner motivation

H2: Relevance of the course or educational activity to the work situation leads to:

a) lower voluntary turnover
b) stronger affective commitment
c) stronger inner motivation

H3: Conscious strategic rooting for courses and educational activities leads to:

a) lower voluntary turnover
b) stronger affective commitment
c) stronger inner motivation

The model in Figure 58.1 illustrates the hypotheses.

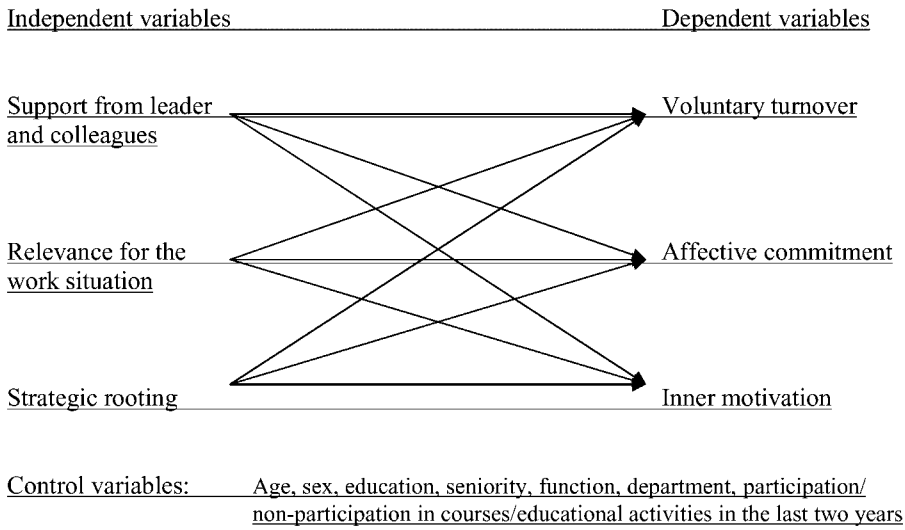


Figure 58.1 An illustration of the hypotheses

Presentation of results

The correlations analysis showed significant correlations between all the independent and the remaining dependent variables and between these and the control variables.

Voluntary turnover

Voluntary turnover proved to have the strongest co-variation (R^2). This means that the independent and control variables are best suited to explain voluntary turnover. The more support leaders and work colleagues give when an employee participates in educational activities, and the more relevance the educational activity has to the work situation, the lower is the employee's voluntary turnover. Vandenput (1973) confirms in his research that human considerations have the largest influence on effective transfer of learning to the work situation. If neither the leader nor work colleagues will support an employee with new knowledge, the work climate can become so bad that the employee does not wish to participate in educational activities. In the worst case, this can contribute to a person's intentions of leaving the job altogether (Argyle 1973). When courses and educational activities correspond with the work situation, the participant will perceive the course as being meaningful. If in addition there is feedback during the course, the participant will develop a belief in his or her ability. When employees in JBV are given feedback and support and in addition perceive the course as relevant for the work situation, then this will contribute to employees remaining in the company.

Age and voluntary turnover

The regression analysis showed a significant correlation between age and voluntary turnover. The younger the employee is, the higher the voluntary turnover. There can be several reasons for this result. Firstly, young employees have not had time to be as closely emotionally connected to the company as older employees who have been employed for several years. There is also a recent tendency that people have several vocations during their lives. Greenberg and Baron (2003:231) use the expression

"boundaryless careers" as a description for the present tendency for people to have several vocational careers in different occupations and different firms during their lives. This tendency is also reflected in JBV. On the other hand, it is possible, in a company of JBV's size, both to have a career in the same occupation and to change occupation through a lifelong employment in the same company.

Education and voluntary turnover

The analysis shows that higher education gives higher voluntary turnover. One reason can be that employees with higher education have greater self-confidence about their own possibilities for employment in other companies. JBV has several employees with technical education. These vocations have always had high turnover, especially among younger employees. Many newly educated engineers choose the railways as their first employer after their education is ended. When they have got some experience from their railway employment, they move on to private companies that often pay better for work experience. Older engineers with longer railway experience can believe that they have less possibility for a job outside the railways, because their competence over the years has become very railway related.

Inner motivation

For inner motivation, the variables "support from leader and colleagues" and "relevance for the work situation" were strongest in the regression analysis. The more relevant the course is for the work situation and the more support the employee receives from leader and colleagues, the stronger the inner motivation will be. Mathieu (1993) shows that people who are strongly involved in their jobs are also most likely to be intrinsically motivated by courses and educational activities, because participation in these activities can increase their abilities, improve their performance and increase their self-esteem. Inner motivation will not last if there is no support from leader and colleagues. Argyle (1973) claims that small groups often have rules and informal structures that work against the changes that occur when an employee has acquired new knowledge. Safety in the environment will for many be an important condition. The thought of having to relate in a new or different way to one's group can create discomfort, and the interest, joy and pleasure of acquiring new knowledge from the course will decrease. Two statements from respondents show how the relevance of the course to the work situation and support from leader and colleagues can have positive and negative effects on inner motivation:

> I go to work to have fun, to have challenges and solve these, and to create good results! – which I may. I am happy in my job!
> My superiors don't give a shit about the work I do. I will retire ..."

Affective commitment

Affective commitment shows weak correlations with all the independent variables for courses and educational activities, but does not give results in the regression analysis. However, the regression analysis shows that the number of years in the company *(seniority)* has a significant correlation to the variable "affective commitment". In this matter it is interesting to see the effects of the railway culture. Earlier, the railway in Norway used to be an important cultural connection for local communities throughout the country. When father, grandfather or uncle worked in the railways, it was often natural for the next generation to choose the railways for a lifelong career. If there was a stationmaster in the family, this gave considerable status in the local community (like priests,

teachers and doctors) in the period from 50 to 150 years ago. To look at the workplace as a lifelong relationship can still be firmly rooted in older employees. This is emphasized by the following comment from a 49 year old respondent who has been employed between 21 and 30 years in JBV:

> Despite the reorganization and the uncertainty around this, I still feel a strong belonging to JBV. I want to continue in this company, and I am willing to adjust to a new situation, and new requirements for competence that the reorganization will demand.

The research shows that *women* have a lower affective commitment than *men*. There are no obvious reasons for this, but we notice that JBV generally is a company with few female employees (16.8% women in January 2005). Women in JBV generally work in professions that can give them good careers outside of JBV (in accounting, HR functions and administrative positions). A larger part of the men are employed in railway-specific professions and may perceive less opportunity to get a new job outside.

Strategic rooting

In the correlation analysis strategic rooting correlates significantly with inner motivation, voluntary turnover and affective commitment, but it does not give results in the regression analysis. One reason is that strategic rooting can give results through the variable work relevance. When course and educational activities are work related, employees may perceive it as if the course also has strategic rooting. But it will only give meaning for the work situation, and not for the organizational goals. When educational activities do not come from the same process as the business plan, then the activities will be on an operative level and miss strategic rooting (Montesino 2000). A connection between the educational goals and organizational goals will give direction and meaning for the participant and the motivation to learn will increase (Fisher 2003:371). The lack of results for strategic rooting may also be caused by the fact that JBV lacks a competence strategy connected to their main goals, but is starting to develop one in 2005. The fact that a competence strategy is missing, leads to frustration and uncertainty about whether the educational activities are truly adequate in relation to the company's main goals and strategy. The following comment from a respondent can emphasize the importance of strategic rooting:

> … I feel that we lack a thorough, basic introduction to how JBV works as an organization. That everybody knows how the organization works as a whole with reason, products and goals, also for each unit, is most important for an efficient and successful organization.

Other respondents have these comments:

> I believe that there are few who know the main goals of competence and education …
> Do job assignments correspond with what JBV needs?

Only about 25% agree that courses and educational activities correspond with JBV's main goals and strategies. The reason for this may be that JBV's strategies and main goals are not sufficiently communicated throughout the organization. JBV has a lot to gain from its work to define a competence strategy that is rooted in the company's main goals and strategies and ensure that this is well communicated in the entire organization.

Limitations

According to the motivational theory "Optimal Level Theories", there is an inner drive in people who are intrinsically motivated to seek out an optimal level of stimulation (Ford 1992:160). From this perspective, the theory that courses and educational activities lead to inner motivation can also be reversed: those who are intrinsically motivated, will seek courses and educational activities as a result of their natural, inner motivation, which is also a consequence of their natural search for optimal stimulation. Huselid (1995) questions whether efficiency is accomplished through HR-activities, or if it is employees who are already efficient who use courses and educational activities. We did not measure whether the employees were intrinsically motivated before they participated in the educational activities or if the activities increased their inner motivation.

Another limitation is that we have only evaluated how the employees *experienced* the dependent variables inner motivation, work effort, voluntary turnover and affective commitment in one company. Due to this, we can't draw general conclusions from the results we have from JBV. At the same time our results have shown consistency between our findings and previous research.

Implications for future research and general recommendations

For future planning and development of educational activities in JBV, we recommend that activities are chosen that correspond with the actual work the employees perform, and that employees are supported before, during and after the educational activities. This can give lower voluntary turnover, which in turn can give lower overall turnover. Lower turnover can have a positive effect on the economic results of the organization with lower costs to recruitment, substitutes and education.

An interesting future research would be to examine association between courses and educational activities with statistics on absence. Lower voluntary turnover and reduced absence can also be achieved by creating loyal employees and affective commitment, ref. May et al (2002) and Kuvaas (2005).

This research only showed results for affective commitment and seniority. The connection to seniority corresponds well with the railway culture where older employees often see employment as a lifelong relationship. The average age in JBV is almost 47 years. Many of the loyal employees will soon retire so that younger employees with "boundaryless careers" mentality will dominate the workforce. Because of this, it is important to start working to influence the affective commitment of younger employees. For a company of JBV's size, the competence strategy should emphasize the possibilities for careers across different trades within the company. Young employees are interested in their future prospects, and clear strategies can identify JBV as an attractive, long-term employer.

The regression analysis supported the hypothesis that the support from leader and colleagues and the relevance of the educational activity to the work situation contribute to increased inner motivation. It seems that JBV has an open work environment with support and room for discussing ideas and is not characterized by individualism and competition. This is also what theories on how to create a learning environment recommend. A learning environment can lead to motivated employees.

The weak correlations between strategic rooting and voluntary turnover, inner motivation and affective commitment can be an important signal to JBV. The descriptive data as well as the comments from several respondents show that the lack of a competence strategy can lead to uncertainty about the educational activities' relevance for the organization and for the individual's work situation. A clearer connection between courses and educational activities and the organization's strategies and main goals can lead to increased motivation to learn, and give the employees an experience of

meaningful learning. Establishing a competence strategy starting from 2005 can be an important way for JBV to increase the motivation among the employees and to strengthen the affective commitment.

Literature

Argyle, M. (1973), *Social Interaction.* London: Tavistock publications.

Fisher, Cynthia, Lye F. Schoenfeldt and James B. Shaw. (2003), *Human Resource Management.* 5th edition. Boston, New York: Houghton Mifflin Company.

Ford, Martin. (1992), *Motivating Humans.* USA. Sage Publications Inc.

Huselid, Mark. (1995), *The Impact of Human Resource Management Practises on Turnover, Productivity and Corporate Financial Performance*. Academy of Management Journal. 38 (3): 635-672.

Kuvaas, Bård. (2005), *A comparison of the relationship between perceptions of Human Resource Practises and employee outcomes.* Department of Leadership and Organizational Management, Norwegian School of Management. *Work in progress. (Reference is cleared with the author).*

May, Tam Yeuk-mui, Marek Korgzynski and Stephen J. Frenkel. (2002), *Organizational and occupational commitment: Knowledge workers in large corporations.* Journal of management studies. 39 (6).

Montesino, Max U. (2002), *Strategic Alignment of Training, Transfer-Enhancing Behaviours, and Training Usage: A Post training Study*. Human Resource Development Quarterly, 13:89-107.

Vandenput, M. A. E. (1973), *The transfer of training. Some organizational variables*. Journal of European Training, 2:251-262.

Yukl, Gary. (2002), *Leadership in Organizations.* 5th ed. Upper Saddle River, N.J.: Prentice Hall.

Chapter 59

Management Toolkits: Solutions for Rule Compliance

Jonathan Berman, Peter Ackroyd, Ann Mills and Toyin Davies

Introduction

A pervasive concern within many industries is rule compliance – why don't people always follow the rules? As with other high-hazard industries, the rail industry is concerned that safety-critical rule compliance appears to fall short of the ideal (whether deliberately with good intent, or inadvertently).

Past research has provided a better understanding of the causes of rule non-compliance, and guidance on their management. However, rule non-compliance continues to be cited as a contributory factor in accidents and incidents. RSSB therefore commissioned research to clarify the nature and extent of the issue within the rail industry and to develop the practical guidance that was being sought.

Two particular issues needed to be addressed. One was similarity with non-compliance in other industries, and hence the extent to which current guidance has relevance to the rail industry (and by extension whether the rail industry can be considered as a single homogeneous group in the context of rule compliance behaviours). The second concerned the need to develop practical and pragmatic guidance that was tailored to the needs of potential rail industry users.

We discuss rule compliance from the perspective of what is known about its causes and mitigations, and how those issues appear within the rail industry. It describes a 'toolkit' of guidance that has been produced for the industry, and the more general issues concerning the dissemination of good practice guidance and solutions. The research was commissioned by RSSB and carried out by Greenstreet Berman in three phases:

- draw together current knowledge on rule compliance
- develop a set of structured guidance to help managers understand non-compliance and reduce its frequency or consequences
- develop and pilot an electronic version of the guidance

What do we know about rule compliance?

Many different studies have been undertaken that consider rule compliance, and its enhancement (eg HFRG, 1995, a guidance document prepared by the Human Factors in Reliability Group for HSE). Additionally, a significant number of studies have focused on the rail industry, whether explicitly addressing non-compliance or whether considering organisational and other factors that affect performance.

Its important to be clear what is meant by rule compliance (and non-compliance). The pragmatic definition that emerged during the course of this work was concerned with the extent to which

people follow formal and agreed rules, procedures and methods of working. Rule Compliance is following those procedures. Consequently, non-compliance is a failure to follow the recognised procedures. However, importantly, non-compliance can arise for a number of different reasons – it can be deliberate but with good intent ('I know a better way of doing this task'), deliberate with malicious intent (sabotage), or inadvertent (misunderstand rule, or error in carrying it out). Malicious non-compliance is a separate issue and not discussed further.

What is important about this definition is that it encompasses a range of different *causes* of non-compliance, even though the consequence might be the same in each case – the failure to follow the procedure. From this flows the equally important issue which is that different causes of non-compliance are likely to be influenced by different solutions. Failure to comply because you think there is a better way of undertaking the task merits a set of solutions that could include better training or making it more difficult to adopt alternative approaches, whereas failure to comply due to poor equipment design (so that you make a simple error) is unlikely to be affected by those same solutions.

For example, when considering speeding by car drivers, prevention by installation of speed cameras and sanctions will only affect those drivers who know the limit and deliberately break it – perhaps because they are running late. It will not influence those who were unaware of the limit (better signage would be more effective) or who misjudged their speed and thought they were compliant (better speedometers or training on the need to read them carefully). In other words, if you wish to reduce the incidence of non-compliance, you need to understand the causes of non-compliance that exist within your organisation, in order to identify effective solutions, and to implement them in a cost-effective manner.

It is clear from previous research that deliberate rule non-compliance is based on judgement by the individual. A balance will be drawn between the benefits of non-compliance (e.g. faster task completion, more interesting work, ease of task performance), and the potential costs (e.g. greater risk of task failure, exposure to sanctions, personal injury). It is important to recognise that the personal judgement may not be soundly based on fact or might have an inappropriate perception of risk which causes them to underestimate the disadvantages of non-compliance. They might over-estimate the value of perceived benefits such as speed of task completion. Thus, many solutions to deliberate non-compliance will include processes that influence those perceptions – such as training to improve risk understanding, increasing the perceived likelihood of sanctions or their severity, or reducing the perceived benefits of non-compliance such as by making the compliant process easier to undertake when compared with the non-compliant.

The objectives of the RSSB work were to understand the range of issues that can affect rule-compliance in the rail industry, and to identify the set of potential solutions. Additionally, and perhaps most importantly, the intention was to provide a guidance package that would enable managers both to understand the causes of non-compliance within their organisation, and to select and adapt relevant solutions.

The current situation in the rail industry

An early part of the research was to review current understanding of non-compliance across a range of industries (Berman et al, 2004). Many of the existing studies have been undertaken within high-hazard industries such as nuclear and onshore/offshore petrochemical. Broadly these studies identified a common set of factors that affect intentional or unintentional rule compliance. In addition, a number of rail industry studies were considered which also identified a similar set of factors:

- Organisational factors (e.g. a participative supervisory style produced greater compliance amongst workers).

- Environmental factors (e.g. SPADs can be influenced by weather and rail conditions, which can affect willingness to comply with driving rules)
- Individual differences (e.g. some drivers believe they can control SPADs, others believe that SPADs are inevitable; such differences can influence whether individuals are likely to attempt to comply with rules or not)
- Cognitive factors (e.g. sometimes tasks can be too cognitively demanding for an individual, and hence force individuals to make a decision to 'cut corners')
- Motivations and behaviour (e.g. there will be a variety of motivators for people when carrying out their actions. Individuals may be influenced by performance pressures or peer pressures)
- Attitudes and beliefs (e.g. individuals may not believe they can comply with some rules, or they may not believe that they need to comply with certain rules)
- Workplace design (e.g. opportunities to use equipment in unintended ways; some workplaces can make it very difficult to comply)

In addition, data were collected from the UK rail industry, using the SMIS database and other sources, in order to identify the extent to which the issue of rule compliance was of concern within the UK industry. Whilst many individuals have expressed concern about rule compliance, it was important both to understand the extent of the issue, and to confirm that the prevailing factors that affect compliance within the UK rail industry are similar to those identified in other industries.

Extensive data collection and analysis supported a number of conclusions. The first was that the factors that emerged were similar to those identified from the literature. Each of these factors suggests a particular approach to compliance management.

Unsurprisingly, the data collection also revealed significant under-reporting of non-compliance and hence reinforced the identified need for practical tools. Whereas the literature tends to report non-compliance as a causal factor in approximately 75% of incidents, the UK rail industry data appeared to suggest a rate of approximately 25%. However, interviews and a questionnaire circulated across the industry provided estimates that were broadly consistent with the literature.

Underlying causes of rule non-compliance were consistent with those identified in other industries. Furthermore, causes varied both across and within organisations, and hence solutions appropriate to, say, a Train Operator may differ from ones effective within a Track Renewals company.

Thus, it was concluded that the set of potential solutions to improve rule compliance were likely to be both generic and similar to those applied in other industries. The original focus of the work was on safety-critical rule compliance, but the evidence from the initial review work was that the emerging guidance would be equally applicable both to safety-critical rule compliance and to compliance with all other rules and procedures, and hence could be of generic value.

Emerging guidance

The emerging guidance was intended to help organisations reduce the incidence of rule non-compliance. To achieve this it was important that it comprised a number of separate, but related processes:

- a process that allowed the user to review non-compliances in order to understand and diagnose underlying causes
- a set of solutions towards which the diagnostic process would direct the user;
- a process for reviewing susceptibility to non-compliance as a proactive means of preventing it (which then directs the user towards the appropriate solution).
- a set of solutions examples to illustrate how they can be adapted and applied.

Although the initial version of this 'toolkit' was paper-based, it was recognised from the outset that it needed to be presented in an interactive electronic format, and hence a CD-ROM version was developed.

A key theme throughout the guidance is the critical importance of gaining a proper understanding of how tasks are carried out in practice, and the real constraints, pressures and influences on task performance. It is not sufficient merely to consider the intent behind the tasks, or the documented processes. Instead it is essential that the user seeks information from people who undertake the tasks on a regular basis. Guidance is provided concerning the sources of such information and how best to elicit it. This is important for proper identification of the causes of non-compliance, proper identification of appropriate solutions, effective tailoring of those solutions, effective delivery of the solutions (including 'buy-in' from the workforce), and reliable proactive identification of non-compliance prone situations.

The solution set

The set of solutions comprise a broad spectrum of human factors interventions, ranging from attitudinal interventions to improvements to workplace ergonomics (depending on the cause of the non-compliance. Figure 59.1 shows the solutions 'set'.

The diagnostic process

The diagnostic process is a simple algorithm that helps the user to consider the known attributes of the observed non-compliance, and hence to consider the causes. It requires the user to answer questions that will distinguish whether the observed non-compliance is:

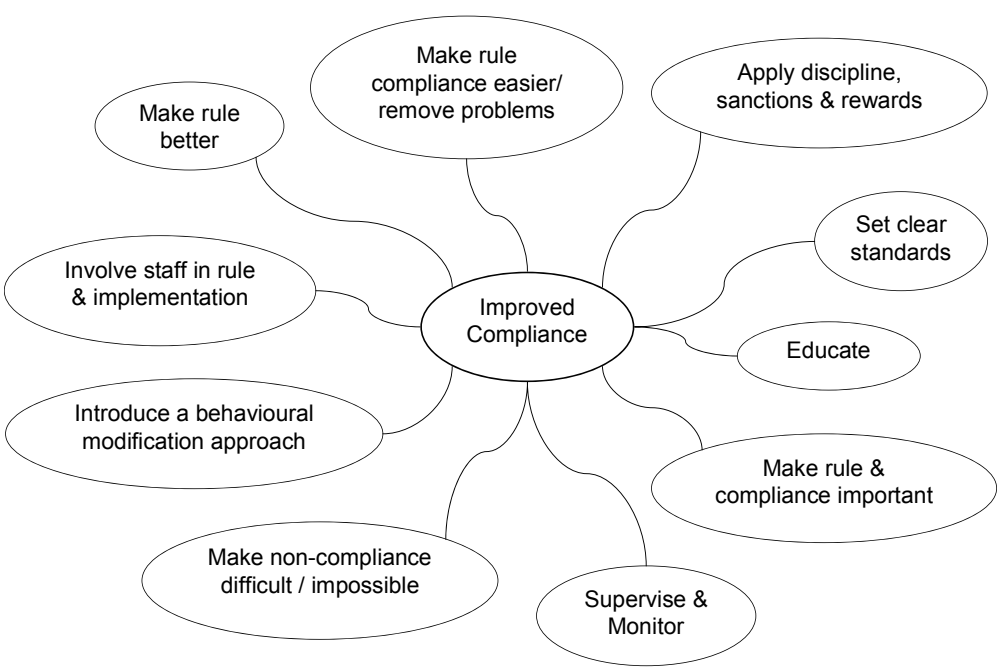


Figure 59.1 Compliance solutions

- committed at random, or by an individual (or a limited number of individuals) or systematic and committed by many people at predictable times
- occurs during unusual circumstances, or whether it is due to unintentional error, or whether it appears to be intentional
- appears to be connected with competence and awareness, or error
- connected with attitudes and culture, or with multiple rule non-compliance
- connected with characteristics of the procedure/rule, e.g.:
  - rule impractical, outdated or obsolescent
  - rule is acceptable but poorly implemented
  - rule is acceptable and well implemented but still not followed

The answers to the questions within the diagnostic process direct the user towards an appropriate pragmatic solution or set of solutions. For example, non-compliances that appear to occur at random and during unusual circumstances are likely to be amenable to solutions that focus on competence and awareness. However, they may also be influenced by steps taken to make it easier to comply. Thus, there is likely to be a set of related solutions which, together, will improve the likelihood of compliance.

Solution guidance

The solutions guidance is structured to provide the user with an understanding of what the solution is trying to achieve, and what steps need to be taken to improve its effectiveness. It sets out when the solution should be used, and why. It identifies the key elements of the process and the things to do or to avoid. It identifies other solutions that should be used in conjunction, and the advantages and limitations as a solution. It describes the likely outcomes or impacts, and provides some examples.

Thus, the solutions information is intended to help the user to understand what is the appropriate solution (or set of solutions) and what is required to make them most effective. It is not intended to be prescriptive. Furthermore, it is expected that users will tailor each solution – every organisation is different.

Proactive assessment

Whereas many users will wish to understand the causes of observed/reported non-compliances, and to identify and implement appropriate solutions, some users may wish to use the process proactively. Instead of waiting until non-compliances start to become apparent, it may be appropriate to attempt to identify and correct factors that could lead to non-compliance before they have an impact on performance. Alternatively, it may be that consideration is being given to changes that could carry a risk of increasing non-compliance, and the user wishes to review the likely impact of those changes.

To cater for this requirement, a proactive assessment process was developed, which provides an alternative route into the solutions. The proactive tool is a checklist that can be completed either by the user, or by a group of appropriate people. It highlights areas where the organisation is vulnerable to non-compliance, and the likely factors that would encourage it. From this, the user can select relevant 'solutions'.

The toolkit: its form and function

The paper-based toolkit was piloted with a small number of organisations within the rail industry, to ensure that it had the capability to address a real issue in a manner that was suitable for the

industry. Positive feedback was received concerning its scope and coverage. As anticipated, there was consensus that the toolkit would be best presented electronically for ease of delivery and, more importantly, to significantly facilitate navigation through a potentially complex process. It was recognised that it would be insufficient merely to translate what was, in effect, a simple text-based document into an electronic form. Instead, the opportunity was taken to consider carefully the potential user population, the manner in which they might wish to use the document and the manner in which the solutions could be delivered.

The approach adopted was to present the guidance as an HTML document, with a web-based interface, but to ensure that it remained simple and focused. This approach facilitates the use of links between elements of the various sections.

It now comprises the following sections:

- A home page
- An introduction that provides both an overview of rule compliance management, and an introduction to the use of the toolkit
- A diagnostic toolkit – which presents the full diagnostic process
- A set of solutions – which can be accessed directly or from the other toolkits
- A proactive toolkit – which comprises the checklists and guidance concerning their use
- Examples and a variety of help functions

One of the key elements of the approach is to acknowledge that users will be taken on a journey from which they will learn both about compliance management, and about their own organisation. Consequently, it is important that the toolkit supports both relatively naïve users and those with significant experience. The expectation is that new users will follow the diagnostic or proactive processes, and be directed from there to relevant solutions, whereas the more experienced user may choose merely to use the solutions section as an aide-memoire.

A further issue concerned the user population – who is likely to be the user of the toolkit, and at what level are they within their organisation? Following consultation with the industry, it was agreed that the toolkit should be aimed at all users, from supervisor upwards. Different organisations are likely to use it in different ways (e.g. a training aide, to support incident investigation, to review procedures, etc) and hence it is important that no user group is excluded. Furthermore, the implementation of solutions and the development of a sound understanding of the manner in which tasks are actually carried out requires active involvement of staff at all levels. This influences the language and terminology, but was found not to influence the nature of the solutions and guidance.

Pilot and feedback

The electronic version of the toolkit was piloted by a number of Railway Group companies. The main outcomes from the pilot process indicated that:

- Generally the toolkit has been very well received and valued. Although the toolkit does not highlight any new solutions, it allows organisations to identify alternative solutions that they might otherwise have overlooked;
- A noted benefit was that it acts as confirmation that the solutions that they have implemented are valid and appropriate;
- It allows for a logical process to be followed, particularly when investigating acts of non-compliance;

- It has potential use as a training tool for less experienced / novice staff in first line management positions;
- All pilot organisations agreed that the content is appropriate and relevant for their specific needs. In essence it is generic and can be applied to all acts of non-compliance, not just safety critical acts, across the rail industry.

A range of positive comments were received from the pilot process, such as:

- 'It highlights the relationship between different solutions. Previously we may have looked for a single solution to an act of non-compliance. This has opened our eyes that there is rarely one solution that sits in isolation.'
- 'It has proved extremely useful for inexperienced and experienced investigators, as it confirms that we are doing it right, and gives alternative solutions that might not have been considered.'
- 'We would recommend the toolkit to everybody involved in safety related investigations and decision making processes because it takes away any ambiguity.'
- 'It will prove particularly useful where there is no 'in house expertise'.'
- 'It allows for a procedure to be followed with good levels of advice. It also allows users to consider all possible angles'.
- 'It is particularly useful as a risk based tool. It allows for potential risk to be minimised by highlighting the precursors to acts of non-compliance to be identified e.g. lack of training, inadequate documentation, procedures etc.'
- 'It can be used as a complementary tool to existing procedures, particularly when investigating incidents. For example; it would be useful to have a line manager complete the form in relation to a specific act of rule non-compliance, then compare this with the non-compliant individual response.'

Further enhancements for the future; include making the toolkit more interactive by providing a tracking process. Some users were concerned that it was too time-consuming for use in incidents, or that it did not highlight new solutions. The majority, however, considered that it significantly improved their approach to rule compliance, and the consistency of that approach.

Examples of the use of the toolkit

Disregarded rule A train entered service in contravention of the rulebook requirements concerning leaving a maintenance depot with defective safety equipment – despite all staff being aware of the instructions. The initial act of rule non-compliance by a single employee led to other employees accepting this as 'normal working practice'. Prior to 2000, this was acceptable working practice.

Following the application of the main diagnostic toolkit, the potential solution that emerged as being appropriate to the nature of the observed non-compliance was to make the rule and rule compliance more important. The solution also emphasised the importance of engaging with the relevant staff, and hence all staff were involved in face-to-face briefings and shown a video highlighting the potential dangers of not following the rule. Internal standards have been reinforced and there have been no further contraventions to date.

Misunderstood train despatch rule An organisation was experiencing continuing failure to carry out train despatch duties in the correct manner as a result of the train guard being in the wrong

despatch matrix position. Following the application of the main toolkit, the process identified that the rule is fit for purpose, but that there was room for interpretation within the rule.

All train guards were involved in the development of new procedures for implementing the rule – this removed any ambiguity. The outcome has been all subsequent despatch duties have been carried out in the correct manner.

Combination causes Over a 12 month period a major issue for a company was the number of cable strikes on their projects. An investigation was undertaken into each incident. In addition they reviewed all incidents to identify trends. They used the proactive toolkit to review the excavation and working around buried services procedure, trying to identify why individuals were not following the rules. Individuals were asked to complete the set of questions raised in the toolkit. The toolkit assisted in this stage of the review by acting as a checklist and ensuring that all potential reasons for non-compliance with the rules were assessed. The toolkit also validated and improved the existing recommendations that had been developed.

The action plan from the Cable Strikes Investigation Report is now in place with recommendations closed out in an ongoing process. The proactive toolkit identified why people were not always following the rules.

Lessons for dissemination of HF good practice

The outcome of the research commissioned by RSSB has been a practical tool that appears to have been well-received by the rail industry. It has been focused on all Railway Group companies, and oriented towards staff at all levels from 1st-line supervisor upwards. A number of issues emerge from this work concerning the manner in which the results of research projects can and should be disseminated:

- Whilst a sound research underpinning is essential, there is a need clearly to distinguish between the research findings, and the guidance that is disseminated
- The level of human factors knowledge and understanding varies widely between companies within the Railway Group
- There is a need for clear, concise and simple human factors guidance
- There is a need for guidance that enables companies to tailor their approach according to the nature of their organisation
- Whilst many organisations may understand the human factors principles that underpin the guidance, they welcome the provision of a clear process that gives them confidence that they have considered the issue comprehensively
- Guidance needs to be tailored to the potential users – and different organisations have different views about who those users might be
- Electronic dissemination of information is entirely acceptable
- Some organisations, including many smaller ones, do not necessarily perceive human factors as embracing issues such as compliance management
- The range of factors that affect compliance, and the extent to which they are amenable to management control, is not immediately apparent
- Smaller organisations do not always recognise the value and relevance of human factors research. Effort must be expended to demonstrate the value.

A major implication from these issues is that there may be a need to inform and educate some organisations within the Railway Group as to the diversity of human factors issues, and the extent to which there is guidance and advice available to assist in their management. A second key

implication is that much that is already known within the human factors community concerning the solutions to human factors issues is not readily accessible to those who are best placed to make use of it.

A third implication is that the industry wishes to be 'solution-focused'. There is less interest in the underlying theory that determines effective compliance management, or in the causes of observed non-compliance. Industry attention is focused on what needs to be done differently in order to improve compliance. Advice and guidance therefore needs also to be solutions focused. This is particularly important for interventions that may take time to have an effect – the guidance needs to provide a process that enables users to start a journey that might need sustained effort – but at least to have made a start.

A similar approach is being adopted for current RSSB work on a safety culture assessment and guidance tool. The entire process will be web-enabled, such that it can present live benchmark data. More importantly, it builds on the existing RSSB assessment tool, to present tailored guidance dependent on the results of an assessment. This guidance is intended to be solutions-focused: it will help to move the organisation upwards in terms of safety culture irrespective of where they are currently. In other words, the key message is that performance will improve – and that users are embarking on a continuing process of improvement. Within limits, it is less important how far they improve, initially, and more important that they start the improvement process.

A key message from both of these projects is therefore that guidance must focus on the user challenge, and be solution-focused rather than problem-focused. The compliance toolkit has been constructed with this intent – although the diagnostic toolkit enables the user to understand the causes of non-compliance, the reason is to enable them to identify and implement effective solutions. Finally, it is important for tools to be seen as 'solutions tools' rather than 'HF tools' – the challenge for organisations is to improve their performance rather than to label their problem.

References

Berman, J., Ackroyd, P., and Turner, D (2004). *Safety Critical Rule Compliance. Task 1 Final Report*. T145 London: Rail Safety and Standards Board.

HFRG (1995). *Improving Compliance with Safety Procedures*. London: HSE.

This page intentionally left blank

Chapter 60

Rules Management: How to Achieve Flexible Guidance for Safe Railway Operations

Gudela Grote

Standardization as key to safety

In most high risk systems, standardization in the form of standard operating procedures has been developed with ever increasing detail in order to streamline human action and to reduce its influence as a risk factor. Procedures are often a direct consequence of incidents and accidents, the analysis of which provides knowledge of unforeseen wrongful courses of action against which new rules are developed as a defence. While generally there is an understanding that rules are useful guides for safe behaviour, there is also an increasing concern that too many rules incrementally developed will not make up a good system to help human actors do the right thing, especially in states of abnormal operation where they would need strong, but also flexible guidance (e.g. Amalberti, 1999; Dekker, 2003; Woods & Shattuck, 2000).

These concerns go back to basic observations on how rules specifying the exact operations to execute can have a detrimental effect on action because they do not allow the performing person to develop an underlying plan of their own, but instead further the atomization of actions and the focus on micro-difficulties (Vermersch, 1985). Another problem with standardization is that reliance on common standards may turn into an overreliance, impeding switches to more explicit coordination and to higher levels of common action regulation, i.e. switches from skill-based to rule-based or to knowledge-based behavior. This problem can be exacerbated by the fact that standardization is a strong force towards shared understanding of a situation and its demands in a team, because it creates a common framework for team behavior reducing the need for explicit coordination. The expectation of shared goals, plans, perspectives, and knowledge bases created by reference to the same set of standard operating procedures, as helpful as it is under most conditions, does involve the risk of not realizing the need for explicit coordination, especially in non-routine situations.

Rasmussen (1997, p. 191, italics in the original) has argued that ”rather than striving to control behaviour by *fighting deviations* from a particular pre-planned path, the focus should be on the control of behaviour by *making the boundaries explicit and known* and by giving opportunities to develop *coping skills at boundaries*”. Rules then would have the function to clarify boundaries and to suggest ways of handling system states close to those boundaries. In line with this approach to rules, some authors have begun to develop typologies of rules in order to help the design of rule systems directly tailored to the needs for guidance as well as for autonomy and control arising in different stages of action regulation (e.g. Hale & Swuste, 1998; LePlat, 1998). From an action regulation perspective, rules can concern goals to be achieved (goal rules), define the way in which decisions about a course of action much be arrived at (process rules), or prescribe concrete actions (action rules). Systematic research into the design and management of safety-related rules has only

recently begun, providing tentative classification schemes mainly based on the rules' relevance for individual action regulation (Hale & Swuste, 1998; Leplat, 1998; Reason, Parker, & Lawton, 1998).

Standardization as a coordination mechanism in organizations

From an organizational perspective, rules should not only be discussed as guidance for individual behavior, but also as elements of the coordination mechanisms operating within and between parts of an organization. As a consequence of the degree and type of division of labour and specialization more or less effort will be required for coordination and different kinds of coordination mechanisms will be more or less successful. Generally, five broad categories of coordination mechanisms are distinguished (e.g. Thompson, 1967; Van de Ven, Delbecq & Koenig, 1976):

- Technologically defined processes
- Central programs and plans, e.g. standard operating procedures
- Personal leadership
- Mutual adjustment via reciprocal team interaction
- Cultural norms

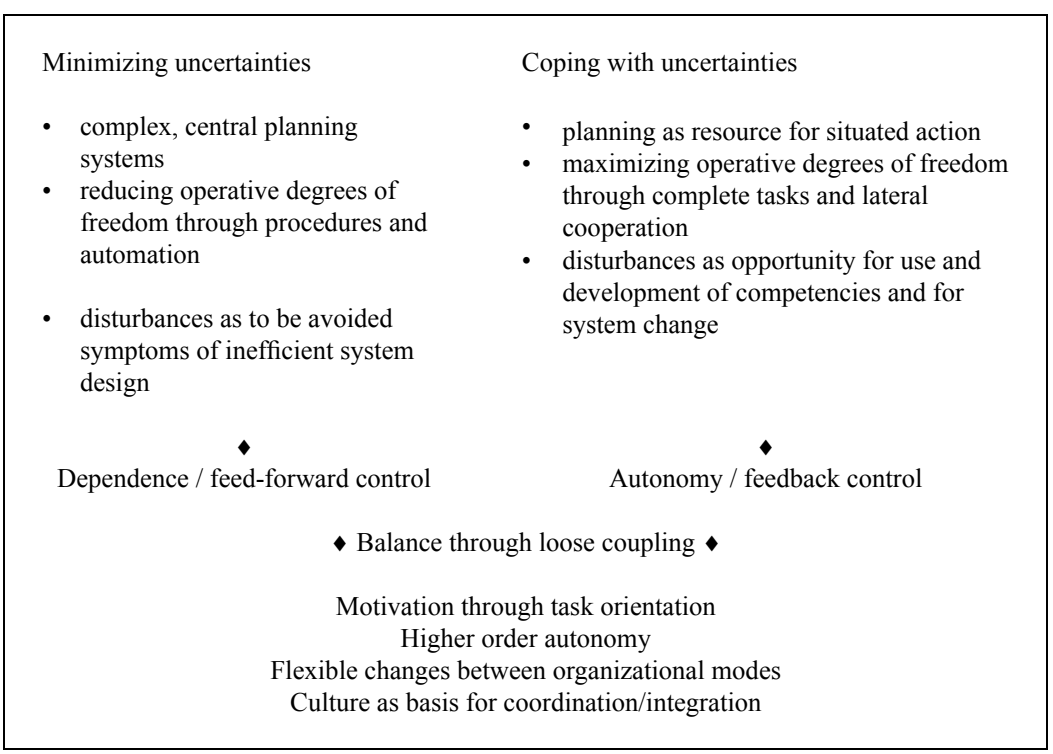


Figure 60.1 Two approaches to managing uncertainties in organizations (from Grote, 2004a)

In order to understand coordination in organizations, it is helpful to conceptualize organizational activities in terms of the management of uncertainties stemming from the transformation processes an organization has to perform and the environment within which these processes take place (e.g. Perrow, 1967; Thompson, 1967; Van de Ven et al., 1976). Two extreme approaches to handling uncertainty can be distinguished (Grote, 2004a; see Figure 60.1). The first one tries to minimize uncertainty or at least the effects of uncertainty in the organization using mainly feed-forward control based on high standardization and programming of work flows. Enormous efforts are put into centralized planning and continuous monitoring of the execution of these plans, providing minimal degrees of freedom to the people in charge of carrying out the plans. The other approach aims to enable each and every member of an organization to handle uncertainties locally and to allow for feedback control. From this perspective, planning is understood primarily as a resource for situated action (Suchman, 1987), not as blueprint for centrally determined and monitored action. Local actors need to be given as many degrees of freedom as possible, achieving concerted action mainly through lateral, task-induced coordination. Disturbances are also regarded as opportunities for use and expansion of individual competencies and for organizational innovation and change.

Standardization can be regarded as the key element in the minimizing uncertainty approach, while the competent coping with uncertainty relies much more on personal and lateral coordination. It can be assumed that standardization will work better in situations with few uncertainties while local autonomy and control are needed when uncertainties are high. In situations with high uncertainty standardization may even be harmful, but so far few concrete suggestions exist as to what coordination mechanisms to use in order to improve the predictability and controllability of a system while at the same time increasing its flexibility (e.g. Perrow, 1984).

Weick (1976) has suggested the principle of loose coupling in order to simultaneously ensure autonomy of actors and sufficient binding forces for all actors to use their autonomy to promote the organization`s objectives. In Figure 1, four examples are given for achieving loose coupling. The concept of motivation through task orientation (Emery, 1959) assumes that tasks allowing for a high degree of autonomy, task completeness and task feedback will further an individual's intrinsic motivation towards fulfilling the goals of the primary task of the organization. The concept of higher order autonomy (Grote, 1997; Klein, 1991) has been suggested to provide autonomy in those situations where in technically tightly coupled systems (Perrow, 1984) little operative autonomy is possible. Higher order autonomy allows the actors in the organization to decide on the restrictions of their own operative autonomy, e.g. through participative design of rules and procedures. In studies of high-reliability organizations (e.g. LaPorte & Consolini, 1991), it has been observed that organizations may also be capable of changing flexibly between the two organizational modes. Weick (1987) has pointed out that culture serves as a strong basis for a form of coordination and integration that incorporates both decentralization of autonomy and centralization of values and norms as binding forces for local action.

Standardization as support for loose coupling in organizations

Rules as support for loose coupling have not been researched much to date (Grote, 2004b). In the words of Woods and Shattuck (2000) the issue is how to avoid Type A problems, where rote rule following persists despite changing circumstances requiring adaptable responses, as well as Type B problems where adaptation happens unsuccessfully due to incomplete knowledge or lack of guidance. In a study by Grote et al. (2004), in which rules in aviation and in anesthesia were compared, it was found that aviation was characterized by generally more rules and more rules concerning the action level, while the generally fewer rules in medicine more frequently concerned the process level. Given the very different nature of uncertainties to be handled in the two

professional settings, these differences seem to reflect adequate adaptations to situational demands. At the same time the study also showed that good cockpit crews were characterized by higher levels of explicit coordination, i.e. very deliberate action planning and implementation even in highly standardized work phases with few uncertainties, which hints at the importance of backing up standards with a constant effort to reassure a common understanding of the situation and the relevance of the standards for the situation in line with the concept of heedful interrelating (Weick & Roberts, 1993). The anesthesia teams coordinated their tasks more implicitly despite overall fewer standards, which may be related to more informal rules and the immediacy of common action in a shared visual field (Heath et al., 2002).

Very rare in both settings were rules that also provide a rationale for the rule. This finding is interesting in light of a study by Wright, Pocock and Fields (1998) in which the annotations in a Quick Reference Handbook by an airline pilot were analyzed. Almost 40% percent of the annotations concerned explanations on why the procedure was the way it was, indicating the need of users of procedures to understand their reasoning, which then can be used to apply the procedure adaptively, as was indicated by another about 25% of the annotations.

Another set of analyses in the study by Grote et al. (2004) concerned patterns of coordination within each professional setting, i.e. aviation and anaesthesia, comparing work phases with different degrees of standardization and task load. One important finding here was that personal leadership seemed to be necessary only in situations with few standards. In highly standardized situations, the standards act as a form of impersonal leadership which does not require additional efforts of personal leadership. To the contrary, high levels of personal leadership in highly standardized situations appeared to be related to worse team performance in line with the leadership substitute theory of Kerr and Jermier (1978).

All these findings point to the importance of helping teams to deal with the demands of adaptive coordination better. Especially process rules could be well-suited for this purpose, but up to now rules usually are not designed with this meta-level in mind, i.e. having rules on when rules should be used as coordination mechanism versus when other coordination mechanisms such as mutual adjustment in a team or leadership are more appropriate.

In the following a study is reported which investigated rule management in a railway organization. This setting was considered particularly suited as railways are among the most proceduralized organizations. They also have a very long history of using rule books as the basis for coordination due to the highly distributed nature of their operation in the face of - until very recently - no means for direct communication between the different actors like train drivers, controllers and maintenance workers (Hale & Heijer, 2006). At the same time, rule violations and inadequacies of rules and rule management have been found to be crucial safety problems in rail organizations (Farrington-Darby et al., 2005; Hale, Heijer & Koornneef, 2003; Lawton, 1998).

Rule management in a railway organization

The general purpose of the study reported here was to provide an analysis of existing rules and the rule management process shared between the regulator and the railway organization. From a research perspective, the adequacy of existing rules and rule management processes for providing optimal support for the management of uncertainties in rail operations was focused on, using the uncertainty management framework described above. For the railway company and the regulator, the study was to provide a basis for reflecting upon four themes: adequate detail of rules, interplay of rules and training, interplay between rule design and rule auditing, and distribution of responsibility between actors involved in rule design, rule following, and rule monitoring.

Methods

The study was based on data stemming from

- interviews with the responsibles for rule management at the railway organization and the regulator,
- analysis of documents on rule management provided by both the railway organization and the regulator,
- analysis of incidents and accidents related to rule following,
- rules analysis of a sample of rules based on two work processes (train departure, coupling and braking of cars during shunting operations),
- observations of the chosen work processes and interviews with shunters, signallers, train drivers and dispatchers

For the rules analysis, categories suggested by Hale and Swuste (1998) were used in a slightly modified form:

1. Level of action regulation
   Goal (the rule only defines the goal without determining the way for achieving the goal)
   Process (the rule describes the processes needed to determine the right course of action)
   Action (the rule prescribes a concrete operative course of action)
2. Obligation (Advice vs. command)
3. Decision latitude (with vs. without decision latitude)
4. Distribution of responsibility (Responsible person(s) mentioned vs. not mentioned)
5. Exceptions (with vs. without exceptions)
6. Reasons (with vs. without reasons)

The interviews and document analysis concerning the rule management process followed the model of rule management suggested by Hale, Heijer & Koornneef (2003), which has been used in previous investigations of rule management in railways.

The following presentation of results will focus on the findings of the rules analysis, using information from observations, interviews, and document analyses to interpret the results and to provide some further context for the findings.

Results

Table 60.1 presents the results of the analysis of rules concerning the coupling and braking of cars during shunting operations and the moving of trains. Regarding the moving of trains, train departure was focused on, for which new rules had recently been devised by the regulator.

The analysis revealed a number of interesting issues, which have to be interpreted cautiously given that only a sample of rules was analyzed that may or may not be representative of all rules in the rail organization studied. Rules for the rail organization are written by different people depending on the topic and the organizational groups affected with very few content-related guidelines provided. In fact, only one content-related statement was found in the existing documents on rule management: "Rules should only contain normative statements and/or recommendations and should define the handling of exceptions; informative statements should be restricted to the absolute minimum." Therefore, it is unlikely that any subset of rules could be representative of all rules in this organization.

Table 60.1 Results of rules analysis

Rule type	Shunting		Moving trains (New rules related to train departure)		Moving trains (Old rules related to train departure)		Moving trains (All rules)	
	Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
Total	28	100	32	100	38	100	155	100
Goal rule	0	0	1	3	0	0	1	1
Process rule	6	21	9	28	2	5	19	12
Action rule	22	79	22	69	36	95	135	87
Advice	0	0	1	3	0	0	3	2
With decision latitude	20	71	15	47	14	37	61	39
With respon-sible person/s mentioned	2	7	17	53	13	34	46	30
With exceptions	7	25	6	19	7	18	34	22
With reasons	1	4	0	0	1	3	5	3

Table 60.1 shows that very few goal rules exist, which on the one hand can be regarded positively as goal rules by their very nature do not offer much guidance for the actors on how to pursue the goal. On the other hand, some goal rules may be helpful in order to stress priorities in certain situations and thereby support the actors in their individual priority setting. This may concern very basic priorities such as punctuality versus safety or more specific priorities such as the timing of own actions and communication to other actors about these actions.

Compared to the old rules for moving trains the new rules included more process rules and also a clearer distribution of responsibility. Especially the latter had been a declared goal of the rule revision in order to avoid diffusion of responsibility. At the same time, the discussion about the stricter definition of responsibility with the people affected indicated the potential problem of too little sharing of respsonsibility, thereby reducing information flow and cross-checking between the various actors. This problem is assumed to gain importance due to different actors like train drivers, maintenance crews, shunters, and signallers belonging to different companies within a holding structure, with fewer and fewer of them having had the experience of belonging to the one previously existing company.

The higher level of process rules in the new rules on train departure were generally considered positive as they provide guidance without restricting flexible action. Signallers and train drivers are well qualified with high professional ethics, who still have to act to a large extent on very defined and prescriptive action rules, but “stretching” these in order to accommodate non-routine situations. So an even higher level of process rules might be appropriate.

The high percentage of process rules and even more of rules with decision latitude was evaluated more critically with respect to shunting operations. The prevalence of more open and less prescriptive rules on how to handle shunting operations has to be seen in the context of the shunters' comparatively low level of qualification and their reluctance to take responsibility for using the decision latitude offered. When discussing this obvious mismatch with the people responsible for rule making, it was pointed out that shunting occurs in so many different contexts (within stations, on shunting yards with varying degrees of automation etc.) that rules have to be less specific. So the issue becomes more whether the current level of qualification and training is sufficient for handling this flexibility. Efforts are underway to improve training for shunting accordingly. One may also raise the issue, though, which has been brought up in other rail organizations as well (Hale & Heijer, 2006), of whether rules for shunting are paid less attention by rule makers as they mainly concern the safety of the individual workers and less the politically more sensitive safety of passengers.

A high percentage of rules with exceptions was found, which can mainly be explained by the fact that a variety of local conditions has to be covered, e.g. regarding the level of automation, the number and distribution of personnel (trains with/without conductors, stations with/without signallers etc.), or the particular geographical layout. The main issue to be discussed with respect to these exceptions is the level of detail the rules should have and whether local knowledge should be incorporated in the rules or rather be provided by complementary training. The fundamental dilemma of not being able to have general, simple, and detailed prescriptions at the same time became obvious here.

Finally, rules very rarely included reasons, raising the question of whether training is designed to provide the necessary background information. Given an increasing pressure for interoperability of rail systems, thereby requiring personnel with more and more diverse educational and training backgrounds having to work together, rules might have to include more explanations in the future.

Discussion

From the perspective of adequately managing uncertainties in the rail organization, it appeared that there is a partially insufficient match between the distribution of uncertainties and resources for coping with them. This seemed particularly prevalent for shunting operations where either uncertainties need to be reduced through more detailed and prescriptive rules and/or individual competencies need to be increased to adequately use the decision latitude provided. Especially for signallers, but also for train drivers, the reverse tendency was observed: they need to be supported more in coping with uncertainties, potentially through more process rules including rules that support adaptive coordination, e.g. switching from proceduralized, implicitly coordinated action to explicit coordination through direct communication and mutual adjustment between different actors. In providing signallers with more scope for action, a critical balance with the decision latitude for other actors has to be maintained. The analysis of the rules and even more so the analysis of some incident reports seemed to indicate that the signallers sometimes handle uncertainties in a way that reduces other actors' possibilities for handling their part of the situation appropriately, e.g. by providing not enough or untimely information about their actions or the actions required by the others. Here also a power differential between different professional groups is apparent, as signallers have the power to shift uncertainties to other actors who sometimes do not even have the adequate resources to handle these uncertainties (Hale & Heijer, 2006; Marris, 1996).

For the practitioners, the rules analysis raised a number of issues which led the rail organization and the regulator to jointly work on defining an explicit philosophy of rule management to be written into guidelines for making rules. The aim is to further a more coherent and conscious

approach to rules management. The guidelines should help to make systematic choices of rule types considering competence profiles of affected actors and the interplay between different coordination mechanisms, e.g. including mutual adjustment through lateral communication more explicitly. Also, the current complexity of layers of general and local exceptions should be reduced. As an overriding goal, rules should be revised to better support local coping with uncertainties, complemented by training and constructive monitoring of rule following.

Conclusions: Improving system safety through rules management

Adequate management of uncertainty is crucial for safe and efficient system operation. Two approaches - minimizing uncertainties versus coping with uncertainties - were presented, arguing for the necessity to push organizations more towards the coping with uncertainties approach whilst maintaining a balance with the minimizing uncertainties approach, however. One way to achieve such loosely coupled systems is systematic rules management focusing especially on the development of process rules. These rules can serve as heuristics for adaptive behavior including swift switches between different coordination mechanisms. The distinctions provided by Hale and Swuste (1998) in their classification scheme for rules have proven very useful for developing a more fine-grained perspective on kinds and levels of standardization operating in a given setting. They could also be good framework for such a systematic rules management. Another important aspect of rules management concerns the ways in which rules are defined and implemented. In line with the concept of higher-order autonomy (Grote, 1997; Klein, 1991) it is very important to have the future users of the rules involved in rule formulation and implementation. This can be understood as a process rule for rules management itself.

Acknowledgements

I thank my project partners at the Swiss Federal Railways and the Swiss Federal Transport Agency for financial support as well as for the stimulating and fruitful cooperation.

References

Amalberti, R. (1999). *Risk management by regulation*. Paper presented at the 19th Myron B. Laver International Postgraduate Course ”Risk Management”, Dept. of Anaesthesia, University of Basel, Switzerland, March 26–27, 1999.

Dekker, S. (2003). Failure to adapt or adaptations that fail: contrasting models on procedures and safety. *Applied Ergonomics*, 34, 233-238.

Emery, F.E. (1959). Characteristics of socio-technical systems. London: Tavistock Document No. 527.

Farrington-Darby, T., Pickup, L. and Wilson, J.R. (2005). Safety culture in railway maintenance. *Safety Science*, 43, 39-60.

Grote, G. (1997). *Autonomie und Kontrolle – Zur Gestaltung automatisierter und risikoreicher Systeme* (Automomy and control – On the design of automated and high-risk systems). Zürich: vdf Hochschulverlag.

Grote, G. (2004a). Uncertainty management at the core of system design. *Annual Reviews in Control*, 28, 267-274.

Grote, G. (2004b). *Organizational measures for achieving loose coupling in high-risk systems: The importance of systematic rules management*. Paper presented at the IEEE-conference on Systems, Man, and Cybernetics, The Hague, October 2004.

Grote, G., Zala-Mezö, E. and Grommes, P. (2004). The effects of different forms of coordination in coping with work load. In R. Dietrich and T. Childress (Eds.), *Group Interaction in High-Risk Environments* (pp. 39-55). Aldershot, UK: Ashgate.

Hale, A.R. and Heijer, T. (2006). Is resilience really necessary? The case of railways. In E. Hollnagel, D.D. Woods and N.G. Levenson (Eds.), *Resilience Engineering – Concepts and Precepts* (pp. 125-148). London: Ashgate.

Hale, A.R., Heijer, T. and Koornneef, F. (2003). Management of safety rules: The case of railways. *Safety Science Monitor*, 7/1.

Hale, A.R. and Swuste, P. (1998). Safety rules: procedural freedom or action constraint? *Safety Science*, 29, 163–177.

Heath, C., Sanches Svensson, M., Hindmarsh, J., Luff, P. and Lehn, D. v. (2002). Configuring awareness. *International Journal of Computer Supported Cooperative Work*, 11, 317-347.

Kerr, S. and Jermier, J.M. (1978). Substitutes for leadership: Their meaning and measurement. *Organizational Behavior and Human Performance*, 22, 375-403.

Klein, J.A. (1991). A reexamination of autonomy in the light of new manufacturing practices. Human Relations, 44, 21-38.

LaPorte, T.R. and Consolini, P.M. (1991). Working in practice but not in theory: Theoretical challenges of "high-reliability organizations". *Journal of Public Administration Research and Theory*, 1, 19-47.

Lawton, R. (1998). Not working to rule: Understanding procedural violations at work. *Safety Science*, 28, 77-95.

Leplat, J. (1998). About implementation of safety rules. *Safety Science*, 29, 189–204.

Marris, P. (1996) *The politics of uncertainty*. London: Routledge.

Perrow, C. (1967). A framework for the comparative analysis of organizations. *American Sociological Review*, 32, 194-208.

Perrow, C. (1984). Normal accidents. Living with high-risk technologies. New York: Basic Books.

Rasmussen, J.(1997). *Risk management in a dynamic society: A modelling problem*. Safety Science, 27, 183–213.

Reason, J., Parker, D. and Lawton, R. (1998). Organizational controls and safety: The varieties of rule-related behaviour. *Journal of Occupational and Organizational Psychology*, 71, 289-304.

Suchman, L. A. (1987). Plans and situated actions: The problem of human-machine communication. Cambridge: Cambridge University Press.

Thompson, J. D. (1967). *Organizations in action*. New York: McGraw-Hill.

Van de Ven, A.H., Delbecq, A.L. and Koenig, R. (1976). Determinants of coordination modes within organizations. *American Sociological Review*, 41, 322-338.

Vermersch, P. (1985). Donées d'observation sur l'utilisation d'une consigne écrite: L'atomisation de l'action. *Le Travail Humain*, 48, 161–172.

Weick, K.E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1-19.

Weick, K.E. (1987). Organizational culture as a source of high reliability. *California Management Review*, 29, 112-127.

Weick, K.E. and Roberts, K.H. (1993). Collective mind in organizations Heedful interrelating on flight decks. *Administrative Science Quarterly*, 38, 357-381.

Woods, D.D. and Shattuck, L.G. (2000). Distant supervision-local action given the potential for surprise. *Cognition, Technology and Work*, 2, 242-245.

Wright, P., Pocock, S. and Fields, B. (1998). *The prescription and practice of work on the flight deck*. Proceedings of ECCE-9, 9th European Conference on Cognitive Ergonomics, pp. 37-42.

This page intentionally left blank

Index

accessibility, passengers' 25–34
  to environments 26–34
  to information 26–34
accident investigation 399, 419–20, 428–30
  and drivers 387–8, 420–25
  lessons learnt 147–52, 429–34
  post-incident interviews 387–97
accidents
  alcohol and drug abuse as causes of 545–52
  causes of 491
  poor communications as cause of 399, 405–7, 424–5, 427–34
  and level 4 vehicle maintenance 439
accidents on the rails
  fatal 103–12
  psychological effects on drivers 103–12
  suicides 103, 110
Acute Stress Disorder (ASD) 104
alarms and alerts 19–20, 126
alcohol abuse
  in Nordic countries 545–52
  and safety critical roles 545–52
alerts, *see* alarms and alerts
Arriva Trains Northern 477
ATLAS 207–8, 237, 239–41, 349
AWS (automatic warning system) 19–20, 190, 205, 208–10, 484, 486
  brake demands 125–32
  and human reliability assessment 466
  interface designs for 125–32

Belgian railways 91, 92
brake demands 125–32, 419–21
British Transport Police 368

cab design
  and CAD (computer aided design) models 142–6
  and drivers 63–70, 137–46, 424, 557, 563
  equipment positioning 137–46, 557, 563
  interfaces 125–32, 424, 557, 563
cab simulators
  design of 156–64
  in design studies 125–32
  for human factors research 155–64
  in safety studies 125–32, 420–21
  in workload studies 92–5, 99
Canada
  anti-trespass initiatives in 38
  RCL operations 79–87
CARMAN (Consensus based Approach to Risk Management) 326–9
CCD Design and Ergonomics 64, 371, 380
CCTV (closed-circuit television)
  and decision-making 63–70
  monitoring 63–70, 371–6, 379–84
CIRAS (Confidential Incident Reporting Analysis System) 461, 465, 496
Clapham Junction incident 536
Cleghorn Automatic Half Barrier Level Crossing 379
Cleghorn Manually Controlled Barrier CCTV Level Crossing 379–84
Common Cause Hypothesis (CCH)
  and SPADs 491, 493–7
communication devices
  positioning in cabs 137–46
communications
  crew in NSW 424–5
  intra-depot 443
  intra-organization 443–4, 446
communications, safety critical
  changing behaviour in 399–407, 432–4
  protocols for 399, 401–3, 405, 427, 431–4
  training in 400–402, 433–4
  verbal 427–34
Communications Working Group (CWG) 399, 405
Community Safety Support Unit (CSSU) 38, 40
compliance, *see* rule and procedure compliance
control systems 301–6
control theory 12–13, 21
Controller of Site Safety (COSS) 334, 335, 427, 429–30, 434
controllers
  and control systems 302–6
  and fatigue 553–6, 559–63
  and mental workload 553–6, 559–63
  and planning 302–6
  tramway 553–6, 559–63
Croatia
  rail network 355
Cross Rail 186
CSR (Cab Secure Radio) 137, 230, 430

CTAC (Computerised Track Access Control System) 344
HFI (Human Factors Integration) 345–51

deadman systems 147–9, 151, 419–21, 424
Denmark DSB 549
Denmark railways
and alcohol and drug abuse 546–7, 549–51
derailment 85, 518
and points maintenance 448, 449, 452
poor communications as cause of 429, 430–31
Waterfall accident 418, 425
design models 175–81
CAD (computer aided design) 142–6, 177–81, 193
human models 138–46, 177
physical 138–46, 175–6, 178–81
Direction 2006 (Canada) 38
DOO(P) (driver only operation (passenger)) 63–70
Double Deck Intercity Cars (DDIC) 149
DRAs (Driver Reminder Appliances) 466
driver behaviour 172
in fatigue and workload study, 554–5
and human reliability assessment 455–63
in route driveability assessment 187–92
in signal/sign co-location assessment 203–12
driver expertise
route knowledge 71–8, 118–20, 121–3
driver fatigue
and accidents 423–4, 565–6, 571
and human reliability assessment 466
management of 151, 152, 566, 570–72
and shiftwork 536, 537, 538–40, 541–2, 554, 557, 562–3, 565–72
and sleepiness 475–80, 529–34, 554
and SPADs 475–80, 566–9
in tramway drivers 553, 554–8, 562–3
driver information processing 63–70, 91–100, 115–23, 186–7, 189, 203–12
driver performance 17–21, 63–70
and alcohol and drug abuse 548–9, 551–2
and brake demands 125–32, 420–21
and human reliability assessment 456, 457
at junctions 115–23
and mental workload 91–100, 553, 554–8, 562–3
prediction of 186–94, 455–63
and psychological trauma 103–12, 150–51
after rest days 475–80
route knowledge 71–8, 116, 118–20, 121–3, 205
in signal/sign co-location assessment 203–12
situation awareness 71–6, 116–17, 122–3
driver training 71, 75–8, 151–2
driver vigilance 91–100, 186
driver workload 63–70, 186–7, 190–94
and brake demands 128, 131–2
mental 91–100, 553, 554–8, 562–3
in signal/sign co-location assessment 203–12
drivers
and accidents 103–12, 147–52, 387–8, 420–25
anthropometric data 137–46, 147–8, 420
and communications 427, 429–30, 431, 434
health of 103–12, 147–52, 423, 425, 529–34
and human reliability assessment 455–63
and mitigation measures for SPADs 483–90
and Obstructive Sleep Apnoea Syndrome (OSAS) 529–34
and sleepiness 475–80, 529–34, 554
drug abuse
in Nordic countries 545–52
and safety critical roles 545–52
Dutch Railways *see* Netherlands Railways, Nederlandse Spoorwegen

Early Human Factors Analysis (EHFA) 78, 345–6, 348
engineering
cab fitments 137–46
and ergonomics 501–9
and HFIPs 511–16
and possessions 331–41
renewals 506–7
standards 506
on tracks 343–51, 427, 429, 430–31
work sites 331–41
see also maintenance
Engineering and Physical Sciences Research Council (EPSRC) 5, 155
engineering models 9–16
engineering possessions
and communications 427
irregularities 427, 430
management of 331–41
engineering supervisors
and communications 427
and possession management 331–41
engineers 430
locomotive 80–87
Enhanced Cognitive Interview (ECI) 388, 389–96
and rail industry 396–7
EPS (Enhanced Permissible Speeds) 203
Epworth Sleepiness Scale (ESS) 531–2, 533
ergonomics
and cab modifications 137–46

and design 137–46, 147, 149, 175–81, 186–7, 501–2, 507–9
and driver's consoles 147, 149
and engineering 137–46, 186–7, 501–9
in fatigue and workload study 553–5
and mechanical signals 257–65
and performance prediction 9–21
standards 506
ERTMS (European Rail Traffic Management Systems) 127, 305, 504, 509
testing 167–73
ETCS (European Train Control Systems) 305
testing 167–73
EURNEX 5
Euro-Balise 169
European Rail Human Factors Conference 3, 6
European Railway Agency (ERA) 496
Eurostar 403, 477
Euston Station (London Underground) 28–34
Euston Station (overland) 28–34
eye movement dynamics 12–15

fatigue
economic implications of 536–7
and health 536, 538, 542, 543, 554, 562–3
and human performance 535–6, 543, 553–63
management of 535–43, 554, 562–3
and safety 535–6, 543, 554, 555–6, 562–3
and safety critical work/roles 535–43, 553–63
and shiftwork 536, 537, 538–40, 541–2, 554, 557, 559–60, 562–3, 565–72
training regarding 537–8, 540–41, 542, 554, 563
and working hours 537, 538–40, 541–2, 554, 557, 559–60, 562–3, 565–72
and workplace 538, 541, 542, 554, 557, 559, 562–3
see also driver fatigue
Fatigue Audit Interdyne (FAID) 151, 423–4, 570
Federal Railroad Administration (FRA) 80
Finland railways
and alcohol and drug abuse 546–7, 549–51
foot pedals 139, 147–9, 420–21, 424

Glenbrook accident 147, 150, 419
GOMS (Goals-Operators-Methods-Selection) 12, 188–9, 240
GRIP (Guide to Railway Investment Projects) 508
and route driveability 185, 187, 192–4
GSM-R (Global System for Mobiles-Railway) 169, 504, 509
HMI (human machine interface) 137–46, 509

Health and Safety Executive (HSE)
Fatigue Index 480, 540, 541, 570–71, 572
and human reliability assessment 460–62, 466
and level crossings 363–70
and mechanical signal levers 260–64
and safety procedures 555, 561
and Selective Door Operations (SDO) systems 460–62
HEART (Human Error Assessment and Reduction Technique) 449, 451, 455, 457, 459–60, 463, 466–7, 470
Hierarchical Task Analysis (HTA) 327, 349, 448, 450, 453
driving activities 187–8
Engineering Supervisors activities 332–3
management of points maintenance activities 447–8, 450
points maintenance activities 447–8, 450, 452
shunting activities 312, 315
signaller activities 239–41, 244, 447–8, 449–50
Highway Code
and level crossings 364, 366
HM Railway Inspectorate 520, 540
and human reliability assessment 461
Inspectors 363, 365–70, 461
and level crossings 363–70
and Selective Door Operations (SDO) systems 461
Hong Kong railways
and human factors 517–25
improvement of 522–3
Inspectorate (KHRI) 518
reliability of 521–4
and safety 518, 521–2, 523–4
and SPADs 518, 521, 523
and UK compared 518–19, 520, 523
Hong Kong Transport Department 518, 521
Human Error Identification (HEI) 455–63
Human Error Quantification (HEQ) 455–63
Human Factors Integration (HFI) 345–51, 501–2, 509
Human Factors Integration Plans/Programmes (HFIPs) 511–16
Human Factors Signal Sighting Framework 483–90
human performance prediction 9–21, 456–60, 462, 465–70
human reliability assessment (HRA)
and aging of workforce 466–70
for driver behaviour/tasks 455–63
for management of points maintenance 447–53
models 447–53
for points maintenance 447–53

and prediction 465–70
and Selective Door Operations (SDO) systems 460–62, 463
for signalling operations 447–53
tools 455–63
inclusive design 25–34
Independent Transport Safety and Reliability Regulator (ITSRR) 150
information accessibility 25–34, 49–58
information presentation
commercial 49–58
travel 49–58
integrated rail operations, human factors model of 275–82
Integrated Workload Scale (IWS)
drivers 128
signallers 217, 240–43, 379–84
Intergo 51, 58

JBV, *see* Norwegian Rail
junction(s)
approach to and passage through 115–23
signalling 115–23

KADS (Knowledge Acquisition and Design System) 288, 289–91, 296, 297
Karolinska Scale of Sleepiness 531–2
KCRC (Kowloon-Canton Railway Corporation) 517–18, 521, 523

Ladbroke Grove accident 71, 387, 523
Inquiry 125, 225, 410
level 4 vehicle maintenance, *see* maintenance
level crossings
and aggressive car drivers 358–61
CCTV systems at 371–6, 379–84
design 366–7
human factors issues at 363–70
poor communications as cause of accidents at 429, 430–31
road traffic flow at 358–61
safety measures 355–61, 363–70, 371–6
Logan's Road Level Crossing 379, 382
London Lines 477
London Underground 28–34, 103
engineering work 343–51, 502
human factors standards 350–51, 520, 523
Track Access Control (TAC) Department 343–51

maintenance, level 4 vehicle
communications and 443, 446
computerized systems for 442–3, 446
and depot design 440, 446
and equipment design 439
and parts 440
procedures for 439, 440, 441–2, 444–5, 446
and training for 441, 446
and vehicle design and modification 440–41, 445
maintenance, points
human reliability assessment for 447–53
MCB circuit breaker trips 127, 129–30
MerseyRail
anti-trespass initiatives 40
Midland Mainline 477
mis-routeings 115, 122
models
engineering 9–16
human performance 9–21
of integrated rail operations 275–82
mathematical 11–16, 21
quantitative 9–16, 21
Motherwell Signalling Centre 379–84
MTRCL (Mass Transit Railway Corporation, Hong Kong) 517–18, 521–3

Nederlandse Spoorwegen, NS
NS Passengers 288
NS Reizigers 49
NS Stations 49
see also Netherlands Railways
Netherlands Railways (Nederlandse Spoorwegen, NS)
Inspectorate 492–6
shunting planning 309–17
and SPADs 491–7
staff planning 287–98
station signage 49–58
Network Rail 4–5, 225
and accessibility 25–34
and accident investigation 387–8
anti-trespass initiatives 38, 40
and communications 402
Community Liaison Team 45
Design and Architectural Team 34
engineering and ergonomics 502–9
Engineering and Operations 337
Ergonomics Group/Team 331, 337, 502–9, 540
Ergonomics National Specialist Team (NST) 3, 4–5, 28, 258
and fatigue management 535–43
Human Factors Standard 512, 520, 523
and human reliability assessment models 447–53
Human Resources System 539

Investigations team 540
and level crossings 372, 379–80, 384
Maintenance group 543
and mechanical signal lever operation 257, 258
National Rostering Application for (NRANR) 539
North-East Region 178
Operations and Customer Services 538–9, 543
and possessions management 331–41
REQUEST questionnaire 247–56
and route driveability assessment 185–94, 197
Safety and Environment Plans 405
Safety Critical Communications Implementation Group (NRSCCIG) 405
Signal Sighting Committee 155
signal/sign co-location assessment 203–12
and signaller workload 215–22, 237–45, 379–80, 384
signalling functions 178–80, 372, 379–80, 384
and simulator design 155–6
SPAD Hazard Checklist 197
Strategic Safety Group 405
Tactical Safety Group 405
New South Wales (NSW) 147, 148, 410–12, 414–15
Norway railways
and alcohol and drug abuse 546–7, 549–51
Norwegian Rail (JBV) 546
effect of educational activities in 575–80

Obstructive Sleep Apnoea Syndrome (OSAS) 529–34
symptoms 529–30
and train drivers 530–34
in UK transport industries 530
in US transport industries 530
on-time running 150, 152
Operation Lifesaver (USA) 38
Operational Demand Evaluation Checklist (ODEC) 215–22, 255, 379, 380, 382–3, 384
OCC (onboard computer control), for RCLs 79, 83–6
OCUs (operator control units), for RCLs 79, 83–7

Paddington Station 28–34
parts for vehicles
design and maintenance of 440–41
supply of 440
passenger information systems 25–34
passenger safety 64–70
passengers
with disabilities 26–34
PICOPs (Person in Charge of the Possession)
and communications 427
and possession management 331, 333, 335–41
planners
reasoning of 287–98
shunting 309–17
planning
shunting 309–17
staff 287–98
systems 287–98, 301–6, 309–17
traffic 301–6
points maintenance
and human reliability assessment 447–53
possessions, *see* engineering possessions
Post-traumatic Stress Disorder (PTSD) 104–7, 111–12, 150
PRDA (perceive-recognize-decide-act) cycle 189–90
procedure compliance, *see* rule and procedure compliance
Procedures Culture Survey 322–5, 329
ProRail 49, 496
prospective memory 92–100

quantitative models 9–16, 21
Queensland railways 411–15
Rail Tram and Bus Union (RTBU) 411
and safety 411–15

Rail Research UK 5
Rail Safety Standards Board (RSSB) 5, 116
and communications 403, 405
Fatigue and Shiftwork project 477, 565–72
Human Factors Group 3–4, 520
and human reliability assessment 455–63
and level crossings 367
library of Formal Investigation reports 428
route knowledge study 71–8
and rule compliance 581–9
and simulators 125, 156
SPAD Hazard Checklist 197
trespass studies 37–47
RailSiTe 167–73
Railtrack 225, 227, 234, 466, 502, 538
Railway Group Standard
driving activities 187
engineering competence 507
Engineering Supervisor defined 331
PICOP defined 337
protections and possessions 238
signals and signs 206, 260, 483
working hours 539, 565, 571
RCLs (remote control locomotives) 79–87
equipment, see OCC, OCUs

RCOs (remote control operators), for RCLs 79–87
training 81, 82–3
RDT (route drivability technique) 187–8, 190–94
Repertory Grid Technique (RGT) 218–22
REQUEST (The Railway Ergonomics Questionnaire) 247–56, 542
administration of 249–50
data and response analysis 250
design and layout of 249
findings 250–55, 341, 541
robot trains 79–87
Route Assessment Checklist (RAC) 197–202
route design 185–94
route driveability 155, 185–94, 197–202
route knowledge 71–8, 205
impact of new technology on 75–8
training 71, 75–8
psychological processes involved in 71–4, 76–8
Route Knowledge Model 72–8
route safety 155, 197–202
rule and procedure compliance
and communications 427, 432
and fatigue 537, 542, 555–6, 561–2
and level crossings 366
management of 321–9, 581–9, 591–8
and vehicle maintenance 441–2, 446
rules management 591–8

safety
and alcohol and drug abuse 545–52
and communications 399–407
and fatigue 535–6, 565, 570–72
in Hong Kong railways 518, 521–2
and human reliability assessment 465–6, 470
and level 4 vehicle maintenance 439
at level crossings 355–61
of passengers 63–70
and RCL operations 79–87
and workforce aging 466–70
safety climate 409–10
in New South Wales rail industry 410–12, 414–15
in Queensland railways 411–15
in UK rail industry 410
in US haulage firms 410
safety culture 152, 409–10
and communications 424–5, 444–5
and level 4 vehicle maintenance 444–5
procedures 321–9, 444–5
Safety Management Information System (SMIS) 459, 583
data related to junctions 116
trespass incidents 37, 45, 46–7
safety systems
brake demands 125–32
deadman systems 147–9, 151
for track access 343–51
ScotRail 477
seat reference point (SRP) 138
Selective Door Operation (SDO) systems
and human reliability assessment 460–62, 463
shunting 445
planning 309–17
rules and procedures for 595–7
signal boxes 218, 221, 364
CCTV systems at 371–6
design of 178–80, 257, 503–4
mechanical levers 257–65
supervision in 225–35
signal detection theory 20, 21
Signal Sighting Committee 155, 483, 489
signallers 137, 178–80
attitudes and opinions on human factors 247–56
and communications 427, 429–30, 431, 433–4
information processing 215–16
and level crossings 364, 371–6, 379–84
operation of mechanical lever signals 257–65
and reliability assessment 447–53
supervision of 225–35
workload 215–22, 230–31, 234, 237–45, 341, 379–84
workload prediction 237–45
signalling equipment
design 178–80, 503–4
at level crossings 356, 364, 371–6, 379–84
mechanical levers 257–65
signalling scheme design 186–7, 192–3, 237, 505
signalling systems 217–22, 503–4
signals
co-location with signs 203–12, 484–5
design 483, 484–8, 503–4
driver response to 115–23, 483–90
junctions 115–23
positioning of 203–12, 484–8
sighting 155, 187–8, 197, 203–12, 483–90
and SPAD reduction 483–90
signs
anti-trespass 37–47
co-location with signals 203–12, 484–5
at level crossings 357–8, 360
positioning of 37–40, 46–7, 49–58, 203–12
Simbase 92–5, 99
simulators
for control systems 301–6
for ERTMS/ETCS conformity testing 167–73

see also cab simulators
Situation(al) Awareness model 71–3, 116–17
Slovenia rail network 355, 359
SMEs, used in research 267–74
on control/controllers 270
on driving/drivers 75, 270, 457, 458–9
on route assessments 197, 199
on safety behaviour of maintenance workers 269–70
on signalling/signallers 218–22, 240, 244, 270–71, 379–84
on simulators 158
on wayfinding 28
Southall accident 565
Southern Trains 128
SPAD Bow Tie model 492–6
SPAD Hazard Checklist (SHC) 197, 198
SPADs 115, 122, 203–4, 422, 505
and alcohol abuse 551, 552
assessment of mitigation measures for 483–90
causes of 491–7
and communications 405, 427, 429, 430–31
evaluation of 197–202
incidence of after driver rest days 475–80
investigations 387–9, 396–7, 566–9
post-incident interviews 387–97
prediction of 16–21, 197–202, 462, 465, 491–7
and rule compliance 583
and safety systems 125, 126, 149, 518, 521
stations
and accessibility 25–34
layout 49–58
signage 49–58
STEG project 303–6
Stockholm 547
Strategic Rail Authority (SRA) 25, 34, 147, 151, 421–2
Swedish Rail Authority 547
Swedish Railways
and alcohol and drug abuse 547, 549–52
Northern Region 103
switching yards 79–87
switchmen 80–87
Sydney 150, 418
system performance prediction 9–16

Tangara (G7) 150, 425
deadman system 147–9, 419–22
Outer Suburban Tangara 149
Technical University of Lisbon
Ergonomics Department 553
Thameslink 186, 190
time-reliability curves 15–16
TPWS (Train Protection Warning System)
HMI (human machine interface) 137
interface designs for 125–32
Overspeed Sensor (OSS) 126, 127, 129–30
'reset and continue' 137
Train Stop (TS) 126, 127, 129–30
TRACEr (Technique for the Retrospective and predictive Analysis of Cognitive Errors) 448, 451, 455–6, 457–9, 463
track access
computerized systems to control 343–51
at London Underground 343–51
and safety 343–51
Track Access Control (TAC) Operators 344–5, 349
traffic control
systems 301–6
train cab design, *see* cab design
train control
automation 79–87
systems 167–73
train design
in Hong Kong 520–21
and maintenance 440, 445
training/educational activities
and affective commitment 575–6, 577–8, 579–80
drivers 71, 75–8, 151–2
fitters 441, 446
guards 151–2
and inner motivation 575–6, 577, 578–80
RCOs 81, 82–3
and voluntary turnover 575–7, 578–80
and work effort 575
trespass 37–47

United States of America (USA)
anti-trespass initiatives 38
RCL operations in 79–87
University of Nottingham 380
Institute for Occupational Ergonomics (IOE) 3, 331, 337
Rail Human Factors Group (RHFG) 3, 5
Uppsala University 301–2, 305

Victoria Department of Infrastructure 148
Victoria suburban railways 148–9
vigilance 91–100
vigilance systems 86, 99, 148–50
VR-group (Finnish state railways) 549, 551

Waterfall accident 147–52, 419
driver 421–5

guard 423–5
human factors 419–25
Special Commission of Inquiry 147, 149, 150, 410–11, 419–20
Watford Junction 28–34
warning systems 19–20, 125–32
see also AWS, TPWS
wayfinding 25–34, 49–58
West Coast
Engineering (WCE) 503–4
and ergonomics 502–4
Project Review Group (WC-PRG) 503
Route Modernisation (WCRM) project 203, 331, 333, 338
Safety Group 331, 337
workforce
age 466–70

youths
and trespass 37–47