Measuring the technical integrity of a complex engineered system

by

Michael Thomas Edwards

B.E., M.Eng.Sci., MBA

A thesis submitted for the degree of

Doctor of Philosophy
Research (Systems Engineering)

Defence and Systems Institute

Division of Information Technology, Engineering and the Environment



University of
South Australia

January 2014

Contents

Contents ii
List of Figures ix
List of Tables xvi
Glossary xx
Summary xxviii
Declaration xxxi
Acknowledgments xxxii
1 Introduction 1
1.1 Background to the research 1
1.2 Research problem 2
1.3 Justification for the research 2
1.4 Methodology 3
1.5 Outline of the thesis 6
1.6 Definitions 7
1.7 Delimitations of scope and key assumptions 8
1.7.1 Delimitations 8
1.7.2 Key assumptions 9
1.8 Conclusion 10
2 The research context 12
2.1 Introduction 12
2.2 Technical regulation 12
2.2.1 The history of technical regulation 13
2.2.2 Theory of regulation 14
2.2.3 Technical regulation in the ADF 19
2.2.4 Technical regulations in other domains 36
2.3 Complex engineered systems 38
2.3.1 Development of complex engineered systems in the ADF 39
2.3.2 Concepts, properties and characteristics of complex engineered systems 41
2.3.3 Definitions of complex engineered systems 47
2.3.4 Design and implementation methods 50
2.4 Conclusion 51
3 Literature survey 54
3.1 Introduction 54
3.2 Support for measuring technical integrity 54
3.3 Extant measures for technical integrity 56

3.3.1 Life cycle value of systems 56
3.3.2 Stakeholder value of a system 57
3.3.3 System quality attributes 57
3.3.4 Dependability assessment 59
3.4 Measurement approaches to be considered 60
3.4.1 General considerations 60
3.4.2 Proxies for Technical Integrity 61
3.4.3 The measurement of risk 62
3.4.4 Consensual Assessment Technique (CAT) 64
3.4.5 Multi-attribute approaches for ranking systems 65
3.4.6 Measuring technical performance 66
3.4.7 The risk-value method; and composite performance model (CPM) 66
3.4.8 System usability scale 67
3.4.9 International nuclear event scales (INES) 68
3.5 Measurement theory 69
3.5.1 Definition and key concepts of measurement 69
3.5.2 Basics of measurement theory 72
3.5.3 Reliability and validity of measurement 75
3.5.4 Aggregation measures 76
3.6 Literature survey conclusions 80
4 Methodology 81
4.1 Introduction 81
4.2 Research phases 85
4.2.1 Concept development 86
4.2.2 Content validity of the technical integrity measurement construct 86
4.2.3 Conjoint measurement aggregation weights 87
4.2.4 Assessment of judgments for estimation of technical integrity 87
4.2.5 Technical integrity theory and measurement instrument 88
4.3 Research methods 88
4.3.1 Survey method 88
4.3.2 Quantitative analysis method 90
4.3.3 Qualitative analysis method 90
4.3.4 Mixed model method 91
4.4 Ethical considerations 92
4.5 Conclusion 93
5 Concept development 94
5.1 Introduction 94

5.2 Defining technical integrity 95
5.2.1 What is technical integrity of a complex engineered system? 95
5.2.2 MIL-STD-882C safety compared to ADF technical integrity. 97
5.2.3 Relationship of emergent properties and technical integrity 97
5.2.4 Candidate concepts for assessing technical integrity 99
5.2.5 A theory of technical integrity 101
5.3 Form of measurement 102
5.3.1 Measurement objective 102
5.3.2 Selecting a form of measurement 103
5.3.3 Conjoint measurement approach 104
5.3.4 Capability measurement approach 109
5.3.5 Dependability measurement approach 110
5.4 Elements of technical integrity 112
5.4.1 Technical goals of regulation 112
5.4.2 Technical goals and system quality attributes 112
5.4.3 Indicators of technical integrity 115
5.5 Conclusion 118
6 Content validity of the technical integrity measurement construct 120
6.1 Introduction 120
6.2 Method 120
6.2.1 Survey instrument 122
6.2.2 Sampling approach 122
6.2.3 Demographics of respondents 122
6.2.4 Analysis of survey responses 129
6.3 Quantitative results 130
6.3.1 Quantitative analysis: technical integrity measurement scale 130
6.3.2 Quantitative analysis: capability measurement scale 135
6.3.3 Quantitative analysis: dependability measurement scale 139
6.3.4 Quantitative analysis: conjoint measurement of technical integrity 143
6.3.5 Summary of quantitative results 147
6.3.6 Additional quantitative analyses and anomalies 149
6.4 Qualitative results 151
6.4.1 Better capability measurement 151
6.4.2 Better dependability measurement 157
6.4.3 Reflections on the measurement of technical integrity 163
6.5 Combined (mixed-method) results 169
6.5.1 Technical integrity measurement proposal validity 169
6.5.2 Technical integrity measurement scale 170

6.5.3 Conjoint measurement for technical integrity ... 171
6.5.4 Measurement of capability ... 171
6.5.5 Measurement of dependability ... 173
6.6 Conclusions ... 174
6.6.1 Findings ... 174
6.6.2 Limitations ... 174
6.6.3 Additional research questions ... 175
7 Conjoint measurement aggregation weights ... 176
7.1 Introduction ... 176
7.2 Method ... 176
7.2.1 Survey instrument ... 177
7.2.2 Analysis of survey responses ... 179
7.3 Quantitative results ... 180
7.3.1 Response data ... 180
7.3.2 Interpretation ... 190
7.3.3 Analysis ... 190
7.3.4 Derivation of conjoint aggregation weights ... 200
7.4 Qualitative results ... 200
7.4.1 General ... 200
7.4.2 Domain of practice or viewpoint ... 201
7.4.3 Services provided by the system ... 209
7.5 Combined (mixed-method) results ... 213
7.6 Conclusion ... 217
7.6.1 Findings ... 217
7.6.2 Limitations ... 218
7.6.3 Additional research questions ... 218
8 Assessment of judgments for estimation of technical integrity ... 219
8.1 Introduction ... 219
8.1.1 Purpose ... 219
8.1.2 Concepts and research questions ... 219
8.1.3 Constraints ... 222
8.2 Method ... 223
8.2.1 Survey instrument ... 223
8.2.2 Indicator models from technical parameters ... 236
8.2.3 Sampling approach ... 238
8.2.4 Demographics of respondents ... 239
8.2.5 Analysis of survey responses ... 246

8.3 Quantitative Results 249
8.3.1 Reliability of using the CAT for technical integrity measurement 249
8.3.2 Reliability of assessing indicators of technical integrity by the CAT 256
8.3.3 Technical integrity scale reliability 258
8.3.4 Correspondence of indicators with expert judgments 261
8.4 Qualitative Results 331
8.4.1 The nature of technical integrity measurement 331
8.4.2 Use of objective data in judgments 333
8.4.3 Judgment approach with no objective data 337
8.4.4 Views on obtaining better ratings 340
8.5 Discussion 343
8.5.1 Reliability of using the CAT for technical integrity measurement 343
8.5.2 Technical integrity scale reliability 345
8.5.3 Correspondence of indicators with expert judgments 346
8.5.4 Safety 347
8.6 Conclusion 348
9 Technical integrity theory and measurement instrument 350
9.1 Introduction 350
9.2 Method 351
9.2.1 Confirmatory factor analysis 352
9.2.2 Exploratory factor analysis 352
9.3 Results 353
9.3.1 Confirmatory factor analysis of theorised model of technical integrity 353
9.3.2 Exploratory factor analysis 356
9.3.3 Confirmatory factor analysis of amended model of technical integrity 375
9.4 Discussion 379
9.4.1 Revision to the theory of technical integrity 379
9.4.2 Assessment of measurement instruments 381
9.4.3 Presentation of technical integrity measurement values 396
9.5 Conclusions 398
9.5.1 Findings 398
9.5.2 Limitations 399
10 Summary, conclusions and implications 400
10.1 Summary 400
10.2 Conclusion about the research questions 405
10.2.1 Conjoint measurement construct for technical integrity 405
10.2.2 Content validity of the technical integrity measurement construct 406

10.2.3 Conjoint measurement aggregation 406
10.2.4 Reliability of using the CAT for technical integrity measurement 407
10.2.5 Reliability of assessing indicators of technical integrity by the CAT 408
10.2.6 Technical integrity scale reliability 408
10.2.7 Correspondence of indicators with expert judgments 408
10.2.8 Components of technical integrity 410
10.2.9 A measurement instrument for technical integrity 411
10.2.10 Summary of conclusions 411
10.3 Conclusion about the research problem 412
10.3.1 Findings and insights 412
10.3.2 Contribution to the research problem and area 414
10.4 Implications for theory 415
10.5 Implications for policy and practice 416
10.5.1 Practical implications and usage of the findings 416
10.5.2 Implications for policies 416
10.6 Limitations 417
10.6.1 Cautions to use of findings 417
10.6.2 Validity of assumptions 418
10.6.3 Threats to validity of findings 419
10.7 Implications for further research 420
References 422
Bibliography 431

Appendix A Survey instrument: validity of the measurement construct A-1
Appendix B Survey instrument: measurement by consensual assessment technique B-1
Appendix C Validity of technical integrity measurement construct: full text extracts referenced in qualitative analysis C-1
Appendix D Conjoint measurement aggregation: full text extracts referenced in qualitative analysis D-1
Appendix E Judgment of technical integrity and associated indicators: full text extracts referenced in qualitative analysis E-1
Appendix F The impact of technical regulation on the technical integrity of complex engineered systems F-1
Appendix G An ontology of technical regulation G-1
Appendix H Is systems engineering necessary for the achievement of technical integrity? H-1
Appendix I A framework for the evaluation of verification evidence supporting the assessment of environmental qualification I-1

Appendix J What is technical integrity and how to measure it? J-1
Appendix K The content validity of a technical integrity measurement construct K-1
Appendix L The consensual assessment technique for the measurement of technical integrity L-1

List of Figures

Figure 1. Research project's phased approach. 5

Figure 2. Reason's 'Swiss Cheese Model' adapted to show technical regulation as one of the layers of defence in the development of a complex system. 19

Figure 3. Total number of ADF fatal aviation accidents 1970-2006 (RAN, 2007) 21

Figure 4. The influence of technical regulation on technical integrity. 30

Figure 5. The main provisions of the ADF's NTRS. 33

Figure 6. Example partial hierarchy of important internal quality characteristics (Firesmith et al., 2009c, Fig. 18.4). 58

Figure 7. Example partial hierarchy of important external quality characteristics (Firesmith et al., 2009c, Fig. 18.5). 59

Figure 8. Magnitude of safety events versus the number of events experienced. 69

Figure 9. Concept of a measurement scale. 77

Figure 10. Concept of a measurement index. 77

Figure 11. Qualitative methods to explain quantitative results. 92

Figure 12. The theory of technical integrity. 101

Figure 13. (i) Product structure for TI. (ii) Mapping of measured attributes to ordinal scale of TI. 105

Figure 14. Expected form of $TI = f(FS,PS)$. 106

Figure 15. Capability and dependability with equal influence ($\alpha=\beta=1,\ TI = C.D$). 108

Figure 16. Extreme cases: (i) *TI* not influenced by capability ($\alpha=0,\ TI = D$). (ii) *TI* not influenced by dependability ($\beta=0,\ TI = C$). 108

Figure 17. Differing relative utilities of capability and dependability. (i) Dependability-biased ($\alpha=1,\ \beta=2,\ TI = C.D^2$). (ii) Capability-biased ($\alpha=2,\ \beta=1,\ TI = C^2.D$). 109

Figure 18. Pie chart of the gender of the respondents (n=116). 123

Figure 19. Histogram of the age of the respondents (n=116). 123

Figure 20. Pie chart of the highest level of technical education of the respondents (n=116). 124

Figure 21. Pie chart of the main discipline of technical education of the respondents (n=116). 125

Figure 22. Pie chart of the current employment sector of the respondents (n=116). 125

Figure 23. Pie chart of the employment sector of most experience gained by the respondents (n=116). 126

Figure 24. Histogram of the years of professional practice of the respondents (n=116). 126

Figure 25. Pie chart of the ADF or other engineering authority delegation held by the respondents (n=116). 127

Figure 26. Pie chart of the system application domain where the respondent has most experience (n=116). 127

Figure 27. Pie chart of the system role viewpoint the respondent believes they represent (n=116). 128

Figure 28. Bar chart of the engineering delegation held by respondents (n=116). 129

Figure 29. Bar chart of the level of agreement ratings of ‘ordinal scale meaningful’. 131
Figure 30. Bar chart of the level of agreement ratings of ‘measurement by a single numeric value’. 132
Figure 31. Bar chart of the level of agreement ratings of measurement on a ratio scale meaningful. 134
Figure 32. Bar chart of the level of agreement ratings of capability measured on a scale of 0 to 1. 135
Figure 33. Bar chart of the level of agreement ratings of capability – more than specified requirement is valued. 137
Figure 34. Pie chart of rating for a better measure of capability. 138
Figure 35. Bar chart of the level of agreement ratings of dependability measured on a scale of 0 to 1. 139
Figure 36. Bar chart of the level of agreement ratings of dependability – better than risk acceptability threshold of stakeholders is valued. 141
Figure 37. Pie chart of rating for a better measure of dependability. 142
Figure 38. Bar chart of the level of agreement ratings of a technical integrity conjoint measure is meaningful. 144
Figure 39. Bar chart of the level of agreement ratings of a technical integrity conjoint measure is meaningful for similar system comparison. 145
Figure 40. Bar chart of the level of agreement ratings of a technical integrity conjoint measure is meaningful for dissimilar system comparison. 147
Figure 41. Scatterplot of ratings of technical integrity: single numeric value meaningful vs conjoint measurement meaningful 150
Figure 42. Number of coded extracts – better capability measurement 152
Figure 43. Number of coded extracts – better dependability measurement 159
Figure 44. Number of coded extracts – reflections on the measurement of technical integrity. 165
Figure 45. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Military – Aviation’. 182
Figure 46. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Military – Naval’. 182
Figure 47. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Military – Land’. 183
Figure 48. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Military – Joint, C4ISR’. 183
Figure 49. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Aviation’ 184
Figure 50. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Maritime’. 184
Figure 51. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Medical’. 185
Figure 52. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Mining’. 185

Figure 53. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Civil – Chemical Process'. 186

Figure 54. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Civil – Industrial'. 186

Figure 55. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Civil – Telecommunications and IT'. 187

Figure 56. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Civil – Transportation'. 187

Figure 57. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Civil – Power Utility – Conventional'. 188

Figure 58. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Civil – Power Utility – Nuclear'. 188

Figure 59. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Civil – Water Utility'. 189

Figure 60. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Civil – Space'. 189

Figure 61. Sample median, mean and interquartile ranges for ratings of relative importance of capability and dependability versus system domains. 191

Figure 62. Model Result for the median of Capability vs Dependability for the Military – Aviation system domain being 'Equally Important' (i.e., 4). 193

Figure 63. Model Result for the median of Capability vs Dependability for the Civil – Aviation system domain being 'Dependability Much More Important' (i.e., 6).194

Figure 64. Difference in distribution of judgements for 'Military – Joint, C4ISR' across the demographic category of Highest Level of Technical Education. 195

Figure 65. Difference in distribution of judgements for 'Civil – Space' across the demographic category of Current Employment Sector. 196

Figure 66. Difference in distribution of judgements for 'Civil – Water Utility' across the demographic category of Sector of Most Experience. 197

Figure 67. Difference in distribution of judgements for 'Civil – Water Utility' across the demographic category of Role View Point. 198

Figure 68. Significant median ratings of relative importance of capability and dependability versus system domains. 199

Figure 69. Number of coded extracts – System Domain or Viewpoint Influences on judgment of Capability versus Dependability. 203

Figure 70. Number of coded extracts: system service influences on judgment of capability versus dependability. 210

Figure 71. Influence diagram of technical integrity and quality attributes of a complex engineered system. 222

Figure 72. Block diagram for System A. 225

Figure 73. Block diagram for System B. 225

Figure 74. Block diagram for System C. 226

Figure 75. Pie chart of the gender of the respondents (n=43). 239

Figure 76. Histogram of the age of the respondents (n=43). 240

Figure 77. Pie chart of the highest level of technical education of the respondents (n=43)..... 240
Figure 78. Pie chart of the main discipline of technical education of the respondents (n=43). 241
Figure 79. Pie chart of the current employment sector of the respondents (n=43). 242
Figure 80. Pie chart of the employment sector of most experience gained by the respondents (n=43). 242
Figure 81. Histogram of the years of professional practice of the respondents (n=43)..... 243
Figure 82. Pie chart of the ADF or other engineering authority delegation held by the respondents (n=43). 243
Figure 83. Pie chart of the system application domain where the respondent has most experience (n=43).. 244
Figure 84. Bar chart of the engineering delegation held by respondents (n=43). 245
Figure 85. Bar chart of the system systems experience of the respondents (n=43)..... 246
Figure 86. Frequency of technical integrity ratings for System A. 250
Figure 87. Frequency of technical integrity ratings for System B..... 250
Figure 88. Frequency of technical integrity ratings for System C..... 251
Figure 89. Median, interquartile range and range of technical integrity ratings. 252
Figure 90. Result of Wilcoxon signed rank test: technical integrity of System B minus System C. 252
Figure 91. Intraclass correlation coefficients for technical integrity ratings. 253
Figure 92. Cronbach's alpha for technical integrity ratings. 253
Figure 93. Scattergram of average of 3 ratings versus range of 3 ratings. 254
Figure 94. Correlation of judgment dimensions with technical integrity. 255
Figure 95. Scattergram of ratings for (i) Fitness for service versus technical integrity ratings; (ii) Safety versus technical integrity ratings. 256
Figure 96. Indicator model and ratings for 'Safety'. 262
Figure 97. Boxplots for judgments of 'Safety'..... 263
Figure 98. Correspondence of ratings for 'Safety'. 264
Figure 99. Indicator model and ratings for 'Impact to the natural environment'. 265
Figure 100. Boxplots for ratings of 'Environment impact'. 266
Figure 101. Correspondence of ratings for 'Impact to the natural environment'. 267
Figure 102. Indicator model and ratings for 'Impact to other systems'. 268
Figure 103. Boxplots for ratings of 'Impact to Systems'. 269
Figure 104. Correspondence of ratings for 'Impact to other systems'. 270
Figure 105. Indicator model and ratings for 'Fit for service'. 271
Figure 106. Boxplots for ratings of 'Fit for Service'..... 272
Figure 107. Correspondence of ratings for 'Fit for Service'. 273
Figure 108. Post facto indicator model and ratings for 'Fit for service'. 274
Figure 109. Correspondence of ratings for 'Fit for Service' with post facto indicator model 274

Figure 110. Indicator model and ratings for ‘Capability’. 275
Figure 111. Boxplots for ratings of ‘Capability’.....276
Figure 112. Correspondence of ratings for ‘Capability’. 277
Figure 113. Post-facto Indicator model and ratings for ‘Capability’. 278
Figure 114. Correspondence of ratings for ‘Capability’ with post facto indicator model. 278
Figure 115. Indicator model and ratings for ‘Effectiveness’.....279
Figure 116. Boxplots for ratings of ‘Effective’.....280
Figure 117. Correspondence of ratings for ‘Effective’. 281
Figure 118. Indicator model and ratings for ‘Efficiency’. 282
Figure 119. Boxplots for ratings of ‘Efficient’. 283
Figure 120. Correspondence of ratings for ‘Efficient’.....284
Figure 121. Indicator model and ratings for ‘Quality’.....285
Figure 122. Boxplots for ratings of ‘Quality’. 286
Figure 123. Correspondence of ratings for ‘Quality’.....287
Figure 124. Boxplots for ratings of ‘Flexible’. 289
Figure 125. Correspondence of ratings for ‘Flexible’.....289
Figure 126. Indicator model and ratings for ‘Durability’.....290
Figure 127. Boxplots for ratings of ‘Durable’. 291
Figure 128. Correspondence of ratings for ‘Durability’. 292
Figure 129. Indicator model and ratings for ‘Survivability’. 293
Figure 130. Boxplots for ratings of ‘Survivable’. 294
Figure 131. Correspondence of ratings for ‘Survivable’.....295
Figure 132. Indicator model and ratings for ‘Dependability’. 296
Figure 133. Boxplots for ratings of ‘Dependable’. 297
Figure 134. Correspondence of ratings for ‘Dependable’.....298
Figure 135. Post facto indicator model and ratings for ‘Dependability’.....299
Figure 136. Correspondence of ratings for ‘Dependable’ with post facto indicator model. 299
Figure 137. Indicator model and ratings for ‘Reliability’. 300
Figure 138. Boxplots for ratings of ‘Reliable’. 301
Figure 139. Correspondence of ratings for ‘Reliability’. 301
Figure 140. Reliability rating vs MTBF – linear relationship. 302
Figure 141. Reliability rating vs MTBF – power law relationship. 302
Figure 142. Indicator model and ratings for ‘Supportability’. 303
Figure 143. Boxplots for ratings of ‘Supportable’. 304
Figure 144. Correspondence of ratings for ‘Supportability’. 305
Figure 145. Indicator model and ratings for ‘Maintainability’. 306
Figure 146. Boxplots for ratings of ‘Maintainable’. 307

Figure 147. Correspondence of ratings for ‘Maintainability’ 308
Figure 148. Indicator model and ratings for ‘Compliance’. 309
Figure 149. Boxplots for ratings of ‘Compliance’. 310
Figure 150. Correspondence of ratings for ‘Compliance’. 311
Figure 151. Judges’ compliance rating vs compliant requirements. 312
Figure 152. Judges’ compliance rating vs number of compliant requirements. 312
Figure 153. Judges’ compliance rating vs full and partial compliant requirements. 313
Figure 154. Indicator model and ratings for ‘Usability’. 314
Figure 155. Boxplots for ratings of ‘Usability’. 315
Figure 156. Correspondence of ratings for ‘Usability’. 316
Figure 157. Boxplots for ratings of ‘Security’. 318
Figure 158. Correspondence of ratings for ‘Security’. 318
Figure 159. Boxplots for ratings of ‘Technical Complexity’. 320
Figure 160. Indicator model and ratings for ‘High Technology’. 321
Figure 161. Boxplots for ratings of ‘High Technology’. 322
Figure 162. Correspondence of ratings for ‘High Technology’. 323
Figure 163. Number of coded extracts – nature of technical integrity measurement. 332
Figure 164. Number of coded extracts – nature of technical integrity measurement. 334
Figure 165. Number of coded extracts – judgment approach with no objective data. 338
Figure 166. Number of coded extracts – views on obtaining better ratings. 341
Figure 167. Influence diagram of the aggregate technical integrity and quality attributes of a complex engineered system. 351
Figure 168. SEM model of technical integrity concepts and indicators. 354
Figure 169. Scree plot of 18 variables. 358
Figure 170. Component plot: 2 components,18 variables, Oblimin rotation. 360
Figure 171. Component plot: 2 components,16 variables, Oblimin rotation. 364
Figure 172. Component plot: 2 components,15 variables, Oblimin rotation. 367
Figure 173. Correspondence of variables to Firesmith’s external quality characteristics. 371
Figure 174. SEM model of amended technical integrity concepts and indicators. 375
Figure 175. Reduced SEM model of technical integrity concepts and indicators. 379
Figure 176. Theory of technical integrity comprising explicit and implicit components. 380
Figure 177. Measurement construct for technical integrity from explicit and implicit components. 384
Figure 178. Values of technical integrity for the three example systems using four measurement instruments. 388
Figure 179. Decision model for best measurement instrument. 394
Figure 180. Results of decision model for best measurement instrument. 394
Figure 181. Sensitivity of best measurement instrument result to cost. 395

Figure 182. Results of decision model for best measurement instrument, cost not discriminated. 396
Figure 183. Plot of technical integrity on axes of capability and dependability. 397
Figure 184. Plot of technical integrity on axes of implicit and explicit technical integrity. 398

List of Tables

Table 1. Rationale for regulating; adapted from (Baldwin and Cave, 1999). 15
Table 2. Provisions of technical regulatory framework and known support for contribution to technical integrity aims. 34
Table 3. Types of emergent properties (Honour, 2007). 45
Table 4. Characteristics of the levels of measurement. 74
Table 5. Hierarchy of Information. 83
Table 6. Tailoring of scale development guidelines for research project. 84
Table 7. Research project phases and aims. 85
Table 8. Research project phases and research methods used. 89
Table 9. Categorisation of system emergent properties and the degree to which they are addressed by technical regulation. 98
Table 10. Candidate concepts and sources of data for assessing elements of technical integrity. 100
Table 11. Technical goals of the ontology associated with quality attributes of a system. 113
Table 12. Postulated indicators of technical integrity. 115
Table 13. Frequency of ratings of the level of agreement of 'ordinal scale meaningful' 131
Table 14. Frequency of support for hypothesis 'Technical integrity can be meaningfully measured on an ordinal scale' 131
Table 15. Frequency of ratings of the level of agreement of 'measurement by a single numeric value'. 132
Table 16. Frequency of support for hypothesis 'technical integrity can be measured by a single numeric value' 133
Table 17. Frequency of ratings of the level of agreement of 'measurement on a ratio scale meaningful' 133
Table 18. Frequency of support for hypothesis 'technical integrity of a system can be measured on a ratio scale' 134
Table 19. Frequency of ratings of the level of agreement of capability measured on a scale of 0 to 1. 135
Table 20. Frequency of support for hypothesis 'Capability of a complex engineered system can be measured on a scale of 0 to 1'. 136
Table 21. Frequency of ratings of the level of agreement of capability – more than specified requirement is valued. 136
Table 22. Frequency of support for hypothesis 'Capability of a system in excess of specified requirements is valued' 137
Table 23. Frequency of ratings for a better measure of capability. 138
Table 24. Frequency of ratings of the level of agreement of dependability measured on a scale of 0 to 1. 139
Table 25. Frequency of support for hypothesis 'Dependability of a complex engineered system can be measured on a scale of 0 to 1'. 140
Table 26. Frequency of ratings of the level of agreement of dependability – better than risk acceptance thresholds is of increased value. 140

Table 27. Frequency of support for hypothesis ‘dependability – better than risk acceptability threshold of stakeholders is valued’. 141
Table 28. Frequency of ratings for a better measure of dependability. 142
Table 29. Frequency of ratings of the level of agreement of a technical integrity conjoint measure is meaningful. 143
Table 30. Frequency of support for hypothesis ‘technical integrity can be measured by a conjoint measurement from the components capability and dependability’. 144
Table 31. Frequency of ratings of the level of agreement of a technical integrity conjoint measure is meaningful for similar system comparison. 145
Table 32. Frequency of support for hypothesis ‘A conjoint measurement of technical integrity allows for comparison of similar systems’ 145
Table 33. Frequency of ratings of the level of agreement of a technical integrity conjoint measure is meaningful for dissimilar system comparison. 146
Table 34. Frequency of support for hypothesis ‘A conjoint measurement of technical integrity allows for comparison of dissimilar systems’. 147
Table 35. Summary of quantitative results. 148
Table 36. Code table – better capability measurement. 151
Table 37. Code table – better dependability measurement. 157
Table 38. Code table – reflections on the measurement of technical integrity. 164
Table 39. Descriptive statistics for ratings of the relative importance of capability versus dependability for each system domain. 181
Table 40. Hypothesis test summaries for test of sample medians differing from a median of 4, 5 or 6. 192
Table 41. Code table: System Domain or Viewpoint influences on judgment of capability versus dependability 202
Table 42. Code table: system service influences on judgment of capability versus dependability. 209
Table 43. Capability versus dependability importance judgements: comparison of quantitative and qualitative data. 215
Table 44. Terminology for information, systems, CAT and models 221
Table 45. Technical parameter provided in the example systems. 229
Table 46. Hypothesised correlation of technical parameters with indicators of technical integrity 236
Table 47. Descriptive statistics for technical integrity ratings. 251
Table 48. Reliability of ratings for system quality attributes/judgment dimensions. 257
Table 49. Technical integrity scale: item-total statistics. 259
Table 50. Technical integrity scale: scale alpha if items deleted. 260
Table 51. Descriptive statistics for judgments of ‘Safety’. 263
Table 52. Descriptive statistics for ratings of ‘Environment impact’. 266
Table 53. Descriptive statistics for ratings of ‘Impact to systems’. 269
Table 54. Descriptive statistics for ratings of ‘Fit for service’ 272
Table 55. Descriptive statistics for ratings of ‘Capability’. 276

Table 56. Descriptive statistics for ratings of ‘Effective’. 280
Table 57. Descriptive statistics for ratings of ‘Efficient’. 283
Table 58. Descriptive statistics for ratings of ‘Quality’. 286
Table 59. Descriptive statistics for ratings of ‘Flexible’. 288
Table 60. Descriptive statistics for ratings of ‘Durable’. 291
Table 61. Descriptive statistics for ratings of ‘Survivable’. 294
Table 62. Descriptive statistics for ratings of ‘Dependable’. 297
Table 63. Descriptive statistics for ratings of ‘Reliable’. 300
Table 64. Descriptive statistics for ratings of ‘Supportable’. 304
Table 65. Descriptive statistics for ratings of ‘Maintainable’. 307
Table 66. Descriptive statistics for ratings of ‘Compliance’. 310
Table 67. Descriptive statistics for ratings of ‘Usability’. 315
Table 68. Descriptive statistics for ratings of ‘Security’. 317
Table 69. Descriptive statistics for ratings of ‘Technical complexity’. 319
Table 70. Descriptive statistics for ratings of ‘High Technology’. 322
Table 71. Summary of correspondence of judgments with indicator models. 324
Table 72. Indicator correspondence versus judgment reliability. 325
Table 73. Indicator correspondence, judgment reliability and selection of measurement approach. 327
Table 74. Code table – nature of technical integrity measurement 331
Table 75. Code table – use of objective data. 333
Table 76. Code table – judgment approach with no objective data. 337
Table 77. Code table – views on obtaining better ratings. 341
Table 78. Standardised regression weights - theorised technical integrity construct. 355
Table 79. Model-fit parameters – theorised technical integrity construct. 355
Table 80. Correlation matrix – 18 variables. 356
Table 81. Components and explained variance: 18 variables. 357
Table 82. Parallel analysis results. 358
Table 83. Pattern and structure matrices: 2 components,18 variables, Oblimin rotation. 360
Table 84. Communalities matrix: 2 components,18 variables, Oblimin rotation. 361
Table 85. Correlation matrix: 16 variables. 362
Table 86. Components and explained variance: 16 variables. 363
Table 87. Pattern and structure matrices: 2 components,16 variables, Oblimin rotation. 364
Table 88. Correlation matrix: 15 variables. 365
Table 89. Components and explained variance: 15 variables. 366
Table 90. Pattern and structure matrices: 2 components,15 variables, Oblimin rotation. 367
Table 91. Variable loading on extracted components. 368

Table 92. Variable loading on extracted components compared to theorised model of technical integrity. 369
Table 93. Correspondence of variables associated with benefit and sacrifice. 373
Table 94. Variables associated with components of technical integrity. 373
Table 95. Interpretations of the nature of components....374
Table 96. Standardised regression weights - amended technical integrity construct. 376
Table 97. Model-fit parameters – amended technical integrity construct. 377
Table 98. Model-fit parameters – amended technical integrity construct (effectiveness removed). 377
Table 99. Model-fit parameters – amended technical integrity construct (effectiveness and survivability removed). 378
Table 100. Standardised regression weights - reduced technical integrity construct. 379
Table 101. Technical integrity measures using the CAT. 381
Table 102. Technical integrity ratings from scale of indicators. 382
Table 103. Technical integrity ratings from a conjoint of capability and dependability. 383
Table 104. Explicit technical integrity ratings from scale of indicators. 385
Table 105. Explicit technical integrity ratings from additive model of indicators....386
Table 106. Implicit technical integrity ratings from scale of indicators. 386
Table 107. Technical integrity conjoint rating from explicit and implicit components. 386
Table 108. “Goodness” of measurement instruments. 390
Table 109. Relative cost of measurement instrument. 393
Table 110. Sensitivity of leading alternative, Instrument 1, to weights. 395

Glossary

Symbols

A_i	Inherent Availability
A_o	Operational Availability
C	Capability
D	Dependability
TI	Technical integrity
α	Cronbach's alpha
r_S	Spearman's rho

Acronyms

ADF	Australian Defence Force
AHP	Analytic Hierarchy Process
ALARP	As Low As Reasonably Practicable
AMOS	Analysis of MOment Structures
ANSI	American National Standards Institute
BBN	Bayesian Belief Network
BIT	Built In Test
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance, Reconnaissance
CAT	Consensual Assessment Technique
CDP	Criterium Decision Plus
CM	Configuration Management
CMMI	Capability Maturity Model Integrated
COI	Critical Operational Issue
COTS	Commercial Off The Shelf
CPM	Composite Performance Measure
CR	Critical Ratio
CVR	Content Validity Ratio
DAR	Design Acceptance Representative
DATUM	Dependability Assessment of safety critical systems Through the Unification of Measurable evidence
DEOP	Defence Explosive Ordnance Publication

DMO	Defence Materiel Organisation
EIA	Electronic Industries Association
EO	Explosive Ordnance
EOC	Explosive Ordnance Certificate
EMC	Electro Magnetic Compatibility
FAR	Federal Aviation Regulation
FMECA	Failure Mode Effects and Criticality Analysis
FSA	Formal Safety Assessment
FTA	Fault Tree Analysis
HAZMAT	Hazardous Material
HMAS	Her Majesty's Australian Ship
HMI	Human Machine Interface
ICT	Information and Communication Technologies
IDA	Information Decision Action
IEC	International Electrotechnical Commission
IEEE	The Institute of Electrical and Electronics Engineers
ILS	Integrated Logistics Support
IMO	International Maritime Organisation
INCOSE	The International Council on Systems Engineering
INES	International Nuclear Event Scale
IPT	Integrated Product Team
ISO	International Organization for Standardisation
IT	Information Technology
LCV	Life Cycle Value
MARPOL	International Convention for the Prevention of Pollution From Ships
MAUT	Multi Attribute Utility Theory
MIL-STD	Military Standard
MOE	Measure Of Effectiveness
MOP	Measure Of Performance
MOS	Measure of Supportability
MOTS	Military Off The Shelf
MTBF	Mean Time Between Failure

MTFF	Mean Time to First Failure
MTTR	Mean Time To Repair
NASA	National Aeronautics and Space Administration
NAVSEA	Naval Sea Systems Command
NTRM	Navy Technical Requirements Manual
NTRS	Navy Technical Regulatory System
OCD	Operational Concept Description
OEM	Original Equipment Manufacturer
PC	Personal Computer
PCA	Principal Components Analysis
PDF	Probability Density Function
PRA	Probabilistic Risk Assessment
RAN	Royal Australian Navy
RMA	Reliability, Maintainability and Availability
SE	Systems Engineering
SEM	Structured Equation Modelling
SE-ROI	Systems Engineering – Return On Investment
SME	Subject Matter Expert
SOLAS	Safety of Life at Sea
SoS	System of Systems
SPSS	Statistical Package for the Social Sciences
SUS	System Usability Scale
TAMM	Technical Airworthiness Management Manual
TAR	Technical Airworthiness Regulator
TI	Technical Integrity
TPM	Technical Performance Measure
TRA	Technical Regulatory Authority
TRAMM-L	Technical Regulation of ADF Materiel Manual - Land
TRF	Technical Regulatory Framework
TRI	Technical Risk Index
UK	United Kingdom
USA	United States of America
USAF	United States Air Force

VFM	Value For Money

Definitions of terms

Certification	The act of issuing a certificate that provides assurance that an entity, including product, service or organisation, complies with a stated specification, standard or other requirement. (Australian Defence Force, 2004b)
Concept	*As part of the information model used in this thesis:* Building blocks of the theory which are usually abstract and cannot be directly measured. (Walliman, 2010, p. 65-69)
Conjoint measure	A combination of component properties which establishes an ordering of the desired property based on measurable component properties (e.g., the establishment of a "comfort" measurement by combining temperature, wind speed and relative humidity measurements).
Content validity	An assurance that the measure includes an adequate and representative set of items that tap the concept. A panel of judges can attest to the content validity of the instrument. (Sekaran, 2003, p. 206)
Cost-benefit analysis	A process for calculating and comparing benefits and costs of an undertaking for two purposes: (1) to determine if it is a sound investment (justification/feasibility), (2) to see how it compares with alternate undertakings (ranking/priority assignment). It involves comparing the total expected cost of the undertaking against the total expected benefits, to see whether the benefits outweigh the costs, and by how much. Benefits and costs are expressed in money terms, and are usually adjusted for the time value of money.
Construct validity	An assurance that the results obtained by the use of a measurement instrument fit the theories around which it is designed. It is assessed through convergent and discriminant validity. (Sekaran, 2003, p. 207)

Convergent validity	An assurance that the scores obtained with two different instruments measuring the same concept are highly correlated. (Sekaran, 2003, p. 206)
Criterion-related validity	An assurance that a measure differentiates individuals on a criterion it is expected to predict. (Sekaran, 2003, p. 206)
Derived measure	A measure formed by establishing a relation between two or more things that can be directly measured (e.g., density which is established from the direct measurement of volume and mass).
Descriptive research	A research approach that is usually quantitative in nature and is characterised by a deductive logic approach' 'for the purpose of ascertaining and describing the characteristics of the variables of interest in a given problem situation'(Cropley and Harris, 2004).
Direct measurement	An instrument that can be directly used to estimate the characteristic of an observable, physical object. One of the two main forms of measurement defined by (Sydenham, 1982, p. 12-17) which includes extensive measurement, matching scales and ranking scales.
Discriminant validity	An assurance that, based on theory, two variables are predicted to be uncorrelated, and the scores obtained by measuring them are indeed empirically found to be so (Sekaran, 2003, p. 207).
Exploratory research	A research approach that has a 'typical goal of ... building of a tentative explanation of the problem situation' and 'is usually qualitative in nature and is characterized by an inductive logic approach' (Cropley and Harris, 2004).
Extensive measurement	A measurement where something like a physical attribute of an object can be directly observed (e.g., a linear measurement).
Face validity	Indicates that the items that are intended to measure a concept on the face of it look like they measure the concept. A minimal form of content validity. (Sekaran, 2003, p. 206)
Fitness for service	'The materiel's ability to satisfy operational requirements, and is hence a subset of technical integrity.' (Australian Defence Force, 2010)
Heuristic	'a guide', 'an abstraction of experience'. (Maier and Rechtin, 2009, p. 20)

Index	A measurement instrument that is a collection of items, without a common cause, combined into a composite score representing the composite effect for a theorised variable (DeVellis, 2011, p. 12).
Indicator	*As part of the information model used in this thesis:* Phenomena which point to the existence of the concepts. (Walliman, 2010, p. 65-69)
Indirect measurement	Where an instrument cannot directly be used to estimate the measurement object's characteristics. The characteristics may be derived from direct measurements or may be quite subjective. Indirect measurement is one of the two main forms of measurement defined by (Sydenham, 1982, p. 12-17) and includes derived, associative, multi-dimensional and conjoint measurement.
Measurement	'The process of empirical, objective assignment of numbers to the properties of objects and events of the real world in such a way as to describe them'. (Sydenham, 1982, p. 2)
Measurement construct	A relationship between variables. At its simplest form it is the relationship between the property to be measured and the mathematical representation to be used as the measure.
Measurement instrument	A practical means by which a number can be obtained.
Multi-dimensional measurement	A collection of component vectors without combination (e.g., a vector that defines the lengths of the 3 sides of a triangle).
Ranking scale	A measurement scale where a measurands is compared to members of a ranked standard set (e.g. Moh's scale of hardness of minerals).
Reliability	The extent to which measurements can be depended upon to secure consistent results in repeated applications. It is an indication of the stability and consistency with which an instrument measures a concept. (Sekaran, 2003, p. 203)
Safety	Freedom from those conditions that can cause death, injury, occupational illness, or damage to or loss of equipment or property, or damage to the environment. (MIL-STD-882C Department of Defense, 1993)

Scale (of measurement)	A categorisation of forms of measurement. There are four categories of scales: nominal, ordinal, interval and ratio (Leedy and Ormond, 2005, p. 25-26). Scales may also be termed as levels of measurement (Allen and Yen, 1979, p. 6-8)
Scale	A measurement instrument that is a collection of items combined into a composite score and intended to reveal levels of theoretical variables not readily observable by direct means (DeVellis, 2011, p. 11)
Standard	A description of a material, product, doctrine or process meant for repeated application by many users. Standards are also a repository of lessons learnt to ensure that mistakes are not repeated and to simplify design of products or systems through the use of known proven building blocks.
System safety	The application of engineering and management principles, criteria and techniques to optimise all aspects of safety within the constraints of operational effectiveness, time and cost throughout all phases of the system life cycle. (MIL-STD-882C Department of Defense, 1993)
Technical integrity	*As defined in ADF technical regulation:* 'An item's fitness for service, safety and compliance with regulations for environmental protection.' (Australian Defence Force, 2010)
Technical regulation	The means employed to assure technical integrity or other technical objectives for a system or product. Regulation may be such that it compels the developers to comply by force of law or other compliance measures.
Theory	*As part of the information model used in this thesis:* Abstract statements that make claims about the world and how it works. Research problems are usually stated at a theoretical level. (Walliman, 2010, p. 65-69)
Validity	'The extent to which an instrument measures what it actually intended to measure.' (Leedy and Ormond, 2005, p. 94)
Value	*As part of the information model used in this thesis:* Actual units of measurement of the variables. These are data in their most concrete form. (Walliman, 2010, p. 65-69)

Variable	*As part of the information model used in this thesis:* Components of the indicators which can be measured. (Walliman, 2010, p. 65-69)
Unprecedented	A specific end product that is not known *a priori*, or the enterprise has limited experience in developing this type of system. (EIA-632 EIA, 1998a)

Summary

Technical integrity is the term used in the Australian Defence Force (ADF) to encompass a system's fitness for service, its safety and its compliance with regulations for environmental protection. Technical regulations in the ADF are explicitly aimed at assuring that technical integrity is achieved during the development of systems and is maintained during their operational life. It would be expected then, in some average sense at least, that the technical integrity of systems developed in compliance with the provisions of technical regulation would be greater than those that are not. This consideration lead to the research problem that initiated this thesis, namely: *How effective are the provisions of technical regulations in assuring the technical integrity of complex engineered systems?*

Establishing a way to measure the technical integrity of a complex engineered system is a necessary prerequisite for objective assessment of the effectiveness of the provisions of technical regulation. The key research question pursued by this thesis was therefore: *How can the technical integrity of a complex engineered system be measured?*

Significant resources are being expended in the technical regulation and associated certification requirements for complex systems with the aim of reducing the consequence and probability of occurrence of mishaps. Technical regulation in the ADF is largely based on the application of heuristics of what technical processes should be applied but with little, if any, empirical evidence for the effectiveness or efficiency of these. The key issue in establishing such evidence is the need for a practical and reliable measure of technical integrity. With a measure it will be possible to establish the benefit (i.e., the amount of additional technical integrity) obtained by the application of technical regulation, or the degree thereof.

The validity and practicality of measuring abstract properties (like technical integrity) of a complex engineered system has no clear consensus of support in the literature. No measurement approach has previously been proposed for technical integrity and technical integrity has not previously been objectively defined such that a particular measurement approach can be seen to be clearly appropriate.

A review of measurement approaches to comparable problems across several domains identified practical approaches including the Consensual Assessment Technique (CAT),

conjoint measurement, scales and indices, and the use of proxy measurements for some variables thought to be indicators of technical integrity. These formed the candidates for investigation in the quest for a practical measurement instrument.

A phased approach that began as *exploratory* research and progressed to *descriptive* research was adopted. In the exploratory research phase a theory, concepts and a measurement construct for technical integrity were developed. In the descriptive research phase, survey research gathered judgments, ratings and qualitative responses that supported the assessment of the reliability and validity of candidate measurement approaches for technical integrity.

A theory of technical integrity consisting of two orthogonal concepts, termed *capability* and *dependability*, arising from the intended, desired emergent properties and the unintended, undesired emergent properties of a complex system was developed. This formed the basis for a conjoint measurement construct for technical integrity for which content validity was established. An associated result was that the relative importance of capability and dependability in the conjoint construct fell into three categories: (i) 'Military systems' and 'Civil ICT systems' where capability and dependability were of equal importance; (ii) 'Important Civil systems' where dependability was considered more important than capability; and (iii) 'Critical Civil systems' where dependability was considered much more important than capability.

Results of an application of the CAT on a set of representative systems indicated that indirect estimation by expert judgment of the technical integrity of systems was reliable. Reliability of judgments for many other theorised indicators of technical integrity was even higher. It was found that indicators that are more subjective in definition, and for which objective data is often not readily available, showed reliability results that were less than what is normally acceptable for use.

Confirmatory and exploratory factor analysis of the same judgment data found support for a theory of technical integrity comprising two components. However, the two components and the indicators that most significantly loaded upon each of them did not completely correspond with the concepts of capability and dependability or with the predicted correlation of indicators with these concepts. A modified theory of technical integrity, grounded in data, was then proposed cognisant of these results. The theory was that technical integrity combines the concepts of: *Explicit technical integrity*,

embodying aspects of a complex engineered system that are usually explicitly required and for which the level of achievement may be objectively verified; and *Implicit technical integrity*, embodying aspects of a complex engineered system that are most often not the subject of explicit technical requirements but are implicitly assumed to be necessary for a critical system. These aspects are often not readily assessable except via an experienced assessor, by logical argument or by actual experience of the system fielded in real, operational conditions.

Using both the developed concepts of technical integrity and the available ratings data for the representative systems, four viable measurement constructs were identified and assessed:

(1) Technical integrity direct measurement using the CAT.
(2) Scale from theorised indicators of technical integrity.
(3) Conjoint construct of capability and dependability concepts.
(4) Conjoint construct of explicit and implicit technical integrity concepts.

The "best" instrument from consideration of reliability, validity and practicality was the direct measurement of technical integrity using the CAT.

Clear contributions to the research problem and area were made by this research thesis. Paraphrasing Lord Kelvin, technical integrity can now be expressed in numbers and in doing so we 'know something about it' and have advanced the understanding of technical integrity 'to the stage of science'. Specifically, to the research area of technical regulation and its effectiveness, a measurement instrument for technical integrity will allow future research to assess the effectiveness of technical regulation on complex engineered systems, allowing for justifiable improvement of regulations thereby improving the technical integrity of systems.

More generally, the CAT (originally developed for applications in the social sciences like *creativity*) can be a reliable measurement approach for other abstract or latent characteristics of an engineered system, providing a basis for evaluating the effectiveness of engineering methods applied to their development leading to improvements in methods and ultimately to the systems produced by these methods.

Declaration

This thesis presents work carried out by myself and does not incorporate without acknowledgment any material previously submitted for a degree or diploma in any university; to the best of my knowledge it does not contain any materials previously published or written by another person except where due reference is made in the text; and all substantive contributions by others to the work presented, including jointly authored publications, are clearly acknowledged.

_____

Michael Thomas EDWARDS

Acknowledgments

My overwhelming debt of gratitude is to my wife Natalie and our children, Maeve and Ciaran. Without their continual encouragement and support this thesis would likely never have been completed.

The regular interest, guidance and exchange of ideas and constructive criticisms from my academic supervisors Associate Professor David Cropley and Mr John Thornton were always greatly appreciated during the course of this research.

The time and diligence of well over one hundred anonymous survey respondents who each provided hours of considered thought in response to my surveys is also acknowledged. You know who you are and my hope is that the results of this research will assist you in some way in your professional practice.

1 Introduction

1.1 Background to the research

A complex engineered system is defined to be a system that has been intentionally engineered from numerous components: technological (hardware and software) as well as human, each of which has several interconnections and is interdependent, yielding a system with both intended and unintended emergent behaviour.

A prevailing view is that the failure of a complex engineered system is unavoidable (Modarres, 2006) and that risk of undesired events is inherent in engineered systems (Lee and McCormick, 2012b). A complex engineered system must contend with both internal causes of failure (e.g., design errors, failure of components, human errors in operation and maintenance) and external causes of failure (e.g., unanticipated malicious acts or states of the environment). Acquirers and users of these systems and the general public are becoming more demanding for high levels of integrity of products and services, but engineering methods required to assure this integrity, like system safety engineering, must be timely and efficient if the developers of systems are to survive amid economic competition (Stephenson, 1991).

Notwithstanding the views of inevitability of system failures, significant resources are being expended in the regulation and certification of complex systems in Australia in order to reduce both the consequences and probabilities of occurrence of mishaps resulting from system failure. While this type of expenditure can be seen to have a positive correlation with outcomes like reduced accident rates in aviation systems, it is unclear if significant benefits have or will accrue from the technical regulation of complex defence systems.

Technical integrity is the term used in the Australian Defence Force (ADF) to encompass a system's fitness for service, its safety and its compliance with regulations

for environmental protection (Australian Defence Force, 2010). Technical regulations in the ADF are explicitly aimed at assuring that technical integrity is achieved during the development of systems and is maintained during their operational life. It would be expected then, in some average sense at least, that the technical integrity of systems developed in compliance with the provisions of technical regulation would be greater than those that are not.

1.2 Research problem

Consideration of the intended impact of technical regulation leads to the research problem that has initiated and subsequently guides this thesis, namely:

How effective are the provisions of technical regulations in assuring the technical integrity of complex engineered systems?

Establishing a way to measure the technical integrity of a complex engineered system is a necessary prerequisite for objective assessment of the effectiveness of the provisions of technical regulation. The key research question pursued in this thesis is therefore:

How can the technical integrity of a complex engineered system be measured?

1.3 Justification for the research

Current and future technological systems are steadily increasing in complexity, both in terms of technology employed and the interactions they have with users and other stakeholders. Defence systems in particular are increasingly required to not only introduce higher organic capabilities within the system itself, but also to integrate into systems-of-systems to enable capabilities with higher performance across a wider spectrum of domains and environments. At the same time, the need for these systems to achieve ever-increasing levels of technical integrity is paramount. Systems are expensive to develop and sustain, therefore there are less of them and each one that is realised must achieve higher and higher levels of availability and dependability. The stakeholders of the system, not least the general public (and consequently the political leaders of the nation), are less and less willing to accept mishaps, be they accidents to personnel, loss or damage of equipment or impacts to the natural environment.

Technical regulation seeks to provide more assurance to the acquirers and ultimately the users of engineered systems than just a reliance on system developers to unilaterally

apply the "best" engineering and management processes. "Best" for a developer does not necessarily lead to a set of practices that will yield the required technical integrity. Engineering practices, like Systems Engineering for instance, do not alone assure technical integrity goals are a primary focus or that the processes will be executed by competent professionals with explicit responsibility for their design decisions and the system outcomes. Technical regulation therefore generally attempts to assure that development is conducted to approved standards, by competent individuals in an organisation that is authorised to do so. There is also an explicit process to achieve certification; at a minimum by the developer formally attesting to the completeness and quality of its own work, but usually also encompassing some independent validation of the developer's claims of compliance with technical standards and processes.

At this time, the intended positive impact of technical regulation on the technical integrity of a complex engineered system remains largely based on the application of heuristics of what mechanisms should be applied but with little if any empirical support for the effective or efficiency of these. One of the key issues in establishing such support is the need to establish some sort of measure for technical integrity. No measurement approach has previously been proposed for technical integrity and technical integrity has not previously been objectively defined such that a particular measurement approach can be readily seen to be appropriate. If a measure were established it would be possible to demonstrate the benefit (i.e., the amount of additional technical integrity) obtained by the application of technical regulation, or the degree thereof. The research is thus justified in focusing on determination of a construct and a measurement instrument for technical integrity.

1.4 Methodology

Theory for technical integrity and technical regulation is not well developed and concepts are not well defined, at least not in a way to enable practical measurement. Similarly, the indicators and variables that may contribute to a measure for technical integrity have not previously been identified. Some of these variables may need to be subjectively assessed as a practical measurement approach.

Consequently, a phased approach that begins as *exploratory* research and progresses to *descriptive* research has been adopted for this thesis. The approach is designed to:

- In the exploratory research phase:
  - Develop theory, concepts and a measurement construct for technical integrity.
- In the descriptive research phases:
  - Assess the content validity of the construct for technical integrity,
  - Assess the feasibility of direct estimation of technical integrity by expert judgement,
  - Establish parameters for the measurement instrument, and
  - Establish the variables to be included in candidate measurement instruments and establish the reliability and validity of those measurement instruments.

The interrelationship of these study phases is illustrated in Figure 1. It shows how the research question is clearly defined as an outcome of the literature survey which also provides the base to ground the concept development phase of the research. The output of concept development, a proposed theory and concepts for a technical integrity measurement, enables the construction of survey instruments which in turn provides quantitative and qualitative data for validation and refinement of the theory and the measurement approach. The outcome of the phases concludes with a recommended measurement instrument for technical integrity that is important to support future research in the area of technical regulation and the engineering of complex systems.

The research methods, being the practical techniques used in the research project, include:

- Evaluation, particularly a systems analysis approach to develop theory and concepts.
- Mixed method designs, where quantitative and qualitative data is collected via questionnaires and analysed to be complementary, is used to:
  - Establish expert support for the validity for the proposed measurement construct
  - Establish variables and combinatorial weightings for the measurement instrument and to validate the resulting measurement instrument.

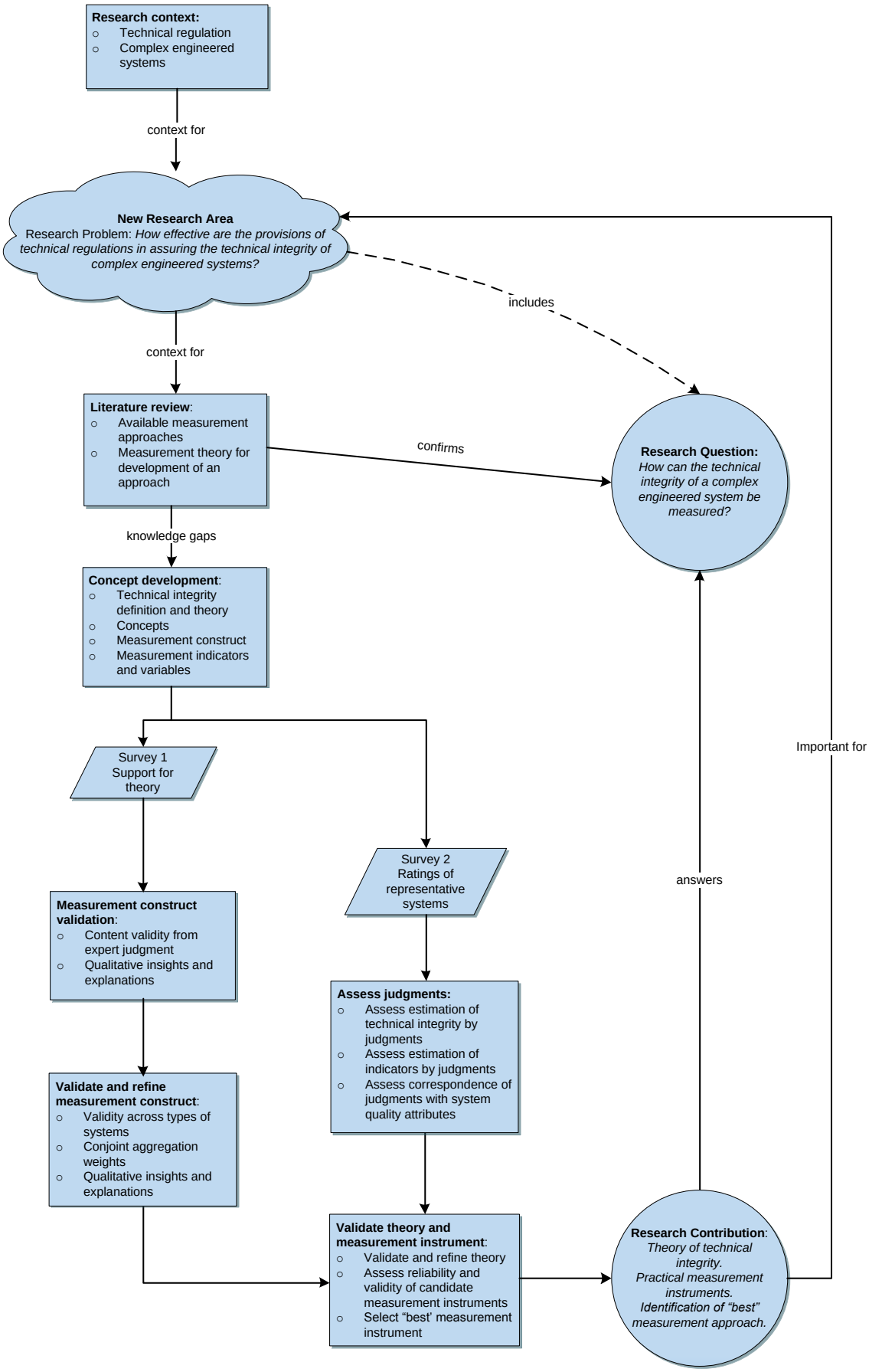


Figure 1. Research project's phased approach.

1.5 Outline of the thesis

This thesis is structured in the following chapters and appendices:

- Chapter 1 – this chapter introduces the research area, research problem, research question, methodology, assumptions and delimitations and provides a preview of the results and conclusions.
- Chapter 2 – The research context: provides a broad background review of literature to establish the context for the research area and the research problem.
- Chapter 3 - Literature survey; reviews three main areas of technical regulation, complex engineered systems and measurement that provide context for and establish knowledge gaps that guide the formulation of the research problem and questions.
- Chapter 4 - Methodology; describes and justifies the phased approach and methods adopted for this research project.
- Chapter 5 - Concept development; inductively develops theory, concepts and indicators for technical integrity and proposes a measurement construct and approach to developing a measurement instrument for technical integrity.
- Chapter 6 - Content validity of the technical integrity measurement construct; presents the results of the phase of research conducted to establish the content validity of the proposed measurement construct.
- Chapter 7 - Conjoint measurement aggregation weights; builds upon the finding of Chapter 6 to determine the relative importance of the concepts of capability and dependability for use in the aggregation by the conjoint measurement construct of technical integrity.
- Chapter 8 - Assessment of judgments for estimation of technical integrity; assesses the viability of using the consensual assessment technique for the direct estimation of technical integrity and related indicators. It compares judgment data with measures of indicators that are constructed from the values of quality attributes of a system, in order to identify pragmatic methods to estimate these indicators.
- Chapter 9 - Technical integrity theory and measurement instrument; validates and refines the theory of technical integrity and assesses the reliability, validity and practicality of the candidate measurement approaches. The outcome is a recommended measurement instrument for technical integrity
- Chapter 10 – Summary, conclusions and implications; summarises the research undertaken, summarises the results and provides conclusions against the research

question and the research problem. Implications for existing theory, policy and practice and for future research are discussed.

Appendices A through E provide additional data directly referenced by the thesis:

- Appendix A - Survey instrument: validity of the measurement construct.
- Appendix B - Survey instrument: measurement by consensual assessment technique.
- Appendix C - Validity of technical integrity measurement construct: full text extracts referenced in qualitative analysis.
- Appendix D - Conjoint measurement aggregation: full text extracts referenced in qualitative analysis.
- Appendix E - Judgment of technical integrity and associated indicators: full text extracts referenced in qualitative analysis.

Refereed papers produced during the course of this research are also provided as appendices F through L. These papers presented an opportunity for peer review of developing elements of this thesis progressively over the course of the research. The appendices are:

- Appendix F - The impact of technical regulation on the technical integrity of complex engineered systems.
- Appendix G - An ontology of technical regulation.
- Appendix H - Is systems engineering necessary for the achievement of technical integrity?
- Appendix I - A framework for the evaluation of verification evidence supporting the assessment of environmental qualification.
- Appendix J – What is technical integrity and how to measure it?
- Appendix K - The content validity of a technical integrity measurement construct.
- Appendix L - The consensual assessment technique for the measurement of technical integrity.

1.6 Definitions

The key terms used in this thesis are defined below. A glossary is included at page xx as a comprehensive reference for all terms and acronyms used within this thesis.

Fitness for service	'The materiel's ability to satisfy operational requirements, and is hence a subset of technical integrity.' (Australian Defence Force, 2010)
Measurement	'The process of empirical, objective assignment of numbers to the properties of objects and events of the real world in such a way as to describe them'. (Sydenham, 1982, p. 2)
Reliability	The extent to which measurements can be depended upon to secure consistent results in repeated applications. It is an indication of the stability and consistency with which an instrument measures a concept. (Sekaran, 2003, p. 203)
Safety	Freedom from those conditions that can cause death, injury, occupational illness, or damage to or loss of equipment or property, or damage to the environment. (MIL-STD-882C Department of Defense, 1993)
Technical integrity	*As defined in ADF technical regulation:* 'An item's fitness for service, safety and compliance with regulations for environmental protection.' (Australian Defence Force, 2010)
Technical regulation	The means employed to assure technical integrity or other technical objectives for a system or product. Regulation may be such that it compels the developers to comply by force of law or other compliance measures.
Validity	'The extent to which an instrument measures what it actually intended to measure.' (Leedy and Ormond, 2005, p. 94)

1.7 Delimitations of scope and key assumptions

1.7.1 Delimitations

This research is primarily focused on the acquisition and development of complex engineered systems. While technical regulation is applicable to the entire system

lifecycle, including operation and maintenance, it is the provisions for design during the concept and development stages of a system, and if they are effective in assuring the technical integrity of a system that is of greatest interest since any improvements to these provisions is likely to have the highest impact on the technical integrity of systems in the future. Therefore, the research is delimited to a consideration of the enforcement of technical regulation during the concept and development phases with technical integrity outcomes manifest in the operation and support phase of the system life cycle. The research is not explicitly examining the maintenance of technical integrity during operation and support or disposal phases of the system life cycle, e.g., maintenance, supply support.

The research is primarily delimited to defence systems and the technical regulations of the ADF. However, it will look more widely at civil and foreign military regulation as a basis for comparison, but will not seek to draw conclusions beyond the ADF context. This delimitation is justified by practical aspects of access to technical information about systems and access to appropriate subject matter experts for survey research.

Included in the considerations of this research are complex engineered systems of the ADF (systems such as naval combat systems, military combat aircraft and their avionics systems, integrated soldier combat systems, and network centric warfare systems of system) but not non-complex materiel (relatively simpler military systems like non-combatant maritime platforms, personal weapons, logistical trucks and aircraft). This delimitation is justified by a focus on systems with significant emergent behaviour, an area where systems science lags the practice of systems engineering, and the effectiveness of many engineering processes to control the intended and unintended emergent behaviour is still maturing and the provisions of technical regulations will consequently likely need to change.

1.7.2 Key assumptions

It is assumed that systems developed for a range of application domains, employing differing architectural or design approaches, and utilising different technologies exhibit some common underlying property, a latent variable, which can be assessed as technical integrity. This assumption is consistent with that implicit in a range of technical regulations that impose similar provisions to assure the technical integrity of disparate types of systems.

The research is reliant on the expectation that there will be some convergence of opinion between diverse stakeholders (e.g. sponsors, regulators, developers, end-users, maintainers) that agrees to differentiate a system with higher technical integrity from one with lower technical integrity.

It is assumed that the imposition of technical regulation has a measureable effect on the technical integrity of a complex engineered system and the effect can be discerned from what could be other drivers of technical integrity (e.g., complexity, level of unprecedented design, aspiration for high performance for competitive advantage).

1.8 Conclusion

A theory of technical integrity, associated measurement constructs and instruments are derived, assessed and refined. An inductive approach posits a theory of technical integrity comprising two concepts or components, capability and dependability, with an associated conjoint measurement construct.

Expert ratings of technical integrity and related indicators of three example systems, allows for confirmatory and exploratory factor analysis that finds support for a theory of technical integrity comprising two components. However, the two components and the indicators that most significantly load upon each of them do not completely correspond with the concepts of capability and dependability or with the predicted correlation of indicators with these concepts. A modified theory of technical integrity, grounded in data, is then proposed cognisant of these results. The theory is that technical integrity combines the concepts of: *Explicit technical integrity*, embodying aspects of a complex engineered system that are usually explicitly required and for which the level of achievement may be objectively verified; and *Implicit technical integrity*, embodying aspects of a complex engineered system that are most often not the subject of explicit technical requirements but are implicitly assumed to be necessary for a critical system.

Using both the original and the amended theory of technical integrity, the previous findings, and the available ratings data for the three example systems, four viable measurement instruments are identified and assessed:

- (1) Technical integrity direct measurement using the CAT.
- (2) Scale from theorised indicators of technical integrity.
- (3) Conjoint construct of capability and dependability concepts.

- (4) Conjoint construct of explicit and implicit technical integrity concepts.

The “best” instrument from considerations of cost, reliability and validity is the direct measurement of technical integrity using the CAT.

Clear contributions to the research problem are made by this thesis, particularly the development of practical measurement instruments ready for use and further validation arising from use. The instruments allow the assessment of the technical integrity of a complex engineered system, thereby allowing the effectiveness of technical regulations to be assessed. Of broader application, is the contribution that has shown that use of the CAT can obtain reliable estimates of the subjective attributes of a complex engineered system. The CAT may therefore be used as an estimation approach in other research areas which require a straightforward and reliable method to estimate a subjective characteristic (latent variables) of engineered entities.

The research commenced with a focus on the regulatory policies and practices of the ADF and the conclusions provide some implications for consideration by regulators and practitioners. Regulations that govern the attainment of technical integrity include in their definition the system attributes of safety and impact to the environment. However, it has been found in this study that those attributes are not judged by practitioners to be strongly correlated with technical integrity. This implies either the intent of technical integrity is not clearly articulated or that they are disparate concepts that may require different regulatory approaches. It was also found in this study that in no instance does the consideration of the capability concept outweigh that of dependability. This is at odds with practice where disproportionate effort is aimed on capability outcomes of system developments while measurable goals for the dependability aspects of a system are often not well defined or pursued with the same vigour.

2 The research context

2.1 Introduction

This chapter establishes the context for the research problem, necessary since the effect of technical regulation on the development of complex engineered systems has not previously been established as a well defined area of research.

A broad review of the literature related to technical integrity is presented, not primarily as a critique of the literature, but rather to clearly establish the main features of technical regulations of the ADF and of regulatory examples across other domains of engineered systems in terms of the goals, provisions and intended operation of these regulations. Similarly, a review of relevant literature to identify the main concepts related to complex engineered systems is reviewed and summarised including an appreciation of the properties and characteristics of complex engineered systems that may be affected by technical regulation and more specifically may be related to technical integrity, the stated goal of ADF technical regulation.

2.2 Technical regulation

In order to understand the factors that have shaped technical regulation in the ADF a broad review of the literature was undertaken. Important elements of the history of development and the theoretical rationale for regulation are identified providing context for a detailed exposition of the objectives and provisions of technical regulations of the ADF. A sample of technical regulation in other domains is reviewed and contrasted with those of the ADF. The provisions which technical regulations impose upon the designers of systems are identified and discussed. The analysis of a range of technical regulatory documents developed an ontology useful to establish the context of this research and for the study of technical regulation more widely.

2.2.1 The history of technical regulation

In order to understand the precedents that exist for technical regulation in the ADF, a brief synopsis of the history of regulation is provided. Technical regulation is not new, stretching back at least as far as the Code of Hammurabi around 1772 BC where, as an example, the integrity of shipbuilding was formally regulated:

> 'If a shipbuilder build a boat for someone, and do not make it tight, if during that same year that boat is sent away and suffers injury, the shipbuilder shall take the boat apart and put it together tight at his own expense. The tight boat he shall give to the boat owner'. (King, 2008)

(Baldwin and Cave, 1999) provide a concise overview of the more recent history of regulation, the salient points relevant to this research are summarised in the following discussion. Regulation is most generally defined as a sustained and focused control enacted by a public agency over activities which are valued by the community at large. Regulation usually acts to restrict behaviour associated with undesirable activities or outcomes. One of the earliest examples in contemporary history is the introduction in the USA in 1887 of an Inter State Commerce Commission to limit discriminatory pricing by the railroads. In the UK, USA and Australia, the 1950s and 1960s saw the introduction of regulation and regulatory bodies to deal with issues such as monopolies, gaming, industrial relations, civil aviation and workplace health and safety. The privatisation of government assets in the 1980s and 1990s led to more regulation as government sought to preserve some control over the public's fair access to, the services of what had until then been public utilities.

Note that there have also been balancing activities by governments seeking to de-regulate, or at least to reduce the red tape and inefficiency often associated with regulation. Since 1981 in the USA all regulations have been subject to cost-benefit analysis and since 1996 in the UK and European Community regulation has been subjected to Cost of Compliance Assessment supported by a risk assessment. (Dobes, 2009) indicates that 'Australia never institutionalised the use of cost-benefit analysis in the way the United States did'. It is a notable difference that cost-benefit considerations have not been required for the introduction of regulations in Australia.

Technical regulation can be considered to be a subset of regulation more generally. Financial regulation, trade regulation and workplace safety are other examples of

particular regulations. Technical regulation is currently defined by the International Organization for Standardization (ISO) as:

> **Technical regulation**: document which lays down product characteristics or their related processes and production methods, including the applicable administrative provisions, with which compliance is mandatory. It may also include or deal exclusively with terminology, symbols, packaging, marking or labelling requirements as they apply to a product, process or production method (from ISO/IEC Guide 59: 1994, Annex A).

2.2.2 Theory of regulation

A synopsis of theory and concepts of regulation is provided in order to understand what underpins, at least implicitly, ADF technical regulations and how the features of the regulations may be characterised.

2.2.2.1 Rationale for regulation

A number of reasons are commonly quoted as leading to a need for regulation. Table 1 indicates the main rationales for regulating. A particular regulation may be driven by more than one rationale (Baldwin and Cave, 1999). Reviewing this model in the context of procuring and developing complex defence systems, leads to the fourth column of Table 1 where the rationale that may justify technical regulation of complex engineered systems for the ADF is indicated. The rationale for technical regulation could include four things:

- Monopolies and natural monopolies – regulation aims to ensure a market or market segment remains technically open. For example, open and interoperable standards may be required by regulation to ensure a national defence force does not get locked into proprietary designs by ensuring that market competition is possible and multiple sources of supply remain viable in times of contingency,
- Externalities – regulation aims to ensure the developer of the system is responsible for the risk-weighted cost of the system posing a hazard to personnel or the environment, i.e. regulation enforces the supplier to assume the risk-weighted costs of harm to personnel and/or the environment,
- Information Inadequacies – regulation aims to ensure the developer of the system is responsible to provide complete and accurate information to safely and reliably operate, maintain and modify the defence system for its long life cycle, and,

- Rationalisation and Coordination - regulation is to ensure the developer of the system is to meet standardisation requirements that will enhance the logistical efficiencies for the owning and operating organisations.

However, this is not to imply that they are the cited rationale or that they were objectively considered during the introduction of ADF technical regulations.

2.2.2.2 Enforcing regulation

Regulation may be enforced by formal or informal approach or a combination of the two. A formal approach is 'insistent' and focuses on deterrence to assure its aims are achieved. It is often associated with compliance measures enforceable under law. An informal approach is 'persuasive' in nature and focuses on seeking compliance though means of education, advice, persuasion and negotiation. When particular regulation aims to intervene is another consideration. Intervention may be preventative, act-based or harm-based (Baldwin and Cave, 1999).

Technical regulation in the ADF context has aspects of formal and informal approaches, but appears to be more biased toward the persuasive approach. Intervention is mostly aimed at being in the preventative stage.

Table 1. Rationale for regulating; adapted from (Baldwin and Cave, 1999).

Rationale for regulation	**Main aims of regulation**	**Example**	**Rationale for technical regulation**
Monopolies and natural monopolies	Counter tendency to raise prices and lower output. Harness benefits of scale economies. Identify areas genuinely monopolistic.	Utilities.	Yes
Windfall profits	Transfer benefits of windfalls from firms to consumers or taxpayers.	Firm discovers unusually cheap source of supply.	No
Externalities	Compel producer or consumer to bear full costs of production rather than pass on to third parties or society.	Pollution of river by factory.	Yes
Information inadequacies	Inform consumers to allow market to operate.	Pharmaceuticals. Food and drinks labelling.	Yes

Rationale for regulation	Main aims of regulation	Example	Rationale for technical regulation
Continuity and availability of service	Ensure socially desired (or protect minimal) level of 'essential' service.	Transport service to remote region.	No
Anti-competitive and behaviour predatory pricing	Prevent anti-competitive behaviour.	Below-cost pricing in transport.	No
Public goods and moral hazard	Share costs where benefits of activity are shared but free-rider problems exist.	Defence and security services. Health Services.	No
Unequal bargaining power	Protect vulnerable interests where market fails to do so.	Health and Safety at Work.	Possible
Scarcity and rationing	Public interest allocation of scarce commodities.	Petrol shortage.	No
Distribution justice and social policy	Distribute according to public interest. Prevent undesirable behaviour or results.	Victim protection. Discrimination.	No
Rationalization and Coordination	Secure efficient production where transaction costs prevent market from obtaining network gains or efficiencies of scale. Standardization.	Disparate production in agriculture and fisheries.	Yes
Planning	Protect interests of future generations. Coordinate altruistic intentions.	Environment.	Possible

2.2.2.3 Standards and the timing of regulatory intervention

The three types of standards enforced by regulation are related to the timing of the intended intervention (Baldwin and Cave, 1999):

- A regulatory approach using specifications or design standards is related to preventative intervention aims. They are relatively easy to enforce and compliance costs are readily estimated. But they are highly intrusive in that the regulator is involved in defining process and design and this may be a disincentive for the supplier to innovate. Normally, this approach should be limited to cases where the regulator has available expertise and information that is not available to private suppliers.

- A regulatory approach using performance or output standards is related to act-based intervention aims. In this approach the standards demand a level of delivery or focus on the level of acceptable risks. The standards are not technologically restrictive.
- A regulatory approach using target standards is related to harm-based intervention aims. Goals are stated directly, e.g. the absolute avoidance of certain harmful consequences. Suppliers are completely free as to how to meet the goals, but it may be socially inefficient if the suppliers do not have the capacity and expertise to understand how to achieve these goals.

ADF technical regulations use both specified requirements (e.g., a 'Certification Basis' (Australian Defence Force, 2005), a 'Materiel Requirements Set' (Royal Australian Navy, 2009)) as well as performance based standards (e.g., conduct systems engineering in accordance with MIL-STD-499A and configuration management in accordance with EIA 649 and AS/NZS 3907 (Royal Australian Navy, 2003a)). ADF technical regulations do not invoke target standards.

2.2.2.4 A perfect regulatory model?

The Law of Requisite variety in cybernetics states that for a system to be stable the control mechanism must have at least as many states as the system being controlled; only variety in the regulator can force down variety or disturbance in the system being regulated (Ashby, 1955). For a technical regulatory system this implies that the scope of regulatory controls will require some minimum variety or complexity in order to effectively regulate the technical integrity of a system. Later, (Conant and Ashby, 1970) developed a theorem which says that:

> The best regulator of a system is one which is a model of that system in the sense that the regulator's actions are merely the system's actions as seen through a mapping.

In cybernetics and engineering control theory this makes immediate sense. But is it reasonable to extend this thinking to a technical regulator? Certainly, if there was a perfect method (i.e., a model) of assuring the technical integrity of a complex engineered system, and if the regulator was knowledgeable of that method, then the coding of that method into regulation by the regulator could certainly be seen as optimal.

(Conant and Ashby, 1970) also observe that the theorem 'leaves open the possibility that there are regulators which are just as successful (just as "optimal") as the simplest optimal regulator(s) but which are unnecessarily complex'. For technical regulations that are considered in this research it may be that the regulatory authority and the regulations imposed on organisations charged with the development, operation and maintenance of complex engineered systems may be unnecessarily complex, notwithstanding some minimum level of complexity is required in keeping with the Law of requisite Variety.

Further, it is pointed out that 'success in regulation implies that a sufficiently similar model must have been built'. Any regulator must model what it regulates to be successful. Here may be the crux of problems with technical regulation in most of its forms; there is no immediately available evidence to suggest that technical regulators have modelled that which they seek to regulate. That is, the technical regulators do not hold a model of the necessarily complex organisations charged with the development, operation and maintenance of complex engineered systems or the characteristics of the systems the regulation is intended to influence.

The concept of technical regulation itself being expressed as a model that seeks to maximise the level of satisfaction of stakeholders and minimise the risk of total losses by all stakeholder has been proposed by (Hurevich and Serenkov, 2009). But their approach does not seek to model the system being regulated or the intended causation of the regulatory system, rather it seeks to model the development of standards which regulations may invoke.

Knowledge gap: There is no model for how technical regulation is meant to achieve its technical aims.

2.2.2.5 Technical regulation as a layer of defence

In complex systems, failure is almost considered inevitable; systems operate on the edge of instability, and often require operators to keep the system from the brink of catastrophic mishaps (Modarres, 2006). Safety mishaps can be thought of as 'inadequate enforcement of safety constraints' (Leveson, 2011) and in very complex systems that opportunity for insufficient or dysfunctional safety controls is significant.

As discussed by (Cook, 2000), ‘interesting complex systems’ are almost always inherently and unavoidably hazardous. The frequency of hazard exposure may be minimised but the presence of hazards remains. The high consequence of failure of complex systems leads to multiple layers of defence being incorporated into the system. These include technological defences (back-up/redundant systems and safety controls) as well as organisational, institutional and regulatory defences (e.g., policies and procedures, certification, work rules, and compulsory training for operators and maintainers).

A model for thinking of technical regulation is as a layer of Swiss cheese in the ‘Swiss Cheese Model’ (Reason, 2000). Figure 2 shows an adaptation of Reason’s model that considers technical regulation as one of the layers of defence against failure in the construction of a complex system. A failure in construction or the application of technical regulation may leave a ‘hole’ in that layer of defence but it would ideally be unlikely to ‘line-up’ with failures in the other defensive layers of the system that would allow a trajectory from hazards to losses to occur.

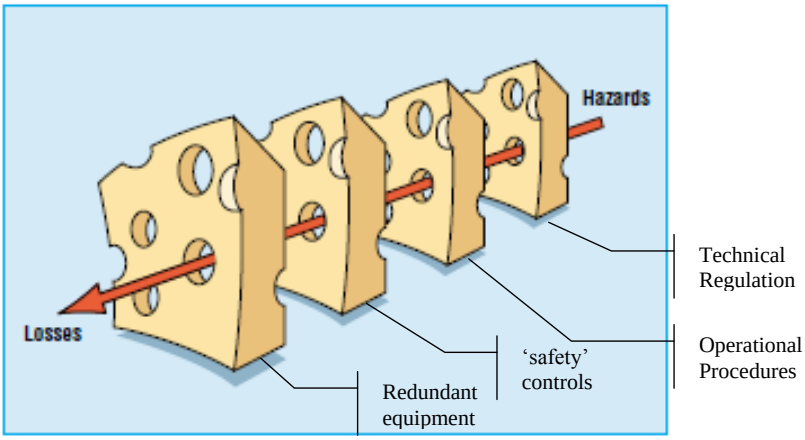


Figure 2. Reason’s ‘Swiss Cheese Model’ adapted to show technical regulation as one of the layers of defence in the development of a complex system.

2.2.3 Technical regulation in the ADF

2.2.3.1 Context

Defence systems are steadily increasing in complexity and are increasingly required to not only introduce higher organic capabilities within the system itself, but also to integrate into systems-of-systems to enable capabilities with higher performance across a wider spectrum of domains and environments. At the same time, the need for these

systems to achieve ever-increasing levels of technical integrity is paramount. Systems are expensive to develop and sustain, therefore there are less of them and each one that is realised must achieve higher and higher levels of availability and dependability. The stakeholders of the system, not least the general public (and consequently the political leaders of the nation) are less and less willing to accept mishaps – be they accidents to personnel, loss or damage of equipment or impacts to the natural environment.

2.2.3.2 The introduction and development of technical regulation in the ADF

Within the ADF, technical regulation commenced in the 1990s led by the Air Force in response to increasing safety concerns with the number of fatal air accidents (Royal Australian Navy, 2007):

> A series of accidents and incidents throughout the 1990s led the then Chief of the Defence Force to commission a Review of ADF Aviation Safety Management. As a result of the review, the Chief of Air Force was appointed as the ADF Airworthiness Authority and tasked to establish, manage and monitor an ADF Airworthiness Management System for aircraft and aviation systems. Since the ADF created a single Airworthiness Management System in the 1990s, ADF accident rates have reduced significantly, despite a much higher operational tempo.
>
> Prior to 1993, the fatal accident rate averaged 4.04 per year. Following the introduction of technical (1993) and operational (1998) airworthiness regulations, the rate of fatal accidents has reduced substantially. Since 1993 the accident rate has averaged 0.38 fatal accidents per year representing a ten fold decrease. Since 1999 the ADF has had to operate at a higher tempo than at any time in the previous twenty-five years. However, our fatal aviation accident rate is at its lowest.

Figure 3 shows the dramatic improvement on the number of fatal accidents experienced by the ADF since the introduction of technical regulation.

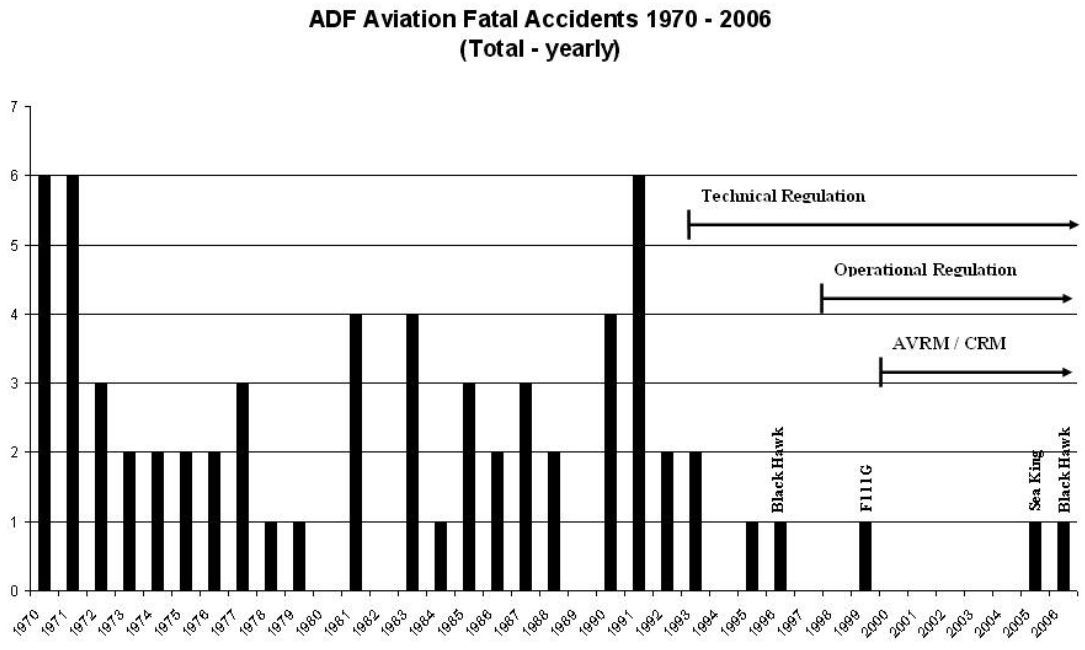


Figure 3. Total number of ADF fatal aviation accidents 1970-2006 (RAN, 2007)

In the late 1990s, the Army and Navy migrated the regulation of aviation assets to the leadership of the Air Force. Maritime materiel was subjected to regulation of technical integrity in 1999 (Royal Australian Navy, 2003b) and land materiel in 2001 (Army, 2003). The ADF imposed overall regulation of ADF materiel in 2002 (ADF, 2002) and appointed regulatory leadership for air materiel, maritime materiel, land materiel and explosive ordnance.

Significant breaches of technical integrity have occurred and inquiries into these accidents indicate that there remains room for improvement of technical regulation and its application in design, operation and maintenance of ADF materiel. For example:

- Sea King helicopter crash on Nias Island on 2 April 2005. Despite a 'sophisticated airworthiness system being in place' the Board of Inquiry found 'that most senior managers did not fully understand their airworthiness management responsibilities, particularly with regard to technical airworthiness' (Royal Australian Navy, 2007).
- Fatal fire in the machinery space of HMAS Westralia on 5 May 1998. The Board of Inquiry revealed shortcomings in configuration management (in particular unauthorised configuration changes with safety implications and inappropriate workarounds), the lack of technical review of work packages by competent engineering personnel, lack of appropriate quality assurance information, and the

level of professional expertise available at acquisition and contractor organisations (Lamacraft et al., 1998).

For ADF technical regulations, there is no evidence identified to show that there has been a cost-effectiveness or cost-benefit analysis undertaken to justify the introduction of, or extent of, regulatory frameworks. Rather, (Dobes, 2009) indicates that related and higher expenditure policy and decision making areas within the ADF, like Force Structure and capabilities, 'appears to be largely based on qualitative analysis by various committees' and 'various sources suggest that relatively rigorous cost-effectiveness analyses – let alone cost-benefit analysis – are seriously lacking'.

> Knowledge gap: There has been no cost-effectiveness or cost-benefit analysis undertaken to justify the introduction of, or extent of, regulatory frameworks.

Change is, however, ongoing. As part of the Australian Government's Strategic Reform Program (Department of Defence, 2010) technical regulation in the ADF continues to evolve. (Schmidt, 2012) commentates that:

> Defence's technical airworthiness regulatory authority has recently taken up the challenge to critically look at what it can do to improve the efficiency of its regulatory set. ... These start with a critical examination as to how well and efficiently its regulatory requirements protect users from airworthiness hazards. Also, moving its resource-intensive audit program to a risk-based compliance assurance program, using predictive techniques and less intrusive methods, to increase assurance with less impact on organisations. ... Work is now focused on improving alignment between respective [ADF] regulatory systems which should reduce compliance costs.

This indicates that the introduction of extensive technical regulation was not subjected to cost-benefit analysis considerations, but it is now recognised as costly and efficiency gains are needed. A decrease in the (presumed) effectiveness of technical regulation would not be desirable, but there is no empirical evidence available to baseline past or current effectiveness.

2.2.3.3 Technical regulations and frameworks in the ADF

This section provides, in some detail, an understanding of the aims and provisions of the technical regulations of the ADF, particularly as the regulations are devolved to sub-domains of defence materiel and this devolution results in both similarities and differences in the regulatory aims and provisions.

Technical regulation in the ADF is enacted by the Defence Instruction DI(G) LOG 08-15 entitled 'Regulation of technical integrity of Australian Defence Force materiel' (Australian Defence Force, 2004b). It introduces itself as:

> Technical regulation is how Service Chiefs, as capability output managers, establish confidence in the processes by which the fitness for service and requisite levels of safety of materiel are achieved.

In the ADF context, technical regulation is the means employed to assure technical integrity of materiel:

> Technical regulation is to identify, analyse, assess, treat, monitor and communicate risk in the areas of safety, performance and environmental compliance.

The principles of technical integrity embodied in DI(G) LOG 08-15 are that materiel must be designed manufactured and maintained:

- to approved standards,
- by competent and authorised individuals,
- who are members of an authorised organisation, and
- whose work is certified as correct.

Responsibility for technical regulation is further devolved to four Technical Regulatory Authorities (TRA) covering the domains of maritime, air, land and explosive ordnance. Each domain's technical regulatory arrangements are enacted through their own defence instruction and supported by regulations, manuals, and other guidance documents issued and maintained by the TRA.

2.2.3.3.1 Maritime domain

Policy and responsibility for the Navy Technical Regulatory System is established by the Defence Instruction DI(G) LOG 08-15 entitled 'Regulation of technical integrity of Australian Defence Force maritime materiel' (Royal Australian Navy, 2009). The policy is supported by the three volumes of the Navy Technical Regulations Manual (NTRM) (Royal Australian Navy, 2003a).

2.2.3.3.2 Air domain

The role and responsibility of the Technical Airworthiness Regulator (TAR) in 'establishing and maintaining the technical airworthiness element of the Aviation Safety

Program' is established in the Defence Instruction DI(G) OPS 02–2 'Defence Aviation Safety Program' (Department of Defence, 2011). Technical regulation sits within the overall airworthiness management system where regulations and guidelines for the latter are established in the 'ADF Airworthiness Manual' (Australian Defence Force, 2001) and technical regulation established in the 'Technical Airworthiness Management Manual (TAMM)' (Australian Defence Force, 2005). Further guidance for design is provided in the 'Airworthiness Design Requirements Manual'(Australian Defence Force, 2004a).

2.2.3.3.3 Land domain

The Army Technical Regulatory Framework (TRF) is the policy, regulations and guidance that define the 'minimum requirements for all engineering, procurement, acquisition and maintenance to ensure the technical integrity of land materiel' is established by the Defence Instruction (Army) LOG 1/2003 'Regulation of the Technical Integrity of Land Materiel' (Army, 2003). It is expanded upon in detail by the Technical Regulation of ADF Materiel Manual - Land (TRAMM-L). 'The TRAMM-L contains the TRF policy, instructions and guidance governing the technical regulatory processes of the TRF' (Department of Defence, 2009).

2.2.3.3.4 Explosive ordnance

The highest level of policy for explosive ordnance (EO) in the ADF is established by the Defence Instruction DI(G) LOG 4–1–006 'Safety of Explosive Ordnance' (Department of Defence, 2011). It is supported by 'DEOP 102 Technical Integrity Of Explosive Ordnance' which provides policy and procedures for the assessment of the safety (and broader technical integrity) of EO items being procured, modified, acceptance tested or introduced into service (Department of Defence, 2007).

2.2.3.4 Aims of ADF technical regulation

2.2.3.4.1 Technical integrity in the ADF

The explicit aim of technical regulation in the ADF is 'ensuring that Australian Defence Force (ADF) materiel is fit for service, and only poses acceptable risk to personnel, public safety, and the environment' (Australian Defence Force, 2010). The implicit assumption is that in the absence of any technical regulation the requisite technical integrity may not be achieved.

Technical integrity is defined in (Australian Defence Force, 2010) as:

> An item's fitness for service, safety and compliance with regulations for environmental protection.

The elements of technical integrity are:

- Fitness for service: 'The materiel's ability to satisfy operational requirements. Hence it is a subset of technical integrity' (Australian Defence Force, 2004b).
- Safety: 'Freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment' (Department of Defense, 2000).
- Environment compliance: Complies with regulations for environmental protection.

The aims of the technical regulations developed across the four domains of maritime, aerospace, land and explosive ordnance domains subtly differ with each other and are described in the subsequent sections. The difference in aims demonstrates the subjectivity associated with the term technical integrity even within the relatively narrow delimitation of the ADF.

2.2.3.4.2 Technical integrity in the maritime domain

Technical integrity of maritime materiel is the stated aim of the Naval Technical Regulatory System (NTRS). The technical integrity definition within the naval technical regulatory system exhibits no substantive deviation from the ADF level definition. But, the definition in the NTRM (Royal Australian Navy, 2003a) goes a little further in defining technical integrity in terms of 'technical regulatory requirements':

> An item's fitness for service, safety and compliance with regulations for environmental protection, i.e., its compliance with Technical Regulatory Requirements.

Where technical regulatory requirements are:

> Requirements imposed by the Technical Regulatory System ensure that ADF Maritime materiel is designed, constructed and maintained throughout its entire life-cycle to standards acceptable to the RAN.

2.2.3.4.3 Technical integrity in the air domain

The Air Force manages the safety and technical integrity all aviation related materiel used by the ADF is consistent with higher level policy of (Australian Defence Force,

2010). The TAMM (Australian Defence Force, 2005) however introduces its own definition of technical integrity:

> Refers to the state of worthiness of a platform, combat system or ancillary item to fulfil its intended mission safely and effectively throughout its planned life. This requires evidence to demonstrate that the material has been designed, constructed and maintained to approved standards by competent and formally approved personnel who are acting as members of an approved organisation and whose work is certified as correct and is accepted on behalf of the ADF.

In general though the technical regulation of the air domain is dominated by the aim termed *technical airworthiness*:

> A concept which defines the condition of an aircraft and supplies the basis for the judgment of its sustainability for flight in that it has been designed, constructed and maintained to approved standards by competent and approved individuals, who are acting as members of an approved organisation and whose work is certified as correct and accepted on behalf of the ADF.

2.2.3.4.4 Technical integrity in the land domain

For land materiel the technical regulation (Army, 2003) aims for technical integrity and defines it as:

> An item's fitness-for-service, safety and compliance with regulations for environmental protection.
>
> Fitness-for-service (and fit-for-service) - the materiel meets an endorsed operational requirement by virtue of its design and manufacture and is maintained to a fully functional state.
>
> Environmental compliance. Following Departmental Environment Instructions to minimise risks and maintain accountability in relation to environmental protection matters.

This contrasts with the previously described domains by being more oriented to using specifications and standards as an objective manifestation of technical integrity and is therefore more biased toward preventative intervention.

2.2.3.4.5 Technical integrity of explosive ordnance

The technical regulation for explosive ordnance (Department of Defence, 2011) defined as its main technical aim 'Safety and Suitability for Service' that is attested to via a EO Certificate (EOC):

> The EOC certifies that, within a defined environment, the EO component of the EO-Weapon-Platform System is
> a. tolerably safe,
> b. will meet specified performance parameters, and

c. is reliable, sustainable, maintainable and environmentally compliant throughout its specified life.

Knowledge gap: The goal of technical regulation is *technical integrity.* However, there is not a clear, objective definition of technical integrity established for the ADF and the subjective definitions vary across the materiel domains.

2.2.3.4.6 Fitness for service

Most of the technical regulations and the definitions for technical integrity across the ADF reference a concept of *fitness for service*. Definitions for fitness for service however are quite broad and are somewhat different across the materiel domains. Fitness for service in each of the domains is covered as follows:

- For maritime materiel: 'the materiel's ability to satisfy operational requirements. Hence it is a subset of technical integrity' (Royal Australian Navy, 2003a).
- For air materiel, the TAMM (Australian Defence Force, 2005) does not include a definition for fitness for service and in the only reference, excludes it as part of technical airworthiness but rather classifies it as 'mission-worthiness':

  While a DAR is primarily concerned with ensuring technical airworthiness of aircraft and aircraft-related equipment, most DMO Chief Engineers and Project Engineers have an additional (but closely related) responsibility for ensuring 'fitness for service', or in other words mission-worthiness.

- For land materiel: 'the materiel meets an endorsed operational requirement by virtue of its design and manufacture and is maintained to a fully functional state' (Army, 2003).
- For explosive ordnance, the term fitness for service is not often used in the regulations and is not defined. However the term suitability for service is used and is defined as 'will meet specified performance parameters and is reliable, sustainable, maintainable' (Department of Defence, 2011).

Knowledge gap: Fitness for service is a key element in the subjective definition of technical integrity as the aim of the ADF technical regulation. But the definitions of fitness for service differ across the devolved regulatory domains. Fitness for service needs to be defined more objectively or a comparable concept introduced to bolster the definition and theory of technical integrity.

2.2.3.5 Provisions of ADF technical regulation

Provisions imposed by regulation in the ADF, or 'technical regulatory requirements', are devolved to each TRA (Australian Defence Force, 2004b):

> Technical regulatory requirements must be promulgated by TRA as:
> a. instructions that define policy;
> b. publications that primarily define process requirements; and
> c. ADF materiel standards (maritime/land/air/EO).

Design development and production of systems for the ADF is covered by acquisition and the TRAs are required to make provisions for certification (Australian Defence Force, 2004b):

> Technical regulation of acquisition must encompass:
> a. certification of the specification as complying with required standards;
> b. certification of supplier competency;
> c. certification by the supplier of compliance with the technical regulatory aspects of the specification; and
> d. if appropriate, certification of verification of compliance with the technical regulatory aspects of the specification.

Competence of the organisation making certifications must also be clearly established against the criteria defined by (Australian Defence Force, 2004b):

> The objective of recognition of the competence of organisations responsible for the design, manufacture or maintenance of ADF materiel is to achieve a satisfactory level of confidence in the certifications provided by the organisation. TRA must only recognise organisations as competent where they meet the following criteria:
>
> a. **Systems.** The organisation must have an established quality system appropriate to the type of work being performed, and any other technical management systems that the TRA may require, such as engineering management plans and configuration management plans.
>
> b. **People.** Individuals within the organisation must have the appropriate training, qualifications, experience, demonstrated competence and integrity to undertake the activities required, and must be authorised to perform those activities within the quality system.
>
> c. **Processes.** The procedures and plans that specify and define technical activities must be controlled and approved by an appropriately qualified individual, nominated within the quality system. The organisation must be able to demonstrate compliance with those procedures and plans.
>
> d. **Data.** The information applied to, and derived from, technical activities must be authoritative, accurate, appropriate and complete. Such data must always be accessible, but need not be retained in-house.

Technical regulation seeks to provide more assurance to the acquirers, and ultimately the users of engineered systems, than just a reliance on system developers to unilaterally apply the best engineering and management processes. ‘Best’ for a developer does not necessarily lead to a set of practices that will yield the required technical integrity. Engineering practices, like Systems Engineering for instance, do not alone assure that technical integrity goals are a primary focus during development, or that the processes will be executed by competent professionals with explicit responsibility for their design

decisions and the system outcomes. Technical regulation therefore generally attempts to assure that development is conducted in compliance with approved standards, by competent individuals in an organisation that is authorised to do so. There is also an explicit process to achieve 'certification'; at a minimum, by the developer formally attesting to the completeness and quality of its own work, but usually also encompassing some independent validation of the developer's claims of compliance with technical standards and processes.

2.2.3.6 The influence of technical regulation on technical integrity

2.2.3.6.1 Causal model

Technical regulation is established as a means to assure technical integrity of developed systems and materiel. Figure 4 shows an influence diagram developed to summarise the key concepts. A technical integrity gap exists where the technical integrity being achieved by systems is less than that desired. It may be that extant systems have some performance or effectiveness (fitness for service) shortfalls, or have incurred accidents or environmental impacts that are beyond what can be tolerated by the users and stakeholders of the system. This gap influences the adoption of technical regulation as a means of achieving the requisite technical integrity. Technical regulation, in an ADF context, requires that approved specifications and standards be in place for development, that development be conducted by competent individuals, working in authorized organisations and that all work is certified to be correct.

Technical regulation is actualised by certifications which attest to the implementation of the regulations:

- The specifications developed to define the acquisition project must be certified as complying with the required standards (the list of required standards is usually maintained by a regulatory body). This certification is normally provided by the regulator (e.g., by endorsement of the specification and the plan for certification of the system).
- Secondly, the supplier must be certified as being competent to undertake the design, development and production of the system being acquired.
- Thirdly, the supplier certifies that the delivered system has achieved the entire technical integrity requirement embodied in the specification.

- For significant mission and/or safety critical systems where the risks to technical integrity are reasonably high, a certification is generally required, usually by a party independent of the supplier, that the compliance claimed by the supplier has been verified (i.e., to provide assurance that the supplier's claim of compliance is valid).

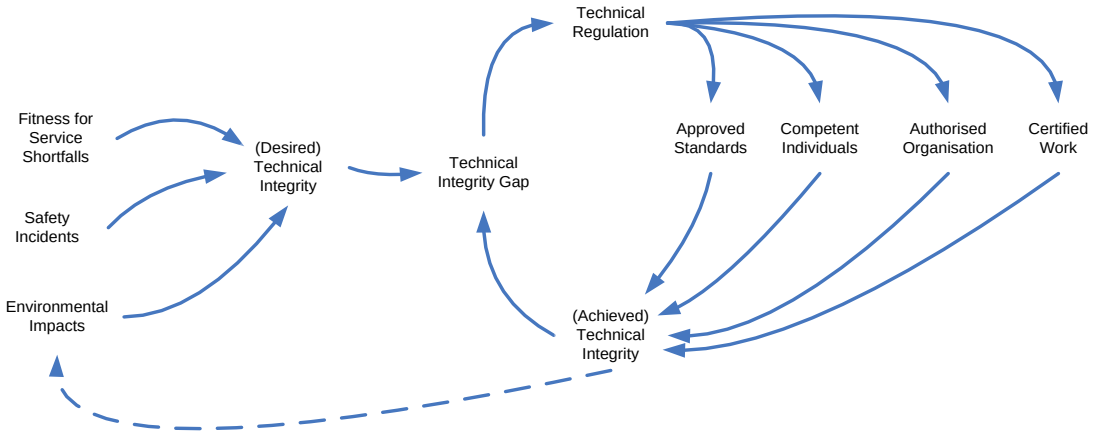


Figure 4. The influence of technical regulation on technical integrity.

2.2.3.6.2 Methods for achieving technical integrity of a complex engineered system

(ISO, 2007) is the current standard concerning the development of man-made systems that include one or more of hardware, software, human processes, procedures, facilities and entities in the environment. It forms the current accepted practice for the engineering of all such systems including the complex engineered systems of interest to this research. While the defined systems engineering processes (requirements analysis and validation, functional analysis and verification, synthesis, design verification, systems analysis and control) are not in dispute as being required for a successful system implementation project, there is no indication in the standard as to their relative priority or the contribution each process may have to the technical integrity of a complex system.

(Elm, 2008) has conducted research that indicates which systems engineering processes have the strongest positive relationship to program performance. Technical Integrity is at least partly a subset of Program Performance (includes customer satisfaction, budget performance and schedule performance) as defined in Elm's research. The systems engineering processes that have a higher positive relationship with project performance than the SE capability as a whole are 'Architecture', 'Trade Studies', 'IPT Capability' and 'Requirements Development and Management'. (Boehm et al., 2008) also report

evidence that the right level of systems engineering applied to a project, increase the likelihood of the project running to cost and schedule budgets, but generally treats technical performance or integrity outcomes as invariant (i.e. implicitly assumes they are achieved at completion of the project). However, this is rarely the case in practice. Many projects are deemed to be 'complete' not having achieved all technical requirements and having to have accepted unresolved risks to technical integrity.

(McDermid, 2001) suggests that rigour in engineering process is perhaps not the dominant contribution to the production of complex systems with the requisite technical integrity, but that relevant experience and competence of the designer is:

> For airborne software developed to DO-178B, there seems little evidence that the more rigorous processes applied in software development for software of higher required Software Integrity Levels (SILs) actually yield a lower hazardous failure rate. The limited data does suggest that the most significant correlated factor in producing software with lower hazardous failure rates is domain experience. This is most linked to the consequent reduction in initial requirements errors that domain experience would bring.

This view is counter to that implied in ADF technical regulations that place substantially more emphasis through provisions on process that they do on assessment of technical competence or expertise. (McDermid, 2001) also casts doubt on the ability of some of the engineering standards to directly contribute to the assurance of technical integrity:

> Do the software safety standards actually address safety issues? They seem to be mostly focused on quality and repeatability. You would think they would aim directly at potentially hazardous failure modes of the software. An evidence-based approach is proposed – e.g. provide analysis to show that data structures can never be corrupted; that a scheduled function can always run on time.

This criticism could be readily extended to systems engineering standards (see for instance (Honour, 2010) who found no correlation between 'technical quality' and 'systems engineering activities') and even to the technical regulations and their supporting manuals. Quality and the repeatability of process seem to be a primary focus rather than the imposition of particular management and engineering measures to attain technical integrity.

A prevalent view in the Australian defence industry at present remains that technical integrity can be assured through the use of systems engineering methodologies and disciplines (Raytheon Australia, 2012) while others think 'The Systems Engineering

discipline remains intrinsically tied to the individual skill, competence and notably the availability of senior systems engineers on projects' (Irving, 2008). This latter view places a focus on experience and competence both at an organisational and an individual level.

(Sheard and Mostashari, 2008) posit a set of 26 principles for complex systems engineering. They include the categories of Systems architecting-type principles, Systems analysis principles, Problem-space-relevant-principles, Configuration management principles, Coordination principles, and Management-related principles. While these principles are generally accepted to be valid heuristics for the development of complex systems and would be expected to have a positive effect on the technical integrity of a complex system, there is currently little empirical evidence to confirm such a conclusion.

2.2.3.6.3 Effectiveness of technical regulatory provisions

The main provisions of the ADF's technical regulation enforce the development of systems to approved standards. The work is to be undertaken by competent and authorised individuals, who are acting as members of an authorised organisation and whose work is certified as correct (Australian Defence Force, 2004b). As an example, Figure 5 shows the main provisions of the NTRS. This distils the main provisions that must be implemented in the acquisition and development of maritime materiel for the ADF.

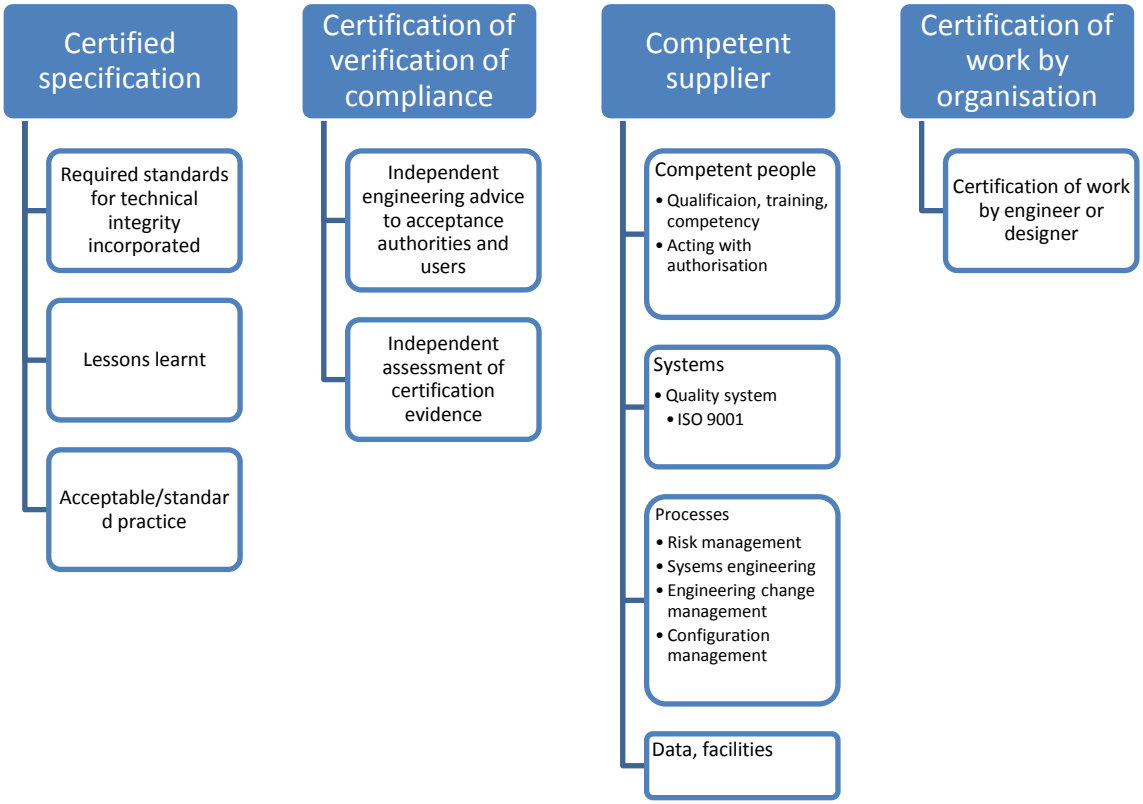


Figure 5. The main provisions of the ADF's NTRS.

Table 2 takes the elements of the NTRS highlighted in Figure 5 and provides an indication as to the extent that a provision is actually known (either through reported research evidence or compiled professional knowledge, e.g., standards) to contribute to enhanced technical integrity outcomes for complex engineered systems.

Table 2. Provisions of technical regulatory framework and known support for contribution to technical integrity aims.

Technical Regulatory Provisions	Contribution to Technical Integrity
Competent Supplier	
Competent people	Competency certifications assure minimum competency in place for significant design decisions. Certainly there are numerous case studies where the lack of competency has been directly linked to lack of required competency (e.g. HMAS Westralia (Lamacraft et al., 1998)). Also anecdotal evidence that the competence and domain experience of key designers is a determinant of project success. No empirical evidence identified that technical integrity is positively related to competent people.
Systems (e.g., Quality System)	Standards like ISO 9001 capture industry-wide knowledge. Certainly focuses on repeatability of processes to produce consistent products. Does not focus specifically on technical integrity. No empirical evidence identified that technical integrity is positively related to quality systems.
Processes	
• Risk Management	The explicit consideration of risks to technical integrity embodied in the ADF Technical regulations does, it would seem, seek to impose a mind-set on the designers involved to minimise the risks. No empirical evidence identified that technical integrity is positively related to technical integrity risk management systems.
• Systems Engineering	Standards like ISO/IEC 26702 capture industry-wide knowledge for current projects. Research is beginning to provide evidence that the application of systems engineering processes do contribute positively to project success and are worth the cost. Other research is indicating that the current processes requiring extension to cater for increasingly complex systems. The evidence for the relationship between these systems engineering processes and the technical integrity of complex systems is not yet established.
• Configuration Management	Heuristics for complex systems point out that CM must prepare for and accommodate design changes and design alternatives. The evidence for the relationship between the configuration management and the technical integrity of complex systems is not yet established.

Technical Regulatory Provisions	Contribution to Technical Integrity
Data, Facilities	Research indicates that good decisions in complex situations are made on data (rather than instinct). It is therefore reasonable to extrapolate that data – correct and readily accessible to the designer – will allow good design decisions to be made. No empirical evidence identified that technical integrity is positively related to data, facilities and other supporting resource.
Certification of Verification of Compliance	
Independent engineering advice to acceptance authorities and user community	No empirical evidence identified that there is support for this contributing to technical integrity.
Independent assessment of verification evidence	Independent assessment is standard practice for high-risk engineering work. Research indicates peer and expert review are effective means of eliminating defects in design.
Certification of Work by Organisation	
Certification of work by engineer/designer	No empirical evidence identified that is support for this contributing to technical integrity.
Certified Specification	
Required Standards for Technical Integrity Incorporated	Prescriptive regulatory regimes impose the list of standards that must be met to assure technical integrity. There are known issues (e.g. standards not able to keep up with pace of technology advances) and known benefits (available standards can embody previous lessons learnt) with the prescriptive approach. No empirical evidence identified that there is support to determine if there is any contribution to the technical integrity of complex engineered systems, which are typically unprecedented by requiring the use of extant standards.
Baseline for technical integrity assessment	
Lessons learnt	
Acceptable/ standard practice	

The overarching view reached by consideration of Table 2 is that most of the provisions of the ADF technical regulations are justified by heuristics as opposed to empirical evidence. There is also no consideration as to if the provisions entail a necessary and sufficient basis for the attainment of technical integrity.

Knowledge gap: The provisions of ADF technical regulations are generally accepted as heuristics, but are not empirically known to be sufficient or effective in assuring the attainment of technical integrity of complex engineered systems.

2.2.4 Technical regulations in other domains

As a point of comparison against ADF technical regulation, a number of other technical regulations and domains of application were considered in order to ascertain what aims were being pursued and what provisions were required and what evidence for the effectiveness of these exists.

The design and operation of significant complex systems is often regulated, to prevent unwanted, unintended consequences, by governments or by international organisations with multi-national government support. The objectives of these regulations are effectively variations on technical integrity objectives. Examples include:

- Lloyds Rules Ship Classification (Lloyd's Register of Shipping): Compliance with a set of rules or other technical standards that have been shown to be appropriate for the objective of safe and compliant ships.
- The U.S. Department of Energy's policies and guidelines for acquisition of capital assets which include the regulations for acquisition of highly safety and security critical nuclear power assets (Department of Energy, 2006, Department of Energy, 2008).
- The United States Navy's platform level and strike force (formally battle force) level Naval Warfare Systems Certification Policy (NAVSEA, 2005) has as its stated technical objectives 'Mature, Safe and Reliable Warfare Systems'.

In these examples, the concepts of safety, compliance, maturity and reliability as technical objectives of regulation are recurring. The following subsections provide some more detailed examples.

2.2.4.1 Civil aviation regulation

The development of civil aircraft and their systems are governed by technical regulation. Airworthiness certification, as it is called by most national regulatory bodies, is 'the systematic process, during the design of an aircraft or airborne system of demonstrating conformance to a set of specific and predetermined airworthiness

regulations (e.g., FAR 25) for a specific type and category of aircraft' (Kritzinger, 2006).

The regulation of airworthiness is the ADF is, to a large extent, equivalent to and even adopts parts of the civil regulatory practice.

2.2.4.2 Safety and environmental compliance of ships

The International Maritime Organisation (IMO), established since 1948, published regulations for the safety of life at sea (SOLAS) and the protection of the marine environment. It produced International Conventions: SOLAS and MARPOL (the Prevention of Pollution in the Marine Environment). The IMO has representation from 168 nations of the world. The IMO has no enforcement powers; rather the member countries enforce the IMO regulations within their own jurisdictions.

Arguably, most of the safety and environmental policy has been developed as a consequence of accidents, e.g., the Exxon Valdez. This reactive approach to setting policy and standards is, in the interest of further improvements and reduction of risks, being transitioned to a proactive approach using Formal Safety Assessment (FSA) as the mechanism (International Maritime Organization, 2007). The intent of using FSA is the 'a rational and systematic process for assessing the risk related to maritime safety and the protection of the marine environment'(Kontovas and Psaraftis, 2009).

Ship classification is process used to assure stakeholders that a ship or marine platform complies with a set of rules or other technical standards that have been shown to be appropriate for the safe and environmentally compliant function of the ship. Lloyds Rules Ship Classification (Lloyd's Register of Shipping).

The regulation of maritime materiel in the ADF is cognisant of, and in some areas (e.g., for hull and platform mechanical systems) moving toward ship classification as a preferred mechanism for assuring technical integrity or seaworthiness.

2.2.4.3 Technical integrity in the commercial industry

Technical integrity or just *integrity* is not a widely used term in commercial industry. One area where it is used is in the area of process plants. For example, Woodside uses the definition by (Hackett, 2008):

> Integrity – achieved when risk of failure that endangers personnel, environment or asset value is tolerable and is reduced to ALARP.

Similarly, (PI Asset Integrity, 2007) state:

> Technical Integrity in any process plant is essential to ensure that there will be no harm done to people or the environment. Assuring Technical Integrity of your facility by regular interval audits makes good business sense.

(Bale and Edwards, 2000) define it as:

> Technical integrity is concerned with the development of the design such that it is carried out by well trained personnel, who have been assessed to be competent, in accordance with recognized, sound practices and procedures and such that there is adequate provision by way of reviews and audits, to ensure the design intent is unimpaired in any way that could cause undue risk or harm to people or damage to the environment.

Notable in these definitions is the direct parallel with the ADF's components of personnel safety and environmental compliance as well as the contrast between the ADF's fitness for service and the concepts of 'asset value' and 'good business sense'. Clearly the latter may be explained by the mission outcomes as required by a commercial entity. That is, creation and maintenance of value to further the business. At a high level, the provisions are also comparable to those of the ADF technical regulations: well trained personnel, assessed compliance, sound practices and procedures.

2.3 Complex engineered systems

The purpose of this section of setting the research context is to

- Establish the characteristics of the type of systems the ADF technical regulations apply to and therefore clarify the delimitation of this thesis to consideration of those types of systems.
- Summarise the properties and characteristics of complex engineered systems in order to understand the way these may be influenced by technical regulation and to identify what may be related to the ADF technical regulatory goal of technical integrity.
- Establish a definition of complex engineered systems useful for establishing the research problem area and specifically to focus the research of this thesis.
- Identify the main historical and contemporary methods for designing and implementing complex engineered systems as a basis for comparing these with the provisions of technical regulation.

2.3.1 Development of complex engineered systems in the ADF

2.3.1.1 System types

The types of complex engineered systems of interest in the scope of this thesis include examples in the ADF like warship combat systems, airborne command and control, sensor and weapon systems, and ground based command and control systems for single service and joint operations. These may be compared and contrasted to similar systems in other services and complex systems like aviation systems, space systems and nuclear power systems in the civil sector.

2.3.1.2 Systems integrated from MOTS/COTS

Systems acquired for the ADF may be MOTS/COTS or an integration of the same into a new system. Such systems projects seek value for money and on-time delivery as an imperative and a justification, in part at least, for a MOTS/COTS approach. Government policy directs that the ADF must consider MOTS/COTS and any decisions to modify these off-the-shelf systems must be with a view to the costs and benefits involved:

> *Defence must also take into account what the market can deliver 'off-the-shelf'. The Government needs to make informed decisions about the appropriate mix of cost, risk and capability. Consequently, as highlighted in the White Paper, the Government has decided that Military-off-the-Shelf (MOTS) and Commercial-off-the–Shelf (COTS) solutions to Defence's capability requirements will be the benchmark. Decisions to modify MOTS or COTS solutions will need to be based on a conscious acknowledgement of the costs and benefits to the military outcomes, and the risks involved.* (Faulkner, 2009)

As systems become larger, more expensive and more complex it is less feasible to design them from scratch. Rather, the current approach in the ADF leans toward the build and integration of systems from MOTS/COTS components in order to meet the objectives for the system. The system solution is therefore constrained by the MOTS /COTS components – not all initial objectives for the system may be met and different design methods need to be applied to assure the risks to technical integrity are identified and mitigated. See, for example, the method described by (Miller, 2010). Integration of a system from COTS/MOTS then, should be done with the customer willing to be flexible in the requirements of the overall system - the vendors of the MOTS/COTS components are really driving a large fraction of the function, performance and technical integrity characteristics of the system.

Difficulties with designing and integrating the emergent behaviour and integrity of this type of system need to be addressed. These include the MOTS/COTS design being relatively unknown and therefore difficult to establish its behaviour, faults and failure modes; addressing incompatibilities between interfacing products; products evolving over the systems lifecycle which may be beneficial or may become prematurely unsupportable.

> Knowledge gap: It is not known how well the technical regulatory system deals with the current trends in the development of more complex systems and systems-of-systems which are increasingly constrained by a philosophy of procuring MOTS components to reduce development risk.

This gap is particularly acute for the establishment of the qualification of MOTS/COTS equipment against requirements relating to its operability and survivability in stressing ambient environments (e.g., high temperature and humidity) and threat environments (e.g., hostile weapon effects). This issue has been partially addressed by the development of a framework to evaluate the evidence provided or available to support qualification against this type of technical integrity requirement (refer Appendix I).

2.3.1.3 Characterising systems by scope and complexity

Another means of delimiting the systems of interest is by scope and complexity. (Kossiakoff et al., 2011a) defined the scope of systems in terms of a system hierarchy from components, subsystems, systems, systems of systems and to the enterprise. (Hitchins, 2007) has addressed scope via five defined levels of systems as follows:

- 5: Societal or socio-economic systems.
- 4: Industry systems.
- 3: Business/enterprise systems.
- 2: Project systems.
- 1: Subsystems, products and technical artefacts.

A system in a layer of this model ‘nests into the layer above it’. So the ADF systems subject to technical regulation are primarily at layer 1 and 2 of this model (e.g., warships, fighter aircraft, command, control and communication systems) and fit into the ADF, a layer 3 enterprise system.

One dimension of complexity of a system that can be considered is the number of technologies involved and the number of technical disciplines involved to bring the

system to fruition. (Royal Academy of Engineering, 2007) identified three levels of complexity of systems:

> Level 1: A sub-system, substantially within one engineering discipline and one organisation. Examples include a PC motherboard, a car gearbox, a sand filter for water treatment, air conditioning, the antenna for an aircraft radio and a secure encryption terminal.
>
> Level 2: A system that involves two or more engineering disciplines and/or requires two or more organizations to design, build, operate or maintain it. Examples include an electricity power station, railway signalling, a car, a waste water treatment plant, a hotel and a fighter aircraft.
>
> Level 3: A system of systems that impacts, or is impacted by, many disciplines and economic, social or environmental factors. Examples include the national rail and roads network, the NHS, military command and control, the telephone network and electricity supply.

The ADF systems subject to technical regulation are primarily at Level 2 in this complexity model.

> Delimitation: Systems of primary interest to the research problem, called complex engineered systems, are, by way of example, systems like warship combat systems, airborne command and control, sensor and weapon systems, and ground based command and control systems, are integrated largely from subsystems and components that are acquired "off-the-shelf" and may be characterised in terms of scope as Hitchins Level 1/2 and by complexity as Level 2.

2.3.2 Concepts, properties and characteristics of complex engineered systems

2.3.2.1 Hierarchies

(Kossiakoff et al., 2011b, p. Ch. 3) has documented the usual approach to the reduction of complex systems through hierarchies of building blocks. The system may be decomposed in a series of physical building blocks comprising subsystems, components, subcomponents and parts. Similarly, the system may be logically decomposed into a hierarchy of functions, functional elements and entities, that latter being the media on which a system operates: information, material and energy.

The ADF technical regulations do not indicate if technical integrity can be considered in a hierarchical terms, i.e., that it may be decomposed into the technical integrity of its subsystems and in turn into the technical integrity of their components.

> Knowledge gap: It is not known if the technical integrity of a system is decomposable to the technical integrity of its subsystems.

2.3.2.2 Life-cycle of a system

ISO/IEC 15288 (ISO, 2008) defines the generic life-cycle stages of an engineered system as:

- Exploratory research stage; where stakeholder's needs are identified and ideas and technologies are explored.
- Concept stage; where stakeholder's needs are refined, feasible concepts are explored and viable solutions are proposed.
- Development stage; where the solution is designed, built, verified and validated.
- Production stage; where multiple instances of the system are produced and verified.
- Utilization stage; where the system is operated.
- Support stage; where the system is sustained to continue to provide its capability (most often happens in parallel with the utilisation stage).
- Retirement stage; where the system is taken out of service and stored, archived, recycled or disposed of.

Variations on these terms do exist but the concepts are consistent. For instance, (Wasson, 2006, p. Ch. 7) defines system life cycle phases as:

- System definition phase.
- System procurement phase.
- System development phase.
- System operations and support phase.
- System production phase.
- System disposal phase.

And, not dissimilarly, (Blanchard, 2008, p. 10-11) identifies the system life cycle as commencing with the identified need, then the system progressing through 'design and development', 'production and/or construction', 'operational use and maintenance support' and 'retirement and material disposal'.

This thesis is primarily focused on ascertaining the technical integrity of a system during the development stage, and to a lesser extent the concept stage (using ISO/IEC 15288 terms) in order to justify the provisions of technical regulation in these stages.

Clearly however, technical integrity outcomes are manifested primarily during the utilisation stage of the system life cycle.

> Delimitation: While technical regulation is applicable to the entire system lifecycle, including operation and maintenance, it is the provisions for design during concept and development stages of the system and if they are effective in assuring the technical integrity of a system that is of greatest interest since any improvements to these provisions is likely to have the highest impact on the technical integrity of systems in the future.

2.3.2.3 Capability, function and performance

The needed or specified characteristics of an engineered system are most often considered in terms of capability, function and performance. (Wasson, 2006, p. Ch. 4) provided a convenient definition of these terms:

> **Capability** An explicit, inherent feature activated or excited by an external stimulus to perform a function (action) at a specified level of performance until terminated by external commands, timed completion, or resource depletion.
>
> **Function** An operation, activity, process, or action performed by a system element to achieve a specific objective within a prescribed set of performance limits. Functions involve work - such as to move a force through a distance, analyse and process information, transform energy or physical properties, make decisions, conduct communications, and interoperate with other OPERATING ENVIRONMENT systems.
>
> **Performance** "A quantitative measure characterizing a physical or functional attribute relating to the execution of an operation or function." (ANSI/IEEE 649-1998, para. 3.0, p.5) "Performance attributes include: quantity (how many or how much), quality (how well), coverage (how much area, how far), timeliness (how responsive, how frequent), and readiness (availability, mission/operational readiness)." (ANSI/IEEE 649-1998, para. 3.0, p. 5).

These terms are not explicitly defined or used in ADF technical regulations, although they would seem to necessarily be related to technical integrity, particularly to the element of fitness for service.

2.3.2.4 Quality attributes

Quality attributes of a complex engineered system may include aspects like reliability, maintainability, security, flexibility and modifiability of the system among others. Often the list of quality attributes is short-handed as the *ilities*. Lack of attention to the required quality attributes at the architectural design phase can readily lead to the failure of the system to achieve its objectives ‘Functionality and quality attributes are orthogonal, ...if not orthogonal, the choice of function would dictate the level of security

or availability... Within complex systems, quality attributes can never be achieved in isolation. The achievement of any one will have an effect, sometimes positive and sometimes negative, on the achievement of others' (Bass et al., 2003, p. 72-74).

This inherent trade-off between quality attributes of a complex engineered system is not recognised by ADF technical regulations. The relationship of the *ilities* to technical integrity is not established in the technical regulations.

> Knowledge gap: There is no established relationship between the quality attributes of a complex engineered system and its technical integrity.

2.3.2.5 Unprecedented systems

Many systems acquired or developed for the ADF are *unprecedented*, that is 'a specific end product that is not known a priori, or the enterprise has limited experience in developing this type of system' (EIA, 1998a). Unprecedented systems are those 'where system operating environments or needs pose new challenges that are unprecedented' (Wasson, 2006, p. 6). Unprecedented systems pose difficulties for technical regulation since there may be no established baseline for the assessment of technical integrity.

The ADF technical regulations do not allow for the tailoring of their provisions based on the level of "unprecedented-ness" of the system to be developed.

2.3.2.6 Emergent behaviour

There have existed alternate views of emergence and emergent behaviour as it relates to complex engineered systems. (Johnson, 2006) identifies the alternate views as follows:

> 'Emergent properties' represent one of the most significant challenges for the engineering of complex systems. They can be thought of as unexpected behaviors that stem from interaction between the components of an application and their environment. In some contexts, emergent properties can be beneficial; users adapt products to support tasks that designers never intended. They can also be harmful if they undermine important safety requirements. There is, however, considerable disagreement about the nature of 'emergent properties'. Some include almost any unexpected properties exhibited by a complex system. Others refer to emergent properties when an application exhibits behaviors that cannot be identified through functional decomposition. In other words, the system is more than the sum of its component parts.

Emergent behaviour of systems or emergence is described by (Hitchins, 2007, p. 7) thus:

> Cartesian Reductionism could not explain why some wholes possess capabilities, have properties, and behave in ways that were not evident from examination of their parts in isolation. This observation was labelled 'emergence' and some wholes were observed to possess or exhibit properties, capabilities and behaviours not exclusively attributable to any of their rationally separable parts.

(Honour, 2007) nominated a categorisation of emergent properties of a system in two dimensions as reproduced in Table 3. Firstly 'Designed' (i.e. intended emergent properties of the system) and 'Surprise' (the potentially unintended emergent properties of the system). The second dimension encompasses 'Useful' (desirable emergent properties), 'Neutral' (neither desired nor undesired properties of the system) and 'destructive' (undesirable properties of the system) emergent properties.

Table 3. Types of emergent properties (Honour, 2007).

	Useful	**Neutral**	**Destructive**
Designed	Desired capabilities	Facts of design	Accepted trade-offs
Surprise	Exploitable features	Facts of existence	Fearful features

(Leveson, 2011) takes the view that safety is an emergent property of systems:

> Safety can be determined only in the context of the whole. Determining whether a plant is acceptably safe is not possible, for example, by examining a single valve in the plant... Safety is determined by the relationship between the valve and the other plant components.

The concept of emergence, or whether technical integrity is or isn't an emergent property of systems, is not evident in the ADF technical regulations.

> Knowledge gap: It is not known if technical integrity is an emergent property of a complex engineering system.

2.3.2.7 Systems of systems

There is debate in literature of the last decade as to if *Systems of Systems* is a distinct new class of engineered entity and therefore requires a new engineering approach or discipline, e.g., (Kasser, 2002, Kasser, 2009). Notwithstanding the academic debate, the US (Department of Defense, 2008) has codified Systems of Systems (SoS) as:

> A set or arrangement of systems that results when independent and useful systems are integrated into a larger system that delivers unique capabilities.

(Kossiakoff et al., 2011b, p. Ch. 3) documents the characteristic properties of SoS as:

- Operational independence of the individual system.
- Managerial independence of the individual system.
- Geographic distribution.
- Emergent behaviour.
- Evolutionary development.
- Self-organization.
- Adaptation.

SoS are not explicitly addressed by technical regulation of the ADF. Technical integrity of such systems can only be assured through the technical integrity of the systems that make up the SoS. By definition, emergent behaviour of the SoS cannot be addressed by the current technical regulations.

2.3.2.8 System complexity and the relationship with system failure

The relationship between complex systems and their relationship to disasters has most conspicuously been brought to attention by (Perrow, 1984). In his book, Perrow speaks to the inevitability of accidents in systems that are characterised as having 'interactive complexity' and 'tight coupling'. Interactive complexity is where a system may have two or more discrete failures associated with subcomponents that interact in unexpected ways. Tight coupling is where subcomponents of the system have prompt and major impacts on each other to the extent that system operators cannot intervene in a timely manner to prevent catastrophe and/or the system is so complex that the system operators cannot readily understand what is going wrong and cannot act correctly to avert an accident.

(Cook, 2000) makes similar observations about how complex system fails, starting with 'Complex systems are intrinsically hazardous systems' and while the likelihood of exposure to the hazard can be reduced through the creation of safety defences, the processes involved remain inherently hazardous. A characteristic of these systems include that they are heavily defended against failure and catastrophic outcomes. Catastrophic failure requires not a single-point failure, but multiple failures that have not been foreseen in the system design. Consequently it can be seen that safety is an emergent property of a complex system. Complex systems are also characterised as normally running in a degraded mode, i.e., one or more components may not be

functioning, but the system continues to run by means or redundant design or workarounds by system operators.

More recently, (Perrow, 2008) has published his thoughts on reducing the possibility of reducing the unexpected interactions within a complex system, and thus the potential for worst case failures, through using modularity to reduce system complexity. The advantages of making a complex system more modular in its design and construction include making it more comprehensible to its operators; making the system more testable and thereby correct; and to reduce tight coupling.

None of these concepts of system failure and complexity appear to have been taken into account in the technical regulations of the ADF.

2.3.3 Definitions of complex engineered systems

2.3.3.1 Systems engineering standards and guides

The practices of systems engineering have developed since World War II, primarily in the aerospace and military domains, to undertake the development of increasingly complex systems. Originally codified in military standards (e.g., MIL-STD-499 (USAF, 1969)), systems engineering processes have been developed and accepted more generally as the means by which to realise complex systems. The contemporary standards and associated guidebooks are therefore a key starting point for definitions of complex engineered systems.

The international standard, Systems and Software Engineering – System Life Cycle Processes (ISO, 2008) defines a system as 'a combination of interacting elements organized to achieve one more stated purposes'. The International Council on Systems Engineering (INCOSE, 2010) expands upon this definition of a system as:

> an integrated set of elements, subsystems, or assemblies that accomplish a defined objective. These elements include products (hardware, software, firmware), processes, people, information, techniques, facilities, services, and other support elements.

As another example, the definition by NASA in its Systems Engineering Handbook (NASA, 2007) is effectively the same, but with the explicit addition of enabling products being considered part of the system:

> (1) The combination of elements that function together to produce the capability to meet a need. The elements include all hardware, software, equipment, facilities,

> personnel, processes, and procedures needed for this purpose. (2) The end product (which performs operational functions) and enabling products (which provide life-cycle support services to the operational end products) that make up a system.

(Blanchard, 2008, p. 3) however, insists that:

> A greater degree of detail and precision is required to provide a good working definition acceptable for describing the principles and concepts of system engineering. To facilitate this objective, a system may be defined further in terms of the following general characteristics:
>
> 1. A system constitutes a complex combination of resources in the form of human beings, materials, equipment, software, facilities, data, money, and so on. To accomplish many functions often requires large amounts of personnel, equipment, facilities, and data (e.g., an airline or a manufacturing capability). Such resources must be combined in an effective manner, as it is too risky to leave this to chance alone.
>
> 2. A system is contained within some form of hierarchy. An airplane may be included within an airline, which is part of an overall transportation capability, which is operated in a specific geographic environment, which is part of the world, and so on. As such, the system being addressed is highly influenced by the performance of the higher-level system, and these external factors must be evaluated.
>
> 3. A system may be broken down into subsystems and related components, the extent of which depends on complexity and the function(s) being performed. Dividing the system into smaller units allows for a simpler approach relative to the initial allocation of requirements and the subsequent analysis of the system and its functional interfaces. A system is made up of many different components, these components interact with each other, and these interactions must be thoroughly understood by the system designer and/or analyst. Because of these interactions among components, it is impossible to produce an effective design by considering each component separately. One must view the system as a whole, break down the system into components, study the components and their interrelationships, and then put the system back together.
>
> 4. A system must have a purpose. It must be functional, able to respond to some identified need, and able to achieve its overall objective in a cost-effective manner. There may be a conflict of objectives, influenced by the higher-level system in the hierarchy, and the system must be capable of meeting its stated purpose in the best way possible.

The ADF technical regulations do not feature a clear definition for systems in either the shorter form of INCOSE or the more detailed form of Blanchard. Rather they more generally refer to ‘materiel’ being a superset of the complex systems of focus in this thesis and less than complex systems (i.e., Hitchins Level 1 or complexity Level 1 as defined in section 2.3.1.3).

2.3.3.2 Definitions in systems research

Despite the standardisation of systems engineering, professionals and academics continue to seek better ways to design complex engineered systems in the face of increasing scope and complexity of the systems that are to be engineered into existence. This work provides more definitions of systems from contemporary research efforts that are germane to this research project, particularly at the more complex end of the spectrum of engineered systems.

The work of (Magee and de Weck, 2002) attempted to classify a complex system as 'a system with numerous components and interconnections, interactions or interdependencies that are difficult to describe, understand predict, manage, design and/or change.' They also define an engineering system as 'a system designed by humans having some purpose'. Engineering systems are further classified as those with designed function that can transform, transport, store, exchange or control operands including matter, energy, information and value.

(DeRosa et al., 2008) assert that the essence of a complexity in engineering of complex systems is interdependence. Interdependence is where a reduction of components into a system's constituent parts is not alone sufficient to understand its behaviour. Behaviour is an emergent property which changes as elements of the system are added, removed or rearranged. A system can be designed for intended emergent properties, but it is very likely that some unintended properties will also be created.

Of particular interest, is the definition of complex engineered systems that are part technology and part human; termed composite-information/decision/action (Composite-IDA) systems by (Hitchins, 2003):

> Such systems operate at high stress levels, making great demands on operators and technologies not only for performance, but also for integrity and reliability in decision making.

In an ADF context, most Command and Control and weapons systems, which are of primary interest to this thesis, fit well within this definition.

(Sheard and Mostashari, 2008) in addressing the principles of complex systems, note that not all systems subject to systems engineering processes are complex. They state that a complex system definition consistent with that being used by researchers in complexity theory is:

> Complex systems are systems that do not have a centralizing authority and are not designed from a known specification, but instead involve disparate stakeholders creating systems that are functional for other purposes and are only brought together in the complex system because the individual "agents" of the system see such cooperation as being beneficial for them.

This is not the type of complexity subject of ADF technical regulation or being investigated by this thesis.

2.3.3.3 Complex engineered system definition

There is no universally agreed definition for a complex engineered system. The following definition for complex engineered systems is adopted to delimit the research area and problem:

> Definition: A complex engineered system is defined to be a system that has been intentionally engineered from numerous components: technological (hardware and software) as well as human, each of which has several interconnections and is interdependent, yielding a system with both intended and unintended emergent behaviour.

Of primary interest therefore is how to determine practical ways to measure the properties and characteristics that may contribute to the technical integrity of the complex engineered system.

2.3.4 Design and implementation methods

Complex engineered systems are the focus of the discipline of systems engineering. The peak professional body for systems engineering, the International Council on Systems Engineering (INCOSE), defines:

> Systems Engineering (SE) is an interdisciplinary approach and means to enable the realization of successful systems. It focuses on defining customer needs and required functionality early in the development cycle, documenting requirements, and then proceeding with design synthesis and system validation while considering the complete problem: operations, cost and schedule, performance, training and support, test, manufacturing, and disposal. SE considers both the business and the technical needs of all customers with the goal of providing a quality product that meets the user needs. (INCOSE, 2010, p. 6)

The ongoing work of (Honour, 2010) and (Elm, 2008) has demonstrated that SE is effective in assuring better cost and schedule outcomes for projects that are developing and producing complex engineered systems, but is less clear as to the effect on technical performance or quality attributes of the system.

Given that the provisions of ADF technical regulations call for the application of SE to assist in the aim of assurance of technical integrity, it would be reasonable to expect more compelling evidence as to the effectiveness of SE for assuring technical integrity. This knowledge gap was explored through secondary research the results of that study are included at Appendix H.

Similarly, the ADF technical regulations also invoke the processes of systems safety engineering in accordance with MIL-STD-882C (Department of Defense, 1993), configuration management in accordance with EIA-649 (EIA, 1998b) amongst other process provisions. There is no empirical evidence available that these provisions positively affect the technical integrity of materiel.

> Knowledge gap: There is no evidence available to support the implicit assumption that Systems Engineering or other processes positively affects the technical integrity of materiel as provisioned for in ADF technical regulations.

2.4 Conclusion

The chapter has established the context for research into the field of technical regulation and its effect on complex engineered systems. It has:

- Clarified and justified the following delimitations:
  - While technical regulation is applicable to the entire system lifecycle, including operation and maintenance, it is the provisions for design during concept and development stages of the system and if they are effective in assuring the technical integrity of a system that is of greatest interest since any improvements to these provisions is likely to have the highest impact on the technical integrity of systems in the future.
  - A complex engineered system is defined to be a system that has been intentionally engineered from numerous components: technological (hardware and software) as well as human, each of which has several interconnections and is interdependent, yielding a system with both intended and unintended emergent behaviour.
- Identified the following knowledge gaps:
  - There is no model for how technical regulation is meant to achieve its technical aims.

- There has been no cost-effectiveness or cost-benefit analysis undertaken to justify the introduction of, or extent of, regulatory frameworks.
- The goal of technical regulation is technical integrity. However, there is not a clear, objective definition of technical integrity established for the ADF and the subjective definitions vary across the materiel domains.
- Fitness for service is a key element in the subjective definition of technical integrity as the aim of the ADF technical regulation. But the definitions of fitness for service differ across the devolved regulatory domains. Fitness for service needs to be defined more objectively or a comparable concept introduced to bolster the definition and theory of technical integrity.
- It is not known if the technical integrity of a system is decomposable to the technical integrity of its subsystems.
- There is no established relationship between the quality attributes of a complex engineered system and its technical integrity.
- It is not known if technical integrity is an emergent property of a complex engineering system. While heuristically supportable, the provisions of ADF technical regulations are not objectively known to be sufficient or effective in assuring the attainment of technical integrity of complex engineered systems.
- There is no evidence available to support the implicit assumption that Systems Engineering or other processes positively affects the technical integrity of materiel as provisioned for in ADF technical regulations.
- It is not known how well the technical regulatory system deals with the current trends in the development of more complex systems and systems-of-systems which are increasingly constrained by a philosophy of procuring MOTS components to reduce development risk.

As a result of this review there is a justified research problem not previously addressed:

Research problem: How effective are the provisions of technical regulations in assuring the technical integrity of complex engineered systems?

The large scope of this research problem necessitates delimiting the scope of this thesis. In particular, a necessary sub-problem that must be addressed first has been selected as the focus of this thesis and is posed as the research question:

Research question: How can the technical integrity of a complex engineered system be measured?

The literature review in Chapter 3 will identify and critically review support for measuring technical integrity or comparable characteristics of complex engineered systems, will seek to identify extant measures for technical integrity, and more general measurement approaches that may be applicable.

3 Literature survey

3.1 Introduction

The literature survey focuses narrowly on the research question: *How can the technical integrity of a complex engineered system be measured?*

The survey began with identifying and critically reviewing support for measuring technical integrity or comparable characteristics of complex engineered systems. Extant measures for technical integrity, and more general measurement approaches that may be applicable were identified and critically reviewed on grounds including practicality, previous reliability, acceptance of results, and similarity of the measurand to technical integrity. Lastly, a review of measurement theory identified salient concepts important to consider in formulating new measurement approaches for technical integrity.

3.2 Support for measuring technical integrity

Lord Kelvin (Thomson, 1883) stated:

> I often say that when you can measure what you are speaking about, and express it in numbers you know something about it, but when you cannot measure it, when you cannot express it in numbers your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely in your thoughts, advance to the stage of science whatever the matter may be.

It has long been posited that there is no single overall measure for the quality or integrity of a complex engineered system (see for example (Boehm, 1978)). 'System Effectiveness' for a complex system, to the extent that it can be thought of as a partial proxy for technical integrity, most often appears as a combination of capability, reliability and availability in a value hierarchy (Parnell et al., 2008). While, a system effectiveness measure is usually constructed for supporting selection decisions between similar, competing systems there is little in the way of comparative measures available to compare the utility of disparate types of systems.

(Hitchins, 2003) nominates a set of common candidates for measuring the value of a system. His candidates include: life cycle cost; efficiency; effectiveness; performance; availability; survivability; process (e.g., simplicity, resource consumption, ...); product (e.g., utility, quality, fitness for purpose, …); entropy (the degree of disorder in the system). (Barbacci et al., 1995) also provides a useful set of software quality attributes encompassing performance, dependability, security, and safety. (Browning and Honour, 2008) take the view that a system's value can only be determined by the stakeholders who receive some utility from the system. These examples all present a view that there is a range of system qualities or attributes that should be able to be measured in some sense, but these qualities may cover a broader or narrower concept than technical integrity.

The increasing focus in recent years on systems architecture as a technique in the development of complex engineered systems often cites as a self justification that the measurement of system attributes akin to technical integrity is, if not impossible, at least not tractable. For example, (Maier and Rechtin, 2009) stated that 'architecture complements science where science is weakest – dealing with immeasurables of highly complex systems', and 'most important factors are not measureable: worth, safety, affordability, political importance, environmental impact ... and have no realistic basis for numerical analysis.' The problem of dealing with the attributes of 'ultra-quality' systems is also salient. For instance how can 100% availability, zero defects, etc. be practically measured? It is posited in this case that measurement can only be indirect, that an analytic design approach alone does not achieve ultra-quality, and heuristic approaches must also be applied.

'Qualities do not have to be expressed "qualitatively" – they should be quantified'. (Gilb, 2005) dedicates a chapter of his book, about the methodology of 'Planguage', to scales of measurement. He tackled concepts often claimed to be immeasurable like 'user-friendly', 'Environmentally-friendly' and 'Customer satisfaction' by specifying scales of measure. He posits principles including 'If you can't define a scale of measure, then the goal is out of control' and 'If you can't quantify it, you can't control it'. It is reasonable to assume that Gilb's views on technical integrity would be that 'defining a scale of measure is a reasonable practical process'.

Knowledge gap: There is no consensus of support for measuring abstract system properties like technical integrity in the literature.

3.3 Extant measures for technical integrity

Measures of technical integrity of a complex engineered system were searched for and no precedented measurement construct or instrument has been identified. This section then identifies and assesses measures determined to have some full or partial similarity to technical integrity and therefore be of some utility to the research question.

3.3.1 Life cycle value of systems

(Browning and Honour, 2008) posit that 'a goal of systems development is to produce enduringly valuable product systems, i.e., systems that are valuable when delivered to their users and which continue to be attractive to their stakeholders over time. However, quantifying the life-cycle value (LCV) provided by a system has proven elusive'. The paper reports on attempts to quantify LCV via 'a conceptual exploration of some of the key ideas surrounding LCV and the genesis of a proposed approach for quantifying a system's LCV'.

There are similarities, or at least similar difficulties, in the measurement of technical integrity and of LCV, in particular the subjective nature of the concepts.

> The value provided by a system, let alone its LCV, is difficult to quantify. Value is largely subjective, and individuals have difficulty articulating exactly what makes a complex system valuable. A particular stakeholder in the development or use of a complex system is typically represented by a group of individuals with conflicting opinions. Plus, there are usually a variety of stakeholders, and thus an even greater likelihood of conflicting points of view on value. Moreover, value is not only absolute, i.e., based on the intrinsic attributes, functions, and performance levels of a system, but also relative, based on perception and depending on the availability and attributes of substitute products, services, or approaches for meeting needs and wants. Finally, stakeholder needs and wants, and the perceived value of a system, change over time, thus making LCV even more amorphous. Clearly, designing for maximum LCV is not easy when even the objective itself is difficult to specify.

(Browning and Honour, 2008) take the view that a system's value can only be determined by the stakeholders who receive some utility from the system. The list of stakeholders they considered is very broad; broader than the group of stakeholders who would directly value technical integrity goals alone. In the end, a conceptual approach only was presented rather than a validated measurement instrument available for use or adaption.

3.3.2 Stakeholder value of a system

If the view is taken that technical integrity can only be approached as a set of value judgments from stakeholders, a means of gathering these judgments is necessary. (Trainor and Parnell, 2007) discuss how to use interviews, focus groups, and surveys for stakeholder analysis. Of interest to the measurement of technical integrity is their view of value versus system requirements: ‘

> Value is a perceived quality stemming from subjective preferences. Stakeholder preferences are distinct and different from requirements. Requirements represent a choice made to achieve a specific level of performance and specify acceptability. A system that meets its requirements will provide different value to different stakeholders, depending on their preferences.

They posited the concept of ‘benefit’, the things the system gives a stakeholder, vice ‘sacrifices’, the things a stakeholder must give up to have the system. Potential approaches for combining stakeholder preferences into an overall value were identified including simple weighted average of parameters, geometric average, analytic hierarchy process (AHP), and multi-attribute utility theory (MAUT), without concluding which of these would be best or practical. Technical integrity may be thought of as some acceptable balance between benefit and sacrifice and therefore there may be some applicability in the approach of combining stakeholder preferences into a measurement construct.

3.3.3 System quality attributes

The technical integrity of a system is likely to be a combination of, or related positively to, one or more of the quality attributes of a complex engineered system and these attributes may be more readily measurable. (Firesmith et al., 2009b) defines a quality model for a system including:

- Quality characteristic: a high-level characteristic or property of a system or architectural component that characterizes an aspect of its quality. Quality characteristics are also known as quality factors, quality attributes, and ‘ilities’. Note that unlike quality attributes, quality characteristics are typically at such a high level of abstraction that they are not measurable along quality measurement scales using quality measurement methods.
- Quality attribute: a major component (aggregation) of a quality characteristic or of another quality attribute that is directly measurable.

- Quality measurement scale (or a quality measure): a measurement scale that defines numerical values used to measure the degree to which a system or architectural component exhibits a specific quality attribute.
- Internal quality characteristics - a quality characteristic that defines an internally visible (i.e., white box) quality of a system or architectural component when it is in the process of being developed, modified, or retired. Internal quality characteristics are visible to and are primarily of interest to the system's developers and maintainers. Refer to Figure 6 for an example of important internal quality characteristics.

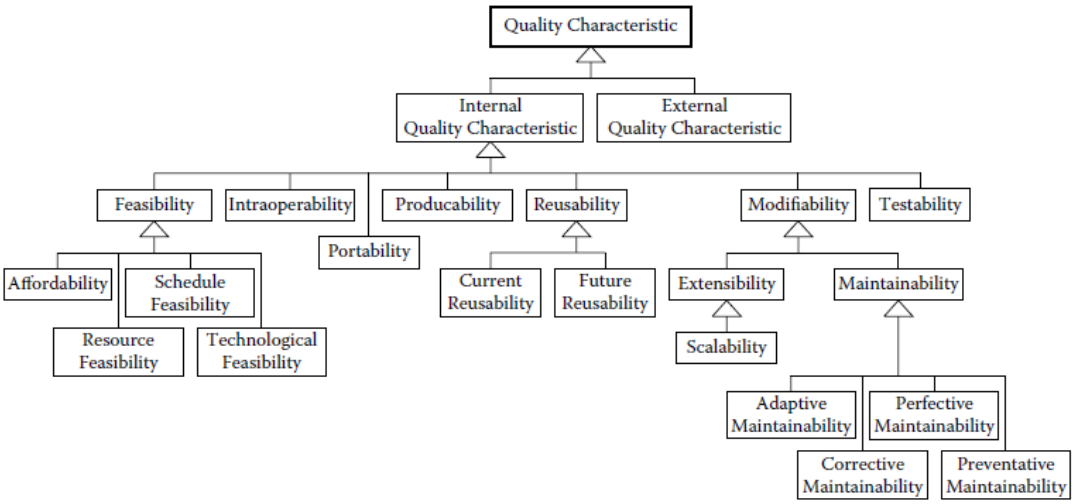


Figure 6. Example partial hierarchy of important internal quality characteristics (Firesmith et al., 2009c, Fig. 18.4).

- External quality characteristic (or a black box quality characteristic): a quality characteristic that defines an externally visible (i.e., black box) quality of a system or architectural component when it is deployed and in service in its operational environment. External quality characteristics are visible to and are primarily of interest to the users, operators, administrators, and sustainers of an existing system or architectural component. Refer to Figure 7 for an example of important external quality characteristics.

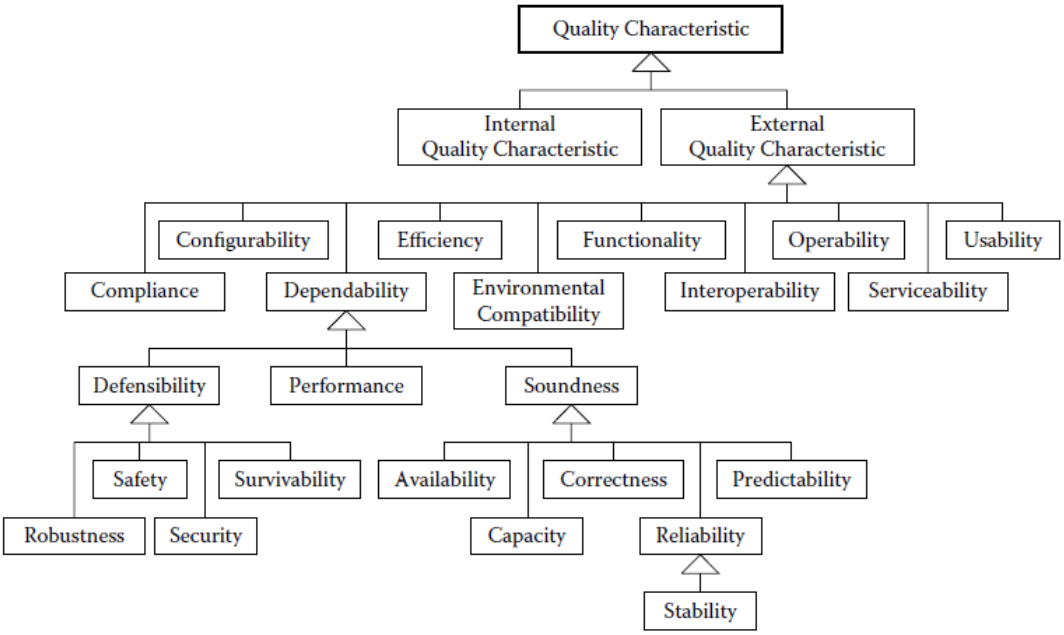


Figure 7. Example partial hierarchy of important external quality characteristics (Firesmith et al., 2009c, Fig. 18.5).

(Firesmith et al., 2009c) included full definitions for each of these characteristics in their model. The definitions indicate some of the expected measures for the characteristic, for example:

> Serviceability is typically measured in terms of financial cost, staff effort, or staff effort per unit time to service the system or component in specified ways under specified conditions.
>
> Reliability is frequently measured in terms of the conditional probability that an architectural element remains functioning for a given duration or the mean time to first failure (MTFF) or mean time between failures (MTBF).

This hierarchical approach and the broad range of definitions available in the work of (Firesmith et al., 2009c) do not provide a viable measurement construct for technical integrity, but do provide a good basis of characteristics and definitions that should be useful in forming such a construct.

3.3.4 Dependability assessment

The system quality attribute of "dependability" is likely to be at least positively correlated with the concept of technical integrity. (Fenton et al., 1998) address the assessment of the dependability of safety critical systems using diverse evidence. They

term this ‘DATUM (Dependability Assessment of safety critical systems Through the Unification of Measurable evidence)’.

A particular difficulty is how to assess reliability in the presence of design faults and human errors in operation. (Fenton et al., 1998) attempted to combine disparate sources of information – actual failure rates, engineering judgments, objective measure of the quality of architecture and design – in order to have a measure of a system’s dependability. They looked at methods to combine the uncertain input data including Bayesian belief networks (BBNs), Dempster-Schafer theory, and fuzzy sets and selected BBN as a preferred approach.

While not a direct solution for the measurement of technical integrity, the approaches within DATUM may be applicable to identify starting points for development of a measurement instrument.

In concluding this section on identifying extant measures for technical integrity, it is apparent that no wholly applicable measurement approach for technical integrity has been identified.

> Knowledge gap: There is no extant measurement construct or measurement instrument for technical integrity.

3.4 Measurement approaches to be considered

This section is focused on identifying and evaluating methods and techniques for estimating and combining quantitative and/or qualitative values of system characteristics that may be used as the basis for the development of a measurement construct and instrument for the technical integrity of a complex engineered system.

3.4.1 General considerations

(Hubbard, 2007) includes a number of insights suitable to begin the consideration of approaches for measuring technical integrity:

- Measurement is a set of observations that reduce uncertainty where the result is measured as a quantity.

- It is usual for the value of an information measurement to be inversely related to the measurement attention it gets – a ‘measurement inversion’.
- In cases of large uncertainty, measure a little bit, if something unexpected shows up it is probably worth refining the measurements.
- Break up uncertain variables into more readily measurable parts – ‘decomposition’.
- Be cognisant of sources of errors:
  - Systemic error/bias; the tendency of measurement technique to favour a particular outcome. *Accuracy* is a measure of how low the systemic errors are.
  - Random error; error that is not predictable or not dependent on the known variables. *Precision* is a measure of how low the random errors are.
- Be cognisant of observation biases:
  - Expectancy bias; seeing what you want to see.
  - Selection bias; arising from inadvertent non-randomness.
  - Observation bias; the act of observing may change what is observed.

3.4.2 Proxies for Technical Integrity

In considering the measurement of technical integrity it is worth identifying any other measures that have been offered as a proxy for technical integrity or at least could be considered as a valid full or partial proxy. One particularly interesting set of data suggests consideration of accident rates of a complex system in service be a proxy for technical integrity. Figure 3 shows the yearly number of accidents being used to demonstrate the positive impact of technical and operational regulation on the fatal accident rate of ADF aircraft. It could be implied that the technical integrity of ADF aircraft has increased markedly since 1992 if it is assumed that the fatal accident rate is negatively correlated with technical integrity. But how much of what is defined as technical integrity is covered by this potential proxy measure?

It is feasible that these numbers could be taken to form a measurement of the accident rate of the entire ADF aviation capability over a defined time frame and that this could have some validity as a proxy measure for technical integrity. However, it is noted that this type of measure may be easily confounded by issues like the numbers of aircraft in service in any particular year, the technology of the aircraft in service, the experience level of aircrews, the type of flying (operations versus training), sustainment spending

and perhaps other variables. All these would need to be carefully controlled, or at least understood, in attempting to use this as a proxy measure. Another consideration discounting this approach is that it would only provide estimates of technical integrity outcomes some considerable time after the system has been in service and therefore would not provide a timely measure that could provide feedback for the effectiveness of regulatory provisions during the development of the system.

3.4.3 The measurement of risk

Risk is most usually characterised in terms of consequence (the undesirable outcome) and likelihood of occurrence of each identified risk. Analysis of, or the assignment of a value to, the risk may be conducted qualitatively, semi-quantitatively or quantitatively (Standards Australia, 2004).

In the context of ADF technical regulations and technical integrity of complex systems, risks to technical integrity can be thought of as faults, flaws, or unintended emergent behaviour of systems that give rise to unintended and undesired consequences. Technical regulation would seek to minimise these risks by either or both of lowering the severity of consequence and lowering the probability of occurrence. In particular, in terms of the component parts of technical integrity:

- Safety – the minimisation of unintended harm to people,
- Environment – the minimisation of unintended harm to the environment, and
- Fitness for service – the minimisation of the chance of the system not working as required in the intended operational environment.

ADF technical regulations tend to operate at the qualitative end of risk measurement, for example the Navy's hazard risk calculator (Department of Defence, 2006). There is no concept of cumulative technical integrity risk, i.e., no defined means to aggregate the value placed on each of the individually identified risks into some overall estimate of the risk associated with a system. Any aggregation is left implicitly for the subjective judgment of the decision makers accepting the system for operational use.

In identifying individual risks, there is clearly a probability that there will be unidentified risks in any complex engineered system. These risks are not estimated by the risk management processes provisioned for in ADF technical regulations.

Increasingly, formal risk analysis techniques based upon quantitative risk analysis approaches are being adopted to improve the safety and reliability of engineered systems without the conservative and more costly approach of regulation for complete risk avoidance that can lead to over design (Modarres, 2006, p. 4). This leads to the development of risk-informed regulations of the sort in place in regulations of the US Department of Energy for nuclear power plants (Lee and McCormick, 2012a). Probabilistic risk assessment (PRA) is an increasingly common quantitative risk assessment technique where uncertainty of both likelihood of occurrence of causal factors and hazards as well as uncertainty in the magnitude of associated consequences are estimated (Modarres, 2006, p. 33-111). Similar approaches called Scenario-Based Hazard Evaluation include the assignment of a risk value to each identified use scenario of a system (Center for Chemical Process Safety, 2008).

Of course, if many risks are estimated they are not likely to be independent. They will be correlated in the sense that if conditions are such that one unintended consequence manifests, it is more likely that other unintended consequences will manifest. Consider a case where failure of a component in a system raises the overall stress in the system making it more likely for other components to fail leading to either a magnified consequence or to multiple disparate consequences, a positive correlation between risks. Of course negative correlation of risks may also occur. These correlations are accounted for in the sophisticated risk modelling of the financial industry, see (Marrison, 2002) for example, but have not been found in the risk management paradigms for engineered system.

Other views as to the tractability of measuring risk of a system exist, for example, (Hubbard, 2009) identifies methods associated with the identification of risks, the quantification of each identified risk in terms of likelihood and consequence, and the assessment of likely outcomes due to the effect of all identified risks through Monte-Carlo modelling. This is in the same vein as approaches like PRA discussed previously, being quite specific to a particular system. However, the gap to be highlighted is the difficulty in measuring the cumulative risk in a complex engineered system in some consistent way. In financial and even most civil systems the consequence of failure may be reasonably converted to dollar values and combined weighted by likelihood of occurrence. But there is not a reasonable dollar or other value conversion for military systems that has been identified or generally accepted (e.g. how many dollars do we

attribute to a design flaw in a missile system that does not work as intended when required?).

Qualitative risk assessment procedures commonly required by ADF technical regulations, typically require analysis quite specific to a particular design and do not lend themselves to provide a generic measurement instrument for the risks inherent in an engineered system. Qualitative risk assessment is not able to be objectively transformed to a cumulative risk assessment for the engineered system. Moreover, quantitative risk analysis is not currently required for systems developed for the ADF, so the data required to support development of a measure of technical integrity would not be available.

The conclusion is that there is little available to practically and generically measure the cumulative risk to technical integrity of a complex engineered system in the ADF.

> Knowledge gap: There is no pragmatic measurement approach for estimation of the cumulative risk (unintended, undesired outcomes) of a system or to compare that risk across a class of similar systems or across dissimilar systems.

3.4.4 Consensual Assessment Technique (CAT)

Teresa Amabile has developed and validated a consensual technique for creativity assessment – termed the Consensual Assessment Technique (CAT) – which establishes the creativity embodied in a sample of products as judged by a range of judges. What Amabile has established, over a large number of studies, is that creativity can be reliably judged by appropriate judges (persons who are familiar with the domain, but not necessarily experts in the domain). Despite the fact that 'creativity may be something difficult to describe, but it is still relatively easy for them [i.e., the judges] to identify [creativity] with a good degree of reliability' (Amabile, 1996).

The main features of the CAT include:

- Judges need to be familiar enough with the domain to have some implicit criteria for 'technical goodness'.
- Judges make their assessment independently.
- Judges are asked to make assessments of dimensions other than creativity of the product (which allows the researcher to check for confounding effects).

- Ratings of products are made relative to one another rather than against an absolute standard.
- Judges consider the products in a randomized order.
- Judges ratings are analysed for interjudge reliability (equivalent to construct validity).
- Specific objectively discernible characteristics of the products are recorded so correlation of these with the judgements may be assessed.

Amabile reports that the CAT has also been applied outside visual and verbal artistic domains, to creativity in problem-solving tasks including computer-programming, high-technology services and design of structures with good reliability results.

While technical integrity of a complex engineered system and the creativity embodied in an artwork or other produced item are not direct analogues, there are enough similar characteristics of the problem space to suggest investigation into if the CAT technique can be successfully applied to establishing the technical integrity of a complex engineered system by a group of appropriate judges. Similarities of the problem space include:

- Both technical integrity and creativity are difficult to define precisely.
- They are not directly measurable.
- There is no complete nor agreed set of measurable components for either.
- Both represent a quality of a product/system that people familiar with the domain are readily willing to subjectively judge in a relative sense (e.g. artwork A exhibits higher creativity than artwork B; system 1 has high technical integrity than system 2).

3.4.5 Multi-attribute approaches for ranking systems

MAUT is a common approach used for ranking alternatives in the management sciences to support decision making. It was established in the late 1960s and was well described and practised by the late 1970s (Humphreys, 1977) and continues to be actively applied and developed (Wallenius et al., 2008).

MAUT could be used to come up with preferences, that is, the rank ordering of systems according to a utility function constructed for technical integrity. Difficulties to overcome include justifying the assumption that technical integrity can really be

modelled as an additive sum of its (presumed) constituent attributes; arriving at common measures for each of the attribute scales and how to convert these to utility scales that would be valid to a range of stakeholders and across a reasonable range of complex system types; arriving at a set of weights that could be common across all, or at least a reasonable range, of complex system types.

A limitation would be that the differences in technical integrity between systems could not be ascertained and more importantly the degree to which the application of technical regulation actually impacts technical integrity could not be quantified.

3.4.6 Measuring technical performance

Consideration is given in (Garvey, 2009, p. 181-193) as to how to arrive at a meaningful measure of the risk of a system not achieving its technical performance. The approach considers how the Technical Performance Measures (TPM) maintained during the development of the system can be combined to measure the overall performance risk of a system. The approach assumes that a sufficient set of TPMs needed to assess overall technical performance are defined and measured. It converts each TPM measured in its natural scale to 'common and dimensionless scale' to form a Technical Risk Index (TRI) and then forms a System TRI by combining all the component TPIs via a simple weighted sum where the sum of the weights is equal to unity.

The method gives no insight in how to objectively determine sufficiency for the set of TPMs to be considered or how to construct the set of weights necessary to form the System TRI.

3.4.7 The risk-value method; and composite performance model (CPM)

(Browning et al., 2002) proposes the risk-value method where the performance parameters of a system are identified and the likelihood of obtaining those performance values is modelled by a probability density function (PDF) and a utility curve is constructed to document the customer's preferences for a range of performance levels for each performance parameter. The gap between the utility of the attained outcome of the performance parameter to the utility of the target performance is used with the PDF to form a measure of the performance risk in that performance dimension of the system. This method proposes to combine the risk values into an overall performance risk

measure by means of a weighted sum. No insight is given to how to objectively obtain the weights.

Appendix A to (Browning et al., 2002) provides a summary of other techniques that have been pursued in the search for a Composite Performance Measure (CPM). The possibilities identified include weighted arithmetic means where weights may be determined by eliciting customer preferences using techniques like Analytic Hierarchy Process; weighted geometric means, MAUT and others. The authors note that a CPM should cover all the requirements values by the customer; complex systems have thousands of requirements so a measure of non-conformances may be a proxy for the CPM; CPMs are necessarily an abstraction and therefore some information is omitted or at least obscured in forming the CPM.

Such an approach to measuring technical integrity via utility curves of stakeholder preferences is not considered practical due to the diverse range of stakeholders for whom high technical integrity is a desired outcome. However, the use of requirements non-conformance as at least a partial proxy measurement that may indicate the level of technical integrity of a system is a practical concept noted for further investigation.

3.4.8 System usability scale

A practical approach to estimate the 'usability' of a system was developed by (Brooke, 1996) and is termed the system usability scale (SUS). He defines usability as follows:

> Usability is not a quality that exists in any real or absolute sense. Perhaps it can be best summed up as being a general quality of the appropriateness to a purpose of any particular artefact.

SUS is a Likert scale where a survey instrument questions users of the system about usability characteristics of the system and their responses combined to form a measure of usability. (Brooke, 1996) claims SUS to be 'robust and reliable' but provides no information as to the reliability of the instrument or steps taken to confirm its validity.

Noting the parallels between the characteristics of usability defined by Brooke and technical integrity, the development of a scale using a similar approach to SUS, may be a means to obtaining a practical measurement instrument for technical integrity and is noted for further investigation. However, the reliability and validity of such a scale would need to be established.

3.4.9 International nuclear event scales (INES)

(Saji, 2003) reviewed the safety concept of the International Nuclear Event Scales (INES). This is a scale adopted in the nuclear power community to consistently report the safety significance of events reported at nuclear power plants. Qualitatively this scale rates events in ranges from Safe (No Safety Relevance), Deviation (No Safety Significance But Safety Related), Incident, Accident, Unsafe. The author introduces the thought that 'more frequent minor events, which challenge safety barriers, can eventually develop into a severe accident. These frequent troubles are inducing fears among the general public against the safety of nuclear power plants ...'

A measure of safety should, arguably, be reflective of the actual or anticipated number of severe safety events that are associated with a system (i.e. those in region A in Figure 8). However the difficulty in trying to form an objective measure is that relatively few (statistically insignificant number) of these severe safety events actually occur for most systems of interest. The view of (Saji, 2003) is that at least the public perception of the safety of a system, is driven by the relatively larger number of minor events. Larger numbers of minor events/deviations (i.e., unwanted, unintended system behaviour that do not result in a severe safety incident) occur. The number of these occurring is a measure of the quality or dependability of the system, i.e., the systems inherent ability to exhibit intended behaviour and not to exhibit errant behaviour under conditions of error or stress. Establishing if this measure can be taken as a proxy measure for the safety of the system as an element of technical integrity is of interest to this research project. It provides a potential solution to the problem of unidentified risks discussed in section 3.4.3.

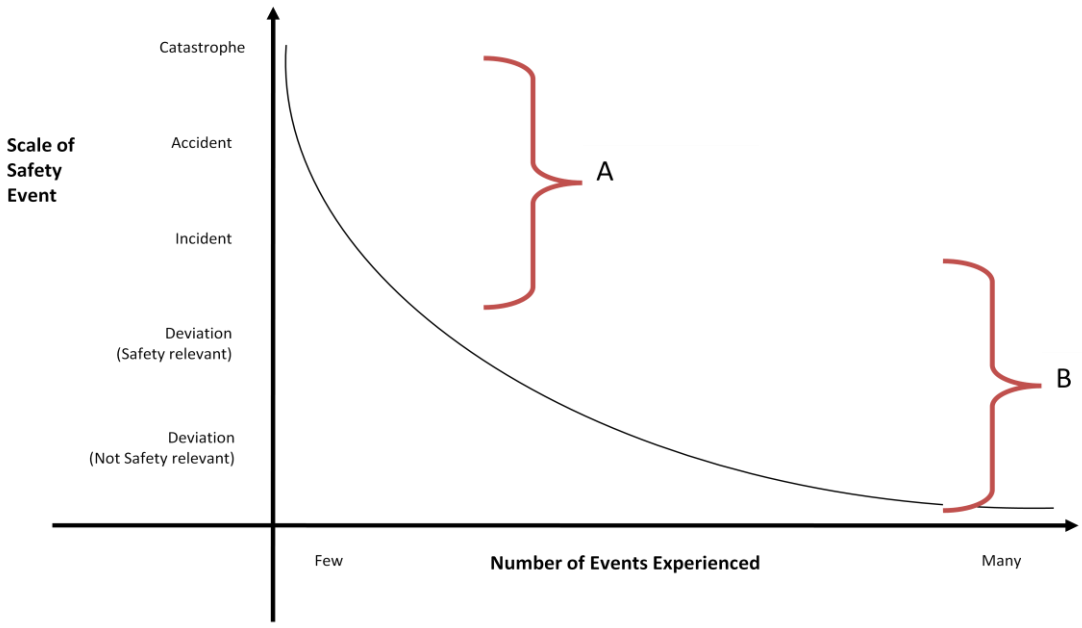


Figure 8. Magnitude of safety events versus the number of events experienced.

3.5 Measurement theory

This section does not narrate the development of the theories of measurement, these being well documented by (Sydenham, 1982, Allen and Yen, 1979, Roberts, 1984) among others. Rather the breadth of measurement theory and its attendant concepts and terminology are summarised to provide a basis for formulating and evaluating candidate measurement approaches for technical integrity.

3.5.1 Definition and key concepts of measurement

3.5.1.1 Definitions

The definition of measurement has been well established and is not in dispute. (Kaplan, 1964, p. 177) defined it thus:

> Measurement, in most general terms, can be regarded as the assignment of numbers to objects (or events or situations) in accord with some rule. The property of the objects which determines the assignment according to that rule is called magnitude, the measurable attribute; the number assigned to a particular object is called its measure, the amount or degree of its magnitude.

Further, (Sydenham, 1982, p. 2) stated:

> Measurement is the process of empirical, objective assignment of numbers to the properties of objects and events of the real world in such a way as to describe them.

Similarly, (Allen and Yen, 1979, p. 2) who focused more on the development of psychological tests and scales, state:

> Measurement is the assigning of numbers to individuals in a systematic way as a means of representing properties of the individuals. Numbers are assigned to the individuals in a carefully prescribed procedure.

(Sydenham, 1982, p. 6) stated clearly that measurement is the assignment of numbers to describe the properties of objects and events and therefore that 'measurement presupposes the existence of a clear concept of a property as an abstract aspect of a whole class of objects'. This indicates a clear concept of technical integrity needs to be established as a precondition to development of a measurement instrument.

It is important to declare and use clear terminology, and there is variation in the literature. For the purpose of this thesis, the following definitions are adapted for use from (Leedy and Ormond, 2005, p. 91-93):

- Measurement *construct* – a relationship between variables. At its simplest form it is the relationship between the property to be measured and the mathematical representation to be used as the measure. A measurement construct could also be the relationship between the underlying physical property to be measured (e.g., temperature of a material) and the physical manifestation of that property than can be directly measured (e.g., the resistivity of the material). The measurement construct can be extended to combine these variables and many others to form more complicated constructs.
- Measurement *instrument* – a practical means by which a number can be obtained. To extend the previous example, a measurement instrument could be a Wheatstone bridge to measure resistance, coupled with analytical equations to convert to resistivity and to temperature of the material. In non-physical areas like psychology, a measurement instrument may be a series of questionnaires, the answers to which are compiled to provide a categorical or numerical measurement.

3.5.1.2 Practical considerations and examples

Considerations for the measurement of the property of an entity are not always straightforward and the consideration of some examples is illuminating (after (Sydenham, 1982)):

- Considering an observable, physical object, it is relatively straightforward to measure its length or its volume. However, consideration of other physical properties soon becomes more problematic. For instance the hardness of a mineral is not measured by extensive means but rather by Moh's ranking scale. Difficulties rise when physical properties that may be considerably more subjective are to be measured: consider the spreadability of butter or the foldability of paper.
- In the behavioural and social sciences, scales are developed for properties that may not have any physical analogue, e.g., measurements for taste, smell and intelligence. More subjective are attempts to define measures like the 'standard of living' where individuals may have wide points of view.
- Some properties of objects may be considered so subjective that they cannot be declared measurable. 'Beauty' is cited by (Sydenham, 1982, p. 11) as a case where 'there is not a objective rule for classifying some aspects of observable objects as manifestations of beauty' and 'the basis for measurement … is thus absent from the outset.'

Technical integrity measurement is likely to not be observable like a physical object in the first example. But it is maintained to be not immeasurable since this would contradict the significant efforts to regulate for it. So it is likely that measurement of technical integrity falls into the second case above, and therefore, the techniques of the social sciences are likely to be applicable.

3.5.1.3 Objectivity

Key to measurement is that it must be an objective process: 'the numbers assigned to a property by measurement must, within the limits of error, be independent of the observer' (Sydenham, 1982, p. 7) and 'the objectivity of measurement makes measures invariant and indisputable in logical discourse' (Finkelstein, 2003). In his 'wide sense of measurement definition' (Finkelstein, 2003) also noted two other key pragmatic elements of objective measurement:

> Measurement is free of the ambiguity and vagueness of other forms of description, say those of natural language.
>
> Measurement is not mere naming, but provides information about the relation of a particular manifestation of a quality and other manifestations of the same quality.

Clearly, objectivity is an important consideration in the selection or development of a measurement instrument for technical integrity.

3.5.1.4 Precision, uncertainty and value

Commonly, many consider measurement to imply preciseness. The 'limits of error' are implied to be small and anything that is not precise cannot be considered to be measurement. (Hubbard, 2007) made the observation that in business many people consider a business value that is 'intangible' to be synonymous with 'immeasurable'. Rather, he pointed out that 'measurement' does not imply 'exactness' or 'certainty'. Rather, measurement should be considered to be a set of observations or estimates that reduce uncertainty and is expressed as a number. The amount of effort or cost in conducting the measurement activity can be justified in terms of how much the reduction in uncertainty is of value to the undertaking or problem at hand.

A measure for technical integrity will be valuable in order to confirm the effectiveness of the provisions of technical regulation and only needs to be precise enough to fulfil this aim. This needs to be considered in development of a measurement instrument for technical integrity; the cost of measurement must be justifiable in terms of value.

3.5.2 Basics of measurement theory

3.5.2.1 Forms of measurement

The following forms of measurement are generally defined (Sydenham, 1982, p. 12-17):

- *Direct measurement*, including:
  - *extensive measurement* where something like a physical attribute of an object can be directly observed (e.g., a linear measurement),
  - *matching scale* where measurands are compared to members of a standard set (e.g., colour matching),
  - *ranking scale* where a measurands is compared to members of a ranked standard set (e.g. Moh's scale of hardness of minerals).
- *Indirect measurement*, including:
  - *Derived:* Establishing a relation between two or more things that can be directly measured (e.g., density which is established from the direct measurement of volume and mass).
  - *Associative*: Measurement of at least one other property and the relationship of that to the desired property (e.g., measure resistivity to determine temperature).

- *Multi-dimensional*: A collection of component vectors without combination (e.g., a vector that defines the lengths of the 3 sides of a triangle).
- *Conjoint*: A combination of component properties which establishes an ordering of the desired property based on measurable component properties (e.g., the establishment of a 'comfort' measurement by combining temperature, wind speed and relative humidity measurements).

3.5.2.2 Levels of measurement

'Levels of measurement', as defined by (Allen and Yen, 1979, p. 6-8, 179-190), or 'scales of measurement' as termed by (Leedy and Ormond, 2005, p. 25-27), include four types of measurement:

- *Nominal* where the measurement is simply ascribing a name to a property.
- *Ordinal* where the measurement is ascribing a symbol or value to a property the symbols allow for rank-ordering.
- *Interval* where the measurement is ascribing a value to a property on a scale with equal units of measurement, but no absolute zero point.
- *Ratio* where the measurement is ascribing a value to a property on a scale with equal units of measurement, and an absolute zero point.

3.5.2.3 Characteristics of measurements

(Allen and Yen, 1979, p. 6-8) define four characteristics of measurements:

- Distinctiveness – 'different numbers are assigned to individuals with different values of the property being measured',
- Ordering in magnitude – 'larger numbers representing more of the property being measured than smaller numbers'.
- Equal intervals – 'equivalent differences between measurements represent the same amount of difference in the property being measured'.
- Absolute zero – 'a score of 0 represents an absence of the property being measured'.

The characteristics of each of the levels of measurements and the allowed mathematical and statistical operations that may be applied to each level of measurement is shown in Table 4 (after (Allen and Yen, 1979, p. 7), (Sekaran, 2003, p. 189) and (Leedy and Ormond, 2005, p. 28)).

Table 4. Characteristics of the levels of measurement.

	Level of measurement			
Characteristic	*Nominal*	*Ordinal*	*Interval*	*Ratio*
Distinctiveness	✓	✓	✓	✓
Ordering in magnitude		✓	✓	✓
Equal intervals			✓	✓
Absolute zero				✓
Mathematical operations	set membership	<,>,=	+, -	÷,×
Central Tendency	mode	median	arithmetic mean	arithmetic or geometric mean
Dispersion		percentile rank, interquartile range	variance, standard deviation	variance, standard deviation, percentage variation
Statistical tests	χ^2	rank order correlation	Pearson product moment correlation *t, F* tests	most inferential statistics *t, F* tests

3.5.2.4 Strongly and weakly defined measurement

There is some dispute in the literature about what truly constitutes measurement with a strong view that the paradigm of measurement as conducted by the physical sciences is normative. To resolve the dispute (Finkelstein, 2003) identified this physical science paradigm as 'strongly defined measurement' that is based upon:

(i) precisely defined empirical operations,
(ii) mapping on the real number line on which an operation of addition is defined,
(iii) well-formed theories for broad domains of knowledge.

He then defines 'weakly defined measurement' as 'measurement that constitutes representation by symbols of properties of entities of the real world, based on an objective empirical process, but lacks some, or all, of the above distinctive

characteristics of strong measurement'. An approach to measure technical integrity may be such a weakly defined measurement when compared with the features of such a measurement as defined by (Finkelstein, 2003):

> (i) it is based on an ill-defined concept of the quality,
> (ii) there is significant uncertainty in the empirical relational system that it represents
> (iii) the symbolic relational system has limited relations defined on it,
> (iv) there is no adequate theory relating the measurement to other measurements in the same domain.

3.5.3 Reliability and validity of measurement

3.5.3.1 Reliability

Reliability of a measurement construct and instrument is an assessment as to the degree to which measured values are correlated with the true value of the property. In many cases, where the value of the property cannot be directly observed or estimated, reliability assessment focuses on the correlation of observed scores on parallel applications of the instrument. That is, does the instrument objectively provide the same value when used by different people and/or at different times?

The reader is referred to (Allen and Yen, 1979, p. Ch 4) and (DeVellis, 2011, p. Ch 3) for a detailed description of measurement reliability.

3.5.3.2 Validity

Validity of a measurement instrument is concerned with establishing the extent to which the instrument actually measures what it purports to measure. Validity is broken down to components that include:

- Content validity that 'is established through a rational analysis of the content of a test, and its determination is based on individual, subjective judgment' (Allen and Yen, 1979, p. 97) and 'is intimately linked to the definition of the measurement construct' (DeVellis, 2011, p. 60).
- Criterion related validity that is established if a measurement has 'an empirical association with some criterion or putative "gold standard." Whether or not the theoretical basis for that association is understood is irrelevant' (DeVellis, 2011, p. 61-64).
- Construct validity that 'is directly concerned with the theoretical relationship of a variable to other variables. It is the extent to which a measure "behaves" the way

that the construct it purports to measure should behave'. (DeVellis, 2011, p. 64-67). The measure should be positively correlated with variables the theory predicts it should be and negatively correlated with variables the theory predicts it should be (convergent validity), and uncorrelated with variables that the theory indicates are unrelated (discriminant validity).

Some differences in terminology are noted to exist. For instance construct validity can also be known as 'congruent validity' and content validity known as 'logical validity' (Sekaran, 2003, p. 204). The reader is referred to (Allen and Yen, 1979, p. Ch 5) and (DeVellis, 2011, p. Ch 4) for a detailed discussion of measurement validity.

3.5.4 Aggregation measures

This section presents the measurement theory applicable to approaches that aggregate other measures in order to form a measurement construct of more complex concepts. These are likely to be applicable to the measurement of technical integrity.

3.5.4.1 Scales and indices

Of particular interest to the measurement of technical integrity is the potential use of measurement scales. SUS identified in section 3.4.8 was a precedent for the use of scales to define otherwise unobservable properties of systems. (DeVellis, 2011, p. 11-12, 17) defines scales thus:

> Measurement instruments that are collections of items combined into a composite score and intended to reveal levels of theoretical variables not readily observable by direct means are often referred to as *scales*. We develop scales when we want to measure phenomena that we believe to exist because of our theoretical understanding of the world but that we cannot assess directly.
>
> The underlying phenomena or construct that a scale is intended to reflect is often called the *latent variable.*

The latent variable is associated with a common cause and may manifest many effects. If these effects can each be measured by indicators, then these indicators can be combined to form a scale. The scale is then the measurement instrument for the latent variable as illustrated in Figure 9.

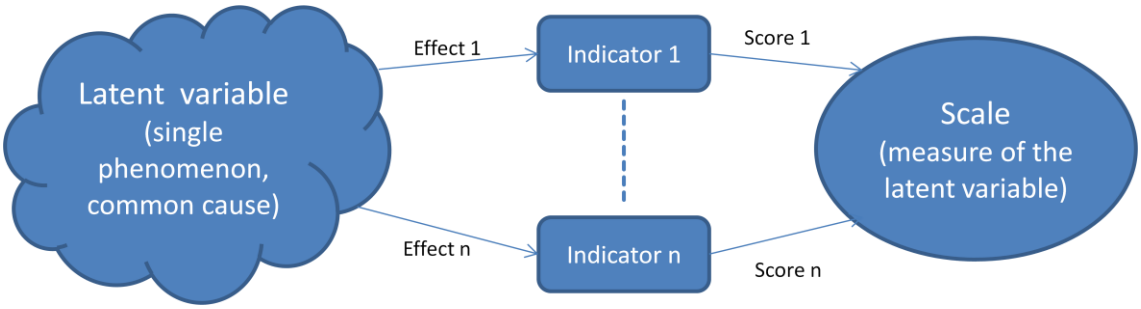


Figure 9. Concept of a measurement scale.

A necessary conclusion is that since the latent variable can be considered to be the cause of all the effects, then an increase in the latent variable should cause an increase in each effect. That is, each indicator will be correlated with the latent variable which in turn means each indicator will be correlated with each other indicator (DeVellis, 2011, p. 19-20). The scale is normally calculated by an additive aggregation of the normalised indicator scores.

In comparison, an *index* is not a measurement of a single phenomenon, but rather, it is the composite of the measures of the effects of two or more phenomena as illustrated in Figure 10. Other terms for index noted in the literature is 'emergent variable' (DeVellis, 2011, p. 12-13) and 'composite indicator' (Nardo et al., 2005).

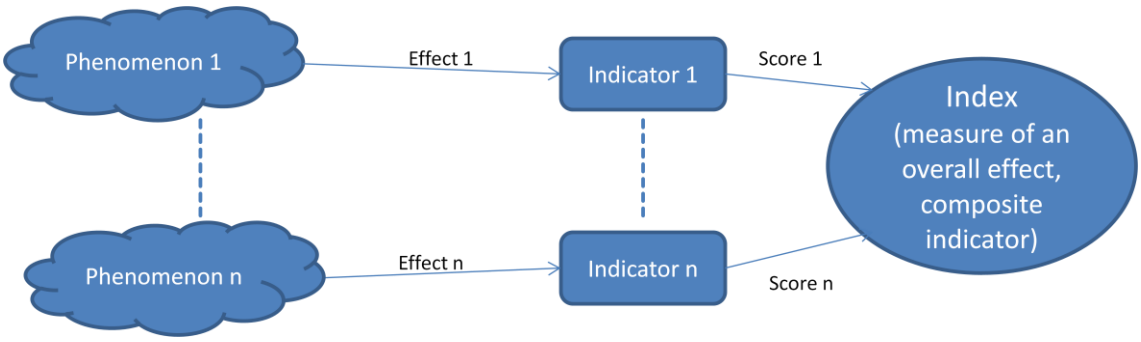


Figure 10. Concept of a measurement index.

In this case, the indicators that contribute to the index are in general not correlated with each other. The index may be arrived at by additive aggregation of the indicators or by geometric aggregation of the normalised indicators. Geometric aggregation allows for the composite index to have a low value when one of the indicators is low in value, irrespective of the other indicators having a relatively high value or not.

3.5.4.2 Conjoint measurement

Conjoint measurement is a measurement approach where objects that can be simultaneously described by several objects are assigned numerical values that preserve ordering of the objects by a qualitative relation such as preference or likelihood. The theory for conjoint measurement is not new, having been established by (Luce and Tukey, 1964) and extended to include polynomial conjoint measurements by (Tversky, 1967). 'Conjoint measurement is considered important for the measurement of non-physical attributes, such as psychological attributes' (Michell, 2005).

'In many kinds of experimental or observational environments, it turns out to be the case that the measurement of a single magnitude or property is not feasible or theoretically interesting. What is of interest is the joint measurement of several properties simultaneously.' (Suppes, 2009).

(Muravyov and Savolainen, 2001) differentiate conjoint measurement from indirect derived measurements by considering a case of an object characterised by multiple properties where the 'manifestation of some property can be determined through manifestations of other properties the connection between which is *not* expressed by a physical law or such a law is unknown as yet' and claim this to be the application field for conjoint measurement. An example of the 'sitting comfort of car seats' is given where the measure could be a conjoint aggregation of 'smooth', 'warm', 'soft', 'thick', etc. Other examples of conjoint measurement structures abound, e.g., 'Loudness depends on both physical intensity and frequency' (Iverson and Luce, 1998, p. 31).

Conjoint structures effectively have 'variables that trade off against one another in influencing the dependent attribute'. 'It is always possible to plot those combinations of the variables that yield equal levels of the attribute. Economists call these indifference curves' (Iverson and Luce, 1998, p. 31).

There are no hard and fast rules for arriving at a conjoint measurement, but the usual method is to form a product structure (Roberts, 1984) and assess that structure against well-defined appropriateness conditions.

A conjoint measurement approach for technical integrity may be appropriate and is a candidate for establishing a measurement instrument.

3.5.4.3 Ranking and multi-attribute utility theory

Ranking approaches, as often used in decision making, may be used to at least successfully rank a list of alternates in the face of competing objectives. Section 3.4.5 identified MAUT a possible approach for developing a measurement instrument for technical integrity. (Clemen, 1996) provides an approach into how to use ranking methods. like MAUT, in decision analysis. The general approach is to:

- Construct a fundamental objectives hierarchy to represent the desired outcome, built from a set of, sometimes conflicting, sub-objectives.
- For each sub-objective, identify a fundamental means to measure the level of achievement ('if something is important, it must be measureable'). This may be termed an 'attribute scale', many of which may have a natural unit of measure (e.g., dollars, hours, percentage). If no natural scale exists, a proxy for the attribute or a constructed scale that converts a qualitative judgement to an interval measure is used.
- Attribute scales that are measured in natural units are converted to an ordered scale (e.g., a 0..100 scale) that can be compared with the constructed scales. This allows for the combination of scores from attributes in accordance with the relative importance of attributes.
- For each attribute, there is generally the need to construct a utility function that expresses the amount of utility increase to the stakeholder for the given amount of increase in the attribute scale. Note that a utility function is rarely linear, although it would be monotonically increasing (e.g., an increase in wealth from $100 to $200 is of much greater utility that an increase in wealth from $1,000,000 to $1,000,100).
- Form an additive utility function based on the fundamental objectives hierarchy. This is also known as MAUT. The overall utility can be expressed as:

$$U(x) = \sum_{i=1}^{m} k_i U_i(x_i)$$

Where x_i is the value on the *i*th attribute scale and $U_i(x_i)$ is the corresponding utility function for the *i*th attribute scale. $U_i(x_i)$ is such that the best outcome is valued 1 and the worst outcome is valued 0. The weights, k_i, that combine the *i* utility functions are such that $\sum_{i=1}^{m} k_i = 1$. Assessment of the weights can be undertaken by various approaches including pricing out, swing weighting and lottery weights.

3.6 Literature survey conclusions

The literature survey has identified the following knowledge gaps:

- There is no consensus of support for measuring abstract system properties like technical integrity in the literature.
- There is no precedented measurement construct or measurement instrument for technical integrity.
- There is no pragmatic measurement approach for estimation of the cumulative risk (unintended, undesired outcomes) of a system or to compare that risk across a class of similar systems or across dissimilar systems.
- No completely applicable measurement approach for technical integrity has been identified.

Therefore, there is a need to develop a measurement construct and associated instrument in order to measure technical integrity. Several measurement approaches warrant investigation:

- The Consensual Assessment Technique.
- Conjoint Measurement.
- Scale and Indices.
- Proxy measurements for variables or attributes that may wholly or partially indicate technical integrity and may be able to be combined in some way to estimate technical integrity or at least rank systems according to their relative technical integrity.

It was noted in section 3.5.1.1 that a clear concept of the property to be estimated must be established as a precondition to measurement. However, the lack of a clear, objective definition of technical integrity was identified in Chapter 2. Together, these findings indicate that the development of theory and concepts for technical integrity is vital in the quest for an answer to the research question. This underpins the research method that was adopted as will be discussed in Chapter 4.

4 Methodology

4.1 Introduction

The literature survey of Chapter 2 revealed that there is no objective definition for technical integrity and Chapter 3 found no wholly applicable measurement approach was available. There is a need to develop a measurement construct and associated instrument to measure technical integrity to support research into the effectiveness of technical regulation. The primary research question to be addressed then is:

How can the technical integrity of a complex engineered system be measured?

The nature of this research problem includes the following aspects which shape the research project:

- The theory of technical integrity and technical regulation is not well developed.
- The concepts of technical integrity are not well defined, at least not in a way to enable practical measurement.
- The indicators and variables needed to contribute to the definition of concepts of technical integrity have not been developed.
- Some of the variables likely to contribute are suitable for objective measurement while other variables may only have subjective measurement as a practical possibility.
- Instruments may need to be developed or at least validated for the variables, particularly those for which subjective means only are available.

Consequently, a phased approach that begins as *exploratory* research and progresses to *descriptive* research has been adopted for this thesis. According to (Sekaran, 2003, p. 118-121):

> An exploratory study is undertaken when not much is known about the situation at hand, or no information is available on how similar problems or research issues

> have been solved in the past. In such cases, extensive preliminary work needs to be done to gain familiarity with the phenomena in the situation, and understand what is occurring, before we develop a model and set up a rigorous design for comprehensive investigation.
>
> A descriptive study is undertaken in order to ascertain and be able to describe the characteristics of the variable of interest in a situation.

Exploratory research has a 'typical goal of ... building of a tentative explanation of the problem situation' and 'is usually qualitative in nature and is characterized by an inductive logic approach' (Cropley and Harris, 2004). In this thesis use of this approach is justified in order to inductively develop theory, concepts and measurement constructs for technical integrity.

Descriptive research 'is usually quantitative in nature and is characterized by a deductive logic approach' 'for the purpose of ascertaining and describing the characteristics of the variables of interest in a given problem situation'(Cropley and Harris, 2004). Within this thesis the theory and concepts developed inductively in the earlier research phase are to be confirmed and refined by seeking expert opinion and by trialling the measurement approaches.

It is acknowledged that it is generally advisable to select a single paradigm (quantitative or qualitative) for research work (Creswell, 2003). However, this research design adopts a paradigm of *pragmatism.* The pragmatist orientation embraces both quantitative and qualitative methods, deductive and inductive logic, and considers both objective and subjective points of view. It acknowledges there may be causal effects underlying the phenomena of interest, but they may never be able to be pinned down exactly (Tashakkori and Teddlie, 1998, p. 22-25). 'Pragmatism appears to be the best paradigm for justifying the use of mixed method and mixed model studies' and 'it presents a very practical and applied research philosophy: Study what interests and is of value to you, study it in the different ways that you deem appropriate, and use the results in ways that can bring about positive consequences' (Tashakkori and Teddlie, 1998, p. 29-30).

The phased approach to this research project is designed to:

- In the exploratory research phase:
  - Develop theory, concepts and a measurement construct for technical integrity.
- In the descriptive research phases:

  - Assess the content validity of the construct for technical integrity,
  - Assess the feasibility of direct estimation of technical integrity by expert judgement,
  - Establish parameters for candidate measurement instruments, and
  - Establish the variables to be included in candidate measurement instruments and establish the reliability and validity of those measurement instruments.

The research methods, being the practical techniques used in the research project, include:

- Evaluation, particularly a systems analysis approach to develop theory and concepts.
- Mixed method designs, where quantitative and qualitative data is collected via questionnaires and analysed to be complementary, are used to:
  - Establish expert support for the validity for the proposed measurement construct,
  - Establish variables and combinatorial weightings for the measurement construct, and
  - Establish the reliability and validity of the candidate measurement instruments.

To structure the research, particularly to abstract the forms of data generated into practical knowledge, the hierarchy of information defined by (Walliman, 2010, p. 65-69) is adopted as shown in Table 5. The first two columns show the term and Walliman's definition while the last column indicates how these terms are adopted in this research.

Table 5. Hierarchy of Information.

Term	**Definition**	**Adoption to this research**
Theory	Abstract statements that make claims about the world and how it works. Research problems are usually stated at a theoretical level.	Theory about the nature of technical integrity and the relationship between concepts leading to a structure for a measurement construct.
Concepts	Building blocks of the theory which are usually abstract and cannot be directly measured.	The underlying abstract concepts of technical integrity that combine to form the theory for technical integrity and its measurement.

Term	Definition	Adoption to this research
Indicators	Phenomena which point to the existence of the concepts.	Perceivable phenomena that give an indication of the concept (e.g. compliance, risk, reliability).
Variables	Components of the indicators which can be measured	Measurable quantities that can form a component measure of technical integrity (e.g., percentage of requirements verified, risk indices, Mean Time Between Failure)
Values	Actual units of measurement of the variables. These are data in their most concrete form	Data and units of measure (e.g., 98%, 'High risk', 2456 hours)

(DeVellis, 2011, p. Ch. 5) provides a set of step by step guidelines for scale development that are utilised and tailored to provide the underlying method and justification for the phasing of this research project. The guidelines and the approach to be taken by this research project are shown in Table 6, noting that the research is not limited to scales as candidate measurement approaches.

Table 6. Tailoring of scale development guidelines for research project.

DeVellis Guideline	Research Project Approach
1. Determine clearly what you want to measure	Inductively develop theory, concepts and indicators for technical integrity.
2. Generate an item pool	Broad content analysis of technical regulations, standards and other documented sources conducted to identify items that wholly or partially reflect technical integrity.
3. Determine a format of measurement	Several measurement approaches warrant investigation: the Consensual Assessment Technique; Conjoint Measurement; Scale and Indices; Proxy measurements for some variables. Outcomes may warrant an aggregation of some or all of these approaches.
4. Have the initial item pool reviewed by experts	Survey research to assess content validity of technical integrity measurement construct. Survey research to gather expert judgments on indicators of technical integrity.
5. Inclusion of validation items	Survey research to collect demographic data in order to assist in the detection of flaws. Survey research to gather expert judgments on indicators not believed to indicate technical integrity.
6. Administer items to development sample	Survey research to cover a sample of expert judges in the acquisition, development and use of complex engineered systems in the ADF. Guideline of 300 respondents is not practical and may limit results.
7. Evaluate the items	Results of the surveys will be analysed to determine the correspondence of expert judgments of indicators of technical integrity with those theorised.

DeVellis Guideline	Research Project Approach
8. Optimise scale length	Results of the surveys will be analysed to determine the most important indicators of technical integrity for inclusion in a final measurement instrument.

4.2 Research phases

The research is conducted over the phases identified in Table 7. Each phase focuses on particular information objectives and has an aim which contributes to answering the research problem.

Table 7. Research project phases and aims.

Phase	Information objective	Phase aim
Concept development	Develop theory, concepts and indicators for technical integrity	Produce a theory, measurement scale and concepts for development of a measurement instrument for technical integrity
Content validity of technical integrity construct	Validate theory and concepts	Establish a consensus of expert opinion for the validity of the technical integrity measurement construct
Conjoint measurement aggregation weights	Develop indicators	Specifically establish parameters for the measurement scale and further refine the measurement instrument.
Assessment of judgments for estimation of technical integrity	Assess reliability of judgment of technical integrity indicators and variables	Assess the reliability and validity of using the CAT for direct estimation of technical integrity and its indicators
Validation of technical integrity theory and assessment of measurement instruments	Validate theory. Assess reliability, validity and practicality of measurement approaches.	Validate a theory for technical integrity. Select a measurement instrument suitable for further research.

The interrelationship of these study phases is illustrated in Figure 1 in the introduction to this thesis. It shows how the research question is clearly defined as an outcome of the literature survey which itself provides the base to ground the concept development phase of the research. The output of concept development, proposed theory and concepts for a technical integrity measurement enables the construction of survey instruments which in turn provides quantitative and qualitative data for validation and refinement of the theory and the measurement approach. The outcome of the phases concludes with a recommended measurement instrument for technical integrity that is important to support future research in the area of technical regulation and the engineering of complex systems.

The detailed research method used for each of these phases is described and justified in the subsequent sections of this chapter. Results from each research phase are presented in the succeeding chapters.

4.2.1 Concept development

The aim of this phase of the research is to develop a definition and construct for technical integrity measurement. The research method is *Evaluation* which aims to 'make sense of the myriad human, political, social and contextual elements involved' in the definition of technical integrity in the context of technical regulation of complex engineered systems. The approach for the evaluation model is 'systems analysis – which is a holistic type of research looking at the complex interplay of many variables' (Walliman, 2010, p. 11-12). The study phase is progressed by:

- Collecting, comparing and contrasting definitions for technical integrity and related system concepts.
- Inductively deriving a general definition for technical integrity (a theory and concepts).
- Determination of objectives for a measurement construct.
- Generation of an initial measurement construct.
- A content analysis of regulations, standards and other documentation related to technical regulation and technical integrity to determine the range of indicators and variables that may contribute to a measurement instrument of technical integrity.

4.2.2 Content validity of the technical integrity measurement construct

The aim of this phase of research is to establish the content validity of the measurement construct proposed by the concept development phase. Content validity refers to the extent to which a measure represents all facets of the phenomenon of interest and is assessed in this phase via a consensus of expert judgments sought by means of survey research.

A survey instrument gathered both rating data to particular questions seeking the level of agreement with various aspects of the measurement construct for technical integrity and asked open-ended questions through which additional insights into the reasons

behind the expert ratings can emerge. Both quantitative and qualitative analysis methods are used to analyse the data.

Chapter 6 presents the details of methods used including the survey instrument, sampling approach and quantitative and qualitative analysis procedures.

4.2.3 Conjoint measurement aggregation weights

The aim of this phase of research is to determine coefficients to use in the aggregation of capability and dependability into the conjoint measurement of technical integrity via quantitative analysis of judges' ratings obtained from survey research. This phase also reports on the elicitation of the reasoning behind the judgements via qualitative analysis of the open-ended survey response data.

The method uses survey research to collect expert judgments in the form of relative importance ratings and to collect written responses from open-ended questions. The rating data is analysed quantitatively in order to arrive at aggregation weightings for the capability and dependability elements of a multiplicative conjoint equation theorised in Chapter 5. Effectively the quantitative analysis procedure is testing the hypotheses that *the ratings of the relative importance of capability versus dependability are the same across the different system domains*. A qualitative analysis procedure is employed on the open-ended response data in order to assist in the interpretation of, and find explanations for, the quantitative results. Finally a mixed-method approach is used to 'triangulate' the quantitative and qualitative results.

Chapter 7 presents the details of methods used including the survey instrument, sampling approach and quantitative, qualitative and mixed-method analysis procedures.

4.2.4 Assessment of judgments for estimation of technical integrity

The aim of this phase of the research is to assess the reliability of judgments for estimating technical integrity and its associated indicators. Three primary techniques are used in this phase of the research: the CAT; scale development, and numerical models of measurable system quality attributes. The CAT and scale development are techniques most often used by the social sciences to measure phenomena that cannot be directly measured so that the application to what could be considered "hard" engineering characteristics of a system is relatively novel. A set of three example system

descriptions is constructed and survey research collects ratings and subjective responses from a range of respondents familiar with the design and operation of complex defence systems. The ratings are compared with numerical models of values from the system descriptions in order to determine feasible measurement approaches.

Chapter 8 presents the details of methods used including the survey instrument, sampling approach and quantitative analysis procedures.

4.2.5 Technical integrity theory and measurement instrument

The aim of this phase of the research is to validate or amend the theory of technical integrity using the survey data from the previous phase. Based upon the results of previous phases of the research, candidate measurement instruments are characterised in terms of reliability, validity and the practicality of implementation in the concept and development phase of a complex engineered system. A "best" measurement instrument for the purpose of future research is recommended.

Chapter 9 presents the details of methods used including the survey instrument, sampling approach and quantitative analysis procedures

4.3 Research methods

This section identifies the broad research methods used and justifies the use of those methods. The specific method used in each phase of this research are detailed in each of the Chapters 6 through 9 inclusive and summarised in Table 8.

4.3.1 Survey method

Surveys have been used in this research project in order to elicit knowledgeable judgments, both ratings and open-ended responses. Two survey instruments were developed and implemented as web-based questionnaires and these are presented in full at Appendices A and B.

Development of the surveys was conducted in accordance with guidance in (Meyer and Booker, 2001, p. 57-184) and (Leedy and Ormond, 2005, p. 183-216). Questionnaires were developed and were piloted with academic supervisors and with a small number of the researcher's professional colleagues in order to ensure that any ambiguity with the questions had been identified and eliminated to the maximum extent practicable.

Table 8. Research project phases and research methods used.

Phase	**Evaluation – systems analysis**	**Survey research**	**Quantitative analysis**	**Qualitative analysis**	**Mixed method analysis**
Concept development	✓				
Content validity of technical integrity construct		✓	✓	✓	
Conjoint measurement aggregation weights		✓	✓	✓	✓
Assessment of judgments for estimation of technical integrity		✓	✓	✓	
Validation of technical integrity theory and assessment of measurement instruments		✓	✓		

The role of respondents to the survey was to provide judgments in the forms of ratings and open-ended responses. The judges were selected to participate based on their familiarity with the definition, design, integration, verification, operation and/or support of complex engineered systems. They did not have to consider themselves to be an expert to participate in this research as a judge. Demographic information obtained in the survey instrument was designed to allow the researcher to identify any significant differences based on education, experience, or other factors and control for these post facto.

Online, web-based implementation of the questionnaire was selected primarily for pragmatic reasons. The experts from whom judgments were to be elicited were geographically dispersed around Australia and overseas and the researcher did not have the means to travel to these destinations for interview based elicitation. The respondents, in general, have little time during regular business hours to conduct face-to-face or telephone based interviews and the researcher was similarly limited, given the part-time nature of the candidature. Implementation and delivery of the questionnaire via the web-based application SurveyMonkey® overcame these limitations and provided for electronic download of the results directly into analysis tools.

The data was downloaded and stored on the researcher's standalone computer and was not accessible to others. The data does not include information that allows the individual identification of the respondents.

Specifics of the survey instrument's rating scales, sampling approach, numbers and characteristics of survey respondents are included in the method sections of Chapters 6 and 8 for each of the two surveys.

4.3.2 Quantitative analysis method

Ratings data obtained from the online questionnaires were imported directly to SPSS. The ratings data obtained are ordinal and therefore non-parametric statistics are used. Frequency data are visualised using pie or bar charts and descriptive statistics (mode, median range and interquartile range) are calculated. Hypothesis testing uses both the chi-square test and Lawshe's Content Validity Ratio (CVR) (Lawshe, 1975).

Analysis of rated questions against the categorical demographic information was conducted by forming a cross-tabulation and using the 'Eta' calculation within SPSS as the appropriate association statistic for ordinal versus nominal data, or by the Independent-Samples Kruskal-Wallis test. This non-parametric test allows for the comparison of 'scores on a continuous variable for three or more groups' (Pallant, 2010, p. 232-235).

The One-Sample Wilcoxon Signed Rank Test provides a test to see if a sample median differs significantly from a hypothesised value. The only assumption for the data needed is that it is an ordinal variable which is true for the ratings data sets (UCLA: Academic Technology Services).

Specifics of the quantitative methods applied in each phase of the research have been included in the method sections of Chapters 6 through 9 inclusive.

4.3.3 Qualitative analysis method

The qualitative analysis follows the categorical strategy as defined by (Teddlie and Tashakkori, 2009, p. 253):

> *Categorical strategies* break down narrative data and rearrange those data to produce categories that facilitate comparisons, thus leading to a better understanding of the research questions.

Textual response data obtained from the open-ended question of the online questionnaires were imported directly to Dedoose. Dedoose[1] is a web-based tool for qualitative and mixed methods research.

The qualitative analysis procedure essentially follows that of (Leedy and Ormond, 2005, p. 140) adapted as follows:

- The researcher becomes familiar with the data by thoroughly reviewing all responses and to identify themes of relevance to understanding reasons for the judgements provided in the ratings responses.
- Evident themes or concepts are defined and codes established to pursue particular understandings that underlie the quantitative results.
- Text is extracted and associated with one or more of the codes.
- The extracts are tabulated for each code.
- Counts of extracts against each code are reported to establish frequency of the associated theme or concept.
- Text extracts within each code group are compared, contrasted and discussed.

Mixed method visualisation and analysis tools within Dedoose were used to further identify and explore underlying relationships and meanings within the response data.

Specifics of the quantitative methods used in each phase of the research have been included in the method sections of Chapters 6, 7 and 9.

4.3.4 Mixed model method

(Tashakkori and Teddlie, 1998, p. 19) define mixed model studies as:

> These are studies that are products of the pragmatist paradigm and that combine the qualitative and quantitative approaches within different phases of the research process.

The primary purposes of using mixed model methods in this research is to seek convergence of results from the quantitative and qualitative analyses, termed *triangulation*, and to examine overlapping or different aspects of the phenomena, termed *complementarity* (Tashakkori and Teddlie, 1998, p. 43). The scenario for combining quantitative and qualitative in this research project is that the quantitative

[1] See http://www.dedoose.com/ for more information on Dedoose.

analysis is dominant in providing the results and the qualitative assists in explaining and confirming the results as illustrated in Figure 11 adapted from (Tashakkori and Teddlie, 1998, p. 44).

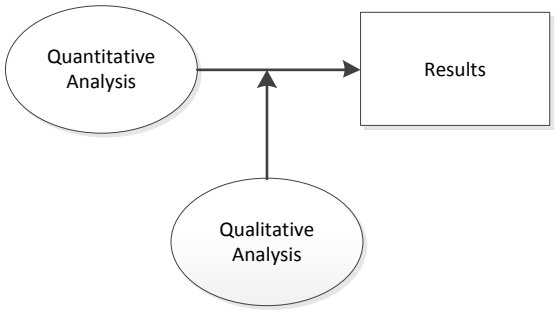


Figure 11. Qualitative methods to explain quantitative results.

In this research, the use of a mixed model method is further justified for content and construct validity analysis, and the assessment of variables to identify any bias or misinterpretation that may not have been evident in the quantitative results.

In general, the mixed model analysis conducted in this research project compares the quantitative ratings data obtained by the survey and compares it to the number of comments received that have been coded to corresponding concepts. The researcher identifies if the trend in the coded text extracts is consistent with the ratings data and further uses the coded text extracts to infer further meaning to the results.

Specifics of the mixed model methods used in the research have been included in the results section of Chapter 7.

4.4 Ethical considerations

All human interaction, including human survey research, has ethical dimensions that must be considered by the researcher. The survey research conducted for this thesis required and obtained ethics approval from the University of South Australia's Human Research Ethics Committee.

In particular the surveys provided each potential respondent with a Research Participant Information Sheet (see Appendices A and B) that included descriptions of the type of data to be requested from the participants, their rights to withdraw from the research, and the researcher’s commitment to preserve anonymity of the response data. This

approach has assured the participants have provided their ‘informed consent’ and that their ‘right to privacy’ has been upheld (Leedy and Ormond, 2005, p. 101-102).

4.5 Conclusion

A research methodology has been proposed to answer the research question:

How can the technical integrity of a complex engineered system be measured?

A paradigm of *pragmatism* has been adopted leading to a phased approach to the research project that begins as *exploratory* research and progresses to *descriptive* research. The guidelines of (DeVellis, 2011) for measurement scale development have been used to formulate the research phases that will seek to develop a theory for technical integrity, a measurement construct and measurement instrument for which validity will be established via the elicitation of expert judgments. Quantitative, qualitative and mixed model research methods are to be utilised in the validation of the theory, construct and measurement instrument.

5 Concept development

5.1 Introduction

This chapter documents the results of the exploratory research phase of the project. In particular it seeks to develop theory, concepts and a measurement construct for technical integrity in order to provide an answer the research question:

How can the technical integrity of a complex engineered system be measured?

In Chapter 2, it was presented that in the Australian Defence Force (ADF), technical integrity is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection, i.e., its compliance with technical standards and the minimisation of unintended consequences. Technical integrity has various different meanings and has synonyms and related concepts occurring in other civilian and military system contexts. The literature survey concluded that:

> The goal of technical regulation is *technical integrity*. However, there is not a clear, objective definition of technical integrity established for the ADF and the subjective definitions vary across the materiel domains.

The literature survey also found for fitness for service, a stated component of technical integrity:

> Fitness for service is a key element in the subjective definition of technical integrity as the aim of the ADF technical regulation. But the definitions of fitness for service differ across the devolved regulatory domains. Fitness for service needs to be defined more objectively or a comparable concept introduced to bolster the definition and theory of technical integrity.

Section 5.2 of this chapter reports the results of secondary research into the various meanings of technical integrity, fitness for service and related concepts in order to establish a working definition of technical integrity and a theory that brings together its

constituent concepts that will be useful across a range of complex engineered systems and can form the basis for the construction of a measurement instrument.

Section 5.3 of this chapter therefore tackles the problem of establishing a form of and construct for the measurement of technical integrity consistent with the theory and concepts established in section 5.2.

Finally section 5.4 of this chapter identifies and assesses a range of indicators and variables that may be directly or indirectly measurable and that when aggregated may provide or contribute to a measurement instrument for technical integrity.

5.2 Defining technical integrity

5.2.1 What is technical integrity of a complex engineered system?

There are many definitions for technical integrity, many related concepts, usages of terminology and stakeholders for the desired and undesired emergent properties of complex engineered systems. Some of the related terms and concepts include: suitability for service; safety; certified products; system effectiveness; airworthiness (in the aerospace domain); systems assurance; fitness for purpose; technical assurance; product quality; compliance with class rules (in the maritime domain); compliance with build standards; mission assurance; and dependability among many others.

Some definitions include concepts of cost or economic value, while others are independent of cost or economic value. Typically all involve the concept of avoidance of unintended negative consequences (particularly safety). Typically most involve the concept of doing what was intended to be done (fitness for purpose/service, quality, mission assurance) and to have that done for most of the time (available, reliable).

5.2.1.1 Technical integrity in the Australian Defence Force

In the ADF, technical integrity is the term used to encompass the concept of a system's fitness for service, safety and compliance with regulations for environmental protection (Australian Defence Force, 2010). Materiel is defined as 'All items including ships, tanks, self-propelled weapons, aircraft etc and related spares, repair parts and support necessary to equip, maintain and support military activities without distinction as to its application for administrative or combat purposes.'

The three main components of technical integrity are further defined as:

- Fitness for service: the system's ability to satisfy operational requirements.
- Safety: freedom from those conditions that can cause death, injury, or occupational illness.
- Environment Protection: poses no hazard to the environment.

5.2.1.2 Concepts of technical integrity in a wider context

The term *technical integrity* is not commonly used across all industries or government enterprises where complex systems are engineered, maintained and put to use. This section initially looks at a few examples where the term technical integrity is used and compares these with the usage of the ADF. Then a number of related concepts and the terminology used in other domains is presented and contrasted.

5.2.1.2.1 Technical integrity in the commercial industry

Technical integrity, or just integrity, is not widely used in commercial industry. One area where it is used is in the area of process plants. For example, Woodside uses the definition 'Integrity – achieved when risk of failure that endangers personnel, environment or asset value is tolerable and is reduced to ALARP' (Hackett, 2008) and (PI Asset Integrity, 2007) state:

> Technical Integrity in any process plant is essential to ensure that there will be no harm done to people or the environment. Assuring Technical Integrity of your facility by regular interval audits makes good business sense'.

(Bale and Edwards, 2000) definition is:

> Technical integrity is concerned with the development of the design such that it is carried out by well trained personnel, who have been assessed to be competent, in accordance with recognized, sound practices and procedures and such that there is adequate provision by way of reviews and audits, to ensure the design intent is unimpaired in any way that could cause undue risk or harm to people or damage to the environment.

Interesting in these definitions is the direct parallel with the ADF's components of personnel safety and environmental compliance as well as the contrast between the ADF's Fitness for Service and the concepts of 'asset value' and 'good business sense'. Clearly the latter may be explained as the mission outcomes as required by a commercial entity. That is, creation and maintenance of value to further the business.

5.2.1.2.2 Objectives in regulations for design and operation of systems

The design and operation of significant and complex systems is often regulated to prevent unwanted, unintended consequences by governments or by international organisations with multi-national government support. The objectives of these regulations may be thought of as technical integrity objectives. Example regulations were identified in section 2.2.4, and from those it is again apparent that the concepts of safety, compliance, maturity and reliability are technical objectives of regulation.

5.2.2 MIL-STD-882C safety compared to ADF technical integrity.

MIL-STD-882C (Department of Defense, 1993) defines *safety* as 'Freedom from those conditions that can cause death, injury, occupational illness or damage to or loss of property or damage to the environment' and identifies 'operational effectiveness, time and cost' as constraints. There is close to an exact parallel when compared to the ADF's *technical integrity:* 'damage to the environment' and 'Freedom from those conditions that can cause death, injury, occupational illness' in MIL-STD-882C is the same respectively as 'environment protection' and 'personnel safety' as defined in the ADF's technical regulations (Australian Defence Force, 2010).

The one point of difference is between the 'freedom from those conditions that can cause ... damage to or loss of property' of MIL-STD-882C and the 'Fitness for Service' concept of the ADF's technical regulations. The latter is fully inclusive of the MIL-STD-882C concept, but as identified in previous sections, also considers the achievement of regulatory requirements and capability requirements. Failure to meet the latter requirements in particular can be treated as a risk to mission achievement or as a mishap where the outcome is the inability to conduct certain missions or aspects of the intended mission.

5.2.3 Relationship of emergent properties and technical integrity

Chapter 2 identified that (Honour, 2007) had nominated a categorisation of emergent properties in two dimensions as shown in Table 3. This framework was extended in Table 9 to examine the relationship of each of the six categories of emergent properties with the extent to which technical regulation address the property.

Technical regulation clearly addresses the 'Designed' range of emergent properties that a complex engineered system can exhibit. Capability is directly addressed by assuring

compliance with requirements. Facts of Design are addressed by a focus on controlled engineering efforts, in particular configuration management and the adoption of systems engineering paradigms to ensure the design is documented, understood and managed. Accepted Trade-Off is addressed by a technical risk management approach and in particular a focus on system safety engineering to ensure that potential compromises to technical integrity are identified and made explicitly known to users and stakeholders of the system.

The 'Surprise' ranges of emergent properties of the complex engineered system are not addressed so comprehensively by technical regulation. Unanticipated hazards and other unwanted features are intended to be identified by system safety engineering efforts. These efforts, to the extent that they are successful, really move identified unintended hazards of the system into the "Accepted Trade-off" category. The categories of Exploitable Features and Facts of Existence are not really addressed by technical regulation. For instance, technical regulation does not explicitly seek to enforce methods or other means to identify, let alone promote, the generation of exploitable features which may create a system of much higher value to stakeholders of the system.

Table 9. Categorisation of system emergent properties and the degree to which they are addressed by technical regulation.

EMERGENT PROPERTIES	**Useful**	**Neutral**	**Destructive**
Designed (i.e. intended)	Capability	Facts of Design	Accepted Trade-Off
Addressed by Technical Regulation	Compliance – assuring that all desired capabilities are achieved.	Design is documented, understood and managed.	Risks to technical integrity are explicitly identified, minimised and residual risk accepted.
Surprise (i.e. unintended)	Exploitable Feature	Facts of Existence	Fearful features (unanticipated hazards)
Addressed by Technical Regulation	Not addressed	Not addressed	Intent is that these are minimised or at least moved into an Accepted Trade-Off.

In summary, technical regulation would seem to be focused on ensuring the intended features are achieved, the design is documented and controlled, risks are accepted and that to the maximum extent possible, detrimental unintended behaviours are mitigated.

There is no focus on measuring or maximising the beneficial, yet unintended, exploitable features of a complex engineered system. Given that the stated goal of ADF technical regulation is technical integrity, it follows that technical integrity could be defined in terms of concepts such as compliance of intended capabilities, controlled and documented design (or quality of the design) and that the risks to safety are identified and mitigated as far as practicable.

It would seem that there is value in systems that have more beneficial features than were originally specified and these would be the features that would lend a complex engineered system to be evolved over its lifecycle. The lifecycle value of a system with more of these beneficial exploitable features should clearly be of more value to stakeholders than one with less. It should be that such a system has a higher technical integrity than one that does not, all other factors being equal.

5.2.4 Candidate concepts for assessing technical integrity

Starting with the elements of technical integrity defined by the ADF, Table 10 presents a list of candidate concepts that may indicate the level of technical integrity achieved by a system. Potential sources of data to evaluate these concepts are also identified as a means of informing the development of theory.

Table 10. Candidate concepts and sources of data for assessing elements of technical integrity.

Technical Integrity elements	Candidate concepts	Sources of Data
Fitness for Service The materiel's ability to satisfy operational requirements	Meeting operational requirements: Compliance Defects Achieved performance Other characteristics of the system: Availability Dependability Robustness Other attributes of a system that could be direct or indirect measures of technical integrity: System operator satisfaction levels Other stakeholder satisfaction levels Public perceptions Rate of problem reports discovered after acceptance	Design data Models and simulation Project verification records Project latent defect claims Results of operational evaluations In service RMA data Calculated RMA data Operator surveys Stakeholder surveys Government committee reports Press articles In service trouble report records Warranty and latent defect records
Safety Freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property	Freedom from hazards: Accident rates Accident severity Incident (near miss) rates	Coverage of formal safety assessments Accident and incident databases Accident and incident investigation reports
Environment Poses no hazard to the environment	Compliance with environmental standards Freedom from environmental impact Environment incident rates Environment incident severity	Compliance statements Environment incident databases. Environment incident investigation reports

It is noted that the majority of data sources readily identified are for a system that is in the operational phase of its lifecycle. These sources of data are generally not available during the design and implementation phase of a system's lifecycle. This implies that measuring technical integrity during that phase will need to rely on other sources of estimates.

5.2.5 A theory of technical integrity

Consideration of the varied definitions of technical integrity, parallels with system safety concepts and the relationship inferred between technical regulation and the emergent properties of a complex engineered system in the preceding sections were used as inputs to an inductive process. The inductive process stoked also by the researcher's own professional experience, and by reflections of the nature of technical integrity with similarly experienced practitioners, and peer review of the developing theory and concepts through peer reviewed conference papers (refer Appendix F and Appendix J) lead to a mature conclusion that there are two separable or orthogonal[2] aspects to technical integrity of a complex engineered system. The bases for the separation are the emergent properties of the system; the 'useful' and 'destructive' emergent properties as indicated in Table 9.

The 'useful' basis indicates that the system meets all the requirements and standards that have been prescribed for it. This property of the system is termed as the concept *capability*. The 'destructive' basis indicates the broad notion of minimising the unspecified, but clearly undesired properties of the system. This property of the system is termed as the concept *dependability*. The theory of technical integrity relating these concepts is illustrated in Figure 12.

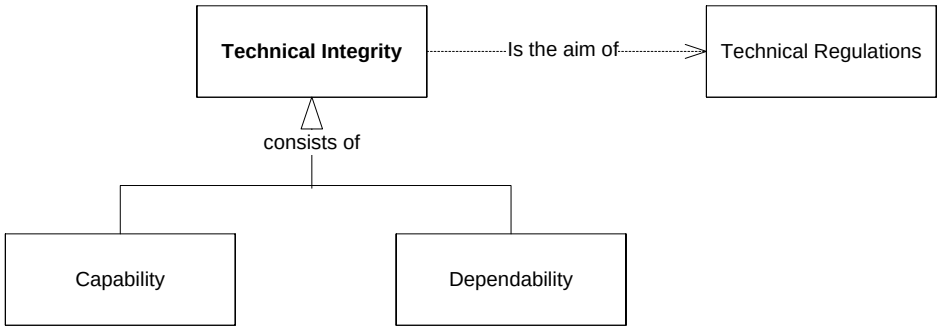


Figure 12. The theory of technical integrity.

More completely, *technical integrity* is defined as consisting of capability and dependability, where:

[2] meaning the characteristic of being independent, not redundant, not overlapping.

- Capability is defined to mean that the system meets all specified requirements for performance, function and quality attributes in the defined environment and when used as intended. Capability is the designed set of emergent properties of a system that are intended and useful (i.e., required); and
- Dependability is defined to mean that the risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. It includes concepts of robustness (system is tolerant of usage that may exceed specified conditions without catastrophic consequences), reliability (failures do not happen more than is acceptable), and system safety (risk of unidentified hazards and failure modes resulting in undesirable outcomes is acceptably low).

It should be noted that these terms may be used differently in different contexts. *Dependability* is often used in the literature but without a universally agreed definition. For instance, (Prasad, 1998) reviews many definitions of *dependability* before arriving at a definition useful for his work, albeit with some differences to that defined above.

The theory and concepts proposed for this research encompass the scope of technical integrity as used in the ADF. Fitness for service is equivalent to capability and safety and environment protection are hyponyms of dependability. Based upon this theory a form of measurement can be determined.

5.3 Form of measurement

5.3.1 Measurement objective

The overarching objective is to establish a measure, a set of measures or some composite measure that directly indicates the technical integrity of a complex engineered system. The measure needs to be general enough so that it is applicable across as large a range of complex engineered systems as practicable. The range of system types for which a measure is valid needs to at least as broad as the scope of application of a particular regulatory framework (e.g., valid for all maritime materiel governed under the NTRS) but ideally larger (e.g., at least valid for all complex defence materiel regulated by the ADF).

The measurement approach must be supported by data that can be practically obtained for those systems during the concept and development phases of the system lifecycle. The measurement aims are:

- The measure is to be valid, i.e., it must measure technical integrity and not some other latent variable of a complex engineered system.
- The measure is to be reliable, i.e., it must yield the same value for each system independent of when or by whom the measurement instrument was used.
- Ideally the measure will form a ratio scale, which would allow percentage changes in technical integrity to be compared with changes in technical regulations.
- The measure will at least form an ordinal scale so that rank order of systems that may have been developed under different technical regulations can be established for study.
- The measure is to be applicable across a wide range of complex engineered systems (at least across the range of systems applicable to a specific technical regulatory system).
- The measure is to be practical to obtain with little additional cost to a system development effort.

5.3.2 Selecting a form of measurement

Measurement theory has been considered in the literature review of Chapter 3 and underlies the consideration of the following forms of measurement:

- Direct measurement, and
- Indirect measurement, including:
  - Derived,
  - Associative,
  - Multi-dimensional, and
  - Conjoint.

Clearly a direct measurement of technical integrity is not possible so direct measurement was ruled out. Consideration of indirect measurement forms concluded that:

- Derived measurement was rejected since there is no known law for combining any measurable properties of a system to evaluate technical integrity.

- Associative measurement was rejected since there is no empirically known relationship between technical integrity and any other directly measureable property of a system.
- Multi-dimensional measurement was rejected since it cannot produce an ordinal, interval or ratio scale that is desired for the research problem.
- Conjoint measurement was accepted as the most appropriate measurement form for technical integrity. It requires some decomposition or at least identification of factors that contribute to technical integrity and the definition of mathematical relationships to combine these measurements into an empirical order.

5.3.3 Conjoint measurement approach

The theory for conjoint measurement is not new, having been established by (Luce and Tukey, 1964) and extended to include polynomial conjoint measurements by (Tversky, 1967). There are no hard and fast rules for arriving at a conjoint measurement, but the usual method is to form a product structure (Roberts, 1984) and assess that structure against well defined appropriateness conditions.

5.3.3.1 Product structure

The product structure for this conjoint measurement is represented by the concepts of technical integrity previously defined, i.e., technical integrity consists of the concepts of capability and dependability. Representing the measure for technical integrity to be *TI*, capability to be *C* and dependability to be *D*, the product structure is illustrated in Figure 13 (i) and defined mathematically as a function by:

$$TI = f(C, D)$$

This function transforms any ordered pair of measurements *(c,d)* into a number on an ordinal, interval or ratio scale for *TI* and is illustrated by Figure 13 (ii).

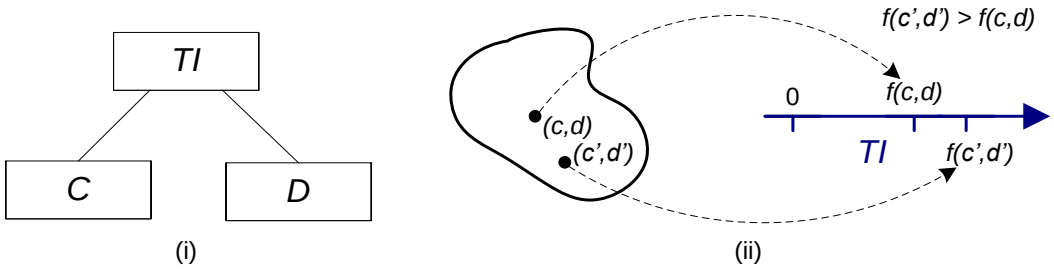


Figure 13. (i) Product structure for TI.
(ii) Mapping of measured attributes to ordinal scale of TI.

5.3.3.2 Appropriateness Conditions

Decision theory provides advice on conditions that should be satisfied by the attributes selected for the product structure (Keeney and Raiffa, 1976). They include *completeness*, *operationality*, *decomposability*, *non-redundancy* and *minimum size*. The *completeness* criterion requires consideration of the adequacy of the measurands as a set in including all the areas of concern for the intended conjoint measurement. This criterion can only be considered as a guide. The *operationality* criterion requires assurance that each measurand has a real contribution to the final measurement value. The *decomposability* criterion requires assurance that the measurands are independent (a necessary mathematical condition for additive conjoint). The *non-redundancy* criterion requires the elimination of double counting. The same parameter should not be measured more than once. The *minimum size* criterion guides a limitation to the number of attributes chosen; more than 5 measurands is considered too many.

In order to assess the proposed product structure against these criteria the expected characteristics of the *TI* conjoint measurement are examined against the criteria.

5.3.3.3 Expected characteristics of TI measure

Consider a system that has met all its specified requirements. This would have a high or maximal value for *C*. However if this same system exhibited emergent behaviour that caused the unintended death or injury of personnel, it would have a low or minimal value of *D*. The expected measurement would be a low value for *TI* since a system that meets its capability objectives yet often unintendedly kills personnel is of no practical value.

Consider now a system that never kills, injures nor poses any health hazard to personnel, equipment or the environment. This system would have a high or maximal

value of *D*. This system however never achieves its service requirements. It may be that it is just largely non-compliant with all its capability requirements. This system would have a minimal value of *C*. The expected measurement would also be a low value for *TI*. A system that does not meet any of its service requirements or performance objectives is of no practical value.

Lastly, consider the desired outcome of a system with a high *D* value and a high *C* value which would exhibit a high or maximum value of *TI*.

These considerations lead to an expectation that on a graph of *D* versus *C* there would be a region of high *TI* values (desired) and a region near the axes of low *TI* values. There are then a range of intermediate values for *TI*. Figure 14 shows what such a graph could look like. The lines on the graph represent isolines of *TI*.

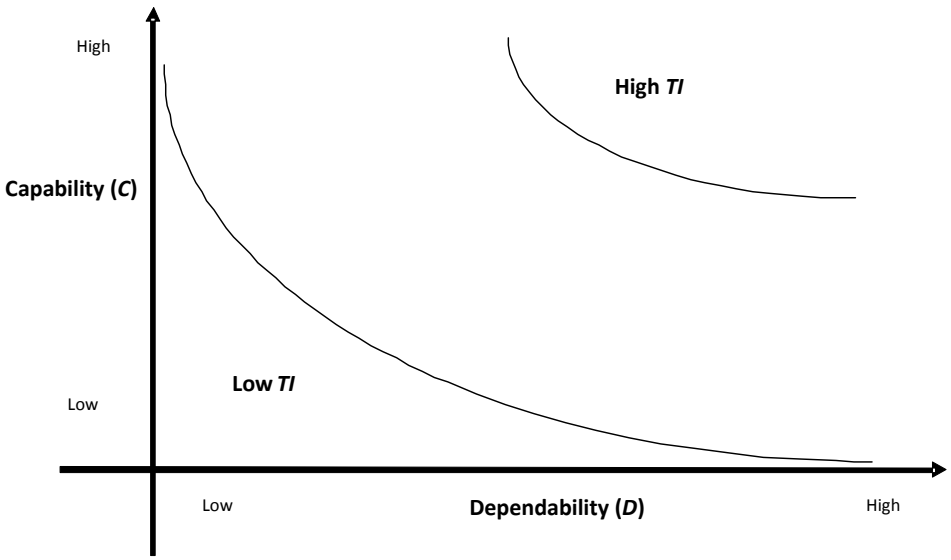


Figure 14. Expected form of *TI = f(FS,PS).*

5.3.3.4 Assessment against appropriateness conditions

The expected measurement outcomes were next assessed against the appropriateness conditions of *completeness*, *operationality*, *decomposability*, *non-redundancy* and *minimum size*. Based on analysis presented previously in this chapter that defined technical integrity, it is concluded that *C* and *D* form a complete set of measurands for *TI*. Each measurand is also clearly operational on the outcome of *TI*, e.g., a low value of either measurand is required to yield a low value of *TI*. At this stage it is asserted that

the measurands are independent based on the intention that they separately assess intended versus unintended properties of the system. Double-counting is a potential difficulty with this approach. For instance, if a reliability requirement is well specified as Mean Time Between Failure in the system's specification then it could be double counting to consider reliability as a component of the dependability measure in this context. This will need careful consideration in the definition of measurements for *C* and *D*. Minimum size has been achieved at this level of the product structure by having simply two contributing attributes. It is concluded that it is appropriate to form a conjoint measurement of *TI* from the components *C* and *D*.

5.3.3.5 Establishment of the conjoint measurement model

Commonly, a conjoint measurement takes an additive form (Green and Rao, 1971, Iverson and Luce, 1998). An additive conjoint measurement for *TI* would be of the form:

$TI = \alpha.C + \beta.D + \gamma$ where α, β, γ are constants.

Considering the boundary conditions of the measurement where it is required that: if *C*=0 then *TI*=0; if *D*=0 then *TI*=0; and if *C*=maximum and *D*=maximum then *TI*=maximum, then the additive form is rejected since it will not yield these values irrespective of the values defined for α, β, γ.

The boundary conditions suggest a multiplicative model to ensure *TI* = 0 if either *C* = 0 or *D*= 0. The mathematical foundations for multiplicative conjoint measures were first established by (Roskies, 1965) as an extension of the additive conjoint model.

The following conjoint measurement model is proposed for *TI*:

$TI = C^{\alpha}.D^{\beta}$ where $\alpha, \beta,$ are constants > 0, values to be determined.

In order to demonstrate this model, *C* and *D* are constrained to be real numbers in the range [0,1][3] which in turn constrains TI to the range [0,1]. Consider when $\alpha, \beta = 1$ as a starting point in order to visualise the resulting mapping of *C*, *D* to *TI* when it is equally influenced by its attributes. This is shown in Figure 15.

[3] A not unreasonable constraint as will be seen when establishing measures for C&D.

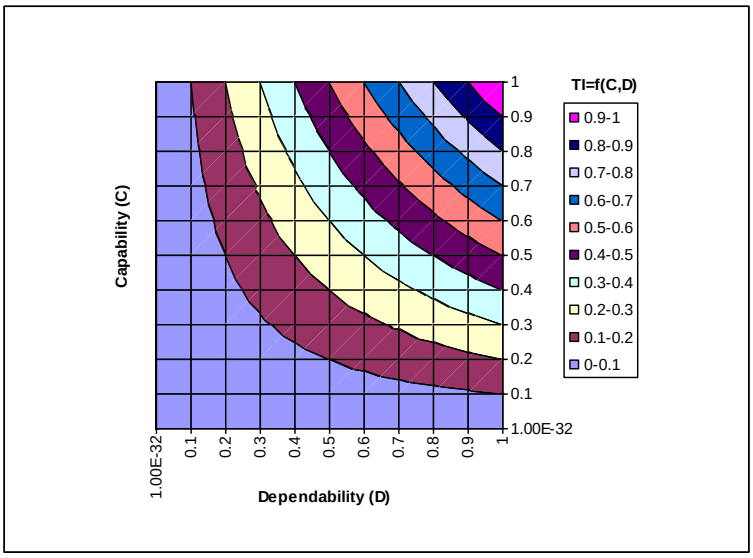


Figure 15. Capability and dependability with equal influence (*α=β=1, TI = C.D*).

To confirm the multiplicative conjoint makes sense at the boundaries, consider the settings for α, β to model extreme conditions where TI is influenced by only one of its attributes as illustrated in Figure 16.

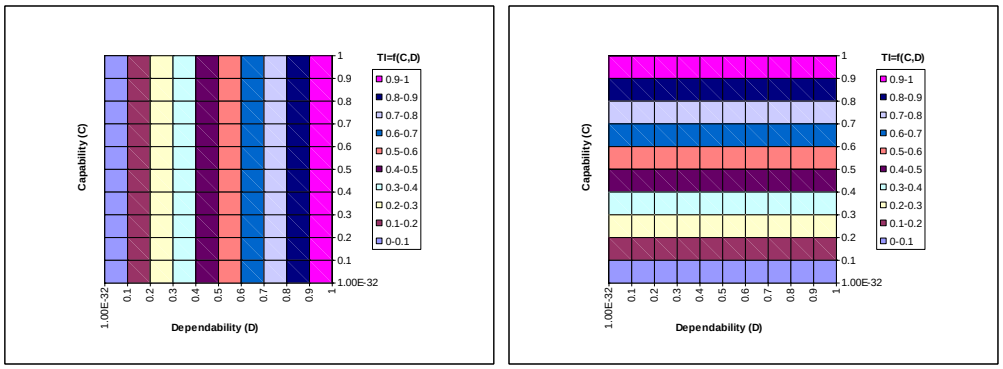


Figure 16. Extreme cases: (i) *TI* not influenced by capability (*α=0, TI = D*). (ii) *TI* not influenced by dependability (*β=0, TI = C*).

The measurement model caters more generally where there may be a different relative utility of *C* and *D*. If this difference in utility is modelled by appropriate selection of values of α, β the shape of the *TI* function can be skewed for a dependability-biased measure of technical integrity or a capability-biased measure of technical integrity. This is illustrated for two selections of α, β a shown in Figure 17.

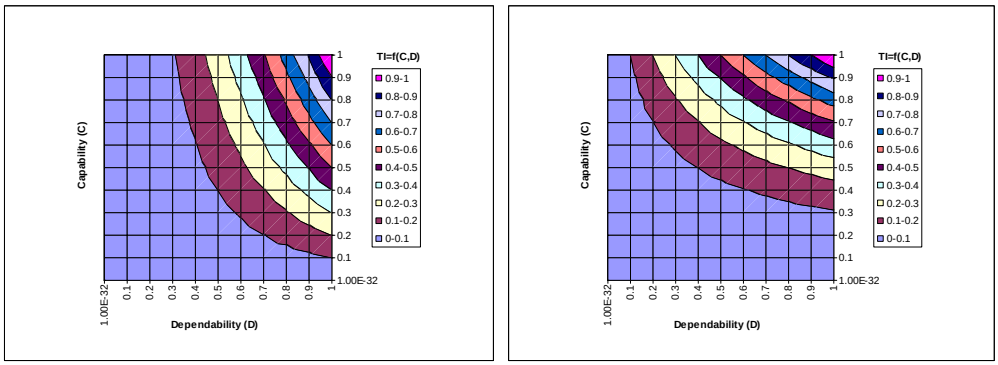


Figure 17. Differing relative utilities of capability and dependability.
(i) Dependability-biased ($\alpha=1, \beta=2, TI = C.D^2$).
(ii) Capability-biased ($\alpha=2, \beta=1, TI = C^2.D$).

If values of α, β across a range of disparate system types are consistent, then the conjoint measurement for *TI* will be useful for comparisons of technical integrity across a range of disparate systems. Establishment of values for α, β and consistency across a range of disparate system domains is the subject of Chapter 7 of this thesis.

Notwithstanding the appropriateness of this measurement construct, the problem of defining measures for capability and dependability remains. How these components are to be measured such that the *TI* of different systems may be compared is considered next.

5.3.4 Capability measurement approach

The definition for capability arrived at earlier is 'that the system meets all specified requirements'. The measure could then be the fraction of the originally intended requirements actually achieved. This would equate to failure to meet any requirements being assigned $C = 0$ and all requirements being achieved as $C=1$ and intermediate values measured on the interval (0, 1) as a ratio scale.

While it is common for detailed metrics to be kept about requirements changes and achievements in most system development projects, these may not be readily disclosed to researchers. It may be that for practical purposes, subjective assessments of *capability* could also need to be mapped to values on the (0, 1) scale.

There is precedent for using an ordinal scale relating to requirement achievement. The Study of Systems Engineering Effectiveness (Elm et al., 2007) uses requirements

verified as the measure for 'Scope', one of the defined elements of 'Project Performance' in that study. The study uses a 4 point ordinal scale to represent a subjective assessment of requirement verification. The SE-ROI project (Honour 2010) considers 'Technical Quality' as being an ordinal scale where satisfaction of requirements is the 'Threshold' level and delivered performance beyond that specified for minimum acceptability, but of value to the stakeholders, is termed the 'Objective' level.

5.3.5 Dependability measurement approach

The definition of dependability derived earlier is equated to the absence of risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. Considerations for measuring the totality of known and unknown risks in a complex engineered system was reviewed in section 3.4.3 and is further addressed in the following sections.

5.3.5.1 Feasibility of measurement

It seems clear that dependability should be measurable, at least on an ordinal scale. It should be valid to say that System A is more dependable than System B. Measurement on a ratio scale should not be beyond reach, e.g., it can be stated how much safer air travel from City A to City B is compared to a car journey between the same cities based on historical accident statistics.

5.3.5.2 Required form of measurement

Consistent with the conjoint measurement structure established previously, a measure of $D \in [0..1]$ is required where 0 indicates the system is not at all dependable and there is certainty in undesired outcomes occurring and 1 indicates the system is 100% dependable and there is no risk of undesired outcomes. Ideally this will form a ratio scale. Conceiving a measurement approach is not as straightforward as the measurement of capability, nor has a precedent for how to measure a concept such as dependability been identified in the literature survey.

5.3.5.3 Approaches for investigation

As found in section 3.4.3, the literature in the area of measurement of overall risk of unintended consequences of a system largely concentrates on the identification of risks

and the qualitative (more often) and quantitative (less often) assessment of each of those risks.

One approach may be to consider dependability as two separately quantified components. One component would be the identified risks and the other the unidentified risks. The identified risks could be combined into a single value where each consequence identified is associated with a cost (i.e., dollar) value and that value weighted by likelihood of its occurrence. All these weighted values may be summed together to form an overall risk. How to quantify the unidentified risks remains an issue. Some proxy measurement that would follow approaches established for the estimation of undetected faults remaining in complex software intensive systems is possible but may not be able to be pragmatically collected across a range of disparate systems. Collection of a range of related measurements and combining them in an assessment tool based on a Bayesian belief network is also possible (Fenton et al., 1998).

Another approach is to look at the number of incidents that breach dependability in the operational life of a system. While this may lead to a more direct form of measurement, there are inherent difficulties that include the infrequent occurrence of high consequence breaches of dependability in critical systems. This makes inference of dependability difficult. A proxy measure where the dependability of the system may be inferred by the more frequent occurrences of low consequence breaches may be a path to further this approach. E.g., (Harris, 2004) finds that in aviation the relationship between frequency of occurrence and consequence magnitude for safety incidents follow a power law indicative of self organising critical systems.

A subjective estimating approach for dependability may be to formulate a survey instrument where stakeholders and/or experts in the system may be questioned about dependability characteristics of the system and their responses combined to form a dependability measure. Effectively this would measure dependability through the means of a scale or index. Precedent for this as a pragmatic approach includes the System Usability Scale (Brooke, 1996). This approach will be pursued in the remainder of this thesis.

5.4 Elements of technical integrity

5.4.1 Technical goals of regulation

The paper at Appendix G developed an initial ontology of technical regulation to form the structure and basis for investigation of the research problem. At the second level of the ontology, the types of technical goals that can be the aim of a broad range of technical regulations were established. They are shown and described in the first two columns of Table 11.

The first seven of these are defined consistent with those identified as 'Aims of standardization' in ISO/IEC Guide 2 (Hunter, 2009). The eighth, Supportability, has been added since regulatory goals associated with reliability, maintainability and other aspects of system support that are evident in many regulations examined are not explicitly covered in the definitions of ISO/IEC Guide 2. These technical goals form a superset that is inclusive of technical integrity.

5.4.2 Technical goals and system quality attributes

Some of the technical regulations considered in development of the ontology express their technical goals in terms of system quality attributes and it is neither always straightforward nor objective to allocate these goals to the technical goals of the ontology. To validate the ontology and to refine the descriptions of each of the technical goals, two generic lists of quality attributes for systems were considered in detail as first reported in the peer-reviewed conference paper at Appendix G (Edwards, 2010). The method in the paper includes the use of (Wikipedia, 2010) as a "crowd source" since it has some of the characteristics of a voluntary survey. It provided a near-exhaustive list[4] of possible quality attributes for a system. The second primary reference was from (Firesmith et al., 2009c) which contains a more constrained hierarchy of quality attributes characterised into focus areas to support system architecture and design.

Cognisant of the limitations of Wikipedia as a source, in particular biases that may arise from a small number of dominant contributors, care was taken to determine that it

[4] "near-exhaustive" since there is likely other quality attributes that could be reasonably asserted and the list in Wikipedia has not necessarily been subjected to a credible review process that is normally associated with a standard, journal or conference paper.

represented a large superset of attributes that was inclusive of other references including (ISO/IEC 9126, 1991), the GE Software Quality Measurement Manual (McCall and Matsumoto, 1980), "Non-Functional Requirements"(Glinz, 2007), and CMU/SEI-95-TR-021 Quality Attributes (Barbacci et al., 1995) amongst others.

The resultant set of attributes is mapped in Table 11 against the technical goals to comprehensively identify the range of attributes that may be considered to be valid variables for the assessment of technical integrity and dependability in particular.

Table 11. Technical goals of the ontology associated with quality attributes of a system.

Technical goal	**Technical goal description**	**Associated quality attributes**
Fitness for Purpose	Ability of a product, process, or service to serve a defined purpose under specific conditions	accuracy affordability compliance correctness dependability deployability effectiveness fidelity functionality mobility operability performance precision relevance responsiveness stability timeliness usability
Compatibility	Suitability of products, processes or services for use together under specific conditions to fulfil relevant requirements without causing unacceptable interactions	compatibility installability interoperability

Technical goal	Technical goal description	Associated quality attributes
Interchangability	Ability of one product, process, or service to be used in place of another to fulfil the same requirements	adaptability configurability customizability flexibility interchangability modularity
Variety Control	Selection of the optimum number of sizes or types of products, processes, or services to meet prevailing needs	capacity configurability efficiency scalability
Safety	Freedom from unacceptable risk of harm	safety
Product Protection	Protection of a product against climatic or other adverse conditions during its use, transport or storage	deployability durability mobility portability robustness security survivability
Supportability	Ability of a product, process, or service to be supported to serve its defined purpose under specific conditions Support may include maintenance and training	administrability availability learnability maintainability manageability predictability recoverability reliability repeatability reproducibility reusability serviceability stability sustainability testability upgradability

5.4.3 Indicators of technical integrity

Consideration of the system quality attributes identified in Table 11 against the established concepts of technical integrity, capability and dependability led to the postulated list of indicators of technical integrity shown in Table 12. This list was refined through iterative review with professional colleagues of the researcher who were knowledgeable in ADF technical regulations and the development of complex engineered systems.

Table 12 lists the name of each indicator, its normal meaning, the concept intended to be encapsulated by that indicator and the expected correlation of that indicator with the concept of either capability or dependability in the conjoint measurement construct for technical integrity. The fourth column identifies variables and attributes of a system expected to be associated with the indicator.

Table 12. Postulated indicators of technical integrity.

Indicator	**Meaning**	**Expected correlation**	**Associated variables and system characteristics**
1. safety	The system will not cause inadvertent loss of life, injury or illness.	depend-ability	safety mishaps number and magnitude of residual safety risks hazardous materials designed safety controls
2. impact to the natural environment	The system meets environmental standards	depend-ability	environmental mishaps design controls for environmental protection percentage of environmental protection requirements met dangerous electromagnetic or acoustic emissions hazardous materials (released during operation or disposal)

Indicator	Meaning	Expected correlation	Associated variables and system characteristics
3. inadvertent impact to other systems	The system does not inadvertently negatively affect other systems	depend-ability	other system interference or damage design controls for impact avoidance integrity of information exchange integrity of energy exchange integrity of material exchange
4. fit for service	The system satisfies operational requirements in the intended operational environment	depend-ability	percentage of technical requirements met end-user evaluation
5. capability[5]	The function and performance required of a system	capability	percentage of function and performance requirements met level of performance in excess of specified requirements
6. effective	The system will perform its intended role with desired effects	capability	percentage of function and performance requirements met multi-function/features in the design end-user evaluation
7. efficient	The system will perform its intended role with a minimum of resources	capability	number of personnel (operating, maintaining) energy required output/input
8. construction quality	The system is “built right”	depend-ability	percentage of build standard requirements met number of non-conformances reported accreditation of quality system numbers in service (market demonstrated quality; mature design and build process)

[5] Performance and function may be entwined in some definitions, e.g., ‘Performance. The degree to which a system ... accomplishes its designated functions within given constraints’ IEEE 1990. IEEE Standard Glossary of Software Engineering Terminology. New York: Institute of Electrical and Electronic Engineers..

Indicator	Meaning	Expected correlation	Associated variables and system characteristics
9. operationally flexible / flexibility	The system's ability to be used for multiple purposes and/or to be adaptable to purposes not originally envisaged for it	capability	multi-function/features in the design end-user evaluation
10. durable	The system will perform its intended role in spite of repeated mechanical handling and stress	depend-ability	percentage of environmental condition requirements achieved materials and finishes MTBF
11. survivable	The system's ability to survive and continue to operate in the presence of defined threat and extreme conditions.	depend-ability	percentage of environmental protection requirements met percentage of survival requirements met (e.g., shock, blast tests) level of design standards threat resistant design features (e.g., electronic protection, shock and blast hardening or isolation, signature management)
12. dependable[6]	The system will perform its intended role when needed; it can be relied upon in the operating environment	depend-ability	A_o, A_i MTBF MTTR number of single points of failure designed redundancy qualification for operating environmental electromagnetic compatibility
13. reliability	The system's ability to continue to operate without fault or failure	depend-ability	MTBF number of single points of failure designed redundancy

[6] In computer systems 'dependability is that property of a computer system such that reliance can justifiably be placed on the service it delivers' and includes attributes of availability, reliability, safety, confidentiality, integrity and maintainability BARBACCI, M., KLEIN, M. H., LONGSTAFF, T. A. & WEINSTOCK, C. B. 1995. Quality Attributes. Pittsburgh, Pennsylvania: Software Engineering Institute, Carnegie Mellon University.

Indicator	Meaning	Expected correlation	Associated variables and system characteristics
14. supportable	The degree to which services and materials are available to allow the system to continue operating	depend-ability	maintenance and logistics support concepts numbers in service MTTR A_o spares delay time stores delay time
15. maintainable	The ease with which the system can be kept in good working order	depend-ability	MTTR condition monitoring features fault identification and recording features hours/cost of preventative maintenance
16. technically compliant	The system meets its specified requirements	capability	percentage of technical requirements met (fully and/or partially compliant)
17. usable	The ease with which operators can interact correctly with the system to achieve its intended purpose.	capability	end-user evaluation HMI design features
18. secure	The system will resist malicious attempts to subvert it (physically or electronically)	depend-ability	security breaches percentage of detected security threats repelled number and magnitude of residual security risks physical security design features computer/information security design features

This list of indicators and associated variables forms the basis for survey research and the analysis presented in the subsequent chapters.

5.5 Conclusion

A theory and definition of technical integrity has been established that is based upon two concepts of a complex engineered system. The concepts reflect the intended, desired emergent properties of a system and the minimisation of unintended, undesired emergent properties of a system. These concepts have been termed capability and dependability respectively and the definition is:

Technical integrity is defined as consisting of capability and dependability, where:

Capability is defined to mean that the system meets all specified requirements for performance, function and quality attributes in the defined environment and when used as intended. Capability is the designed set of emergent properties of a system that are intended and useful (i.e., required); and

Dependability is defined to mean that the risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. It includes concepts of robustness (system is tolerant of usage that may exceed specified conditions without catastrophic consequences), reliability (failures do not happen often), and system safety (risk of unidentified hazards and failure modes resulting in undesirable outcomes is minimised).

A conjoint measurement construct has been formulated from the concepts of capability and dependability. Approaches to measure capability and dependability have been considered and a list of indicators that may be used in a measurement instrument for technical integrity have been identified.

The concepts developed in this chapter were presented for critical review at INCOSE Symposium 2011 and is included at Appendix J. Peer review of the paper and discussions at the symposium presentation provided confidence that these concepts are sound and research to further validate and develop the concepts is warranted.

6 Content validity of the technical integrity measurement construct

6.1 Introduction

Technical integrity of a complex engineered system is not measurable directly and no means for measurement via other indirect means has previously been determined. Chapter 5 developed a theoretical model and definition of technical integrity and from this theory proposed a measurement construct involving a conjoint measurement based upon component measures of *capability* and *dependability*.

This chapter presents the results of the phase of research conducted to establish the content validity of the proposed measurement construct. Content validity refers to the extent to which a measure represents all facets of the phenomenon of interest and is assessed in this research via a consensus of expert judgements sought by means of an online survey (Allen and Yen, 1979, p. 95-96) (Tashakkori and Teddlie, 1998, p. 83).

A survey instrument has gathered both rating data to particular questions seeking the level of agreement with various aspects of the conjoint measurement construct for technical integrity and asked open-ended questions through which additional insights into the reasons behind the expert ratings can emerge. Both quantitative and qualitative analysis methods have been used to analyse the data gathered from 116 respondents.

6.2 Method

The method used a survey research strategy to collect expert judgements in the form of ratings reflecting the levels of agreement with aspects of the proposed measurement construct. Each judgement item uses a Likert-type response format, but they do not

form a Likert scale (Carifio and Perla, 2007). The survey also collects written responses from open-ended questions.

The rating data is analysed quantitatively in order to determine the level of support for the measurement construct. Effectively, the quantitative analysis procedure is establishing the content validity of the proposed measurement construct by testing the following hypotheses:

- Technical integrity can be measured on an ordinal scale.
- Technical integrity can be measured by a single numeric value.
- Technical integrity can be measured on a ratio scale.
- Capability of a complex engineered system can be measured on a scale of range 0 to 1.
- Capability of a system in excess of specified requirements is valued.
- The proposed capability measure is best.
- Dependability of a complex engineered system can be measured on a scale of range 0 to 1.
- Dependability of a complex engineered system that is better than the risk acceptability threshold of stakeholders is valued.
- The proposed dependability measure is best.
- Technical integrity can be measured by a conjoint measurement from the components capability and dependability.
- The conjoint measurement of technical integrity allows for comparison of similar systems.
- The conjoint measurement of technical integrity allows for comparison of dissimilar systems.

A qualitative analysis procedure is employed on the open-ended response data in order to assist in the interpretation of and find explanations for the quantitative results. This is essentially ‘theory guided research’(Cropley, 2002, p. 101) where the analysis begins with the response data and develops an understanding from it.

Finally a mixed-method approach is used to ‘triangulate’ the quantitative and qualitative results in order to provide a ‘better indicator of the true situation than either qualitative or quantitative findings alone’(Cropley, 2002, p. 86-87).

6.2.1 Survey instrument

The survey instrument was implemented in Survey Monkey. The full survey instrument is at Appendix A. The survey was piloted, as recommended by (Walliman, 2010, p. 98), with academic supervisors and professional colleagues of the researcher in order to identify and resolve ambiguities in the instrument prior to its full launch.

6.2.2 Sampling approach

The identification of target respondents is best described as 'purposive sampling' (Tashakkori and Teddlie, 1998, p. 76). Survey respondents have been requested to participate based on their familiarity with the definition, design, integration, verification, operation and/or support of complex engineered systems. Intentionally in the survey design, they did not need to be considered to be an expert to participate in this study as a judge since demographic data collected allows examination of variations in judgements against criteria that may be proxies for expertise.

To some extent, the approach could also be considered to be 'convenience sampling' (Tashakkori and Teddlie, 1998, p. 76) where the researcher utilised his professional network in the Australian military and defence industries in order to identify suitable respondents. While this is not inconsistent with the delimitations for the overall research (particularly the focus on ADF technical regulations) potential bias would need to be considered in generalising findings beyond the ADF technical regulatory environment.

The survey responses were collected in July through September 2012. Data is stored as non-identifiable records (i.e., no personal identification data is stored and the demographic data supplied cannot be attributed to a particular person).

207 requests were made, with 125 of those commencing the survey and 116 completing the survey, representing a 56% completion rate.

6.2.3 Demographics of respondents

This section presents the raw demographic information and broad observations about the respondents as a group only. The results sections will use this data to identify if there are any significant differences in judgements and other responses that may be attributable to or explained by demographic groupings.

The charts in this section summarise the demographic data provided by the respondents. It is noted in Figure 18 that there is a preponderance of male respondents in the group. This is not particularly unrepresentative in the engineering of defence systems.

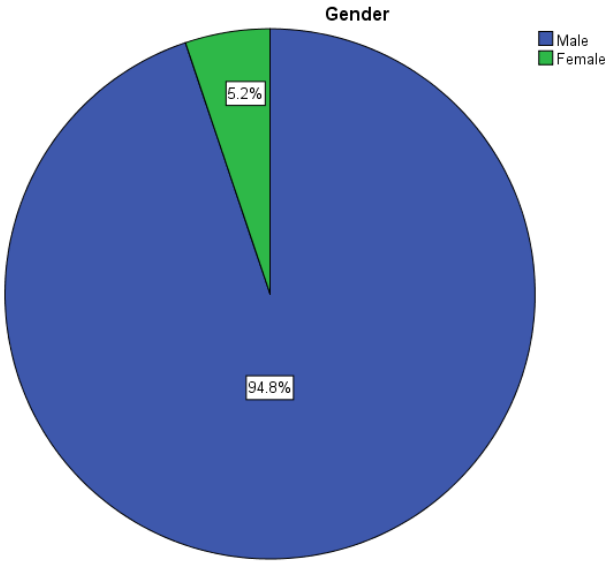


Figure 18. Pie chart of the gender of the respondents (n=116).

The distribution of ages of the respondents is shown in Figure 19. Along with Figure 24 which shows the years of professional experience, it indicates that most of the survey respondents are mature and experienced.

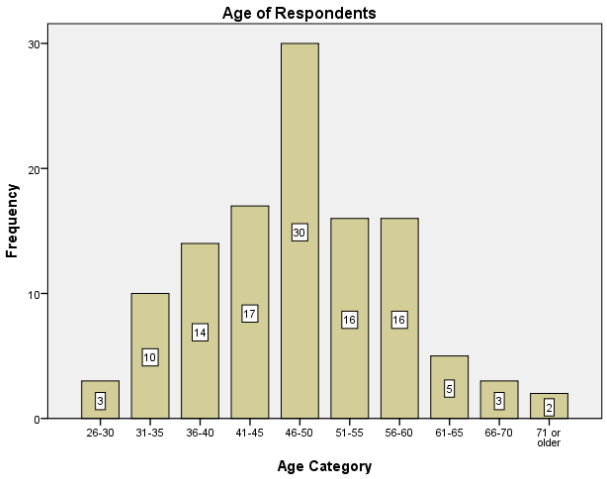


Figure 19. Histogram of the age of the respondents (n=116).

Figure 20 shows the highest level of technical education achieved by the respondents. 94.8% of the respondents have a university level education. This may be compared to

the general Australian population where 24% people aged between 15 and 64 years hold a bachelors level degree or higher (Australian Bureau of Statistics, 2011) and demonstrates that the respondents as a group are significantly more educated than the general population as expected in the engineering and usage of complex systems.

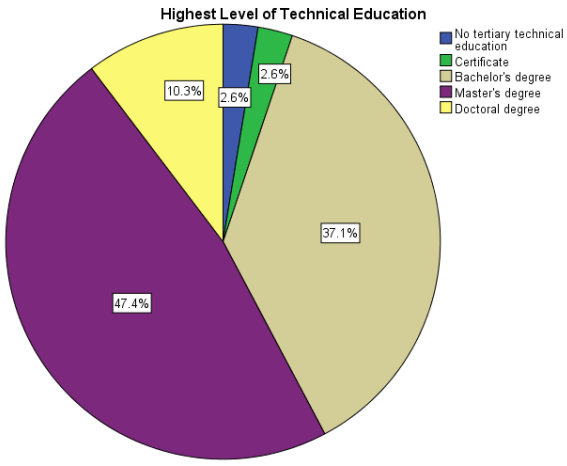


Figure 20. Pie chart of the highest level of technical education of the respondents (n=116).

Figure 21 shows there is a majority of respondents who have identified their main technical education as being in electrical/electronic engineering. This is somewhat unexpected since most technical disciplines are needed in the realisation of complex engineered systems. It may be partly explained in that most, if not all, of these systems embody electrical systems whereas, for example, aircraft based systems do not use the services of naval architects.

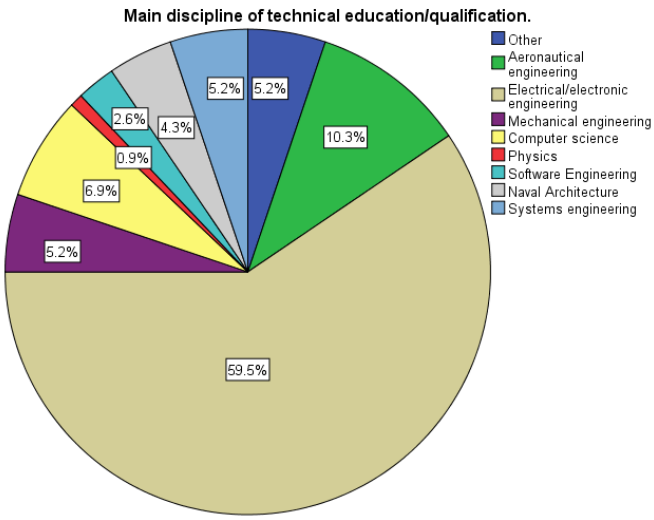


Figure 21. Pie chart of the main discipline of technical education of the respondents (n=116).

Figure 22 and Figure 23 on employment sector reflect a good spread among the defined sectors. Note the larger number of respondents who are currently employed as consultants compared to where most experience had been gained. This could be explained by persons having long careers in the military or government and then taking on consulting work later in their careers.

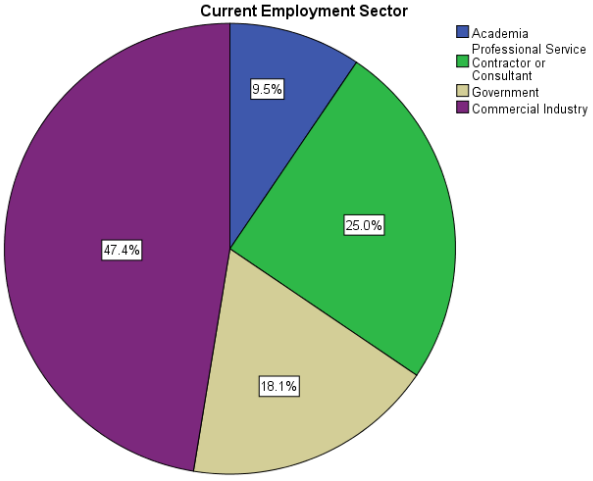


Figure 22. Pie chart of the current employment sector of the respondents (n=116).

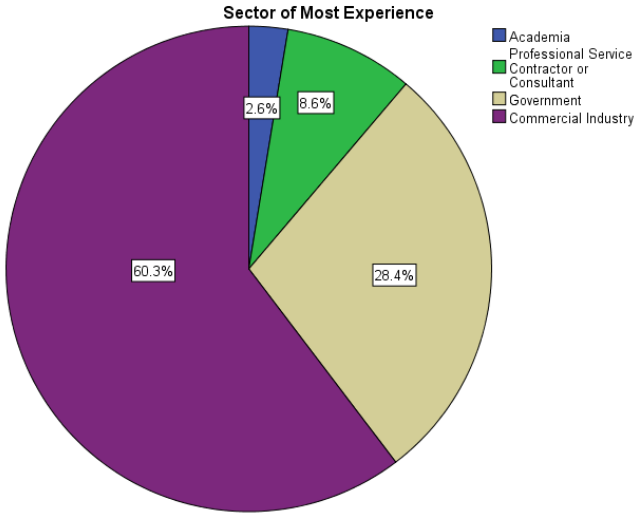


Figure 23. Pie chart of the employment sector of most experience gained by the respondents (n=116).

The group exhibits significant professional experience (most with greater than 10 years as shown in Figure 24) and 52% with formal engineering delegations in ADF regulatory frameworks (refer Figure 25 and Figure 28); an experienced and responsible group of professionals.

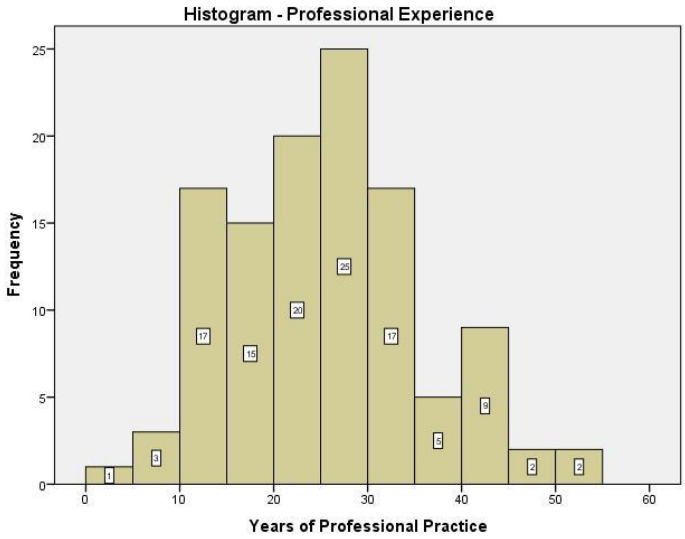


Figure 24. Histogram of the years of professional practice of the respondents (n=116).

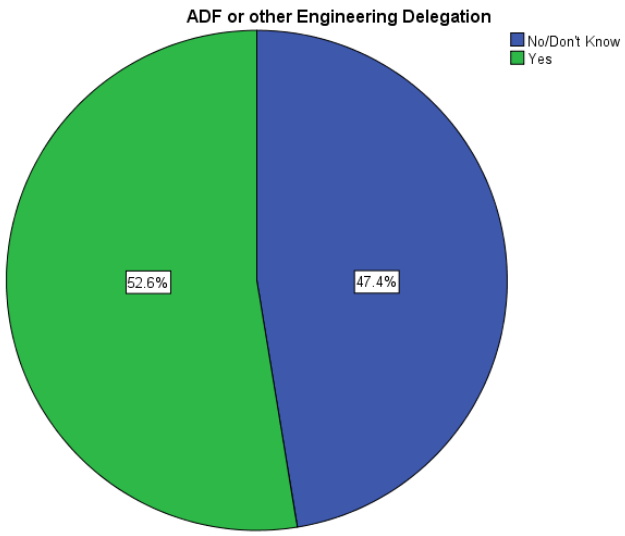


Figure 25. Pie chart of the ADF or other engineering authority delegation held by the respondents (n=116).

Figure 26 shows the system application domain where the respondents have their main experience. A majority of respondents have their main experience in the application domains of Maritime Defence Systems and Aerospace Systems. It is possible that this reflects the researcher's own professional network from which most respondents were invited and also that the area of technical regulation and technical integrity is of high importance and is currently in active debate amid regulatory reform initiatives within the Australian Air Force and Navy particularly. The relatively small number of respondents, about 11%, in 'Other' or Commercial ICT may limit any generalisations that can be made across non-military application domains.

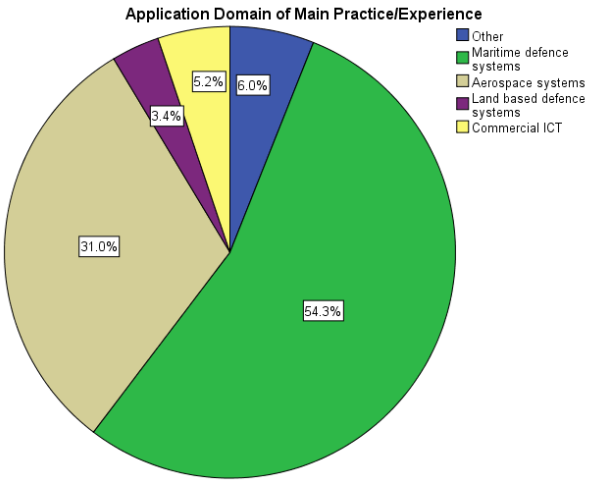


Figure 26. Pie chart of the system application domain where the respondent has most experience (n=116).

Figure 27 demonstrates a reasonable spread in the group between roles defined as Capability Manager, Acquisition Manager and System Developer. There are relatively few respondents who have identified as System Operators. This will limit any generalisation of results to be inclusive of stakeholders with an operator viewpoint.

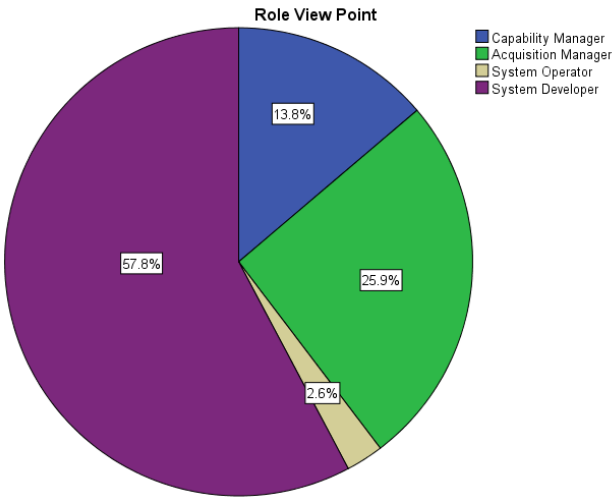


Figure 27. Pie chart of the system role viewpoint the respondent believes they represent (n=116).

Sixty one (52.6%) of the respondents indicated that they hold or have held formal delegations of Engineering Authority in organisations governed by ADF technical regulations or other regulatory systems. The delegations held are shown in Figure 28. This demonstrates that many of the respondents represent a group of very experienced engineers and that they are, or have previously been, entrusted to make significant engineering decisions for the ADF. There is also representation across the three technical regulatory domains of the ADF (maritime materiel– regulated under the Naval Technical Regulatory System (NTRS); aerospace materiel – regulated by the Technical Airworthiness Regulator (TAR); and land materiel under the Technical Regulation of ADF Materiel – Land (TRAM-L)).

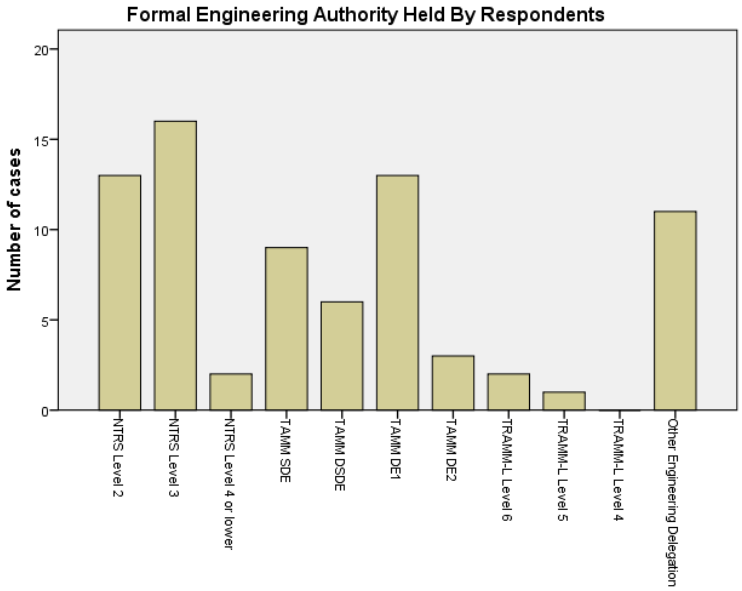


Figure 28. Bar chart of the engineering delegation held by respondents (n=116).

6.2.4 Analysis of survey responses

6.2.4.1 Quantitative analysis procedure

The research methodology overarching this procedure is survey research, that is, 'acquiring information about one or more groups of people ... about their opinions ... by asking them questions and tabulating their answers' (Leedy and Ormond, 2005, p. 183-186). Additionally, ratings data is used directly to test support for the hypotheses defined previously in this chapter. The analysis procedure for each set of ratings data is:

- Tabulate the frequencies of the level of agreement ratings and produce a bar chart to assist in visualising the data.
- Re-code the data into categories to allow testing of the hypothesis.
- Significance of the hypothesis testing result is established using both a chi-square test and Lawshe's Content Validity Ratio (CVR) (Lawshe, 1975) with significance values as amended by (Wilson et al., 2012).

6.2.4.2 Qualitative analysis procedure

The qualitative analysis procedure essentially follows that of (Leedy and Ormond, 2005, p. 140) adapted as follows:

- The researcher becomes familiar with the data by thoroughly reviewing all responses and to identify themes of relevance to understanding reasons for the judgements provided in the ratings responses
- Evident themes or concepts are defined and codes established to pursue particular understandings that underlie the quantitative results.
- Text is extracted and associated with one or more of the codes.
- The extracts are tabulated for each code.
- Counts of extracts against each code are reported to establish frequency of the associated theme or concept.
- Text extracts within each code group are compared, contrasted and discussed.

The qualitative analysis commenced by importing the responses into Dedoose. Dedoose[7] is a web-based tool for qualitative and mixed methods research. Mixed method visualisation and analysis tools within Dedoose were used to further explore for underlying relationships and meanings within the response data.

6.3 Quantitative results

6.3.1 Quantitative analysis: technical integrity measurement scale

6.3.1.1 Technical integrity – measurement on an ordinal scale

The survey instrument posed the assertion 'It is meaningful to say system A has higher technical integrity than system B' and the respondent was requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 13 and Figure 29.

[7] See http://www.dedoose.com/ for more information.

Table 13. Frequency of ratings of the level of agreement of 'ordinal scale meaningful'.

Technical Integrity: Ordinal Scale Meaningful

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	2	1.7	1.7	1.7
	Disagree	10	8.6	8.6	10.3
	Neither Agree or Disagree	6	5.2	5.2	15.5
	Agree	78	67.2	67.2	82.8
	Strongly Agree	20	17.2	17.2	100.0
	Total	116	100.0	100.0	

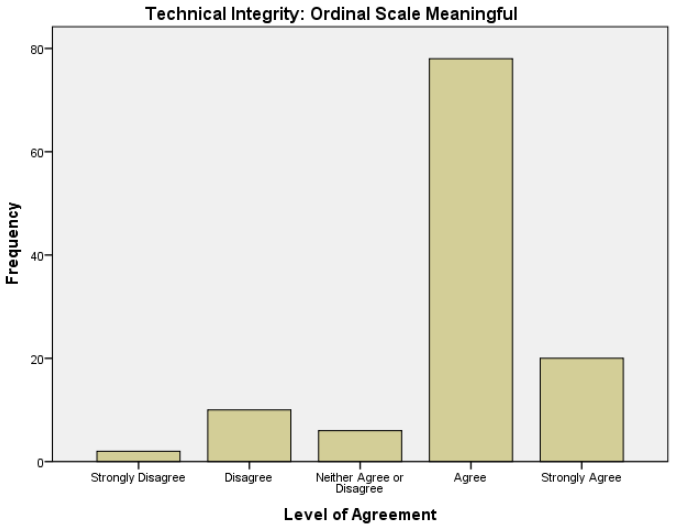


Figure 29. Bar chart of the level of agreement ratings of 'ordinal scale meaningful'.

To find if there is support for the hypothesis 'Technical integrity can be meaningfully measured on an ordinal scale', the 5 point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 14.

Table 14. Frequency of support for hypothesis 'Technical integrity can be meaningfully measured on an ordinal scale'.

Support for Ordinal Scale of TI

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	18	15.5	15.5	15.5
	Supported	98	84.5	84.5	100.0
	Total	116	100.0	100.0	

A chi-square test indicates that the 84.5% of judgements in support of the ordinal scale for technical integrity is significant, χ^2 (1, n=116) = 55.2, p<0.001.

Lawshe's CVR is calculated at 0.69 which indicates that at least half the respondents agree and is significant at the 0.005 level (one tailed test).

6.3.1.2 Technical integrity - measurement by a single numeric value

The survey instrument posed the assertion 'It is meaningful to ascribe a single number value to indicate the technical integrity of a system' and the respondent was requested to rate their agreement on a 5-point scale. Results obtained are as shown in Table 15 and Figure 30.

Table 15. Frequency of ratings of the level of agreement of 'measurement by a single numeric value'.

Technical Integrity: Single Numeric Value Meaningful

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	12	10.3	10.3	10.3
	Disagree	51	44.0	44.0	54.3
	Neither Agree or Disagree	21	18.1	18.1	72.4
	Agree	29	25.0	25.0	97.4
	Strongly Agree	3	2.6	2.6	100.0
	Total	116	100.0	100.0	

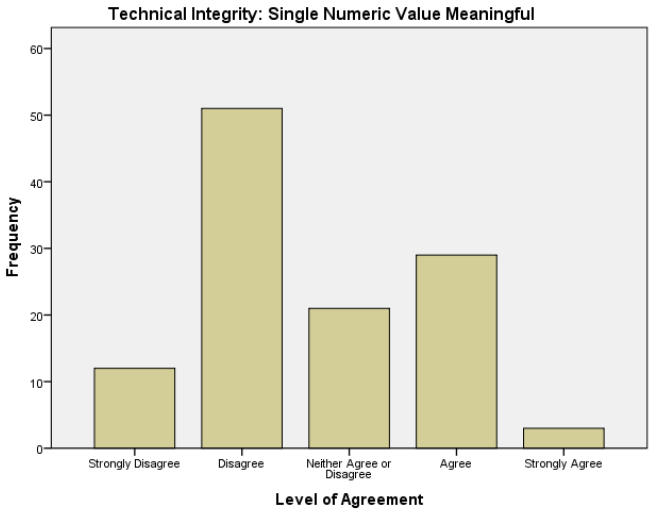


Figure 30. Bar chart of the level of agreement ratings of 'measurement by a single numeric value'.

To find if there is support for the hypothesis 'technical integrity can be measured by a single numeric value', the 5 point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 16.

Table 16. Frequency of support for hypothesis 'technical integrity can be measured by a single numeric value'.

Support for TI Measure as a Numerical Magnitude

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	84	72.4	72.4	72.4
	Supported	32	27.6	27.6	100.0
	Total	116	100.0	100.0	

A chi-square test indicates that the 72.4% of judgements **not** in support of the measurement of technical integrity by a single numeric value is significant, χ^2 (1, n=116) = 23.3, p<0.001.

Lawshe's CVR is calculated at -0.45 which indicates less than half the respondents agree and is significant at the 0.005 level (one tailed test).

6.3.1.3 Technical integrity – measurement on a ratio scale

The survey instrument posed the assertion 'It is meaningful to say that system A has, for example, twice the technical integrity of system B' and the respondent requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 17 and Figure 31.

Table 17. Frequency of ratings of the level of agreement of 'measurement on a ratio scale meaningful'.

Technical Integrity: Ratio Scale Meaningful

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	12	10.3	10.3	10.3
	Disagree	68	58.6	58.6	69.0
	Neither Agree or Disagree	19	16.4	16.4	85.3
	Agree	13	11.2	11.2	96.6
	Strongly Agree	4	3.4	3.4	100.0
	Total	116	100.0	100.0	

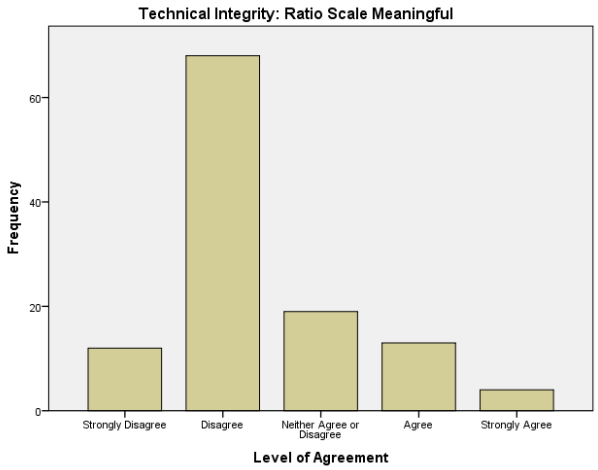


Figure 31. Bar chart of the level of agreement ratings of measurement on a ratio scale meaningful.

To find if there is support for the hypothesis 'technical integrity of a system can be measured on a ratio scale', the 5 point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 18.

Table 18. Frequency of support for hypothesis 'technical integrity of a system can be measured on a ratio scale'.

Support for TI Measure as Ratio Scale

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	99	85.3	85.3	85.3
	Supported	17	14.7	14.7	100.0
	Total	116	100.0	100.0	

A chi-square test indicates that the 85.3% of judgements **not** in support of the measurement of technical integrity on a ratio scale is significant, χ^2 (1, n=116) = 57.9, $p<0.001$.

Lawshe's CVR is calculated at -0.71 which indicates less than half the respondents agree and is significant at the 0.005 level (one tailed test).

6.3.2 Quantitative analysis: capability measurement scale

6.3.2.1 Capability - measurement on a scale of 0 to 1

The survey instrument posed the assertion ‘Capability of a complex engineered system can be meaningfully measured on a scale of 0 to 1’ and the respondent requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 19 and Figure 32.

Table 19. Frequency of ratings of the level of agreement of capability measured on a scale of 0 to 1.

Capability: [0..1] Measurement Scale Meaningful

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	5	4.3	4.4	4.4
	Disagree	20	17.2	17.5	21.9
	Neither Agree or Disagree	12	10.3	10.5	32.5
	Agree	66	56.9	57.9	90.4
	Strongly Agree	11	9.5	9.6	100.0
	Total	114	98.3	100.0	
Missing	System	2	1.7		
Total		116	100.0		

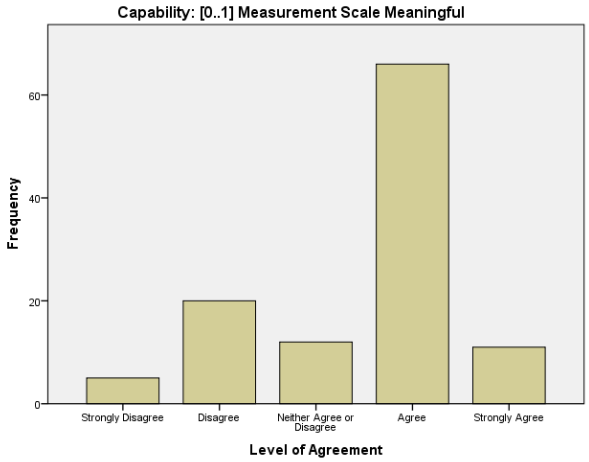


Figure 32. Bar chart of the level of agreement ratings of capability measured on a scale of 0 to 1.

To find if there is support for the hypothesis ‘Capability of a complex engineered system can be measured on a scale of 0 to 1’ the 5 point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 20.

Table 20. Frequency of support for hypothesis 'Capability of a complex engineered system can be measured on a scale of 0 to 1'.

Support for [0..1] Capability Scale

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	37	31.9	32.5	32.5
	Supported	77	66.4	67.5	100.0
	Total	114	98.3	100.0	
Missing	System	2	1.7		
Total		116	100.0		

A chi-square test indicates that the 67.5% of judgements in support of the measurement of Capability on a scale of 0 to 1 is significant, χ^2 (1, n=114) = 14.0, p<0.001.

Lawshe's CVR is calculated at 0.35 which indicates more than half the respondents agree and is significant at the 0.005 level (one tailed test).

6.3.2.2 Capability – more than specified requirement is valued

The survey instrument posed the assertion 'Capability of a complex engineered system that exceeds the specified requirements is not of value' and the respondent requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 21 and Figure 33.

Table 21. Frequency of ratings of the level of agreement of capability – more than specified requirement is valued.

Capability Over Specification is Not Valued

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	18	15.5	15.8	15.8
	Disagree	74	63.8	64.9	80.7
	Neither Agree or Disagree	9	7.8	7.9	88.6
	Agree	13	11.2	11.4	100.0
	Total	114	98.3	100.0	
Missing	System	2	1.7		
Total		116	100.0		

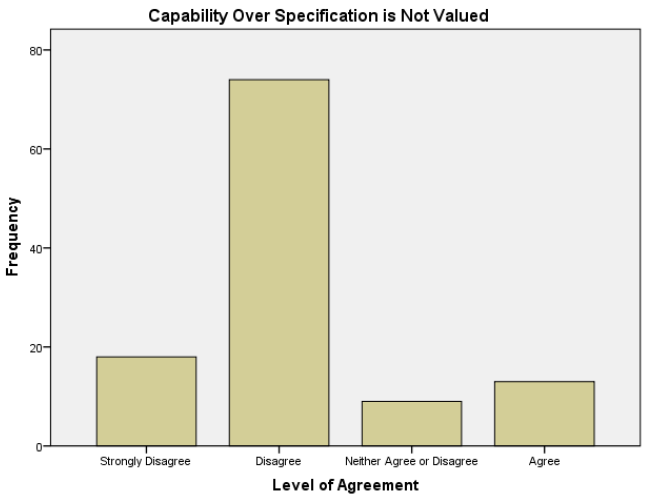


Figure 33. Bar chart of the level of agreement ratings of capability – more than specified requirement is valued.

To find if there is support for the hypothesis 'Capability of a system in excess of specified requirements is valued', noting the question was intentionally inverted, the 5 point scale was re-coded into two categories: Supported (Disagree, Strongly Disagree) and Not Supported (Neither Agree or Disagree, Agree or Strongly Agree) with results as shown in Table 22.

Table 22. Frequency of support for hypothesis 'Capability of a system in excess of specified requirements is valued'.

Support for Extra Capability is Valued

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	22	19.0	19.3	19.3
	Supported	92	79.3	80.7	100.0
	Total	114	98.3	100.0	
Missing	System	2	1.7		
Total		116	100.0		

A chi-square test indicates that the 80.7% of judgements in support of Capability of a system in excess of specified requirements in valued is significant, χ^2 (1, n=114) = 43.0, p<0.001.

Lawshe's CVR is calculated at 0.61 which indicates more than half the respondents agree and is significant at the 0.005 level (one tailed test).

6.3.2.3 Capability – better measures

The survey instrument asked the question 'What, if anything, do you think may be a better measure of Capability for a complex engineered system?' The respondent could reply that the proposed measure was OK, or if they thought there was a better measure. In the latter case they were expected to provide their suggestion in response the open-ended question which followed (results presented in the qualitative section of this chapter). Results obtained are as shown in Table 23 and Figure 34.

Table 23. Frequency of ratings for a better measure of capability.

Better Capability Measure

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes - Better Measure Suggested	72	62.1	63.2	63.2
	No - Measure OK	42	36.2	36.8	100.0
	Total	114	98.3	100.0	
Missing	System	2	1.7		
Total		116	100.0		

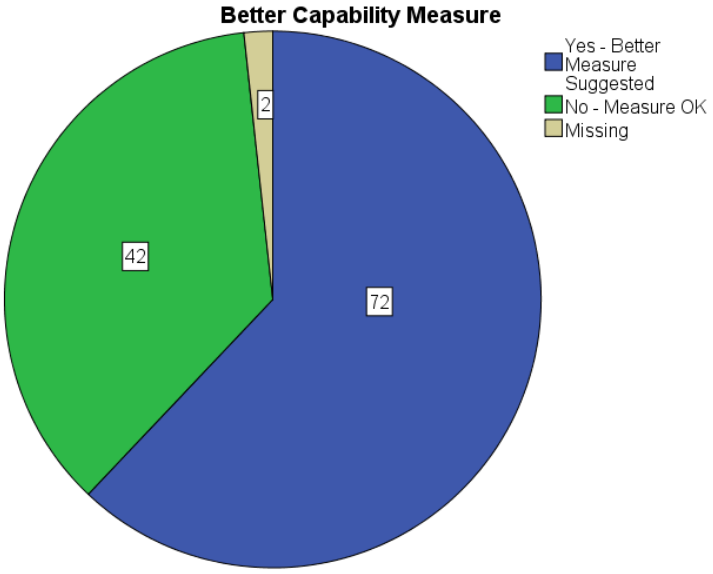


Figure 34. Pie chart of rating for a better measure of capability.

A chi-square test indicates that the 63.2% of judgements in support of a better Capability measure is significant, χ^2 (1, n=114) = 7.9, p<0.005.

Lawshe's CVR is calculated at -0.28 which indicates less than half the respondents agree that the proposed measure is the best and is significant at the 0.005 level (one tailed test).

6.3.3 Quantitative analysis: dependability measurement scale

6.3.3.1 Dependability - measurement on a scale of 0 to 1

The survey instrument posed the assertion 'Dependability of a complex engineered system can be meaningfully measured on a scale of 0 to 1' and the respondent requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 24 and Figure 35.

Table 24. Frequency of ratings of the level of agreement of dependability measured on a scale of 0 to 1.

Dependability: [0..1] Measurement Scale Meaningful

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	6	5.2	5.4	5.4
	Disagree	21	18.1	18.9	24.3
	Neither Agree or Disagree	15	12.9	13.5	37.8
	Agree	59	50.9	53.2	91.0
	Strongly Agree	10	8.6	9.0	100.0
	Total	111	95.7	100.0	
Missing	System	5	4.3		
Total		116	100.0		

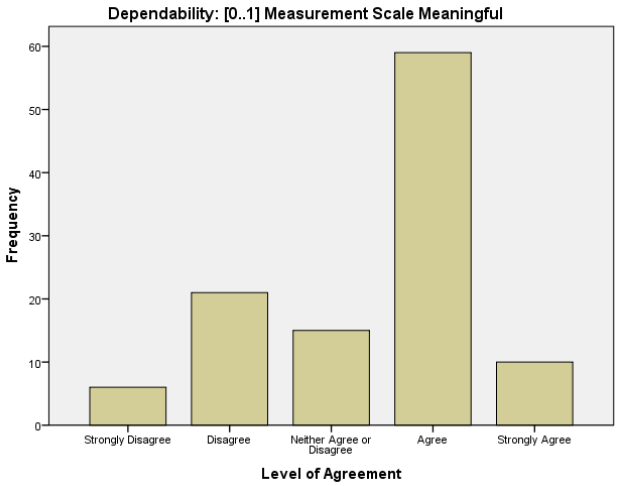


Figure 35. Bar chart of the level of agreement ratings of dependability measured on a scale of 0 to 1.

To find if there is support for the hypothesis 'Dependability of a complex engineered system can be measured on a scale of 0 to 1' the 5 point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 25.

Table 25. Frequency of support for hypothesis 'Dependability of a complex engineered system can be measured on a scale of 0 to 1'.

Support for [0.1] Dependability Scale

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	42	36.2	37.8	37.8
	Supported	69	59.5	62.2	100.0
	Total	111	95.7	100.0	
Missing	System	5	4.3		
Total		116	100.0		

A chi-square test indicates that the 62.2% of judgements in support of the measurement of Dependability on a scale of 0 to 1 is significant, χ^2 (1, n=111) = 6.6, p<0.010.

Lawshe's CVR is calculated at 0.24 which indicates more than half the respondents agree, but is **not** significant at the 0.005 level (one tailed test).

6.3.3.2 Dependability – more than risk acceptance threshold is valued

The survey instrument posed the assertion 'A complex engineered system with dependability that is better than the risk acceptance threshold of the system's stakeholders is of increased value' and the respondent was requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 26 and Figure 36.

Table 26. Frequency of ratings of the level of agreement of dependability – better than risk acceptance thresholds is of increased value.

Dependability better than the risk acceptance threshold is of Increased Value.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	3	2.6	2.7	2.7
	Disagree	10	8.6	9.0	11.7
	Neither Agree or Disagree	10	8.6	9.0	20.7
	Agree	76	65.5	68.5	89.2
	Strongly Agree	12	10.3	10.8	100.0
	Total	111	95.7	100.0	
Missing	System	5	4.3		
Total		116	100.0		

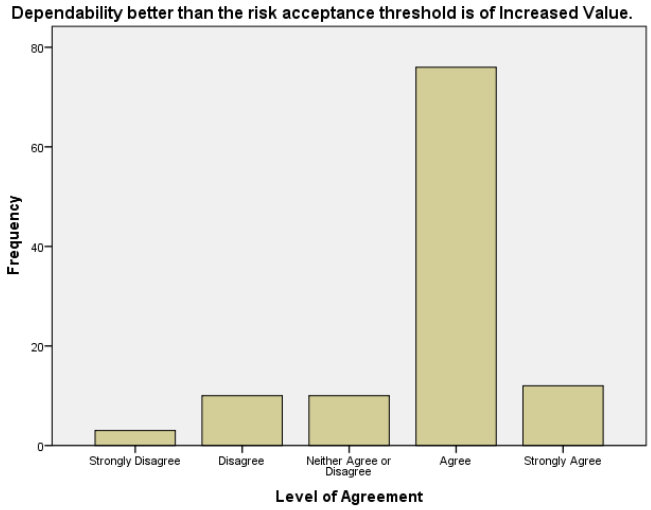


Figure 36. Bar chart of the level of agreement ratings of dependability – better than risk acceptability threshold of stakeholders is valued.

To find if there is support for the hypothesis 'Dependability of a complex engineered system that is better than the risk acceptability threshold of stakeholders is valued' the 5 point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 27.

Table 27. Frequency of support for hypothesis 'dependability – better than risk acceptability threshold of stakeholders is valued'.

Support for Extra Dependability is Valued

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	23	19.8	20.7	20.7
	Supported	88	75.9	79.3	100.0
	Total	111	95.7	100.0	
Missing	System	5	4.3		
Total		116	100.0		

A chi-square test indicates that the 79.3% of judgements in support of the value of Dependability better than the risk acceptability threshold is valued is significant, χ^2 (1, n=111) = 38.1, p<0.001.

Lawshe's CVR is calculated at 0.59 which indicates more than half the respondents agree and is significant at the 0.005 level (one tailed test).

6.3.3.3 Dependability – better measures

The survey instrument asked the question 'What, if anything, do you think may be a better measure of Dependability for a complex engineered system?' The respondent could reply that the proposed measure was OK, or if they thought there was a better measure. In the latter case they were expected to provide their suggestion in response the open-ended question which followed (results presented in the qualitative section of this chapter). Results obtained are as shown in Table 28 and Figure 37.

Table 28. Frequency of ratings for a better measure of dependability.

Better Dependability Measure

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes - Better Measure Suggested	44	37.9	39.6	39.6
	No - Measure OK	67	57.8	60.4	100.0
	Total	111	95.7	100.0	
Missing	System	5	4.3		
Total		116	100.0		

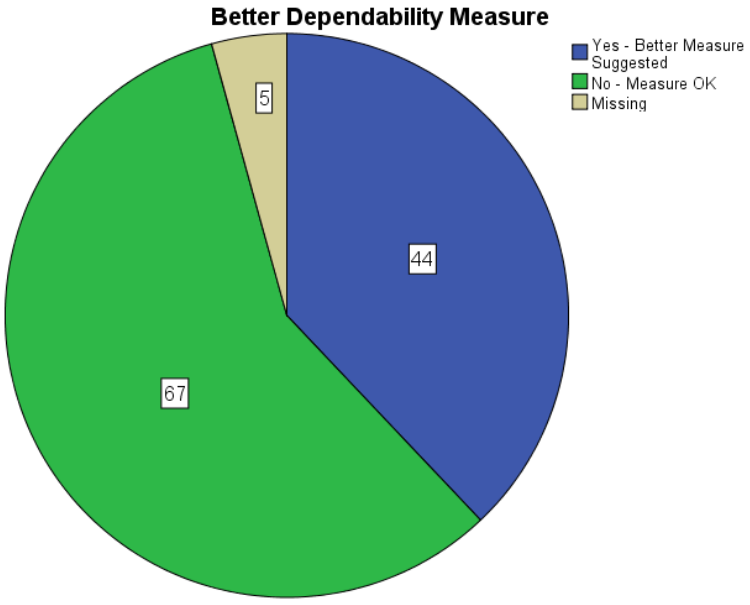


Figure 37. Pie chart of rating for a better measure of dependability.

A chi-square test indicates that the 60.4% of judgements in support of the proposed Dependability measure is significant, χ^2 (1, n=111) = 4.8, p<0.029 (or is **not** significant at p<0.005 level).

Lawshe's CVR is calculated at 0.21 which indicates more than half the respondents agree the proposed measurement approach is the best but is **not** significant at the 0.005 level (one tailed test).

6.3.4 Quantitative analysis: conjoint measurement of technical integrity

6.3.4.1 Technical integrity - conjoint measurement based on capability and dependability

The survey instrument posed the assertion 'Constructing a conjoint measure for technical integrity from the component measures of Capability and Dependability is meaningful' and the respondent requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 29 and Figure 38.

Table 29. Frequency of ratings of the level of agreement of a technical integrity conjoint measure is meaningful.

Technical Integrity Conjoint Measure Meaningful.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	3	2.6	2.7	2.7
	Disagree	19	16.4	17.3	20.0
	Neither Agree or Disagree	13	11.2	11.8	31.8
	Agree	71	61.2	64.5	96.4
	Strongly Agree	4	3.4	3.6	100.0
	Total	110	94.8	100.0	
Missing	System	6	5.2		
Total		116	100.0		

To find if there is support for the hypothesis 'technical integrity can be measured by a conjoint measurement from the components *capability* and *dependability*' the 5- point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 30.

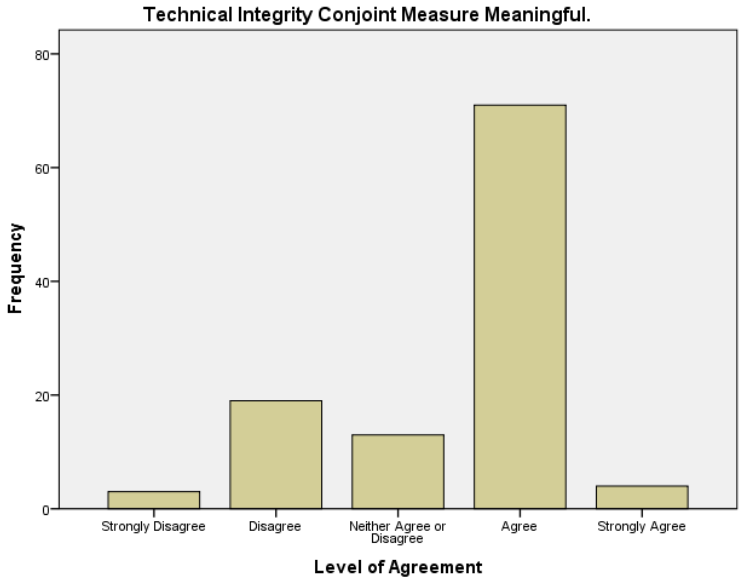


Figure 38. Bar chart of the level of agreement ratings of a technical integrity conjoint measure is meaningful.

Table 30. Frequency of support for hypothesis 'technical integrity can be measured by a conjoint measurement from the components capability and dependability'.

Support for the Conjoint Measurement of TI

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	35	30.2	31.8	31.8
	Supported	75	64.7	68.2	100.0
	Total	110	94.8	100.0	
Missing	System	6	5.2		
Total		116	100.0		

A chi-square test indicates that the 68.2% of judgements in support of a conjoint measurement for technical integrity is significant, χ^2 (1, n=110) = 14.5, p<0.001.

Lawshe's CVR is calculated at 0.36 which indicates more than half the respondents agree and is significant at the 0.005 level (one tailed test).

6.3.4.2 Technical integrity - conjoint measurement for comparison of similar systems

The survey instrument posed the assertion 'A conjoint measure for technical integrity will allow meaningful comparison of the actual technical integrity of systems that are similar in type and application' and the respondent requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 31 and Figure 39.

Table 31. Frequency of ratings of the level of agreement of a technical integrity conjoint measure is meaningful for similar system comparison.

Technical Integrity Conjoint Measure Meaningful for Similar System Comparison

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	3	2.6	2.7	2.7
	Disagree	16	13.8	14.5	17.3
	Neither Agree or Disagree	15	12.9	13.6	30.9
	Agree	73	62.9	66.4	97.3
	Strongly Agree	3	2.6	2.7	100.0
	Total	110	94.8	100.0	
Missing	System	6	5.2		
Total		116	100.0		

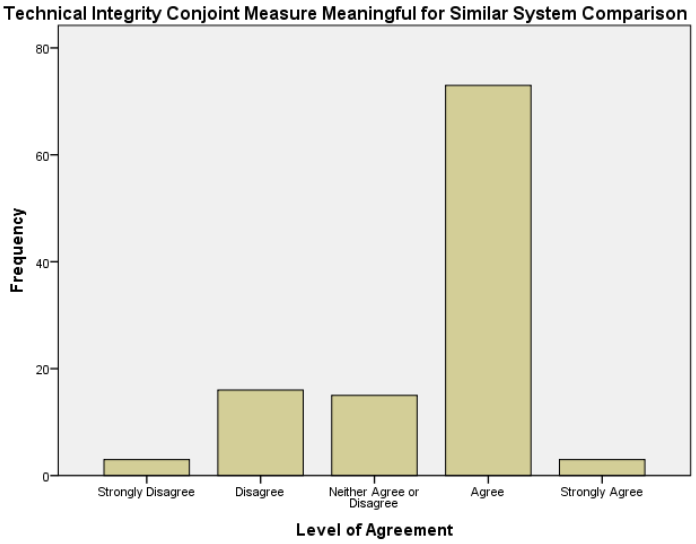


Figure 39. Bar chart of the level of agreement ratings of a technical integrity conjoint measure is meaningful for similar system comparison.

To find if there is support for the hypothesis 'A conjoint measurement of technical integrity allows for comparison of similar systems' the 5-point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 32.

Table 32. Frequency of support for hypothesis 'A conjoint measurement of technical integrity allows for comparison of similar systems'.

Support for TI Conjoint Meaningful for Similar Systems

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	34	29.3	30.9	30.9
	Supported	76	65.5	69.1	100.0
	Total	110	94.8	100.0	
Missing	System	6	5.2		
Total		116	100.0		

A chi-square test indicates that the 69.1% of judgements in support of the conjoint measurement for comparison of similar systems is significant, χ^2 (1, n=110) = 16.0, p<0.001.

Lawshe's CVR is calculated at 0.38 which indicates more than half the respondents agree and is significant at 0.05 (one tailed test).

6.3.4.3 Technical integrity - conjoint measurement for comparison of dissimilar systems

The survey instrument posed the assertion 'A conjoint measure for technical integrity will allow meaningful comparison of the actual technical integrity of systems that are different in type and application' and the respondent requested to rate their agreement on a 5-point scale. Results obtained are shown in Table 33 and Figure 40.

Table 33. Frequency of ratings of the level of agreement of a technical integrity conjoint measure is meaningful for dissimilar system comparison.

Technical Integrity Conjoint Measure Meaningful for Dissimilar System Comparison

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Strongly Disagree	12	10.3	10.9	10.9
	Disagree	40	34.5	36.4	47.3
	Neither Agree or Disagree	27	23.3	24.5	71.8
	Agree	30	25.9	27.3	99.1
	Strongly Agree	1	.9	.9	100.0
	Total	110	94.8	100.0	
Missing	System	6	5.2		
Total		116	100.0		

To find if there is support for the hypothesis 'A conjoint measurement of technical integrity allows for comparison of dissimilar systems' the 5-point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with results as shown in Table 34.

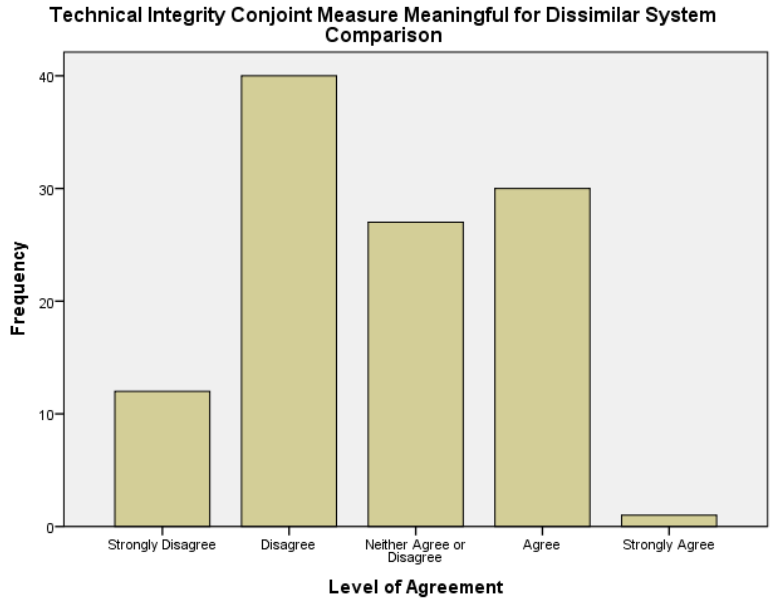


Figure 40. Bar chart of the level of agreement ratings of a technical integrity conjoint measure is meaningful for dissimilar system comparison.

Table 34. Frequency of support for hypothesis 'A conjoint measurement of technical integrity allows for comparison of dissimilar systems'.

Support for TI Conjoint Meaningful for Dissimilar Systems

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Not Supported	79	68.1	71.8	71.8
	Supported	31	26.7	28.2	100.0
	Total	110	94.8	100.0	
Missing	System	6	5.2		
Total		116	100.0		

A chi-square test indicates that the 71.8% of judgements **not** in support of conjoint measurement for comparison of similar systems is significant, χ^2 (1, n=110) = 21.0, p<0.001.

Lawshe's CVR is calculated at -0.44 which indicates less than half the respondents agree and is significant at 0.005 (one tailed test).

6.3.5 Summary of quantitative results

Table 35 summarises the quantitative results in terms of Lawshe's CVR where this has been calculated.

The results for the measurement construct appropriate for technical integrity indicate support for an ordinal scale but they are not in support of a ratio scale. While the data

also indicates non-support of measuring technical integrity using a single numerical value, it is at odds with results obtained on conjoint measurement. This anomaly is explored more in the subsequent section.

Table 35. Summary of quantitative results.

		Not Supported	Indeterminate	Supported
Technical Integrity Scale	Ordinal			0.69
	Numeric (single value)	-0.45		
	Ratio	-0.71		
Capability Scale	[0..1] capability measurement meaningful			0.35
	Capability beyond specification is valued			0.61
	Proposed capability measurement is Best	-0.28		
Dependability Scale	[0..1] dependability measurement meaningful		0.24	
	Dependability greater than risk acceptance threshold is valued			0.59
	Better dependability measurement possible		0.21	
Conjoint for Technical Integrity	Conjoint from capability and dependability is meaningful			0.36
	Measure useful for comparison of similar systems			0.38
	Measure useful for comparison of dissimilar systems	-0.44		
Values are Lawshe's CVR.				

The following subsections present exploration of apparent anomalies in the quantitative data and other quantitative analyses.

6.3.6 Additional quantitative analyses and anomalies

6.3.6.1 Judgment variation across demographic categories

Most of the demographic information collected is in the form of categorical variables. The ratings data was explored to see if any of the responses were significantly associated with demographic categories. Each of the rated questions was analysed against the categorical demographic information by forming a cross-tabulation and using the 'Eta' calculation within SPSS as the appropriate association statistic for ordinal versus nominal data. With Eta, values close to 1 represent an association and values close to 0 represent little association.

10 demographic categories were analysed (questions 2, 4, 5, 7, 8, 9, 10 and 11 from the survey instrument) for associations across the 8 judgement questions from the survey (questions 13, 14, 15, 16, 17, 19, 20, 22, 23, 24 from the survey instrument) – a total of 80 associations were examined. The highest value from Eta obtained was 0.379 with most falling well below 0.3 indicating no significant association of ratings across these demographic categories.

The other demographic category information is the age of respondents and the years of professional practice, both ordered values. We analysed each of the rated questions against this demographic information by forming a cross-tabulation and using the 'Correlations' calculation within SPSS as the Spearman's correlation coefficient, rho is the appropriate association statistic for rank orders.

2 demographic categories were analysed (questions 3 and 6 from the survey instrument) for associations across the 8 judgement questions from the survey (questions 13, 14, 15, 16, 17, 19, 20, 22, 23, 24 from the survey instrument) – a total of 16 associations were examined. None of the rho values were large with most below 0.1 and none statistically significant.

It can therefore be concluded that none of the ratings results are significantly associated with any demographic category.

6.3.6.2 Numeric measure versus meaningful conjoint measurement

It appears anomalous that a significant majority of respondents do not support that ascribing a single number to measure technical integrity is meaningful (CVR = -0.43),

yet a significant majority think that it is meaningful to form a conjoint measure from capability and dependability (CVR=0.36).

The anomaly is clearly illustrated in the scatterplot of Figure 41. In particular, note the cluster of responses at 'Agree' on the horizontal axis and 'Disagree' on the vertical axis.

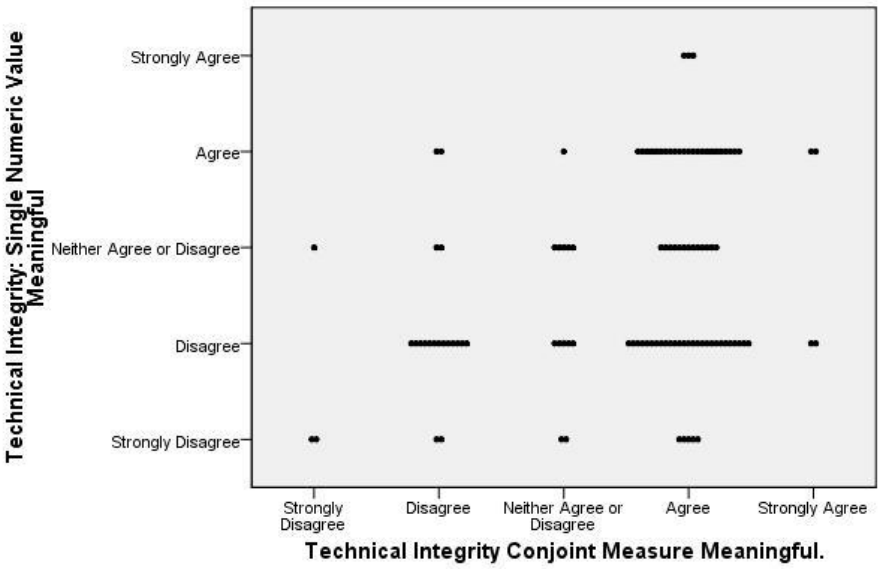


Figure 41. Scatterplot of ratings of technical integrity: single numeric value meaningful vs conjoint measurement meaningful

Possible explanations for this anomaly may be:

- Respondents didn't understand a conjoint measurement produces a single numerical value. This seems unlikely as this was explicitly defined in the information prior to the rating question.
- Respondents may not have been cognisant of conjoint measurement when they answered the first question and may have only been thinking of a direct measurement approach. Questions subsequent to that question and the lead-in explanation to the conjoint question may have enabled them to move toward the view that a single numerical value could be meaningful if it had been derived conjointly from components.

The strength of the response to the technical integrity conjoint measurement question (judgements in support of a conjoint measurement for technical integrity provided a CVR= 0.36) and the fact that it had been well explained in the survey instrument at that

part of the survey indicates that the overall response to the numerical scale question should be considered invalid.

6.4 Qualitative results

This section presents the qualitative analysis of the open-ended survey responses. Generally the analysis proceeded by importing the responses into Dedoose[8]. Codes were defined to pursue particular understandings that underlie the quantitative results. Mixed method visualisation and analysis tools within Dedoose were used to further explore for underlying relationships and meanings within the response data.

6.4.1 Better capability measurement

The open-ended response question, posed immediately after the associated rating question, requested an alternate measurement construct be proposed if the respondent had indicated they thought there was a better measure. 75 respondents answered in this way. The responses have been coded in accordance with Table 36.

Table 36. Code table – better capability measurement.

Code Name	Criteria for Coding
BETTER CAPABILITY MEASURE	Answer provided when the respondent indicated they thought there was a better Capability Measure than that proposed.
No Alt Capability Measure Proposed	Coded when there is no real measure proposed; where a significantly different definition is proposed; or concerns about tractability of measurement are expressed.
Set of Components	Coded when the response indicates that Capability cannot be measured by a single number; a set of capability measures is required (either by capabilities or capabilities of subsystems); vector and matrix approaches suggested.
Capability Measure > 1	Coded when the response includes the concept that the proposed [0..1] measure should be extended to reflect when more than the specified requirements are achieved by the system.
Qualitative Assessment	Coded when response includes the concept of qualitative assessment by end-users, experts or others.
Operational Capability Measure	Coded where the response reflects assessment against operational performance, operational concept documents, e.g., COI, MOE, MOP, MOS.

[8] Dedoose is a web-based tool for qualitative and mixed methods research. See http://www.dedoose.com/ for more information.

Code Name	Criteria for Coding
Weighted Components	Coded when the response indicates a weighted combination of constituent capabilities or requirements is considered better; weighted hierarchy of capabilities or of system elements.
Other	Other measurement concepts not covered by other codes.

All responses have been coded to 'BETTER CAPABILITY MEASURE'. This represents the total number of responses to this question. Each of these responses has had one or more text extracts made and coded to one of the other codes in accordance with the above criteria. The number of coded extracts is shown in Figure 42.

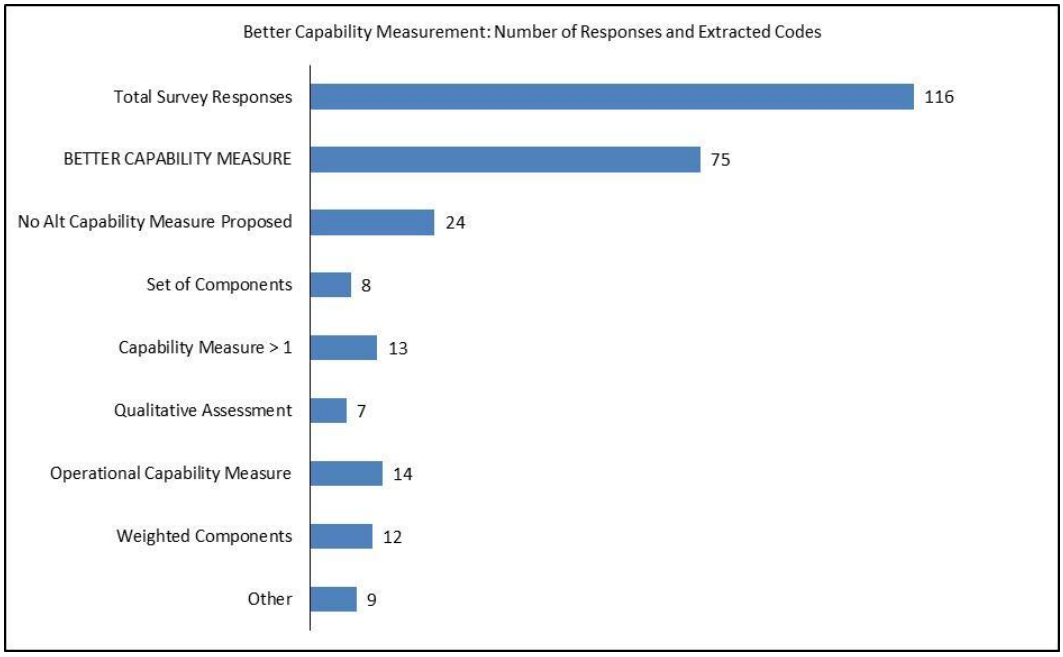


Figure 42. Number of coded extracts – better capability measurement

It is noted that the large number of respondents (26) who did not really offer an alternative or better measure for capability despite answering that they thought there was one. If these 26 responses were added back into the quantitative analysis as supporting the proposed capability measure, then 60% of respondents would be in support (CVR = 0.19, not statistically significant), compared with the proposed measure (63% of respondents (CVR = -0.28, a statistically significant majority) indicating they do not support the proposed measure).

6.4.1.1 No alternate capability measurement proposed

Extracts that have been coded as 'No Alt Capability Measure Proposed' are presented in Attachment 1. 32% of respondents who indicated they thought there was a better

measure of capability (or 21% of total respondent set) indicated that they thought capability would be better evaluated by other means.

However, the extracts in general do not provide any measurement proposal and often seek to impose a different definition of *capability,* for example, 'Working definition is NOT that required for a "capability", i.e., the power to achieve a desired operational effect in a nominated environment, within a specified time, and to sustain that effect for a designated period'. This identifies that the term 'capability' may be overloaded, that is, some respondents already have a view as to what capability is and that is different to the definition of *capability* proposed for *technical integrity* measurement. An alternate term may be considered for future research.

Another theme in this set of extracts regard the system's requirements as exemplified by these responses: 'I have concern about the quality of the documented required capabilities' and 'It is generally intended that requirements should specify characteristics that allow the system to meet its intended purpose(s), but these are generally an approximation to the intended purpose'. These types of responses express doubt as to the validity of the specified requirements being an adequate basis for measuring capability. The doubt seems to stem from experiences where capability requirements may be incorrectly specified or omitted from the capability specifications, or capability not being fulfilled for the end-user despite system technical requirements being achieved.

Several respondents were concerned that a single value for capability is too abstract and loses too much information or that the measure could only be subjective and therefore its value is questionable: 'The single number scale you have proposed is too lossy in information to be useful to a capability evaluator. Context has to be maintained, and as each solution will achieve each requirement in a potentially different way, then the single number measure loses insight into this' and 'from a concern that the combination of constituent capability assessments (e.g. a 0.9 for deployment of troops, but a 0.6 for carriage of equipment) is not meaningful to represent as a single number'. However, an alternative measurement approach was not proposed by these respondents.

6.4.1.2 Set of components

Extracts that have been coded as 'Set of Components' are presented in Attachment 1. 11% of respondents who indicated they thought there was a better measure of capability

(or 7% of total respondent set) indicated they did not believe that a single number for capability would be sufficient. Rather there are indications that a set of measures in perhaps up to 3 dimensions should be considered:

- by the capabilities of subsystems ('measurement system should go down to subsystem level');
- by the components of capability ('use this 0-1 scale, against a small profile of the systems key capabilities');
- By environment/mission domain ('there should be a measure for each mission set that the system is intended to perform in'; 'a simple capability "profile" against a small set of environments/modes').

The comments also include suggestions for visualisation of a multi-component capability measure, by 2D plots or radar charts: 'It can a 2D plot/bar chart of sub-system and level of effectiveness' or 'a range of measures (shown on, for example a radar chart) would facilitate display of multiple dimensions of capability'.

These approaches would not lead to a generic measurement approach. The specificity implied would drive a unique measurement instrument for each system for which we wish to evaluate its Capability. This is not the objective for technical integrity measurement in this research.

6.4.1.3 Capability measure > 1

Extracts that have been coded as 'Capability Measure > 1' are presented in Attachment 1. 17% of respondents who indicated they thought there was a better measure of capability (or 11% of total respondent set) indicated that since they thought capability above specified requirements was valued, then the measure for capability should be allowed to be bigger than 1 (i.e., >100% where 100% is total satisfaction of all specified capabilities).

Despite this conceptual agreement, there is little of substance proposed on how to actually measure this capability above the specified capabilities. For example, three respondents stated: 'plus the additional capabilities that are inherent in the system even though not stated as a requirement', 'plus other capabilities that are inherent to the design/technology or as a result of unintended but beneficial behaviour/properties', and

‘overall measure needs to take into account the emergent system level behaviour provided by the underlying capabilities’.

Specifically, if the [0..1] part of the scale is estimated using requirement compliance numbers, then what would be used to measure the requirements that have not been specified? Two of the comments at least recognise this difficulty: ‘If it cannot be contained within the definition of the system requirements, any value cannot be meaningfully attributed within the context of the capability as a whole’ and ‘the evaluation cannot exceed 1 because the secondary features/benefits are not described in scope’.

6.4.1.4 Qualitative assessment

Extracts that have been coded as ‘Qualitative Assessment’ are presented in Attachment 1. 9% of respondents who indicated they thought there was a better measure of capability (or 6% of total respondent set) indicated that they thought capability would be better evaluated by qualitative assessments by end-users, experts or others. Concepts within these extracts include:

- direct evaluation (respondents suggested ‘a “fit for purpose” assessment of the complex system by the operators’, ‘input from a team of SME's from various disciplines with detailed knowledge of the system’ and ‘ratings given by a potential user group’), and
- end-user satisfaction (suggestions include ‘fulfilment of client need’, ‘the perceived usefulness of that behaviour in terms of the system operators/owners’, and ‘the System User's view on whether the system provides the expected experience in achieving the desired capability outcomes. (end-user experience)’).

This approach could be generalised to a wide range of systems and therefore worth additional consideration. Although it needs to be remembered, this line of thinking is evident in only around 7% of respondents (c.f. 37% that thought the proposed measure was acceptable).

6.4.1.5 Operational capability measure

Extracts that have been coded as ‘Operational Capability Measure’ are presented in Attachment 1. 19% of respondents who indicated they thought there was a better measure of capability (or 12% of total respondent set) indicated that they thought

capability would be better evaluated by operational capability measures including Measures of Effectiveness (MOE), Measures of Performance (MOP) and effectiveness in addressing Critical Operational Issues (COI). Responses along these lines include 'Measures of Effectiveness (MOE) against the various stakeholders' criteria of benefits of the system' and 'Compliance against Measures of Performance and Sustainment (MOP and MOS)'.

This approach is not distinctly different from that of the proposed capability measure. Statements like 'Any measure of Capability should be measured against the COIs and MOE/MOS' and 'a measure of capability should also explicitly refer to the "Need" or "Operational" aspects' perhaps indicate that some respondents were thinking about the requirements that would appear in a system specification and do not acknowledge these as being adequate to define *capability* or do not believe they effectively represent a translation of operational requirements expressed in OCDs.

Responses of 'Does the system perform in the intended manner in its intended environment?' and 'Achieving the desired operational effect' reflect a view that evaluation of the *capability* of a complex engineered system can only be evaluated in the operational environment. No clear suggestion has been proposed for this evaluation, but it could not produce a *technical integrity* measure until the operational test and evaluation phase of the system lifecycle.

6.4.1.6 Weighted components

Extracts that have been coded as 'Weighted Component' are presented in Attachment 1. 16% of respondents who indicated they thought there was a better measure of capability (or 10% of total respondent set) indicated that they thought capability would be better evaluated by weighting the contribution of each constituent capability or capability requirement, for example 'explore a weighting factor in the components (set of requirements) of the desired capability'.

The responses indicate that they judge that not all requirements or capabilities of a system could be judged equally and therefore some weighting would need to be applied to them in order to construct a valid, overall measure for capability: 'you must have a clear, measurable baseline (with appropriate weightings etc.) as all requirements are not equal and interdependencies may have significant impacts on overall capability beyond a binary measure of achievement per requirement' and 'Importance (weighting) of each

required capability should be factored into measurement of System Capability rather than simple proportion measure'.

While one respondent stated 'Some sort of weighting of the capabilities to reflect user preferences would help to discriminate', no proposal has been suggested as to how to derive these weightings. It would appear that any derivation of weights would necessarily need to be specific to each system and its application domain. The objective of a generalised measurement instrument would therefore be lost if this approach were pursued.

6.4.1.7 Other

Extracts that have been coded as 'Other' are presented in Attachment 1. 12% of respondents who indicated they thought there was a better measure of capability (or 8% of total respondent set) indicated that they thought capability would be better evaluated by other means not fitting the previous analysis codes. There is no reliable conclusion to be drawn from these contributions.

One respondent suggests that 'a range of at least four measures would seem appropriate: fails to meet required level of capability, achieves required level of capability with work-arounds, achieves required level of capability, and exceeds required level of capability'. Such a qualitative scale may be more readily assessed that the proposed measure of the proportion of specified capability requirements but at the loss of some objectivity. The idea, perhaps on a larger range of values that may allow for more objectivity, is worthy of further consideration.

6.4.2 Better dependability measurement

The open-ended response question requested an alternate measurement construct be proposed if the respondent had indicated they thought there was a better measure. 47 respondents answered in this way. The responses have been coded in accordance with Table 37.

Table 37. Code table – better dependability measurement.

Code Name	Criteria for Coding
BETTER DEPENDABILITY MEASURE	Answer provided when the respondent indicated they thought there was a better Dependability measure than that proposed.

Code Name	Criteria for Coding
No Alt Dependability Measure Proposed	Coded when there is no real measure proposed; where a significantly different definition is proposed; or concerns about tractability of measurement are expressed.
(D) Set of Components	Coded when the response indicates that Dependability cannot be measured by a single number; a set of dependability measures is required (either by sub-elements of dependability or dependability of subsystems); vector and matrix approaches suggested.
Dependability Measure > 1	Coded when the response includes the concept that the proposed [0..1] measure should be extended to reflect when the residual risk of the system is below the stakeholder's risk acceptance criteria.
(D) Qualitative Assessment	Coded when response includes the concept of qualitative assessment by end-users, experts or others.
Risk Assessment	Coded when the response reflects a measurement based on the assessment of risk, quantitative or subjective; where the measure is proposed to reflect uncertainty; probabilistic structures or suggestions.
(D) Weighted Components	Coded when the response indicates a weighted combination of constituent elements of dependability is considered better; weighted hierarchy of dependability elements or dependability of system elements.
(D) Other	Other measurement concepts not covered by other codes.

All responses have been coded to 'BETTER DEPENDABILITY MEASURE'. This represents the total number of responses to this question. Each of these responses has had one or more text extracts made and coded to one of the other codes in accordance with the above criteria. The number of coded extracts is shown in Figure 43.

It is noted that the large number of respondents (20 out of 47) who did not really offer an alternative or better measure for dependability despite answering they thought there was one. If these 20 responses were added back into the quantitative analysis as supporting the proposed dependability measure, then 78% of respondents would be in support (CVR = 0.57, statistically significant at the 0.05 level), compared with the previous response (60% of respondents supported the proposed measure (CVR = 0.21, but not a statistically significant majority)). However, the validity of this conclusion is threatened by other credible interpretations of the qualitative data:

- Given time, these 20 respondents could have derived an alternative measurement approach.

- The non response or response not specific to the question may indicate a belief that there isn't a sensible single measure.

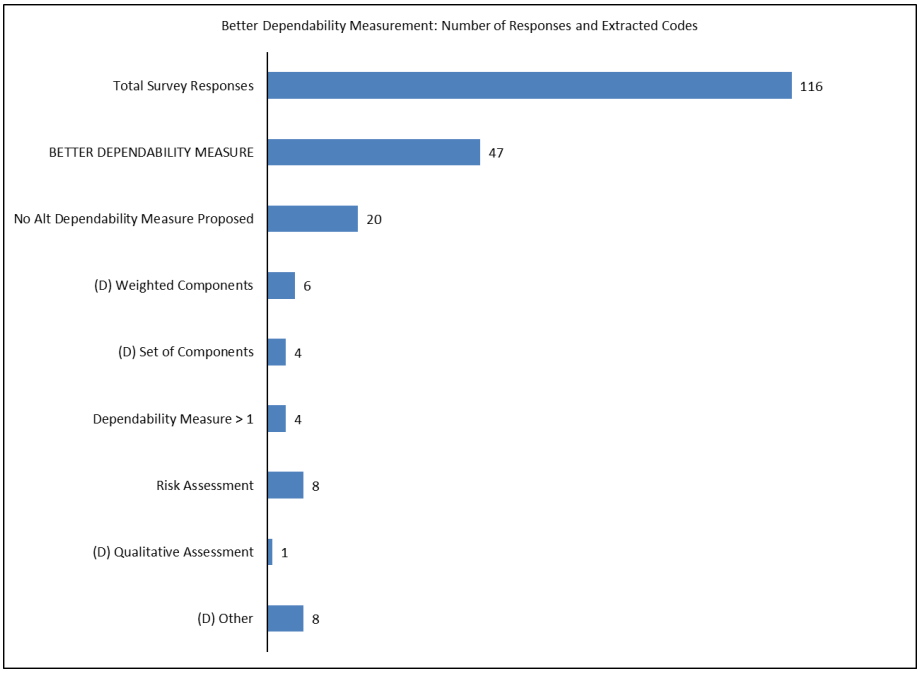


Figure 43. Number of coded extracts – better dependability measurement

6.4.2.1 No alt dependability measure proposed

Extracts that have been coded as 'No Alt Dependability Measure Proposed' are presented in Attachment 1. 43% of respondents who indicated they thought there was a better measure of dependability (or 17% of total respondent set) indicated that they thought dependability would be better evaluated by other means.

However, the extracts in general do not provide any measurement proposal and often seek to impose a different definition of dependability.

Three responses indicate doubt as to the tractability of measuring dependability on an ordinal scale using a risk acceptance paradigm: 'Dependability … can never be simplified to a single ordinal measure'; 'multi-faceted "system stakeholders" all have different risk acceptance paradigms/models, which of itself makes the problem non-linear in trying to work out who is "right"' and 'I am not sure that it is meaningful to measure dependability on a numerical scale when the benchmark is something as soft and variable and risk acceptance threshold of system stakeholders.'

Three respondents correctly point out that 'Dependability is always dependent on context', 'based on priorities and that these priorities are subjective and can change over time' and 'the operating environment needs to be addressed' but provide no alternate concept for measurement based on these observations.

Two respondents indicate that in their view the definition of dependability should not include safety*:* *'*By including reliability and safety in one measure you are mixing apples and oranges. Whilst related, these 2 aspects are often in conflict; by raising the reliability (availability) of a system it may be at the expense of safety but by having one measure there may be no overall change in value. In effect, you are then hiding the true level of unsafeness' and 'safety should be a separate section and not included with dependability as whether it is safe or not has nothing to do with whether it is reliable or not'. The measurement construct we have proposed does not preclude breaking out, for example, safety and reliability as lower level elements of dependability. Clearly however, safety and reliability are related with system safety standards and texts including what would be regarded as key reliability engineering techniques (for example, Fault Tree Analysis (FTA) and Failure Mode Effects and Criticality Analysis (FMECA) as indicated in (Stephenson, 1991) chapters 14 and 15 and (Bahr, 1997) chapters 7 and 8) and speciality engineering disciplines of system safety and reliability engineering often being seen as a common discipline (for example in the journal 'Reliability Engineering and System Safety'[9]). Related, one respondent questions the mix of what he considers to be 'measurable/calculable things like availability and reliability, and risk probability values. Does it work to put them together or should they be two different measures?'. Of course system characteristics like availability and reliability are commonly measured by probabilistic quantities like mean time between failure and criticality of the failure modes that is not conceptually different from characteristics like safety that may be measured by the likelihood and consequence of hazards and accidents. However, what is a valid point is that we cannot lose sight of the need to communicate specific safety risks to the users and stakeholders of this system,

[9] Reliability Engineering and System Safety is an international journal (published by Elsevier in association with the European Safety and Reliability Association, and the Safety Engineering and Risk Analysis Division) devoted to the development and application of methods for the enhancement of the safety and reliability of complex technological systems.

and the proposed measure of *technical integrity* and *dependability* is not intended as a substitute for this.

6.4.2.2 Set of components

Extracts that have been coded as 'Set of Components' are presented in Attachment 1. 9% of respondents who indicated they thought there was a better measure of dependability (or 3% of total respondent set) indicated they did not believe that a single number for dependability would be sufficient.

There is no common theme among the four respondents in this grouping. One refers to dependability needing to be 'a multi-dimensional concept'. Other views suggests the system's components should be the basis of a measurement ('reliability must link to the components of the system, and have a scale to indicate how good the components of the system'); a set of dependability measures for each mission area ('a matrix should be used to characterise the dependability against the mission sets') and a set of measures for each undesirable outcome ('the probabilistic expectation of an adverse outcome, due to a wide range of undesired system attributes, which may be expressed as a function of multiple variables rather than a single value. This could be also communicated as the likelihood of mission failure due to those undesired system attributes').

These approaches would not lead to a generic measurement approach. The specificity implied would drive a unique measurement instrument for each system that is to have dependability estimated. This is not the objective for technical integrity measurement in this research.

6.4.2.3 Dependability measure >1

Extracts that have been coded as 'Dependability Measure >1' are presented in Attachment 1. 9% of respondents who indicated they thought there was a better measure of capability (or 3% of total respondent set) indicated that since they thought dependability better than the system stakeholder's risk acceptance threshold was valued, then the measure for capability should be allowed to be bigger than 1 (i.e., >100% where 100% is all residual risk at the stakeholder's risk acceptance threshold).

Despite this conceptual agreement, there is little of substance proposed on how to actually measure above the threshold dependability. The single clear suggestion is that

dependability above 100% could be measured ‘using the same proportionality as for 0–1’.

Two of the four respondents indicate that cost should be taken into account in measuring dependability above 1 or 100%: ‘better than risk acceptance threshold that is generally good unless it cost a lot’ and ‘Value for Money (VFM) must come into the equation. If I can get acceptable dependability for what I'm prepared to pay and additional dependability comes at a much higher price, then it has less value to me.’

6.4.2.4 Qualitative assessment

Extracts that have been coded as ‘(D) Qualitative Assessment’ are presented in Attachment 1. One respondent representing 2% of respondents who indicated they thought there was a better measure of dependability (or 1% of total respondent set) indicated a qualitative measure was appropriate: ‘Willingness of the operator to utilise a complex engineered system to its full capability’. In comparison, eight respondents indicated qualitative measurements should be considered for *capability* measurement. This is somewhat unexpected since the proposed measure for *capability* (measure of achieved requirements) is arguably much more tractable than that proposed for *dependability* (assessment of overall risk), or perhaps that proposed *dependability* measure is already perceived as qualitative to the respondents.

6.4.2.5 Risk assessment

Extracts that have been coded as ‘Risk Assessment’ are presented in Attachment 1. Eight respondents representing 17% of respondents who indicated they thought there was a better measure of dependability (or 7% of total respondent set) indicated a measure that assessed risk or uncertainty was appropriate.

The responses include observations that the criticality of undesired events must be taken into account in the measure: ‘Frequent catastrophic failures would likely imply low availability. Some unintended behaviours are more serious than others. The measure of this seriousness should take into account the impact on ability of the users to perform their mission.’ Two respondents mention that the measure should include reference to the As Low As Reasonably Practicable (ALARP) concept or at least attempt to measure those risks that are not ALARP: ‘Perhaps the percentage of hazards which are not mitigated to ALARP within the hazard log’ and ‘defining that a system is safe should match the ALARP process once a level of acceptable risk is defined’.

One respondent also identifies the issue of how to include in the measurement an estimate of the unidentified risks: ‘problem with dependability is that there could be unknown unknowns’. This is an issue we have already identified and will be addressed in further development of the measurement approach.

6.4.2.6 Weighted components

Extracts that have been coded as ‘(D) Weighted Components’ are presented in Attachment 1. Six respondents representing 13% of respondents who indicated they thought there was a better measure of dependability (or 5% of total respondent set) indicated a measure that included some sort of concept of weighting dependability components in order to come up with a dependability measure was appropriate.

For example, one respondent commented ‘What is dependability - reliability, adaptability, resilience, safety? Each term has a meaning and requires a weighting for the project to be developed’ and another ‘perhaps weighting factors could be used in determining the level of dependability’. These are typical of the responses in this category and there is no firm measurement proposal made. It would seem that any weighting scheme may need to be specific to the particular system (similar to the suggestions made by respondents for weighted component measurement of *capability*) which would negate the objective of a generalised measurement instrument.

6.4.2.7 Other

Extracts that have been coded as ‘(D) Other’ are presented in Attachment 1. Eight respondents representing 17% of respondents who indicated they thought there was a better measure of dependability (or 7% of total respondent set) indicated other measures may be appropriate. There is no insight or conclusion that has been drawn from these disparate responses.

6.4.3 Reflections on the measurement of technical integrity

A final open-ended question was asked at the conclusion of the survey:

> This page is not mandatory, but if you would like to provide any of your observations having completed this judgements task, please do so here.
>
> Any other comments or observations? In particular, how has your understanding of "technical integrity" been changed by undertaking this survey?

This question was posed, without any specific a priori intention in mind, rather it was intended to provide a last chance for the respondent to offer any other comment they felt like making at the survey conclusion. The responses were reviewed carefully, looking for themes, these themes leading to the definition of codes against which response extracts have been made. Responses have been coded in accordance with Table 38.

Table 38. Code table – reflections on the measurement of technical integrity.

Code Name	Criteria for Coding
Valid Research Problem	Coded when the response indicates concurrence on the existence of a 'researchable' problem; clear benefits to arise from the research.
Tractability and Usefulness of Measurement	Coded when a response indicates doubt if the measurement of technical integrity or its components is tractable; doubt as to the usefulness of the measurement.
No Cap vs Dep Prioritisation Possible	Coded when any response suggests they are unable or unwilling to prioritise Capability versus Dependability.
Valid TI Definition and Structure	Coded when the response indicates agreement with the Definition and Structure proposed for technical integrity and its measurement.

All 63 responses to this question have been coded to 'OTHER COMMENT OR UNDERSTANDING OF TI CHANGE'. Each of these responses has had zero or more text extracts made and coded to one of the other codes in accordance with the above criteria. The number of coded extracts is shown in Figure 44.

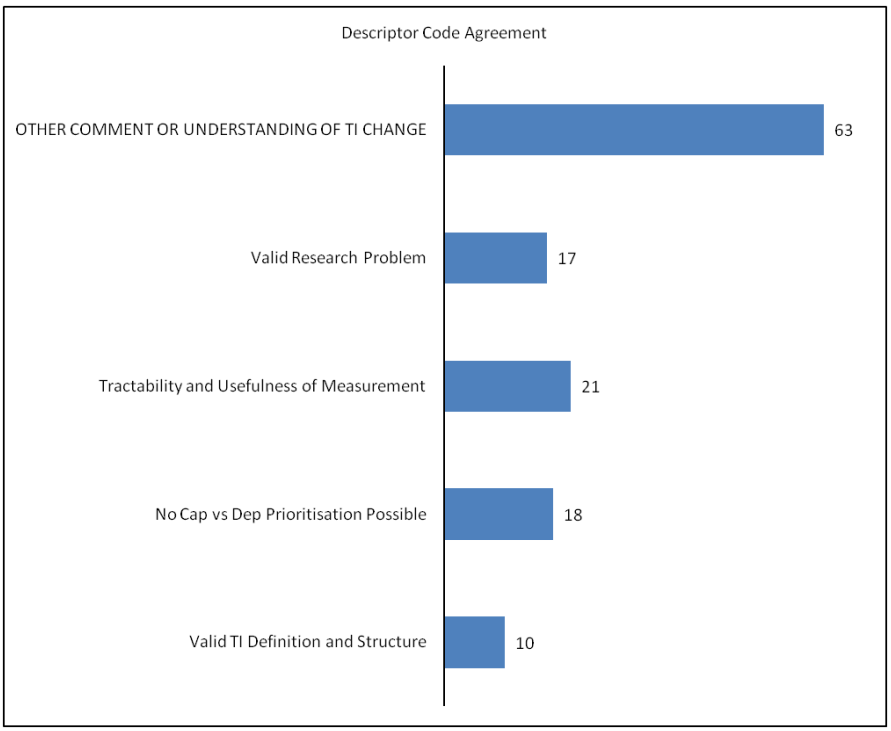


Figure 44. Number of coded extracts – reflections on the measurement of technical integrity.

It is noted that more than half (63 of 115) of the respondents provided comment at the end of the survey. Several have made quite strong comments for and against the measurement construct and these are discussed in the following sections.

6.4.3.1 Valid research problem

Extracts that have been coded as 'Valid Research Problem' are presented in Attachment 1. 15% of the total respondent set has made a comment related to the validity of the research and the benefits that may be derived from it.

Support for the validity of the research includes comments like 'I support the research work being undertaken as a valid research subject', 'the reference paper "What is Technical Integrity and How to Measure it" is very well presented and the concepts are definitely worthwhile exploring further', and 'a very important field of enquiry for today's complex systems of systems'.

Several respondents clearly stated a belief in the usefulness of developing a measurement instrument for technical integrity: 'I think that the research aim (if met) will have considerable benefit in quantifying an outcome that often can (at the moment) only be truly appreciated by highly experienced practitioners (who are most likely not

the key decision makers!)'; 'it is desirable that we can estimate/assign a comparative 'value' to technical integrity so that we can identify both the required Integrity and when a particular mission system falls above or below that level' and 'A subject/approach that will help future large complex projects be more successful both technically and programmatically.'

Both these sets of views are taken as support for the research effort to establish a measurement for technical integrity and the validity of the documented approach.

6.4.3.2 Tractability and usefulness of measurement

Extracts that have been coded as 'Tractability and Usefulness of Measurement' are presented in Attachment 1. 18% of the total respondent set has made a comment related to doubting the tractability or usefulness of the proposed measurement construct.

One respondent commented 'I find pinning down dependability to one number difficult and therefore pinning both dependability and capability to one number even more challenging.' This indicates there is some doubt as to the tractability of assigning one number to measure technical integrity. It is commented by some respondents to be difficult at best: 'While I have never really considered assessing technical integrity numerically it would appear to be most difficult.'

Two respondents indicate doubt the instrument will work at all: 'Not sure if a combined TI index will work' and 'Never thought of "measuring" TI before and am sceptical that a meaningful measure can be derived.'

Doubts about technical integrity being quantifiable at all are expressed by two respondents: 'the implication that Capability and Dependability are equally quantifiable is perhaps not true' and 'technical integrity is a difficult area with a lot of subjective elements.'

Doubts are expressed about the usefulness of the measurement. One comment reflects doubt as to the comparableness of the basis of measurement of capability, i.e., requirements, between systems: 'I think there is far too much variability in how the requirements were developed.' Others comment the proposed instrument may lose information: 'I think the approach you are proposing is too lossy to be useful. Actually it has a very high chance of being misleading.'

It must be remembered, that the main objective of the research is to gain a measurement of technical integrity so that the impact of technical regulation may be assessed. There are responses that indicate that many of the respondents are thinking of the measurement instrument as primarily being used as a measure of a system's acceptability or other decision making. The usefulness of the measurement instrument for this latter purpose may be problematic as highlighted in these comments:

> There is a real risk in simplifying such a difficult subject as technical integrity to a single number where such a number is used by those who really don't understand the impact of experience with complex systems by organisations and individuals.
>
> Compressing complex information (in this case down to a single number) is a lossy process, but rarely do such approaches ever highlight this to the consumer, which leaves them with a false impression that this single number is sufficient basis for decision making.
>
> Reducing technical integrity to a numerical score can never substitute for a deep understanding of the inherent dynamics and complex structures of complex systems. All that the parameters can do is provide a quick look insight that should assist in prioritising work to address necessary changes required to capability and dependability to meet a given mission.

6.4.3.3 No capability vs dependability prioritisation possible

Extracts that have been coded as 'No Cap vs Dep Prioritisation Possible' are presented in Attachment 1. 16% of the total respondent set has made a comment related to their inability or unwillingness to provide a prioritisation between capability and dependability or to state that they are always equal.

Several respondents commented to the effect that capability and dependability are not comparable quantities ('I can't answer the above question because it assumes that capability and dependability are the same types of thing') or conversely are too entwined to be separable ('I doubt that capability and dependability can be regarded as 'orthogonal' aspects, as in my experience they may related to one another').

Several comments were made that indicated that both capability and dependability are absolutely needed: 'Operational effectiveness needs both', 'I feel that capability is meaningless if it is not dependable' and 'they are not in competition, one wants robust, dependable capabilities'. These comments do not reflect an understanding of the relativity or weighting that can be applied to components in a conjoint measurement. Perhaps this indicates the questions may have been better posed to allow a value based

prioritisation (e.g., within a given system domain, would you prefer a 10% increase in dependability or capability?).

Another respondent commented that 'Without dependability the complex system will be unable to deliver the intended capability therefore equals no capability' reflecting an inability to disassociate the two components for the purpose of prioritising them in a conjoint measurement. However, the comment actually reflects the properties of the theorised measurement instrument – if dependability approached zero then so too would technical integrity irrespective of the level of capability.

Several respondents reinforced their quantitative rankings of the equal priority between capability and dependability: 'Defence must be capable and dependable. These qualities are equally as important', 'The warfighter demands capability in order to be effective and competitive but always at the expense of being safe and reliable. Therefore equal.', and 'In all systems I have delivered including Naval military and civil Transport both Capability and Dependability are equally important.' However, other respondents justify that the relative priorities should be different in different contexts: 'what matters more for one mission may be quite different for a different mission', 'this trade off isn't able to be consistently resolved within a broad domain. It depends on the platform, capability and the operational situation', and 'Weighting needs to be considered between capability and dependability for different systems, for same systems in different operational environments, etc.'

6.4.3.4 Valid technical integrity definition and structure

Extracts that have been coded as 'Valid TI Definition and Structure' are presented in Attachment 1. 9% of the total respondent set has made a comment indicating agreement with the definition and structure proposed for the measurement of technical integrity or has provided some alternate definition for consideration.

Agreement implying validity of the proposed definition of the technical integrity measurement construct includes the following comments from respondents: 'I think this is a very good approach to the issue'; 'I read the two white papers in your link to understand the meaning. Of course in my professional life as an aeronautics complex system developer I have been dealing with all these concepts but not structured in such a way. Good job!', 'The approach mentioned in the reference paper 'What is Technical Integrity and How to Measure it' is very well presented and the concepts are definitely

worthwhile exploring further' and 'I feel relatively comfortable viewing technical integrity as a culmination of a range of subordinate, "engineered" aspects of a complex system from the perspective of capability and dependability'.

The predominate alternate view expressed is that safety should be considered a separate element in the conjoint measurement: 'I still maintain that "safety" is a specific measure of 'integrity' which does not lend itself to be 'rolled up' into other integrity measures' and 'Add Safety in as a separate box and you may find that your proposal works better'.

6.5 Combined (mixed-method) results

Recall that the main objective of this element of the research project is to establish the content validity of the proposed technical integrity measurement construct. Content validity means that experts agree that the instrument will meaningfully measure what it purports to measure (Tashakkori and Teddlie, 1998, p. 83).

6.5.1 Technical integrity measurement proposal validity

The quantitative results summarised in Table 35 showed that the conjoint measurement structured from measures of capability and dependability was meaningful and useful for the comparison of similar systems as discerned from a statistically significant majority of the respondents. The qualitative responses analysed in section 6.4.3 provide some further insight into the reasoning behind the quantitative results. A dichotomous response was observed in the qualitative coding where:

- 24% of respondents were positively dispositioned to the proposal: 15% of respondents indicated the research objective (i.e., pursuit of a measurement construct for technical integrity) was appropriate and it would be of benefit to practice and another 9% commented that they agreed with the theorised definition and structure; and
- 18% made comments that indicated doubts as to the tractability of the measurement proposal or its usefulness.

It is noteworthy that many of the negatively dispositioned comments seem to consider an application of the measurement instrument beyond that of the primary application intended. Recall that the primary intent is to validate a measurement instrument in order to evaluate the impact of technical regulations on the technical integrity of complex

engineered systems. Many of the doubting comments reflect an application where the measurement may be used by decision makers to determine the acceptability of a system. The latter would be an additional benefit of the measurement construct if it were found to be valid and useful for that application.

It is concluded from these considerations that the overall proposal to measure technical integrity is valid for the primary purpose, but extension to other applications would need to be the subject of further validation.

6.5.2 Technical integrity measurement scale

The scale that technical integrity is to be measured on is now considered. The ideal objective for the technical integrity measure is for measurement on a ratio scale, but it needs to be at least measurable on an ordinal scale to contribute to the resolution of the research problem.

The quantitative analysis indicates that an ordinal scale for a technical integrity measurement is valid but that a ratio scale is not valid. This means that the majority of respondents agreed that complex engineered systems could be ranked according to a measure of their technical integrity (i.e., it is valid to say that system A has greater technical integrity than system B). They do not agree that the technical integrity of complex engineered systems can be measured on a ratio scale (i.e., it is not valid to say that system A has twice the technical integrity of system B).

Based on the quantitative analysis, no significant view as to the representation of the technical integrity measurement as a single numeric measurement emerged. The section ‘Numeric measure versus meaningful conjoint measure‘ examined the anomaly between this outcome and positive results from the validity of the conjoint measurement question and concluded that the former result should be discounted in favour of the significant majority that considered a conjoint approach to technical integrity, and therefore implicitly a numeric value, to be valid.

It is conservatively concluded that an ordinal measurement scale for technical integrity is appropriate and that technical integrity should not be defined on either an interval or ratio scale.

6.5.3 Conjoint measurement for technical integrity

The proposed conjoint measurement approach is now discussed. The objective was to confirm that the combination of the defined capability and dependability measures form a meaningful and valid measure. The concept postulated for a conjoint measurement of technical integrity is shown diagrammatically in Figure 13.

The quantitative analysis of section 6.3.4 indicates that a conjoint measurement constructed from components capability and dependability is considered appropriate and the use of this measure to compare the technical integrity of similar types of systems has appropriate content validity. However, the majority of respondents considered that such a measure would not be appropriate in comparing the technical integrity of dissimilar systems.

It had been hoped that the derivation of different relative weightings for capability and dependability for different types of systems may allow for comparison across dissimilar system domains. Perhaps, in hindsight, a limitation of the survey instrument is that it did not directly pose a question to the respondents to see if they would consider the use of appropriate weightings to each category of system types would allow for comparison of dissimilar systems. In any event, accepting the result as obtained, it can only be considered that the measure is valid for the comparison of similar systems. This may not be a severe limitation for the purpose of this overall research. For example, it is to be considered valid to use a conjoint measure to compare the technical integrity of different surface combatant ships and in doing this assess the impact of the Naval Technical Regulatory System. Likewise we can consider it valid to use a conjoint measure to compare the technical integrity of different aircraft in the air force and in doing this assess the impact of regulations for Technical Airworthiness Management.

6.5.4 Measurement of capability

Capability was defined in the survey instrument to be ‘The achievement of the intended capabilities and attributes of a system (those that are explicitly required and become the main focus of system design activities).’ The survey posed a measurement construct where capability would be measured on a scale from 0 (indicating no required capabilities or attributes achieved) to 1 (indicating all required capabilities or attributes achieved).

A statistically significant majority of respondents supported the concept that capability could be meaningfully measured on such a 0 to 1 construct (CVR = 0.35). Further, the significant majority supported the concept that capability that exceed the specified requirements was of value (CVR = 0.61). This combined with qualitative responses indicated an extension of the capability scale beyond 1 is warranted but there is no basis suggested as to how to approach the evaluation of these capabilities beyond those defined by capability and/or technical requirements.

But, a significant majority of respondents indicated they did not think the capability 0 to 1 measurement construct was the best measure available (CVR = -0.44). The open-ended responses provided with these judgements showed a large number of respondents could not indicate what the better measurement approach may be.

Many suggestions for an alternate measurement indicated using evaluation against operational criteria (i.e., detail in an OCD like MOPs, MOEs or capability in the operational environment as evaluated by the end-users). The open responses indicate that some respondents do not see and/or have confidence in the linkage between end-user capability measures and the system's more technically based requirements and there is doubt about how well requirements for a system represent the capability actually required by the system's stakeholders.

Many of the open responses indicated the respondent was not satisfied with the definition of Capability given or had defaulted to their own pre-conceived idea of capability. This led to responses that did not propose an alternate measurement scale but rather an alternate definition for capability. This indicates there may be a need to re-label c*apability* to another term that does not carry such preconceptions. 'Fitness for Service', for instance, had been considered (particularly since in the ADF's Regulation of Technical Integrity (Australian Defence Force, 2010) it is defined as 'fitness for service is the materiel's ability to satisfy operational requirements') but comes with other different preconceptions in different domains. Often, 'Fitness for Service', is not considered to be tightly linked to the system requirements which a developer is focused on realising in the complex engineered system.

6.5.5 Measurement of dependability

Dependability was defined in the survey instrument to be 'the minimisation of unintended and/or undesired behaviours or characteristics of a system'. The survey posed a measurement construct where dependability would be measured on a scale from 0 (indicating a complete lack of system dependability (e.g., the system never works as intended, it frequently causes catastrophic accidents, etc.)) to 1 (indicating all risk of unintended and undesired behaviour of the system has been identified and either eliminated in the system or are acceptable to the owners, operators, and regulators (e.g., failure rates are tolerable, no failure ever cause catastrophic accidents, etc.)).

A majority of respondents supported the concept that dependability could be meaningfully measured on such a 0 to 1 measurement construct (Chi-square test significant ($p<0.010$), but CVR = 0.24 not significant at 0.05 level).

A significant majority supported the concept that dependability that was better than the stakeholders' risk acceptability threshold was of value (CVR = 0.59). This combined with qualitative responses from four respondents indicated an extension of the dependability scale beyond 1 is warranted but with only one suggested measurement approach for this: 'using the same proportionality as for 0 – 1'.

60% of respondents indicated they did think the dependability 0 to 1 measurement construct was the best measure available (CVR = 0.21, not a statistically significant majority). The open responses provided with these judgements indicated an additional 18% (20 respondents) who did not really offer an alternative or better measure for dependability despite answering they thought there was one. Adding these responses back into support for the proposed scale gives a CVR = 0.57, a statistically significant result at the 0.05 level that adds confidence to the claim of validity for the proposed measure of dependability.

To summarise, it is concluded that for dependability:

- The proposed 0 to 1 measurement construct is valid (a majority of respondents concur that it will be meaningful) and is the best scale available (a majority of respondents concur that there is no better measurement construct available).
- Extension of the construct beyond 1 to reflect additional value is considered desirable, but there is no consensus as to how that should be measured.

6.6 Conclusions

Previously a conjoint measurement construct for technical integrity was inductively developed and published for critical review. This chapter has documented the data and results from survey research that has aimed at further establishing the content validity of that same conjoint measurement construct through a consensus of expert opinion.

6.6.1 Findings

Most generally, the conjoint measurement construct for technical integrity is considered valid for the ranking of similar systems for the purpose of assessing the impact of technical regulation. More specifically it has been found that:

- Measurement of *technical integrity* on an ordinal scale is valid.
- Constructing a conjoint measurement from measures of *capability* and *dependability* is valid.
- Measurement of *capability* on a 0 to 1 range is valid.
- *Capability* in excess of that specified is valued, but no practical basis for extending the measurement of *capability* above 1 has been identified.
- Measurement of *dependability* on a 0 to 1 range is valid.
- *Dependability* better than the acceptability threshold of the system's stakeholders is valued, but there is no consensus opinion offered as what to base measurement of *dependability* above 1 upon.

6.6.2 Limitations

It is not considered valid to measure *technical integrity* on a ratio scale. This may limit future research goals that will use this construct (e.g., it may never be able to produce results that indicate x% more spent on technical regulation has increased *technical integrity* of systems by y%).

Based on the results in this chapter, it is not considered valid to use the measure of *technical integrity* to compare systems across different domains and this will be a focus of the next chapter in this thesis.

Other applications of the conjoint measurement construct for *technical integrity* may not be valid (e.g., for the purpose of judging the acceptability of a system). This will not impact the planned research, but does detract from the potential value of the approach.

The obvious starting point for the measurement of *capability* is to measure the fraction of the originally intended requirements actually achieved in the system. Doubts and scepticism as to the validity of the system's written requirements as a basis for measurement have been expressed in the survey results. These views may limit the acceptance of the *capability* measurement by some stakeholders.

6.6.3 Additional research questions

As anticipated by the nature of exploratory and descriptive research, additional questions have been raised within this chapter that are to be addressed in the subsequent phase of research including:

- Should *safety* be a separate element of the conjoint measurement construct for *technical integrity*, or will the estimation of safety risk dominate the estimation of *dependability*?
- Will the measurement of *dependability* lend itself to calculating values greater than unity?

7 Conjoint measurement aggregation weights

7.1 Introduction

The research reported in Chapter 6 found that the conjoint measurement construct for technical integrity is considered valid for the ranking of similar systems for the purpose of assessing the impact of technical regulation.

This chapter builds upon that finding to determine coefficients to use for the aggregation of capability and dependability into the conjoint measurement of technical integrity via quantitative analysis of judges' ratings obtained from survey. This chapter also reports on the elicitation of the reasoning behind the judgements via qualitative analysis of the open-ended survey response data.

7.2 Method

A specific hypothesis is generated for the purpose of verification: the testing of the validity of theories within a real-world context (Leedy and Ormond, 2005, p. 134-135). The hypothesis is formulated to verify the applicability of the theorised measure across different classes of systems, noting that ideally it would be applicable across all types of system domains.

The method used a survey research strategy to collect expert judgments in the form of relative importance ratings and to collect written responses from open-ended questions. The rating data is analysed quantitatively in order to arrive at aggregation weightings for the capability and dependability elements of the multiplicative conjoint equation theorised in Chapter 5. Effectively the quantitative analysis procedure is testing the hypothesis that *the ratings of the relative importance of capability versus dependability are the same across the different system domains*.

A qualitative analysis procedure is employed on the open-ended response data in order to assist in the interpretation of, and find explanations for, the quantitative results. This is essentially 'theory guided research' where the analysis begins with the response data and develops an understanding from it (Cropley, 2002, p. 101).

Finally a mixed-method approach is used to 'triangulate' the quantitative and qualitative results in order to provide a 'better indicator of the true situation than either qualitative or quantitative findings alone'(Cropley, 2002, p. 86-87).

7.2.1 Survey instrument

A survey instrument, part of the instrument as described in Chapter 6, has gathered data to particular questions seeking the rating of the relative importance of capability and dependability. It also asked open-ended questions through which additional insights into the reasons behind the expert ratings can emerge. Both quantitative and qualitative analysis methods were used to analyse the data gathered from 116 respondents.

The survey instrument was implemented in Survey Monkey. The full survey instrument is at Appendix A. The survey responses were collected in August and September 2012. Data is stored as non-identifiable records (i.e. no personal identification data is stored and the demographic data supplied cannot be attributed to a particular person).

The question asked: 'In each domain in which you feel competent to judge, please indicate the relative importance of Capability or Dependability to systems within that domain.' The 16 domains for which judgments were requested were:

- Military – Aviation
- Military – Naval
- Military – Land
- Military – Joint, C4ISR
- Civil – Aviation
- Civil – Maritime
- Civil –Medical
- Civil - Mining
- Civil – Chemical Process
- Civil – Industrial
- Civil – Telecommunications and IT

- Civil – Transportation
- Civil – Power Utility – Conventional
- Civil – Power Utility – Nuclear
- Civil – Water Utility
- Civil – Space

The respondents were free to not provide a rating to each and every domain. The ratings for each domain were requested on a seven point scale labelled from left to right on the survey as:

- Capability Significantly More Important
- Capability Much More Important
- Capability More Important
- Equally Important
- Dependability More Important
- Dependability Much More Important
- Dependability Significantly More Important

The order of the system domains presented to each respondent was randomized in order to minimize the chance of cognitive bias arising due to the order of consideration. The primary cognitive bias of concern is *anchoring*: the failure to adequately adjust a judgment due to the consideration of the previous judgment made (Meyer and Booker, 2001, p. 136-139).

In addition the following open-ended questions were asked after the ratings were completed:

- 'Is there something specific to your domain or viewpoint that has influenced the way you have judged the relative importance of Capability and Dependability?' and
- 'Do you think the service(s) provided by the system(s) you were thinking of, has an impact on the way you have rated the relative importance of Capability and Dependability?'

The method of selecting judges and the resulting demographic data is the same as that reported in Chapter 6.

7.2.2 Analysis of survey responses

7.2.2.1 Quantitative analysis procedure

The research methodology overarching this procedure is survey research, that is 'acquiring information about one or more groups of people ... about their opinions ... by asking them questions and tabulating their answers' (Leedy and Ormond, 2005, p. 183-186). The analysis procedure follows the following steps:

- For each system domain, descriptive statistics (mode, mean[10], median range and interquartile range) and bar charts of the frequency of each rating are generated to characterise the response data.
- The mean, median and interquartile range is exported to a spreadsheet for further inspection, ordering and visualisation in order to assist in interpretation of the results.
- The significance of the sample median values against set rating values is conducted using the non-parametric One-Sample Wilcoxon Signed Rank Test which provides a test to see if a sample median differs significantly from a hypothesised value. The only assumption for the data needed is that it is an ordinal variable which is true for this data set (UCLA: Academic Technology Services).
- Variations across groups, constructed from the demographic data, are explored using the Independent-Samples Kruskal-Wallis test. This non-parametric test allows for the comparison of 'scores on a continuous variable for three or more groups' (Pallant, 2010, p. 232-235).

7.2.2.2 Qualitative analysis procedure

The qualitative analysis procedure essentially followed that of (Leedy and Ormond, 2005, p. 140) adapted as follows:

- The researcher became familiar with the data by thoroughly reviewing all responses and identified themes of relevance to develop an understanding of the reasons for the ratings provided.
- Evident themes or concepts were defined and codes were established to pursue particular understandings that underlie the quantitative results.

[10] Note, the rating scale is ordinal, and therefore the mean is not strictly valid, but it is useful for initially assessing trends in the results.

- Text was extracted and associated with one or more of the codes.
- The extracts were tabulated for each code.
- Counts of extracts against each code were reported to establish frequency of the associated theme or concept.
- Text extracts within each code group were compared, contrasted and discussed.

The qualitative analysis commenced by importing the responses into Dedoose. Dedoose[11] is a web-based tool for qualitative and mixed methods research.

Mixed method visualisation and analysis tools within Dedoose were used to further explore for underlying relationships and meanings within the response data.

7.3 Quantitative results

7.3.1 Response data

Table 39 presents the descriptive statistics for ratings of the relative importance of capability versus dependability for each system domain. Figure 45 through Figure 60 present a summary of response data showing the frequency of ratings of relative importance of capability and dependability in each of the system domains.

[11] See http://www.dedoose.com/ for more information on Dedoose.

Table 39. Descriptive statistics for ratings of the relative importance of capability versus dependability for each system domain.

Statistics

	N						Percentiles		
	Valid	Missing	Mean	Median	Mode	Range	25	50	75
Cap v Dep Rating: Military - Aviation	81	35	4.57	4.00	4	6	4.00	4.00	6.00
Cap v Dep Rating: Military - Naval	89	27	3.90	4.00	4	5	3.00	4.00	5.00
Cap v Dep Rating: Military - Land	53	63	3.74	4.00	4	6	3.00	4.00	4.00
Cap v Dep Rating: Military - Joint, C4ISR, ..	57	59	3.74	4.00	4	6	3.00	4.00	4.00
Cap v Dep Rating: Civil - Aviation	54	62	5.61	6.00	6	6	5.00	6.00	7.00
Cap v Dep Rating: Civil - Maritime	42	74	4.86	5.00	5	5	4.00	5.00	6.00
Cap v Dep Rating: Civil - Medical	35	81	5.40	6.00	7	6	4.00	6.00	7.00
Cap v Dep Rating: Civil - Mining	35	81	4.86	5.00	5	6	4.00	5.00	6.00
Cap v Dep Rating: Civil - Chemical Process	30	86	5.27	5.50	6	5	4.00	5.50	6.00
Cap v Dep Rating: Civil - Industrial	29	87	4.62	4.00	4	4	4.00	4.00	5.00
Cap v Dep Rating: Civil - Telecommunications and IT	48	68	4.04	4.00	4	6	3.25	4.00	5.00
Cap v Dep Rating: Civil - Transportation	35	81	4.91	5.00	4[a]	5	4.00	5.00	6.00
Cap v Dep Rating: Civil - Power Utility - Conventional	32	84	5.13	5.00	6	5	4.00	5.00	6.00
Cap v Dep Rating: Civil - Power Utility - Nuclear	32	84	6.00	7.00	7	6	6.00	7.00	7.00
Cap v Dep Rating: Civil - Water Utility	31	85	5.19	5.00	6	5	4.00	5.00	6.00
Cap v Dep Rating: Civil - Space	35	81	5.2286	6.0000	7.00	6.00	4.0000	6.0000	7.0000

a. Multiple modes exist. The smallest value is shown

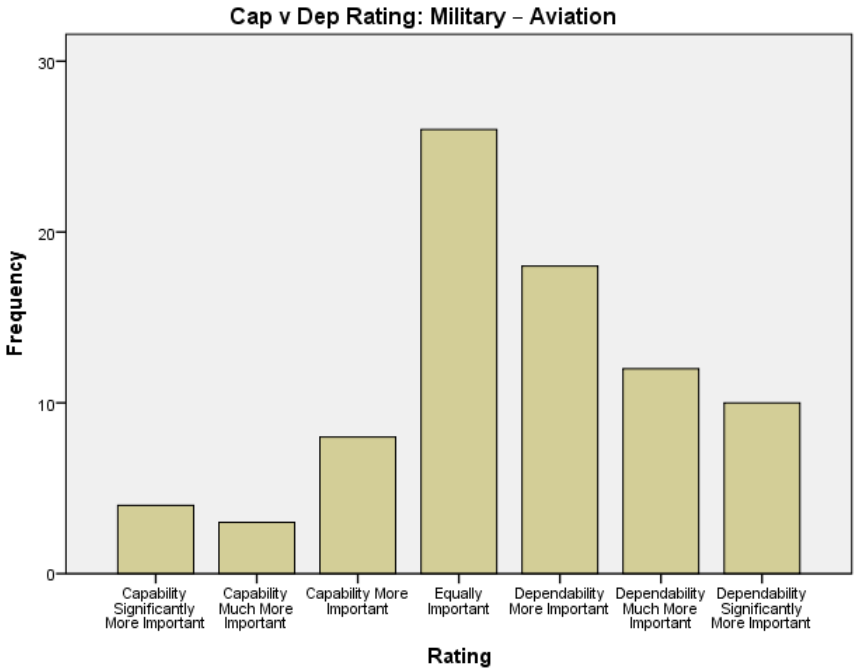


Figure 45. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Military – Aviation’.

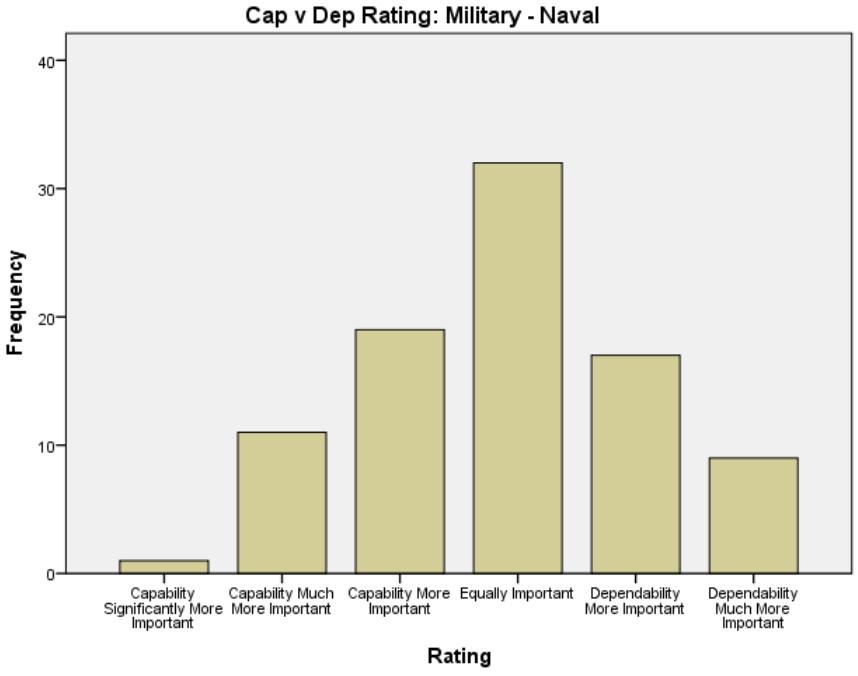


Figure 46. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Military – Naval’.

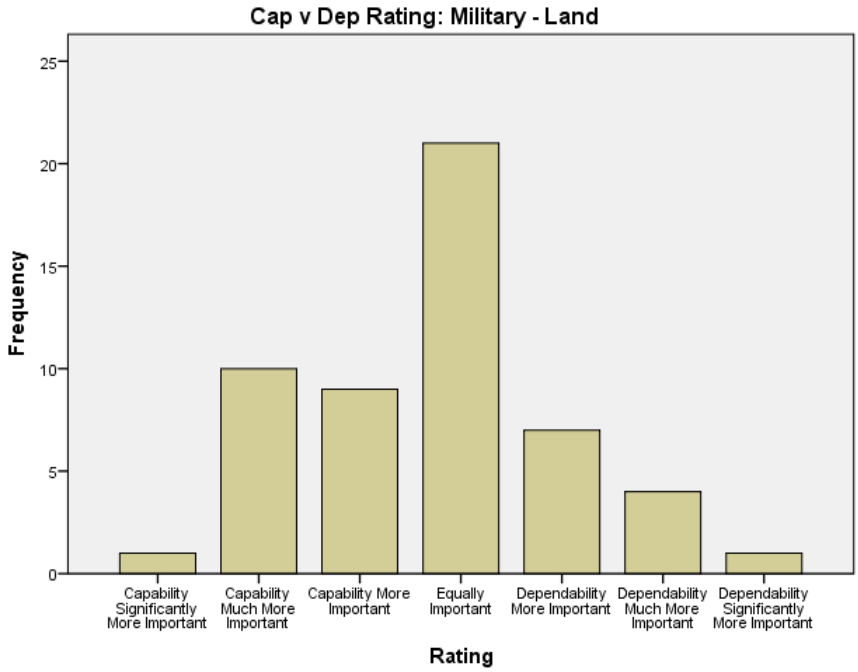


Figure 47. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Military – Land’.

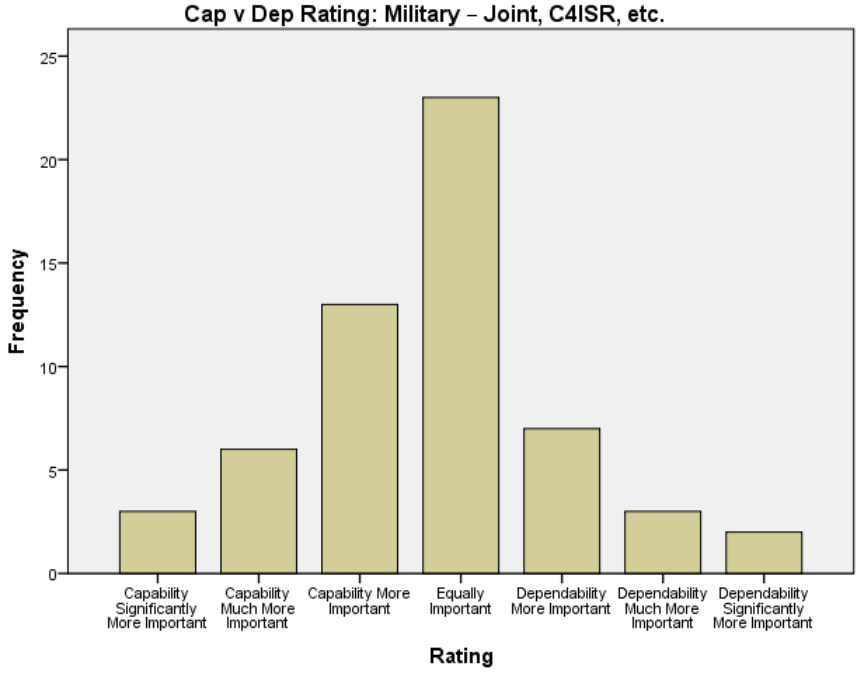


Figure 48. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Military – Joint, C4ISR’.

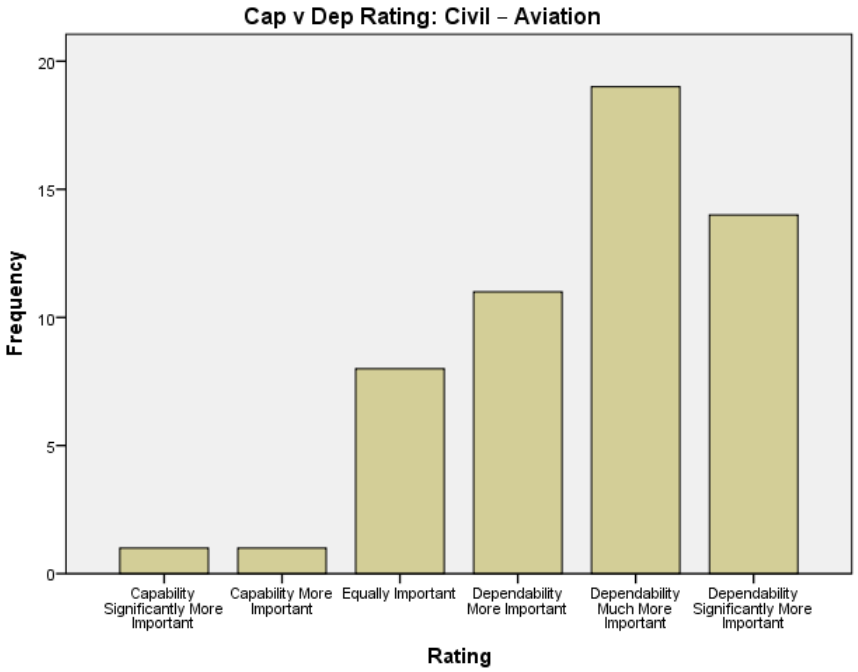


Figure 49. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Aviation’.

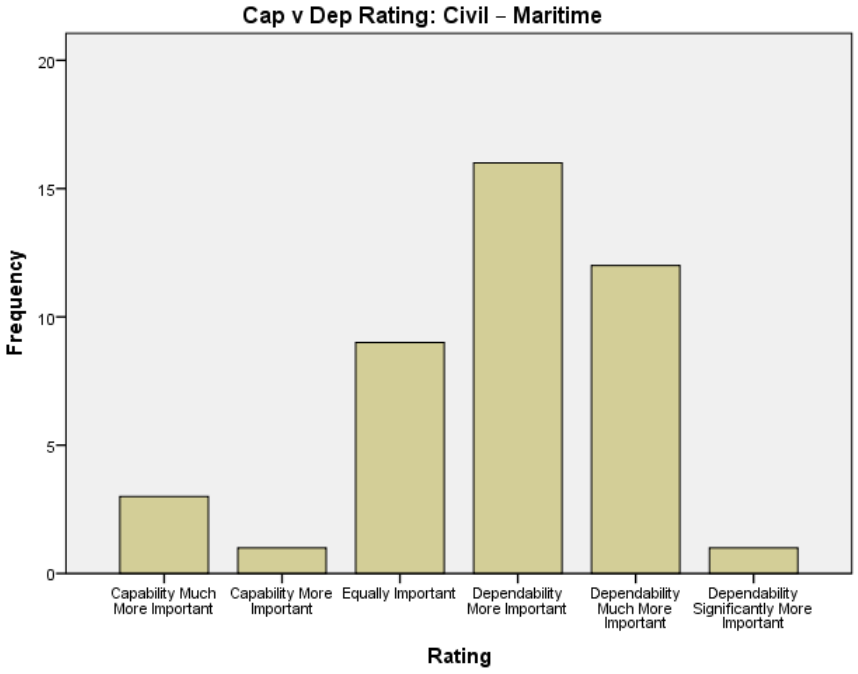


Figure 50. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Maritime’.

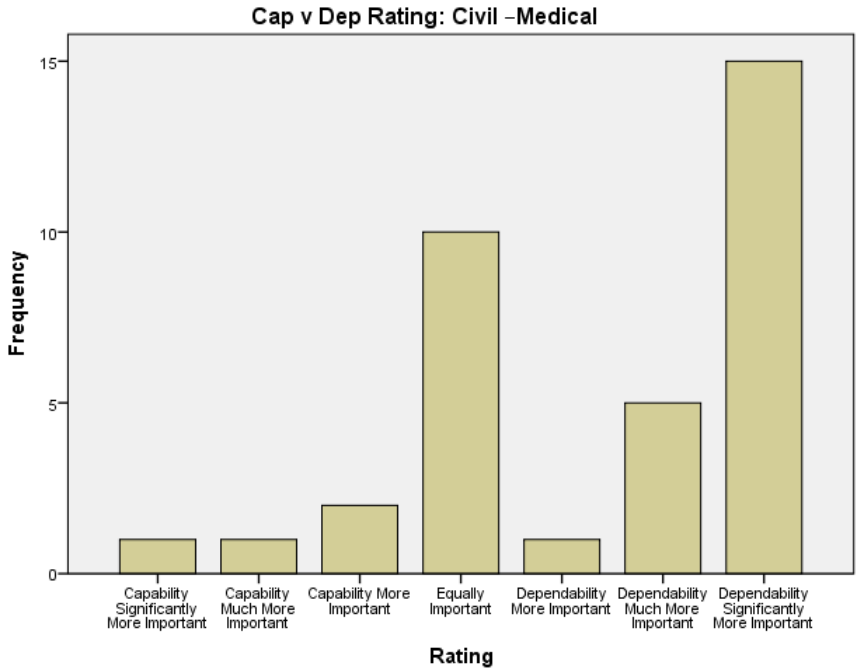


Figure 51. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Medical’.

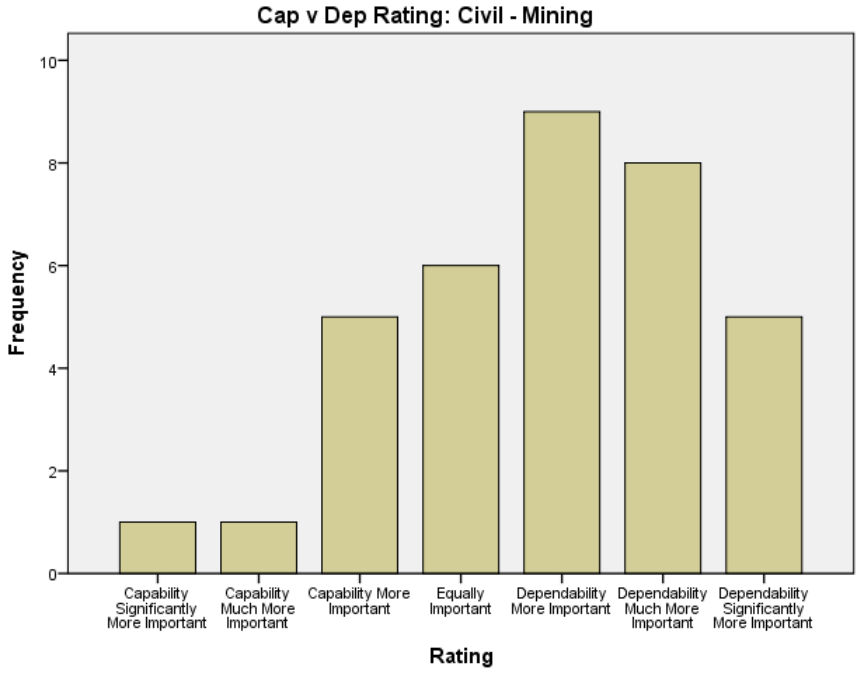


Figure 52. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Mining’.

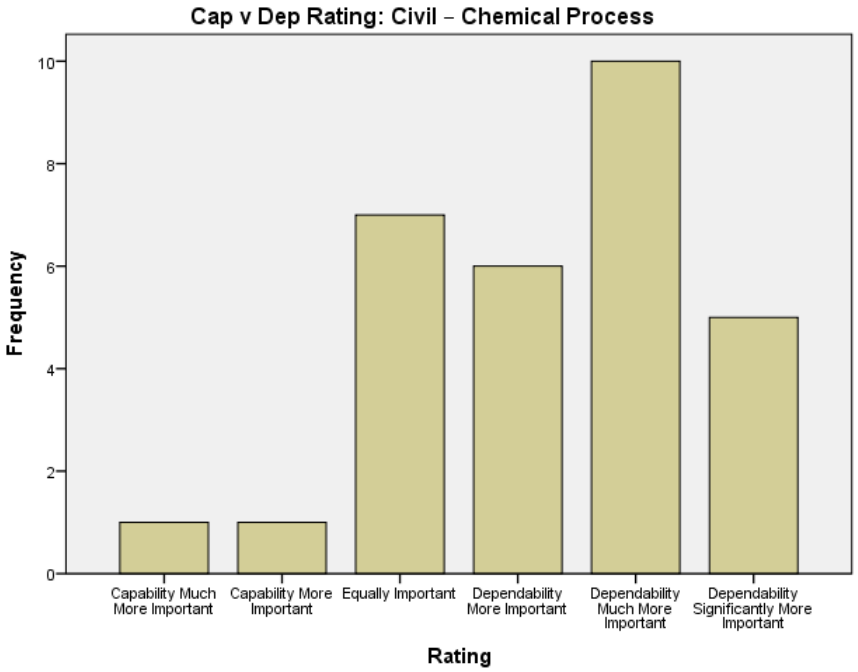


Figure 53. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Chemical Process’.

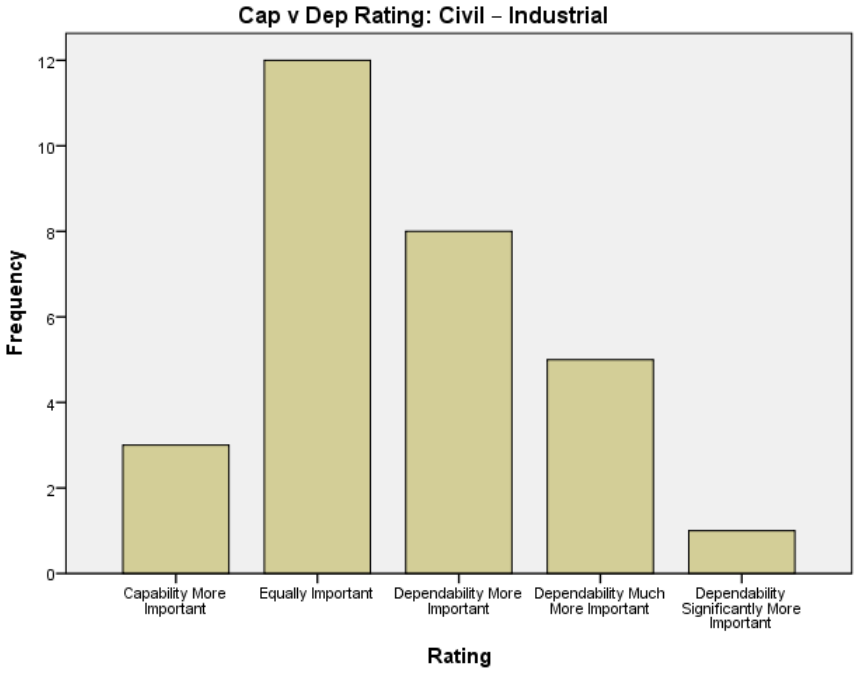


Figure 54. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Industrial’.

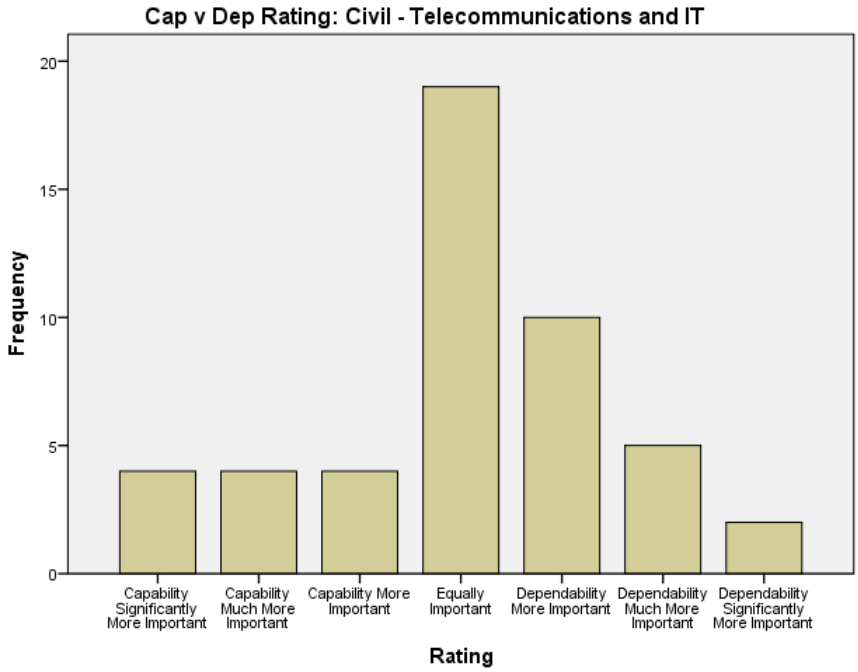


Figure 55. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Telecommunications and IT’.

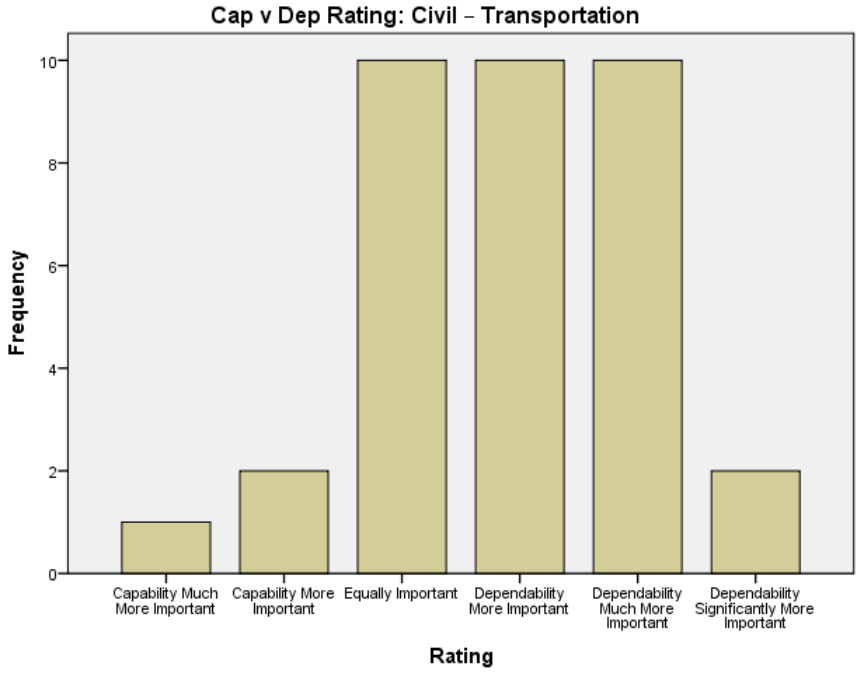


Figure 56. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Transportation’.

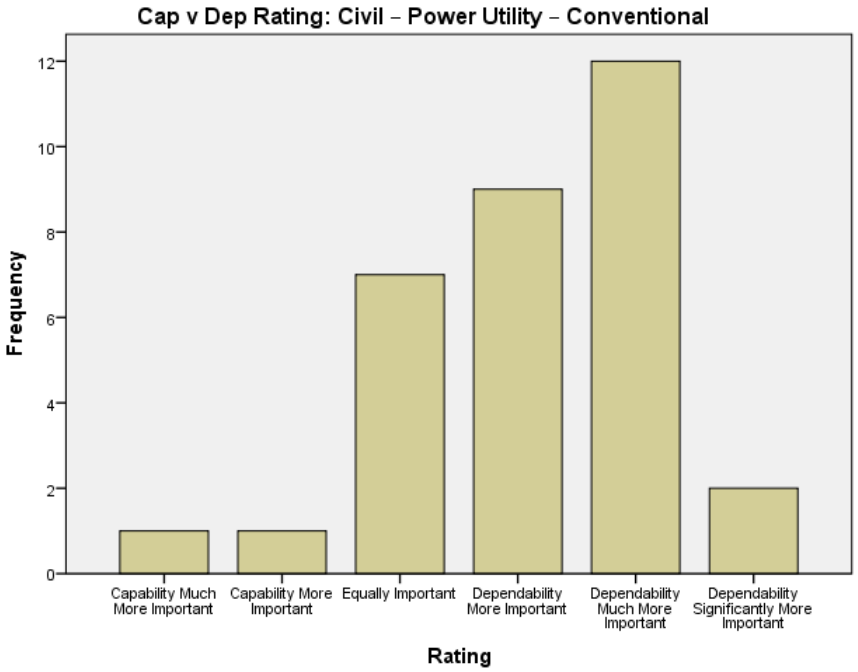


Figure 57. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Power Utility – Conventional’.

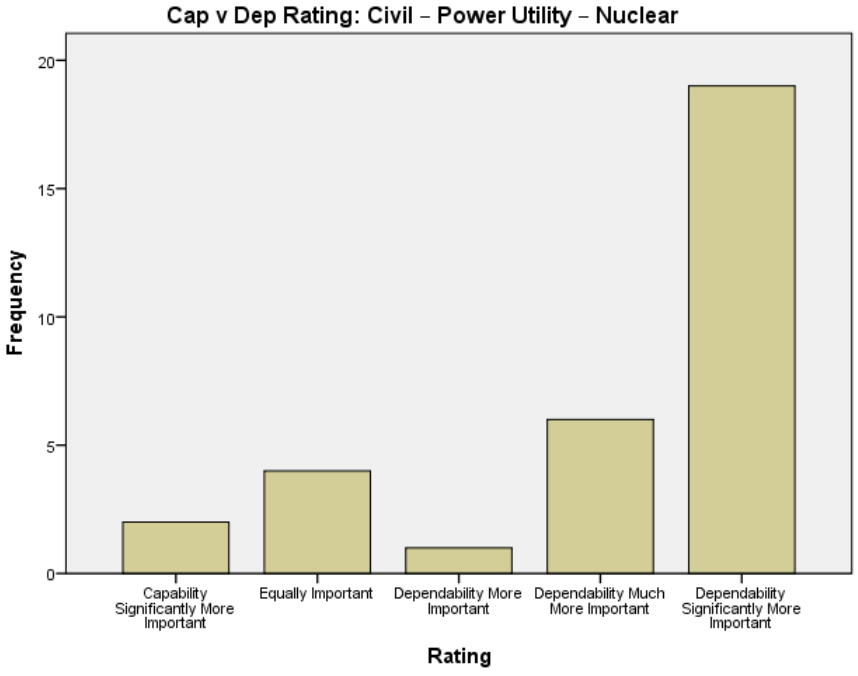


Figure 58. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Power Utility – Nuclear’.

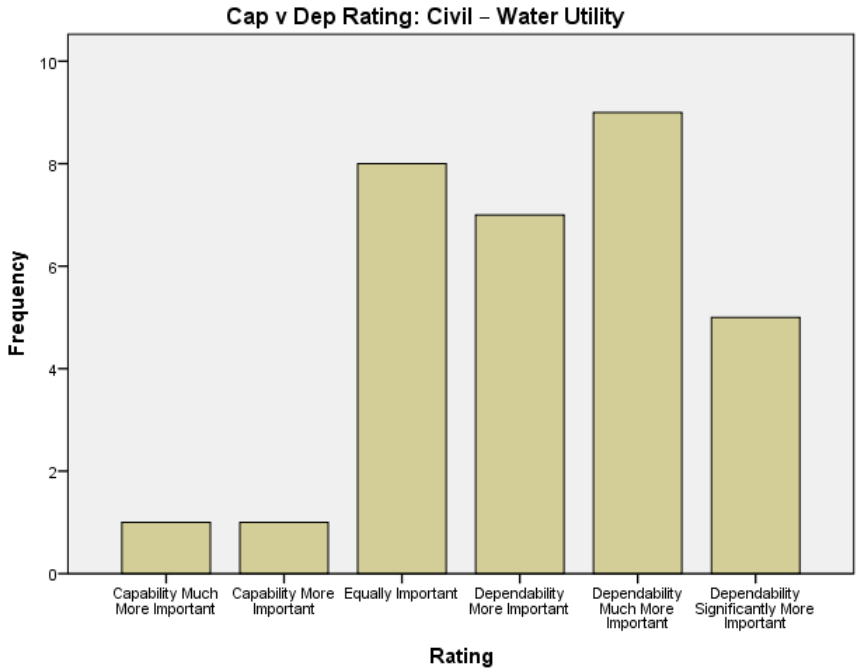


Figure 59. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Water Utility’.

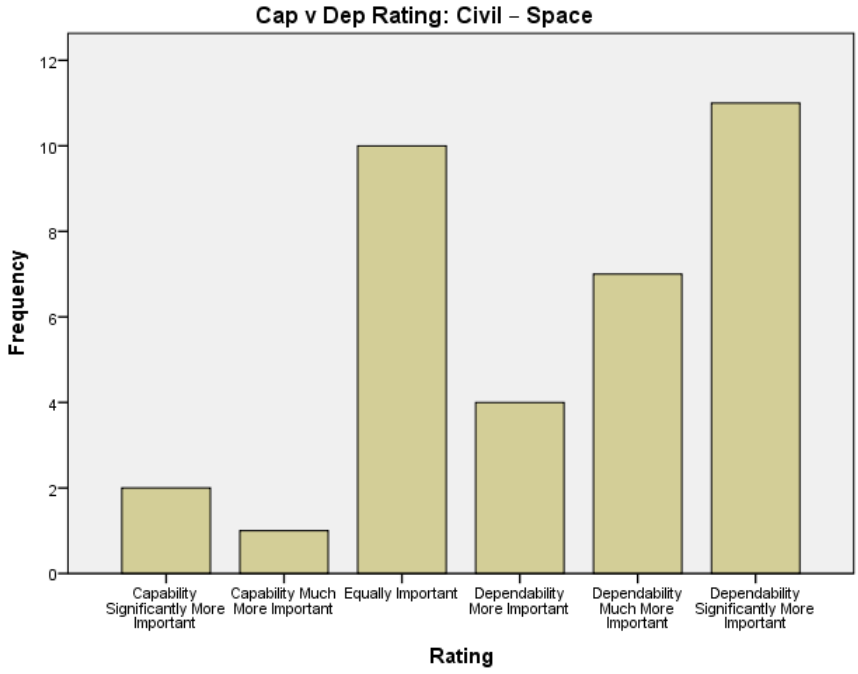


Figure 60. Frequency of ratings of relative importance of capability and dependability in the system domain of ‘Civil – Space’.

7.3.2 Interpretation

Most of the response data shows the ratings distributed closely around a modal value. This is generally to be expected, that is, the judgements are reasonably consistent.

However the following response data indicate unexpected bi-modal responses:

- Civil – Medical (refer Figure 51): where one mode is at "Equally Important" and one at the extreme of "Dependability Significantly More Important".
- Civil – Space (refer Figure 60): with the same bi-modal response as Civil-Medical.

It is plausible that these results may have occurred from an inconsistent interpretation of the system domain. For Civil-Medical, one modal response may have been from those considering monitoring systems while the other modal response may have been considering life support equipment. For Civil-Space, one modal response could reflect manned space systems and the other unmanned space systems which would have significantly different requirements related to dependability. This poses a threat to the validity of these results and more generally to the other results where although no unexpected results are apparent, inconsistent interpretations of the system domain could exist.

7.3.3 Analysis

Initially, the quartile percentages for the judges' ratings in each system domain have been calculated. The percentile data was then used to sort the system domains in descending order by median, then by 75th percentile, then by 25th percentile. These values are then plotted in Figure 61. The mean (although strictly not meaningful since not a true interval scale) is calculated and plotted for comparison.

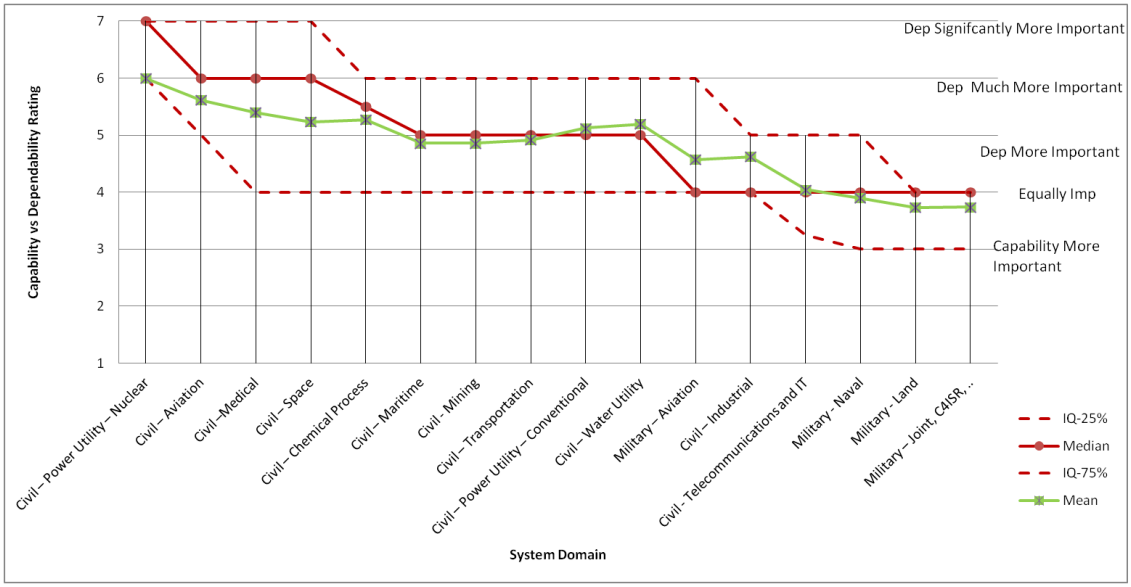


Figure 61. Sample median, mean and interquartile ranges for ratings of relative importance of capability and dependability versus system domains.

The median values suggest the judgments on the relative importance of capability and dependability may be clustered around 3 values. In order to see if this is statistically valid, a Wilcoxon Signed Rank Test is conducted to test if the sample median of each variable differs significantly from 4, 5, and 6 (Capability and Dependability Equally Important, Dependability More Important, and Dependability Much More Important). This requires the construction of 16 operationalised hypotheses for the purpose of executing the test in SPSS. Note that the Wilcoxon Signed Rank Test does not assume the data is interval and normally distributed (which is the case with this data set). The test was conducted at a Confidence Interval of 95% at significance of 0.05. The hypothesis test results from SPSS are presented in Table 40.

Table 40. Hypothesis test summaries for test of sample medians differing from a median of 4, 5 or 6.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of Cap v Dep Rating: Military – Aviation equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.002	Reject the null hypothesis.
2	The median of Cap v Dep Rating: Military - Naval equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.441	Retain the null hypothesis.
3	The median of Cap v Dep Rating: Military - Land equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.130	Retain the null hypothesis.
4	The median of Cap v Dep Rating: Military – Joint, C4ISR, .. equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.135	Retain the null hypothesis.
5	The median of Cap v Dep Rating: Civil – Aviation equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
6	The median of Cap v Dep Rating: Civil – Maritime equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
7	The median of Cap v Dep Rating: Civil –Medical equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
8	The median of Cap v Dep Rating: Civil - Mining equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.003	Reject the null hypothesis.
9	The median of Cap v Dep Rating: Civil – Chemical Process equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
10	The median of Cap v Dep Rating: Civil – Industrial equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.004	Reject the null hypothesis.
11	The median of Cap v Dep Rating: Civil - Telecommunications and IT equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.904	Retain the null hypothesis.
12	The median of Cap v Dep Rating: Civil – Transportation equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
13	The median of Cap v Dep Rating: Civil – Power Utility – Conventional equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
14	The median of Cap v Dep Rating: Civil – Power Utility – Nuclear equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
15	The median of Cap v Dep Rating: Civil – Water Utility equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
16	The median of Cap v Dep Rating: Civil – Space equals 4.00.	One-Sample Wilcoxon Signed Rank Test	.001	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of Cap v Dep Rating: Military – Aviation equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.035	Reject the null hypothesis.
2	The median of Cap v Dep Rating: Military - Naval equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
3	The median of Cap v Dep Rating: Military - Land equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
4	The median of Cap v Dep Rating: Military – Joint, C4ISR, .. equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
5	The median of Cap v Dep Rating: Civil – Aviation equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
6	The median of Cap v Dep Rating: Civil – Maritime equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.569	Retain the null hypothesis.
7	The median of Cap v Dep Rating: Civil –Medical equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.072	Retain the null hypothesis.
8	The median of Cap v Dep Rating: Civil - Mining equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.639	Retain the null hypothesis.
9	The median of Cap v Dep Rating: Civil – Chemical Process equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.204	Retain the null hypothesis.
10	The median of Cap v Dep Rating: Civil – Industrial equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.058	Retain the null hypothesis.
11	The median of Cap v Dep Rating: Civil - Telecommunications and IT equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
12	The median of Cap v Dep Rating: Civil – Transportation equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.720	Retain the null hypothesis.
13	The median of Cap v Dep Rating: Civil – Power Utility – Conventional equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.433	Retain the null hypothesis.
14	The median of Cap v Dep Rating: Civil – Power Utility – Nuclear equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.001	Reject the null hypothesis.
15	The median of Cap v Dep Rating: Civil – Water Utility equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.317	Retain the null hypothesis.
16	The median of Cap v Dep Rating: Civil – Space equals 5.00.	One-Sample Wilcoxon Signed Rank Test	.172	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Hypothesis Test Summary

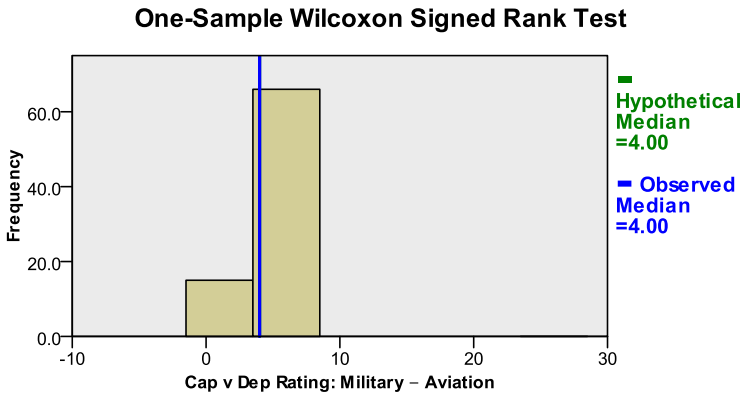
	Null Hypothesis	Test	Sig.	Decision
1	The median of Cap v Dep Rating: Military – Aviation equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
2	The median of Cap v Dep Rating: Military - Naval equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
3	The median of Cap v Dep Rating: Military - Land equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
4	The median of Cap v Dep Rating: Military – Joint, C4ISR, .. equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
5	The median of Cap v Dep Rating: Civil – Aviation equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.022	Reject the null hypothesis.
6	The median of Cap v Dep Rating: Civil – Maritime equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
7	The median of Cap v Dep Rating: Civil –Medical equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.027	Reject the null hypothesis.
8	The median of Cap v Dep Rating: Civil - Mining equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
9	The median of Cap v Dep Rating: Civil – Chemical Process equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.004	Reject the null hypothesis.
10	The median of Cap v Dep Rating: Civil – Industrial equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
11	The median of Cap v Dep Rating: Civil - Telecommunications and IT equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
12	The median of Cap v Dep Rating: Civil – Transportation equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
13	The median of Cap v Dep Rating: Civil – Power Utility – Conventional equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.000	Reject the null hypothesis.
14	The median of Cap v Dep Rating: Civil – Power Utility – Nuclear equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.519	Retain the null hypothesis.
15	The median of Cap v Dep Rating: Civil – Water Utility equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.002	Reject the null hypothesis.
16	The median of Cap v Dep Rating: Civil – Space equals 6.00.	One-Sample Wilcoxon Signed Rank Test	.007	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

The only System Domains for which there is not a significant median are the Military – Aviation and the Civil – Aviation. More detail of the Wilcoxon Signed Rank Test model from SPSS for these tests is shown in Figure 62 and Figure 63 respectively.

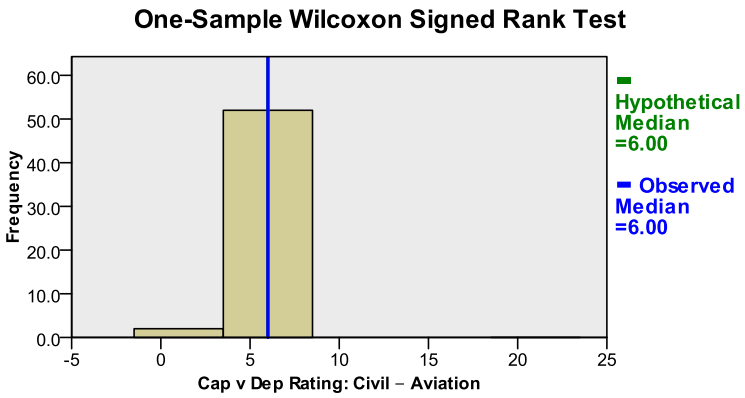
For Military – Aviation the observed median is 4 and the mode is 4. Neither the hypothesis that the median is 4 or that the median is 5 is statistically significant at the 0.05 level (0.002 and 0.035 respectively). However, for the purposes of this study it is reasonable to assume the median to be 4 and note the lack of significance for future confirmatory research.

For Civil – Aviation the observed median is 6 and the mode is 6. Neither the hypothesis that the median is 5 or that the median is 6 is statistically significant at the 0.05 level (0.000 and 0.022 respectively). As above, for the purposes of this study it is reasonable to assume the median to be 6.



Total N	81
Test Statistic	1,136.000
Standard Error	117.271
Standardized Test Statistic	3.121
Asymptotic Sig. (2-sided test)	.002

Figure 62. Model Result for the median of Capability vs Dependability for the Military – Aviation system domain being ‘Equally Important’ (i.e., 4).



Total N	54
Test Statistic	182.000
Standard Error	58.241
Standardized Test Statistic	-2.284
Asymptotic Sig. (2-sided test)	.022

Figure 63. Model Result for the median of Capability vs Dependability for the Civil – Aviation system domain being 'Dependability Much More Important' (i.e., 6).

The distributions for each of the systems domains have been tested using the non-parametric Independent-Samples Kruskal-Wallis Test from SPSS across the following demographic groups:

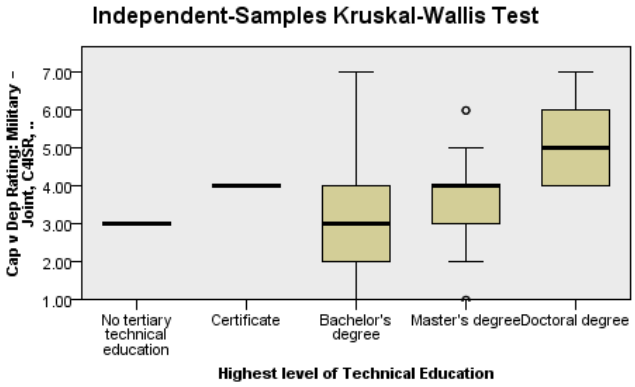
- Role View Point
- Sector of Most Professional Experience
- Current Employment Sector
- Application Domain of Most Professional Experience
- Main Technical Discipline of Education
- Highest Level of Technical Education
- Age Category
- Gender

Of these 128 tests, 8 were unable to be computed (all in the Gender group since no females provided judgments for those 8 System Domains), 4 indicated there was a difference in distribution (significant at the 0.05 level), and the vast majority, 116,

indicated there was no statistical difference in the way the groups of judges assessed the relative importance of capability and dependability within each of the system domains.

The 4 where there was a significant difference are:

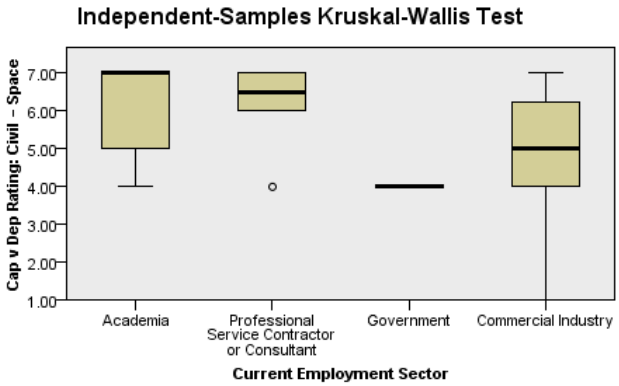
- Military – Joint, C4ISR across the categories of Highest Level of Technical Education (significance of 0.023 as indicated in Figure 64)
- Civil – Space across the categories of Current Employment Sector (significance 0.027 as indicated in Figure 65)
- Civil – Water Utility across the categories of Sector of Most Experience (significance 0.029 as indicated in Figure 66)
- Civil – Water Utility across the categories of Role View Point (significance 0.045 as indicated in Figure 67)



Total N	57
Test Statistic	11.318
Degrees of Freedom	4
Asymptotic Sig. (2-sided test)	.023

1. The test statistic is adjusted for ties.

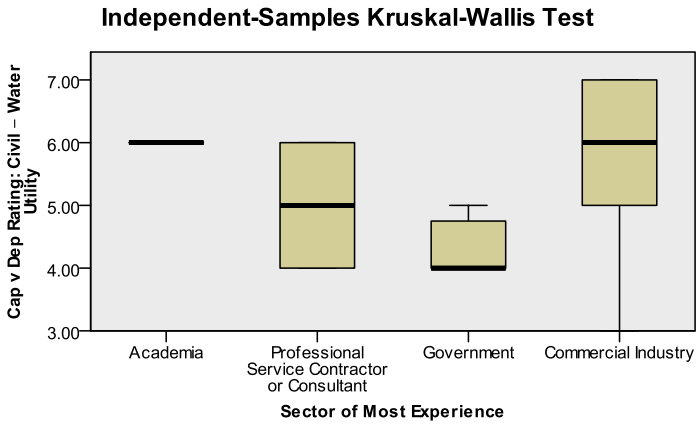
Figure 64. Difference in distribution of judgements for 'Military – Joint, C4ISR' across the demographic category of Highest Level of Technical Education.



Total N	35
Test Statistic	9.167
Degrees of Freedom	3
Asymptotic Sig. (2-sided test)	.027

1. The test statistic is adjusted for ties.

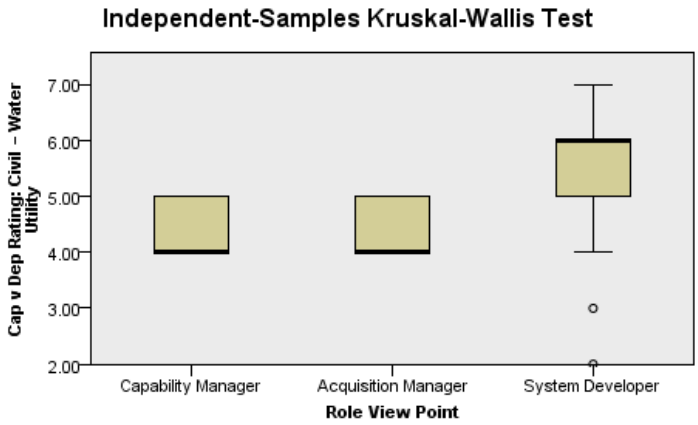
Figure 65. Difference in distribution of judgements for ‘Civil – Space’ across the demographic category of Current Employment Sector.



Total N	31
Test Statistic	9.024
Degrees of Freedom	3
Asymptotic Sig. (2-sided test)	.029

1. The test statistic is adjusted for ties.

Figure 66. Difference in distribution of judgements for ‘Civil – Water Utility’ across the demographic category of Sector of Most Experience.



Total N	31
Test Statistic	6.188
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.045

1. The test statistic is adjusted for ties.

Figure 67. Difference in distribution of judgements for 'Civil – Water Utility' across the demographic category of Role View Point.

Given the very small number of significant differences across all the system domains and demographic groups, there seems little to be gleaned from these 4 exceptions.

Plotting the statistically significant median values now produces an amended version of Figure 61. The system domain types now clearly fall into three significant categories – A, B and C as shown in Figure 68.

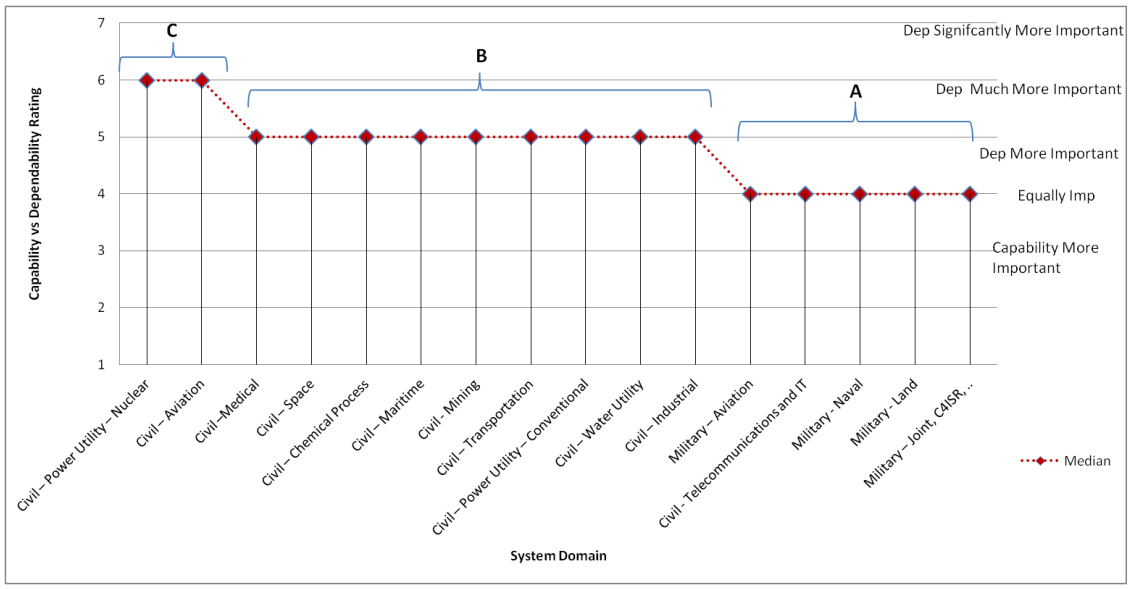


Figure 68. Significant median ratings of relative importance of capability and dependability versus system domains.

The following categories and characteristics are evident from this analysis:

- Category A: for all the Military system domains and the Civil ICT system domains, capability and dependability are rated of equivalent importance.
  - Although for Military – Aviation domain many ratings start emphasising the increased importance of dependability.
- Category B: in the Civil system domains of Water Utility, Power Utility, Transportation, Maritime and Mining, dependability is considered more important than capability.
- Category C: in these domains we a marked increase in the relative importance of dependability over capability through the domains of Chemical Process, Space, Medical, Civil Aviation and Nuclear Power was observed. It is likely that this reflects a perception by the judges as to the magnitude of the catastrophe that may arise in these system domains.

The hypothesis that *the ratings of the relative importance of capability versus dependability are the same across the different system domains* is rejected by these results.

7.3.4 Derivation of conjoint aggregation weights

The multiplicative conjoint equation $TI = C^{\alpha}.D^{\beta}$ was selected as an appropriate form of measurement in section 5.3 for the technical integrity theory inductively derived in section 5.2.5. The conjoint measurement approach follows the method of (Roberts, 1984) in forming a product structure and assessing that structure against well defined appropriateness conditions. The objective of this phase of the research was to derive information from the survey data that would allow the establishment of the aggregation weights (α, β) for the conjoint equation.

The results illustrated in Figure 68 indicate that for Category A system domains (Military system domains and the Civil ICT system domains) the aggregation weights of the conjoint equation should be equal, i.e., α = β, to reflect the equal importance of the capability and dependability components of technical integrity. Further, it is reasonable to set α = β = 0.5 to preserve the linearity on the (0,1) interval.

Further, the results suggest that β > α for Category B Civil system domains and β >> α for Category C Civil system domains (civil aviation and nuclear power systems) reflecting the increased importance of dependability over capability in these categories of system domains

As with any empirically derived relationship or form of measurement, while this multiplicative conjoint measurement model fits this data set, it will need to be tested by further research in order to build greater support.

7.4 Qualitative results

7.4.1 General

This section presents the qualitative results from the open-ended survey responses.

Recall that two open-ended questions were asked after the respondents had ranked the relative importance of capability and dependability across a range of system application domains:

- What is there about your domain of practice or viewpoint that has influenced the way you have judged the relative importance of Capability and Dependability?

- What is there about the service(s) provided by the system(s) you were thinking of, that has an impact on the way you have rated the relative importance of Capability and Dependability?

There were 102 answers provided to the first question and 92 answers provided to the second. In many cases the answers did not correlate well with the question. For example, respondents provided feedback about the systems they were considering in response to the first question and then realising this when confronted with the second question, referred back to their former answer. To compensate for this, both sets of answers have been considered together in the qualitative analysis. Many answers did not provide a direct response to either question and did not provide any other information or insights of use to our research and therefore have not been coded in the qualitative analysis.

Also noticeable in the responses, is a difficulty some respondents seem to have had in judging the relative importance of capability and dependability. The difficulty seems to be in trying to consider capability and dependability independently. That is, they implied they had to give up one in favour of the other. This misses one of the key objectives of using a conjoint measurement, that is, capability and dependability both contribute to the measurement of technical integrity rather than the dependability level directly influencing the capability level or vice versa. Comparing the example of apparent temperature given in the survey instrument, we are not requesting the judge trade humidity for air temperature or that humidity can exist without temperature, rather, that the relative importance of those quantities in forming a measure of the apparent temperature be assessed.

It could be thought that these difficulties should be seen as detracting from the validity of a conjoint measurement approach. But rather, the comments may be attributed as supporting the multiplicative conjoint model which assures that a poor dependability irrespective of capability level will yield a poor technical integrity level (and similarly for poor capability).

7.4.2 Domain of practice or viewpoint

Responses related to a respondent's stated Domain of Practice or Viewpoint have been coded in accordance with Table 41.

Table 41. Code table: System Domain or Viewpoint influences on judgment of capability versus dependability

Code Name	Criteria for Coding
DOMAIN OR VIEWPOINT INFLUENCES	Coded to entire response to the question regarding domain of practice or viewpoint that influenced judgements.
Domain of Practice Influences	Not coded.
• **Defence Systems Practice Influences**	Coded when a comment reflecting that the respondent indicated their experience in, or view of, Defence Systems influenced the rating of Capability vs Dependability.
○ **Aerospace Defence Practice Influences**	Coded when a comment reflecting that the respondent indicated their experience in or view of Aerospace Defence Systems influenced the rating of Capability vs Dependability.
○ **Maritime Defence Practice Influences**	Coded when a comment reflecting that the respondent indicated their experience in or view of Maritime Defence Systems influenced the rating of Capability vs Dependability.
○ **Land Defence Practice Influences**	Coded when a comment reflecting that the respondent indicated their experience in or view of Land Defence Systems influenced the rating of Capability vs Dependability.
• **Civil Systems Practice Influences**	Coded when a comment reflecting that the respondent indicated their experience in or view of Civil Systems influenced the rating of Capability vs Dependability.
○ **Civil Aerospace Practice Influences**	Coded when a comment reflecting that the respondent indicated their experience in or view of Civil Aerospace Systems influenced the rating of Capability vs Dependability.
○ **Other Civil Practice Influences**	Coded when a comment reflecting that the respondent indicated their experience in or view of Civil Systems (other than Aerospace) influenced the rating of Capability vs Dependability.
Role in System Viewpoint Influences	Coded when a comment is made reflecting the respondent's role in the System's lifecycle has influenced the rating of Capability vs Dependability.
• **System Developer Viewpoint Influences**	Coded when a comment is made reflecting the respondent's role as a System Developer has influenced the rating of Capability vs Dependability.
• **Capability Manager Viewpoint Influences**	Coded when a comment is made reflecting the respondent's role as a Capability Manager has influenced the rating of Capability vs Dependability.
• **Acquisition Manager Viewpoint Influences**	Coded when a comment is made reflecting the respondent's role as an Acquisition Manager has influenced the rating of Capability vs Dependability.
• **System Operator Viewpoint Influences**	Coded when a comment is made reflecting the respondent's role as a System Operator has influenced the rating of Capability vs Dependability.

All 102 responses to the first question have been coded to "DOMAIN OR VIEWPOINT INFLUENCES". This represents the total number of responses to this question, but as noted previously some extracts have been coded for responses made from the extracts coded as "SYSTEM SERVICE INFLUENCES" (see Table 42). Each of these responses has had one or more text extracts made and coded to one of the other codes in accordance with the above criteria. The number of coded extracts is shown in Figure 69.

It is noted that the large number of extracts (32 + 16 + 13 + 6 = 67) relate to an experience or viewpoint based in defence systems. Not many extracts (only 13) are able to be coded to system roles, although we know from the demographic data that the majority (58% of responses) were from respondents that indicated they have a System Developer viewpoint (see demographics in Chapter 6).

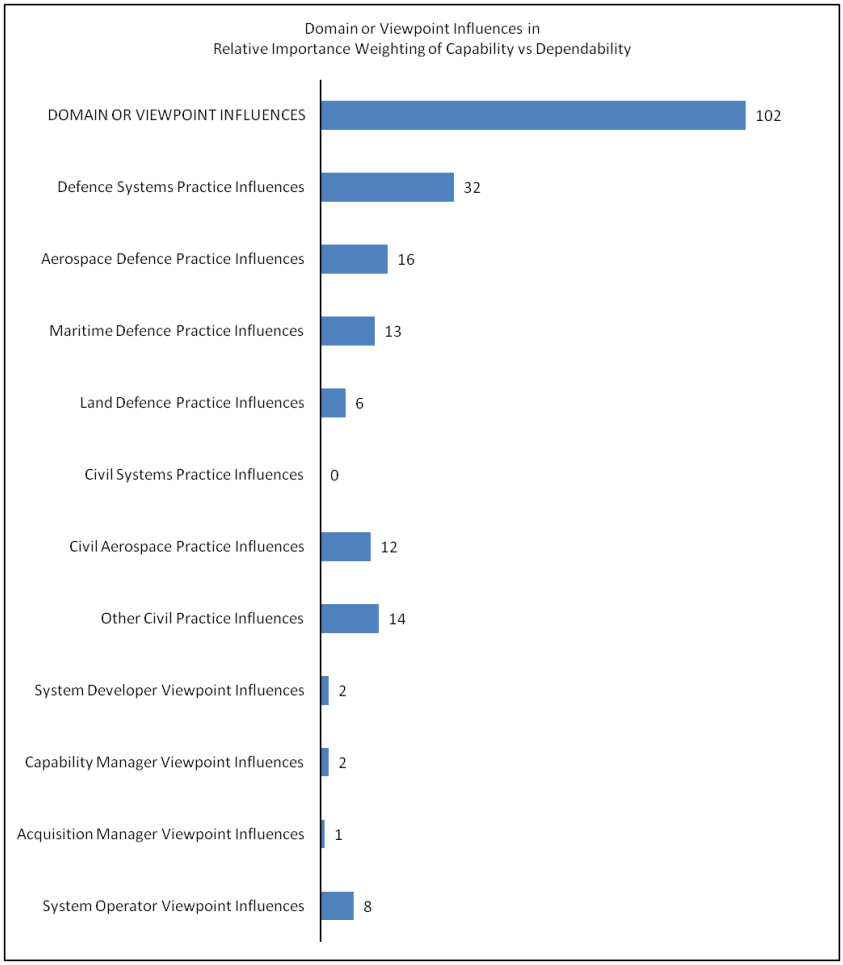


Figure 69. Number of coded extracts – System Domain or Viewpoint Influences on judgment of Capability versus Dependability.

7.4.2.1 Domain of Practice Influences

It is to be determined what domain experience of the respondents they considered to be influencing their judgments on the relative importance of capability and dependability.

7.4.2.1.1 Defence systems practice influences

27 comments have been coded as Defence Systems Practice Influences and are presented in Appendix D. Views expressed that reflect both capability and dependability are considered to be very important in the Defence systems domain.

9 comments clearly reflect judgements that capability is more important than dependability. Some examples include: 'Military systems tend to be accepted more on the capability presented, even if there is some level of undependability remaining in the system'; 'Capability determines survivability in a threat environment. Whilst dependability is also very important, I lean towards capability as being of higher importance'; and 'From a military systems perspective I think that they are inherently risky and therefore providing better capability should be more of a focus to maintain a technical edge.' The judgments in the main seem to reflect a view that capability is what is required to survive in the military environment.

5 commented that they considered capability and dependability equal, for example: 'In the military environment, there may be a perception that Capability is more important but the requirements of technical integrity related to safety and reliability are now judged of equal importance, i.e., that Dependability is of equal importance in the greater Government outcomes perspective' and 'From a military perspective, capability and dependability go hand-in-hand. In theatre, you want the best capability you can get; however, it cannot fail (must be dependable)'.

7 commented that dependability is more important than capability: 'It could be argued dependability is more important as a known capability that is known to behave as expected is more important than a great capability that may or may not work when required.'

Another interesting point of view is that it is possible for the users to work around capability shortfalls but not around dependability shortfalls: 'My view point is influenced by experience which suggests that the military rarely acquires exactly the

capability they desire, but can develop workarounds to adapt to this. However, it is very difficult to adapt to poor dependability.'

7.4.2.1.2 Aerospace defence practice influences

14 comments have been coded as Aerospace Defence Practice Influences and are presented in Appendix D.

1 comment clearly reflected a judgement that capability is more important than dependability: 'Dependability is important, but accomplishment of the mission is paramount. Within limited high-risk scenarios such as war or emergency, the capabilities of a system will be of paramount concern and over-ride concerns with reliability.'

2 commented that they considered capability and dependability equal, for example: 'My military (combat) aircraft experience dealt with vehicle and major systems that needed to be highly capable and dependable to provide a reasonable probability of a 'win'.'

9 respondents commented that dependability is more important than capability. One comment reflects a common theme within the nine that while capability needs are important, that there is a stronger importance on dependability: 'For military aviation, there are some circumstances where the achievement of the mission (the capability) far outweighs the necessary dependability of the asset. This is a rare occasion though, and I think that on balance, particularly in peacetime the dependability argument will always win.'

7.4.2.1.3 Maritime defence practice influences

13 comments have been coded as Maritime Defence Practice Influences and are presented in Appendix D. Many of these comments reflect the theme that the respondent wants both capability and dependability and are not able or willing to prioritise (this theme was analysed further in Chapter 6).

Two comments clearly reflect judgements that capability is more important than dependability. Some examples include 'Capability for naval projects has been more important than dependability as there has been availability of spares and trained personnel at sea - it is part of the naval business' reflecting an ability to cope with poorer than desired dependability.

One comment considered capability and dependability equal: ‘In all systems I have delivered including Naval military and civil Transport both Capability and Dependability are equally important.’

Two commented that dependability is more important than capability. One response indicates dependability is most valued - ‘Most operations involving naval platforms are conducted during peace time operations where focus is on safe operations - these activities are conducted over extended periods where access to support is limited and dependable operation is valued’ – conflicting with the aforementioned comment that dependability shortfalls can be worked around.

7.4.2.1.4 Land defence practice influences

Six comments have been coded as Land Defence Practice Influences and are presented in Appendix D. Two of these comments reflect judgements that ‘in the land environment, there is often greater flexibility/redundancy of systems than enables a greater focus on capability vs dependability’.

7.4.2.1.5 Civil aerospace practice influences

Twelve comments have been coded as Civil Aerospace Practice Influences and are presented in Appendix D. Eight of these comments reflect a strong bias to the relative importance of dependability. Comments like ‘Dependability is the driving influence for civil aviation applications given the use of it by the public’ and ‘Civil aviation has airworthiness as the paramount aim - even at the exclusion of payload/range (i.e.: redundancy is provided at the expense of capability)’ reflect these judgments.

Two respondents indicated an equal rating is justified, e.g., ‘Safety first, in particular in the aviation environment, but capability always. As such, I think they are equally as important to achieve’. But as expected from the qualitative results, no respondent made a case for the higher importance of capability in their thinking in this domain.

7.4.2.1.6 Other civil practice influences

Fourteen comments have been coded as Other Civil Practice Influences and are presented in Appendix D. While not a large set of comments, some insights can be gained from comments made for particular civil domains.

Seven comments were made reflecting the importance of dependability over capability in the domains the respondents were familiar with. These domains all share the characteristic that failure may lead to great public safety or poor commercial outcomes:

- In the mining domain it was commented that ‘Mining is dominated by safety concerns and predictability of system performance’.
- In Commercial industrial domains ‘industrial monitoring and control systems have relatively high dependability requirements as compared with capability’.
- As expected in the nuclear system domain one respondent has ‘rated nuclear regarding high dependability due to my understanding of the safety aspects of that industry’ and another for space and medical domains as well: ‘In space you only get one chance (this then also applies to medical, from the perspective of the individual, and for nuclear from the perception of the community). They have to work, there is no trade off of capability against dependability that would satisfy those involved’.
- For the civil transportation domain too we have the comment that ‘Safety has been the most critical aspect for transportation experience (rail, aircraft etc), hence dependability has been the primary driver’.

One comment has been made in this domain that reflects a view that in ‘civil Transport both Capability and Dependability are equally important’.

Two comments have been made that reflect a view that capability is more important, both these comments though reflect a viewpoint of civil ICT systems. For example ‘From an IT and Telecoms perspective, technology is fast moving and capability is required to stay ahead of competitors’.

7.4.2.2 Role in system viewpoint influences

What characteristics of the respondents’ system viewpoint that they considered to be influencing their judgments on the relative importance of capability and dependability are determined in this section

7.4.2.2.1 System developer viewpoint influences

Two comments have been coded as System Developer Viewpoint Influences and are presented in Appendix D.

An interesting comment from a developer perspective is that ‘the customer / stakeholders are predominantly interested in the fulfilment of the specified requirements

- which translates as the capability. Dependability, being the areas defined by ILS disciplines often take second place'. If this view is accepted then it points to a de facto preference for capability over dependability.

7.4.2.2.2 Capability manager viewpoint influences

Two comments have been coded as 'Capability Manager Viewpoint Influences' and are presented in Appendix D.

Both comments reflect a balanced judgement between capability and dependability in the mind of the Capability Manager as illustrated in this comment: 'capability and dependability go hand-in-hand. In theatre, you want the best capability you can get; however, it cannot fail (must be dependable)'.

7.4.2.2.3 Acquisition Manager Viewpoint Influences

One comment has been coded as Acquisition Manager Viewpoint Influences and is presented in Appendix D. The respondent makes the comment that 'Consequences of dependability short falls' 'are high as customer' 'is highly sensitised to the risks professionally they perceive to their jobs'. This is in conflict with other views that the acquisition customer is only focused on achieving contracted requirements, and may be an outlying view only.

7.4.2.2.4 System operator viewpoint influences

Eight comments have been coded as System Operator Viewpoint Influences and are presented in Appendix D.

The only theme drawn from these comments is that two respondents indicated real or perceived shortfalls in capability in systems they had operated: 'Many of the systems I worked with although dependable had performance that was significantly inferior to those that I perceived the potential enemies had' and 'Effort from the capability acquisition organisations was highly focussed on dependability at the expense of rigorously examining performance and putting in place high performance systems'. Note again that the last quote here is in conflict with those presented earlier that expressed the view that the acquisition customer is only focused on achieving contracted requirements.

7.4.3 Services provided by the system

Responses related to a respondent's stated consideration of the services provided by the system have been coded in accordance with Table 42.

Table 42. Code table: system service influences on judgment of capability versus dependability.

Code Name	Criteria for Coding
SYSTEM SERVICE INFLUENCES	What is there about the service(s) provided by the system(s) you were thinking of, that has an impact on the way you have rated the relative importance of Capability and Dependability?
Dependable	Coded when a comment reflects the importance of Dependability or bias thereto.
Importance of or Loss of Capability	The impact of losing the capability is considered in the judgement; or the overall importance of Capability for the system.
Safety	The safety consequences of the system impacts judgements.
Support Concept	Need of the system to operate away from support - reliance on organic support.

All 92 responses have been coded to "SYSTEM SERVICE INFLUENCES". This represents the total number of responses to this question, but as noted previously some extracts have been coded for responses made from the extracts coded as "DOMAIN OR VIEWPOINT INFLUENCES" previously. Each of these responses has had zero or more text extracts made and coded to one of the other codes in accordance with the above criteria. The number of coded extracts is shown in Figure 70.

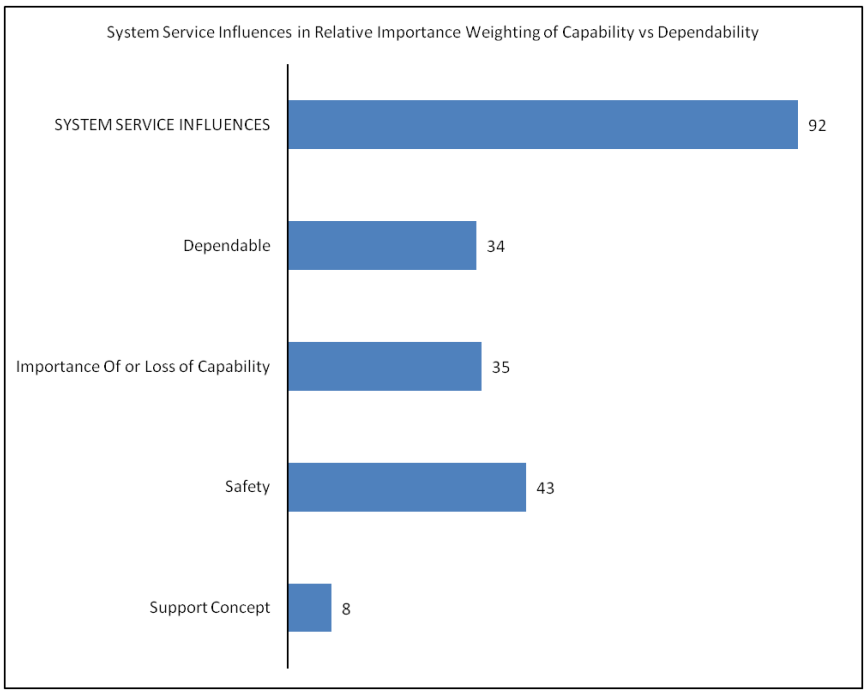


Figure 70. Number of coded extracts: system service influences on judgment of capability versus dependability.

A significant number of comments received reflected on the importance of a system's dependability, the importance of the capability of function of the system and the largest number of comments about the safety consequences of a system and how that influences the judgements of capability versus dependability.

7.4.3.1 Dependable

Thirty-four comments have been coded as Dependable and are presented in Appendix D. Three main themes are evident in these comments:

- Dependability becomes a primary consideration of a system when the consequence of failure of the system is high or there are no second chances. Cited examples include aviation, submarines and space systems:
  - 'In the aviation domain, the consequences are potentially far more significant if undesired consequences remain in the system and therefore reliability and redundancy are more important'.
  - 'Within the military domain, aviation and submarines would place a high importance on dependability as they are vulnerable when failures do arise, whereas surface ships less so', and
  - 'In space you only get one chance'.

- ‘If customer/stakeholder confidence in the system is paramount then dependability becomes a primary consideration:
  - ‘A dependable system is vital for confidence in the operation of a complex system’,
  - ‘For systems that can kill people, it is more important that they do so only in a well known and controlled manner’, and
  - ‘A less capable system that they can really rely on would be far more desirable than a more capable system that often let them down. Predictability is more important’.
- Dependability is more important since ‘the rarity of actual operational need makes dependability important - it has to be available when the moment arrives. Being able to operate at the limit of technological capability is a very infrequent requirement’ or because of ‘experience which suggests that the military rarely acquires exactly the capability they desire, but can develop workarounds to adapt to this. However, it is very difficult to adapt to poor dependability.’

7.4.3.2 Importance of or loss of capability

Thirty-five comments have been coded as Importance of or Loss of Capability and are presented in Appendix D. Two main themes, different than identified previously under Dependability, are evident in these comments:

- The importance of being capable of defeating a threat in a military environment.
  - ‘Within limited high-risk scenarios such as war or emergency, the capabilities of a system will be of paramount concern and over-ride concerns with reliability’,
  - ‘In the Military Naval domain, the failure of a Capability can have severe consequences to the mission - e.g. not being able to defend against an attack’, and
  - ‘Capability determines survivability in a threat environment’.
- Dependability may be ‘assumed’ or its shortfall readily accepted by stakeholders:
  - ‘Capability always tends to drive the market. Dependability tends to be assumed’.

7.4.3.3 Safety

Forty-three comments have been coded as Safety and are presented in Appendix D. Three strong themes are evident in these comments:

- A belief that safety should be considered as an explicit or separate element of technical integrity:
  - ‘Each complex system I have been involved in has emphasized the importance of Capability, Dependability and SAFETY’,
  - ‘I therefore have a great issue with rolling safety into dependability’, and
  - ‘I still maintain that 'safety' is a specific measure of 'integrity' which does not lend itself to be 'rolled up' into other integrity measure’.
- Safety figures much more highly than other risks of unintended consequence in the consideration of dependability:
  - ‘From other civil applications, safety of the public users of systems is paramount’, and
  - ‘Safety has been the most critical aspect for transportation experience (rail, aircraft etc), hence dependability has been the primary driver’
- Consideration of potential personnel safety consequences drives the rating of dependability:
  - ‘ensuring that there no undesirable consequences are more important in endeavours where there is a significantly higher risk to life (i.e. aviation and medical)’,
  - ‘Dependability is more important where the consequences of failure are life threatening or would lead to a catastrophe, i.e., safety critical systems’, and
  - ‘The consequences of unexpected or undesired behaviour especially with regard to major personal/population safety risks, should be considered first in all domains.’

7.4.3.4 Support concept

Eight comments have been coded as Support Concept and are presented in Appendix D. Strong themes are not apparent in such a small set of comments. Some views are evident that the organic support available to the mission system can change the relative importance of dependability: ‘I feel that the capability is preferential to the dependability for there is time to address the dependability with stores and technical

support onboard to manage the dependence issues' but not 'where there is an overhead (e.g., because the system is remotely deployed) in repair of the system.'

7.5 Combined (mixed-method) results

Recall the objective is to determine valid weightings for the combination of the defined capability and dependability measures in the conjoint technical integrity measure and these are to be derived by establishing the relative importance of capability and dependability.

The quantitative analysis results indicate that there are three sets of aggregation weights of relative importance of capability and dependability. Each set of weights covers three categories of systems:

- Category A ('Military systems' and 'Civil ICT systems') where capability and dependability are of equal importance,
- Category B ('Important Civil systems' including the domains of Water Utility, Power Utility, Transportation, Maritime and Mining) where dependability is considered more important than Capability, and
- Category C ('Critical Civil systems' including the domains of Chemical Process, Space, Medical, Civil Aviation and Nuclear Power) where dependability is considered much more important than Capability.

It is noteworthy that for no class of system do the consensus judgements indicate that capability is more important than dependability. This is a key outcome considering the practice in systems development over the recent decades.

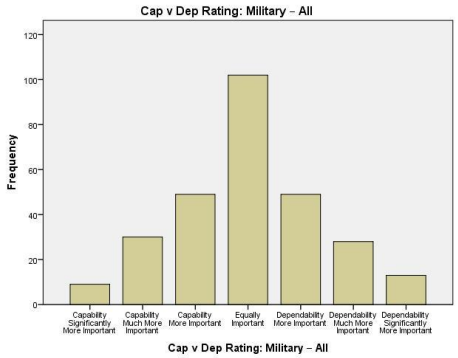
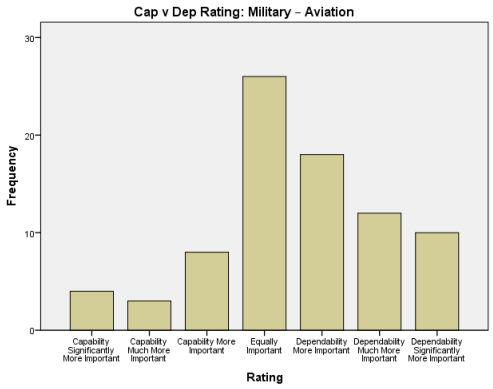
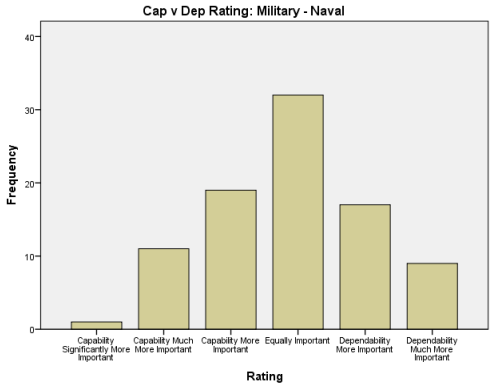
In the researcher's professional experience, predominately in the design and production lifecycle phases of military systems, there has clearly been a bias in process and behaviour toward achieving specified requirements (capability in our definition) over the minimisation of unintended consequences (dependability in our definition). Consider that the Military acquisition processes in Australia make payments to contractors (usually the designers and producers of these systems) when requirements have been verified for system acceptance and that there is often little or no payment associated with design evidence supporting a claim of the dependability of the system (Defence Materiel Organisation, 2007, p. 46-79). However it is evident that the introduction of technical regulation over the last decade has had some effect on

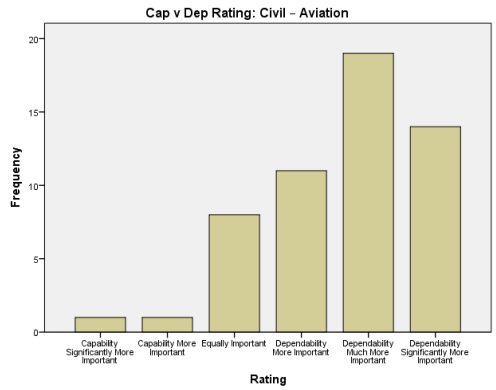
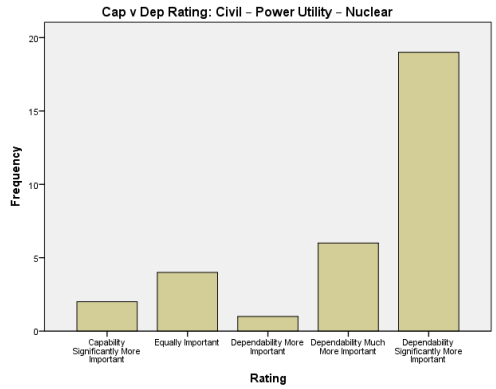
increasing the importance of dependability in design by increasing emphasis on techniques like system safety and product. However, there is still room for improvement since it is believed ‘that the organisations as a whole do not place technical integrity at the same level of importance as short-term operational imperatives’ (Rizzo, 2011, p. 67).

The result that has been termed ‘Critical Civil systems’ likely reflects a perception by engineering and risk management more generally and ensuring overall technical integrity outcomes are explicitly tied to professional responsibilities of designers through formal certification of design the judges of the magnitude of the catastrophe that may arise from these system domains and the public perception of risk for these type of systems (nuclear power systems being the prime example).

Table 43 compares the ratings data obtained by the survey and compares it to the number of comments received that have been coded to corresponding domains of practice as analysed in “Domain of Practice Influences”. The table indicates that the spread of qualitative comments made in justifying relative importance judgements is consistent with the distribution of the ratings data.

Table 43. Capability versus dependability importance judgements: comparison of quantitative and qualitative data.

Quantitative Ratings Data	**Qualitative Comments**			**Comments**
	C>D	**C=D**	**C<D**	
	9 **14**	5 **8**	7 **18**	Military - All is combined from ratings of Military - Aviation, Military - Naval, Military - Land, Military – Joint, C4ISR. Number of comments made on defence systems generally. Number of comments made generally plus specific comments from Military - Aviation, Naval and Land. Even spread of comments made is consistent with distribution quantitative responses, i.e., symmetric about capability and dependability being equally important.
	1	2	9	The larger number of comments made on importance of dependability is consistent with the skew in the distribution of quantitative responses, toward dependability.
	2	1	2	Even spread of comments made is consistent with distribution quantitative responses, i.e., symmetric about capability and dependability being equally important.

Quantitative Ratings Data	Qualitative Comments			Comments
	C>D	C=D	C<D	
	0	2	8	The larger number of comments made on importance of dependability is consistent with the distribution of quantitative responses centred around dependability being much more important.
	0	0	1	The quantitative ratings of Civil - Power Utility - Nuclear exhibit the most extreme response in favour of the priority of dependability. Only one comment has been made that correlates with this result: “rated nuclear regarding high dependability due to my understanding of the safety aspects of that industry”

The analysis of comments presented in the quantitative results section “Services Provided by the System” supports expected reasoning for the relative importance of capability and dependability. Dependability becomes a primary consideration of a system when the consequence of failure of the system is high, if there are no second chances, or if stakeholder confidence in the system is paramount. For military systems the importance of meeting capability requirements for defeating a threat in a military environment is a key consideration but dependability is equally important, i.e., having the capability available to use when required. Some viewed that dependability may be “assumed” or dependability shortfalls readily accepted since workarounds such as increased maintenance staff or on-hand spares can be applied to achieve an acceptable ‘operational dependability’ but there is no such ready fix considered for capability shortfalls.

A theme that is evident in several comments that may impact on the validity of the conjoint measurement construct is that “Safety” should be considered as an explicit or separate element of technical integrity. This view, if accepted, would have us change the

construct of technical integrity to have a third element in the conjoint measure. Related comments indicate a view that Safety figures much more highly than other risks of unintended consequence in the consideration of dependability and thus safety consequences drives the rating of dependability. Acceptance of this view may see that the estimation of dependability is adequately approximated by an estimation of safety related risks alone. Judgement data that may support one or other of these approaches is being obtained in the subsequent phase of our research.

7.6 Conclusion

7.6.1 Findings

This chapter has extended the finding of Chapter 6 that the conjoint measurement construct for technical integrity is considered valid for the ranking of similar systems for the purpose of assessing the impact of technical regulation.

In this chapter, the relative importance of capability and dependability for the purpose of establishing aggregation weights in a multiplicative conjoint equation, has been demonstrated to fall into three categories:

- 'Military systems' and 'Civil ICT systems' where capability and dependability are of equal importance,
- 'Important Civil systems' where dependability is considered more important than capability, and
- 'Critical Civil systems' where dependability is considered much more important than capability.

Significantly, it is also noted that in no system domain considered does the consensus view indicate that capability outweighs dependability.

For the purpose of using a conjoint measurement construct for military systems developed under the provisions of ADF technical regulations, the conjoint equation may be set to:

$$TI =$$

7.6.2 Limitations

Numeric weightings for the multiplicative conjoint measure have been established only for the category of ‘Military systems’ and ‘Civil ICT systems’ where we can set both the weighting parameters to unity since the consensus opinion is that capability and dependability are of equal importance. Numeric weightings for the other categories of systems cannot be estimated from the available data. This does not affect this research which is delimited to focus on ADF military systems.

Statistical significance was not established for the system domains of Military-Aviation and Civil-Aviation and further data is required to confirm the assumed result.

7.6.3 Additional research questions

As anticipated by the nature of exploratory and descriptive research, additional questions have been raised within this chapter that are to be addressed in the subsequent phase of research including:

- Is safety an indicator of the concept of dependability that may be so dominant that it is in effect a suitable proxy for dependability in the conjoint measurement construct?
- Should safety form a third, additional concept of the technical integrity theory?

Note that these questions are effectively the same as that identified in the previous chapter (see section 6.6.3).

8 Assessment of judgments for estimation of technical integrity

8.1 Introduction

8.1.1 Purpose

Technical integrity of a complex engineered system is not directly measurable. No direct measures that may in whole or part be used as proxies for technical integrity have previously been established. The findings of Chapter 6 indicated that while the intent to measure the concepts of capability and dependability on a numeric scale were valid, there was not overwhelming support for the intent to measure dependability via some estimation of residual risk. Nor were other pragmatic objective means of estimation identified. These findings lead to the need to assess other pragmatic means to estimate technical integrity or its component parts. To this end, the CAT was identified in section 3.4.4 as a candidate approach for the measurement of technical integrity. In this research phase the CAT is evaluated both as

- a direct measurement instrument for technical integrity, and
- as a measurement instrument for indicators of technical integrity where these estimates may be combined in the form of a scale, or a conjoint measure, or some combination of these approaches to provide a numerical value for technical integrity.

8.1.2 Concepts and research questions

This section outlines the specific concepts which underpin the approach to this phase of the research, and the research questions arising from these considerations.

A complex engineered system will exhibit a range of system quality attributes. The system quality attributes may be indicators of the concepts of capability and

dependability and in turn of technical integrity. Some of the system quality attributes or indicators will be related to zero or more variables that may be objectively ascertained and are normally readily available as technical parameters of a system. As an example, it would be expected that the system quality attribute, reliability, would be an indicator of the concept dependability. The variable, MTBF is the technical parameter of the system that can be used as a measure of reliability.

Other quality attributes may not be empirically quantifiable and may have little or no relationship to the readily available technical parameters of a system. An example of this is the quality attribute usability, which despite being the focus of the specialist engineering activities of Human Engineering, is not readily represented by a technical parameter of the system. More usually, usability is subjectively assessed to be at least acceptable to the users and acquirers of the system.

Other quality attributes of the system lie somewhere between these examples. They may have some relatively weak correlation with objective variables of the system than reliability does with MTBF, and may require some level of subjective judgment involved in arriving at its "true" value. An example of this would be the quality attribute Fit for Service which would be expected to correlate strongly with objective indicators of compliance with requirements (particularly when there is confidence that the requirements comprehensively capture the purpose of the system and the conditions under which it is to be used) but is also expected to be influenced by the end-users perceptions of how they intend to use the system in realistic conditions.

Table 44 provides a reference to the concepts and terminology used within this chapter. The information hierarchy established in Table 5 is shown in the first column with examples of corresponding terms in the second column. The third column aligns commonly used system terminology with the information hierarchy. Hence it can be seen that 'quality attribute' of a system is synonymous with 'indicator' in the information hierarchy. The CAT has its own extant terminology. In this thesis the judgment dimensions of the CAT are synonymous with indicators of the information hierarchy. The fifth column defines terms used by indicator models introduced later in this chapter.

Table 44. Terminology for information, systems, CAT and models

Information Hierarchy	Examples in this research	System terminology	CAT terminology	Model terminology
Theory	Technical integrity		-	
Concepts	Dependability		-	
Indicators	Reliability	Quality attribute	Judgment dimension	Indicator model
Variables	MTBF	Technical parameter		Quantity variables Quality variables
Values	762 hours	Parameter value		Numerical values Values or information that enables ranking

For a complex engineered system in the concept and development phase of its lifecycle, its characteristics are defined by its technical description. When realised, the system will exhibit quality attributes and will exhibit a level of technical integrity for which a measure is required. These concepts are captured in the influence diagram at Figure 71. Further, the diagram shows that the technical description enumerates numerical technical parameters and enunciates other non-numerical technical parameters of the system. Zero or more of these technical parameters should form a full or partial proxy measurement for a system quality attribute. The technical description may also provide the information that allows a suitable judge or expert to estimate the quality attributes of the system and may also allow the estimate of the technical integrity of the system. Estimates or measures of the quality attributes may allow a scale for technical integrity to be developed.

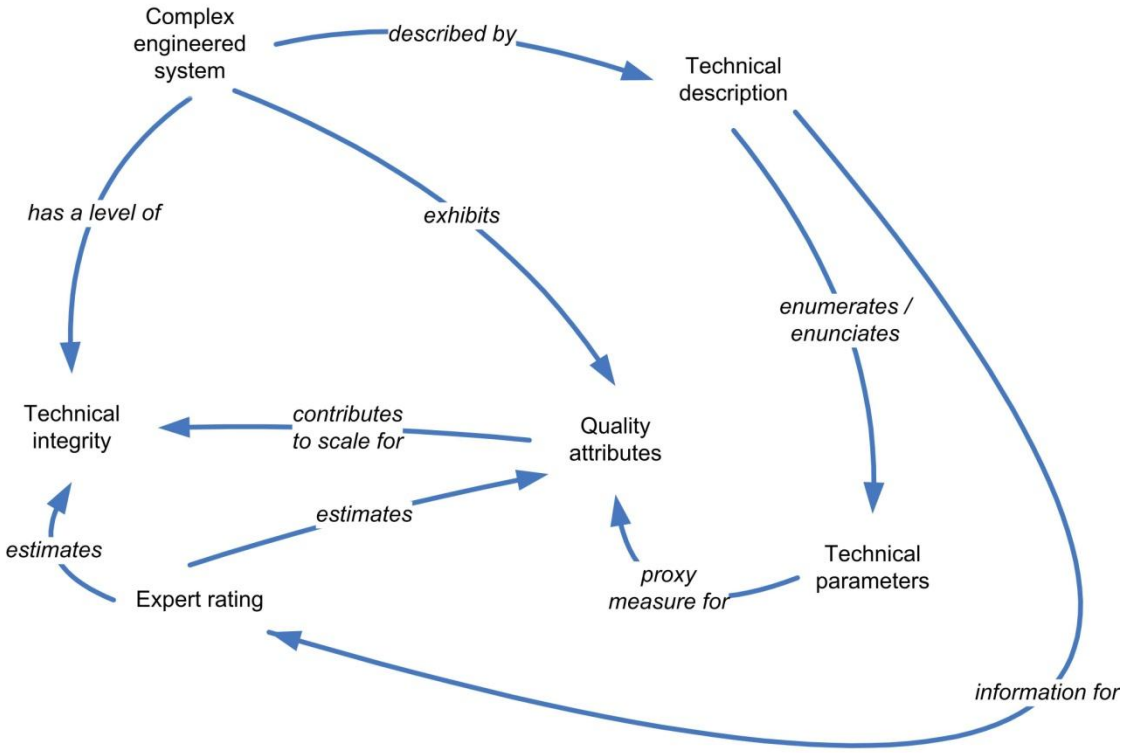


Figure 71. Influence diagram of technical integrity and quality attributes of a complex engineered system.

This model provided a structure for formulating the detailed research questions for the study reported in this chapter. The research questions were:

- How reliable can direct expert ratings of technical integrity be?
- How reliable can expert ratings for other system quality attributes be?
- Can the judgments of the quality attributes form a reliable scale for technical integrity?
- Where values and information are available for technical parameters of the system, how well do these values and expert ratings correspond?
  - Do the expert ratings show a rating or ranking that differentiates the attributes of the example systems in a statistically significant sense and do they correspond to that given by the system technical parameters?
- What is the best measurement instrument to be used for each indicator (via empirical estimates of technical parameters or via expert judgment)?

8.1.3 Constraints

Necessarily the study is subject to the following assumptions and constraints:

- For this study, no material manifestation of the system is available for informing the judges by inspection or verification activities (this is consistent with research intent of assessing if technical integrity can be reliably assessed or predicted during the development phase of the system's lifecycle).
- The technical description of the system is correct and valid (in practice this is never completely true).
- There is no objective, empirical process by which to measure the technical integrity of a complex engineered system. Judgments may therefore be assessed for consistency of rating and ranking across systems but there is no "ground-truth".
- A reasonable number of respondents was required in order to gain statistical significance in the results. It was considered that an impost of more than an hour of time per respondent would severely limit participation in the research, so the scope of the example systems provided was kept reasonably small. Consequently the results of this research may not hold for systems of greater scope and complexity.

8.2 Method

Three primary techniques underlie the research presented in this chapter: CAT; Scale Development, and weighted sum indicator models of quantity and quality variables. The CAT and scales are techniques most often used by the social sciences to measure phenomena that cannot be directly measured.

8.2.1 Survey instrument

The CAT was identified as a possible means of measuring technical integrity and its main features were described in section 3.4.4. The CAT involves having three aspects developed or available. There are *judgments dimensions* that the judges will be asked to rank the *products* on, as well as *objective measures* that may be determined for the *products.*

8.2.1.1 Products

The products in this investigation are complex engineered systems. Ideally, it would be desirable to focus on very large, very complex systems, but it is considered not tractable, for an initial investigation at least, to document the relevant characteristics of a number of very large systems and to brief these to the judges in a time effective manner or in enough depth for them to reasonably make judgments.

Rather an approach of defining a limited number of representative products was pursued. A set of naval/maritime radar systems was selected based on a consideration that these systems were reasonably complex, but their main function, operation and construction would be readily understood by most experienced system practitioners. The products for this study are three descriptions (akin to specification/product data sheets) of fictional but realistic radar systems that allow for the variation of the technical parameters of the systems which are expected to influence technical integrity and its indicators. The product descriptions were developed by consideration of openly available data on this type of radar systems coupled with the researcher's own professional experience in the field. The descriptions were refined through review by similarly qualified colleagues of the researcher that identified additional parameters that were included and errors in some parameter values. It was recognised that these systems may not be optimal in terms of testing the full potential range of a technical integrity measurement scale since the construction of the model erred toward realistic rather than more artificial examples. This was considered appropriate since the objective is for the measurement instrument to be able to be able to differentiate between real systems. No *a priori* expectation for the absolute or relative technical integrity of these three example systems was established.

The three products are included in Appendix B to this report and are identified as:

- System A: an older technology, full military qualified navigation radar system with secondary air search capabilities.
- System B: a new technology, lightweight, commercially qualified navigation radar system.
- System C: a recent technology, part-military qualified navigation and surface search radar system.

Block diagrams for each of these systems, as provided in the system technical descriptions at Appendix B, are reproduced here as Figure 72, Figure 73 and Figure 74 for ease of reference.

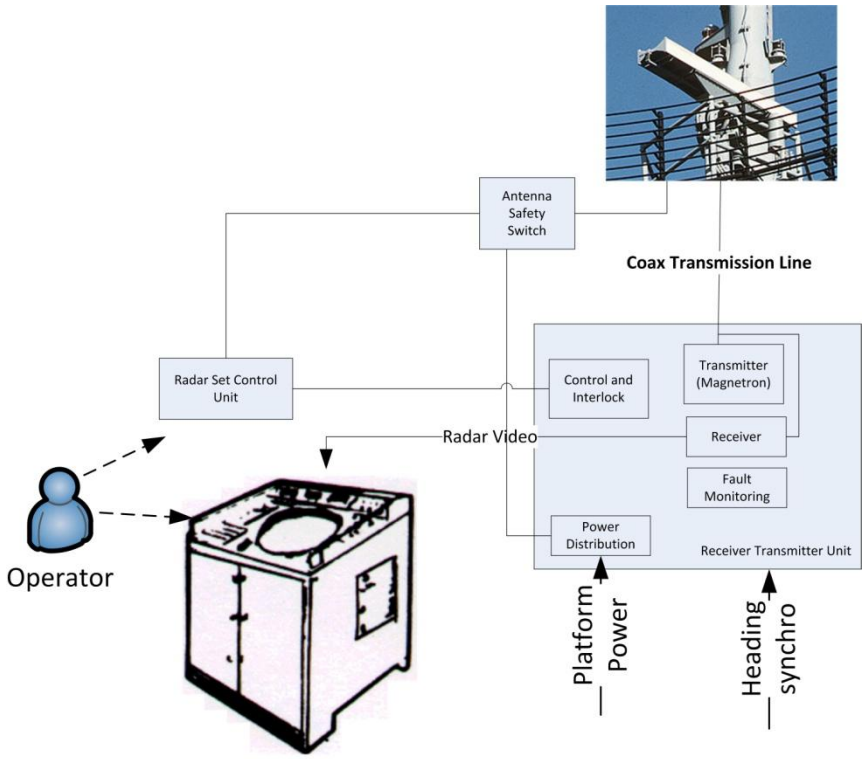


Figure 72. Block diagram for System A.

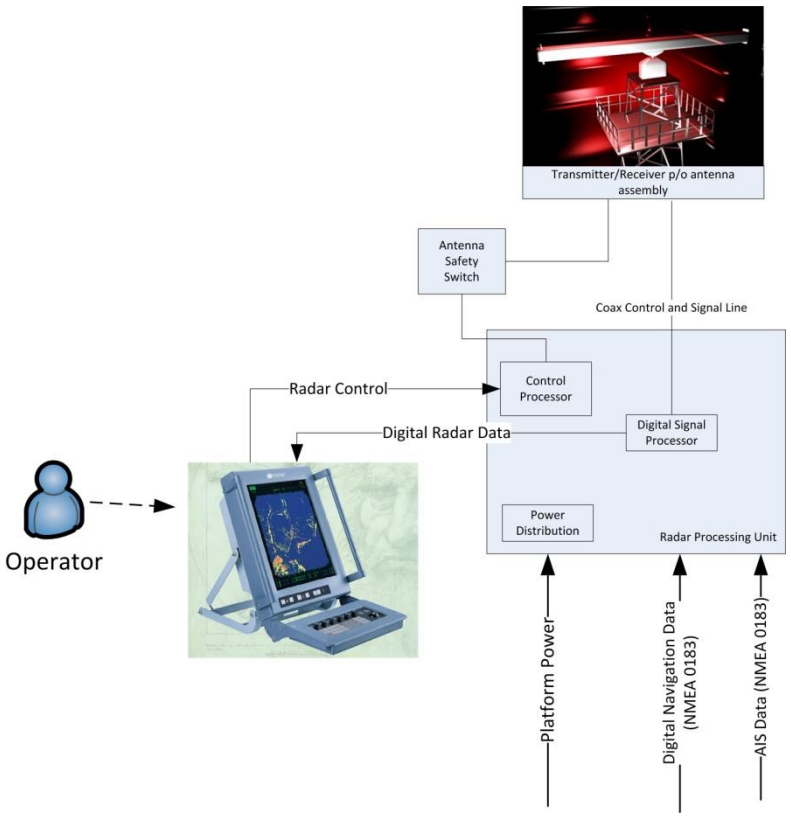


Figure 73. Block diagram for System B.

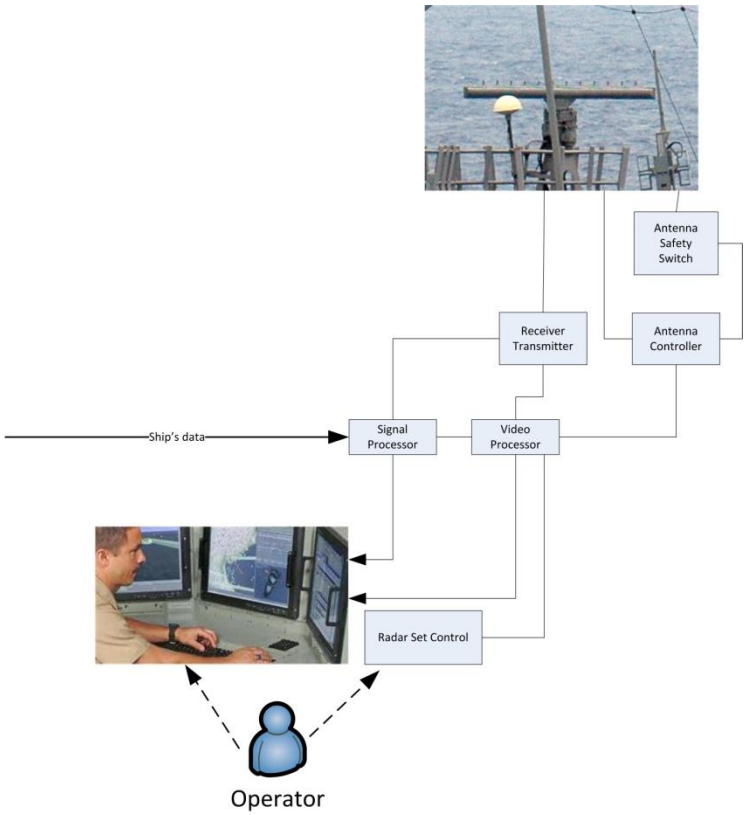


Figure 74. Block diagram for System C.

8.2.1.2 Judgment dimensions

An essential element of the CAT procedure is to ask for judgments using 'non-definitional definitions' (Amabile, 1996, p. 74). As an example, for technical integrity it is asked "Using your own subjective definition of technical integrity rate the degree to which each system exhibits technical integrity relative to the other example systems".

The following dimensions were requested to be evaluated by the judges for each system:

1. safety,
2. impact to the natural environment,
3. inadvertent impact to other systems,
4. fit for service,
5. capability,
6. effectiveness,
7. efficiency,
8. construction quality,
9. operationally flexible,

10. durability,
11. survivability,
12. dependability,
13. reliability,
14. supportability,
15. maintainability,
16. specification/standard compliant,
17. usability,
18. security,
19. technical complexity,
20. high technology, and
21. technical integrity.

On each judgment dimension, the judge was requested to rate each system on a 16 point scale with intervals to be interpreted as equi-spaced. The ratings for each system were enforced to be different in order to be able to extract definite ranking data. The judge was required to rate all products (systems) on the one judgment dimension before moving on to consider the next judgment dimensions. This was designed to assist in maintaining consistency in the judge's subjective criteria for a particular dimension and allows separability of the different dimensions. The order in which the dimensions were presented to the judge was randomised to minimise the chance of the order of evaluation unintentionally biasing the results.

Note, that technical complexity and high technology were not expected to be indicators of technical integrity. It would not be expected that these would correlate with either concept of capability or dependability. They were intentionally included and were used to assess the discriminant validity.

8.2.1.3 Objective measures

By way of example, in one of Amabile's studies, a paper mosaic product, she collected objective data about the number of pieces of mosaic used, the number of colours used, and the number of shapes used. These were shown to be correlated significantly with the judges rating of creativity.

Identifying objectively measurable variables, i.e., technical parameters, for each of the systems and then correlating these measurements with results obtained from the related

judgment dimensions allows correlation analysis that will indicate those measurable variables that correlate with the judgment of technical integrity or with other indicators for which judgments are obtained. The technical parameters for each system are included in the system technical descriptions at Appendix B. The survey respondents were free to use this data as they saw fit.

8.2.1.3.1 Technical parameters provided in the example systems

Table 45 summarises the value of each technical parameter for each system that was included in the system technical descriptions. Where the values of the technical parameter can give a clear ranking of the systems this is also presented. It is noted that many of the technical parameters provide only ordinal information therefore allowing a ranking only. These technical parameters may correlate with one or more of the indicators that are used as judgment dimensions. Note that the judges did not see the data in the format of this table when making their judgments.

Table 45. Technical parameter provided in the example systems.

Technical parameter	Technical parameter's value in survey example systems			Implied ranking of systems	Comments
	System A	System B	System C		
Compliance (25 requirements)					
Not Compliant	7 requirements / 28%	4 requirements / 16%	4 requirements / 16%	B=C >A	Note differences in ranking if compliance or non-compliance is considered.
Partially Compliant	4 requirements / 16%	2 requirements / 8%	4 requirements / 16%		
Compliant	14 requirements / 56%	19 requirements / 76%	17 requirements / 68%	B > C > A	
Capability and Performance					
Range	190 km	200 km	104 km (56 NM)	B>A>>C	Overall System B would be considered the most capable system and differentiating between A and C would be determined by relative importance of capability attributes.
Range accuracy	22 m	1% of range (< 20 m)	30 m	B>A>C	
Bearing accuracy	1.5º	1.2º	1.5º	B>A=C	
Data rate	3.75 s	1 s	2 s	B>C>A	
Automatic tracking capacity	0	20	128	C>B>A	

Technical parameter	Technical parameter's value in survey example systems			Implied ranking of systems	Comments
	System A	System B	System C		
Military threat resistance					
Electronic protection	tuneable frequencies, polarisation selection	matched filtering	coherent processing, sector based CFAR processing	A≈C>B	Not expected to be able to be objectively ranked by all judges.
Shock qualification	Yes, qualified to MIL-S-901D	No qualification	Yes, qualified to MIL-S-901D	A=C>B	
Supportability					
Built in test (BIT)	Low (some, but no details given)	High (full go/no-go testing)	Very High (Full testing to LRU level)	C>B>A	
MTBF	1200 hrs	4000 hrs	3000 hrs	B>C>A	
MTTR	20 mins	20 mins	30 mins	C>A=B	
Inherent availability (Ai)	99.972%	99.988%	99.983%	B>C>A	Not presented directly to judges. Calculated from MTBF and MTTR here.
Maintenance concept	On board repair	Return unit to OEM for repair	On board repair	A=C>B	Assume that ability to repair on-board would be superior for a shipboard system that may be a long way from the OEM.

Technical parameter	Technical parameter's value in survey example systems			Implied ranking of systems	Comments
	System A	System B	System C		
System heritage					
Year first deployed	1972	2003	2003	None	Subjective - System A may be considered "old" which may imply it is not considered as well as newer systems. However, newer systems may be considered "less mature" and therefore more prone to exhibit technical issues.
Numbers in service	140	30	50	A>C>B	Taken alone, more instances of the system in service indicate the system is likely to be more supportable and more likely to have had technical issues rectified.
Manufacturer's quality system	ISO 9001	ISO 9001	ISO 9001	A=B=C	No differentiation provided in this study (as is very often the case in practice, all manufacturers of complex systems will have quality accreditation).
Safety					
Hazardous materials (HAZMAT)	Yes / Cadmium plated connectors	No	Yes / Cadmium plated connectors and use of zinc chromate primer.	B>A>C	No significant health hazards when controlled properly, but cadmium plating and chromate primers highlighted in ADF technical regulations as requiring special authorisation for use.

Technical parameter	Technical parameter's value in survey example systems			Implied ranking of systems	Comments
	System A	System B	System C		
Antenna safety switch	Yes / No key	Yes / No key	Yes / No key	A=B=C	No distinction across the example systems for this indicator. Note that in ADF technical regulations there are important, but not mandatory, requirements for key switches so this may be seen as less than ALARP for some judges.
Electrical power interlocks	Yes	No	Yes	A=C>B	
Security					
Physical security	None	None	None	A=B=C	No differentiation across these three example systems.
Computer/information security	N/A	Password protection	Password protection		
Output Radiated Power	160 kW peak at 9.5 GHz	100 kW peak at 9.5 GHz	280 kW at 5.5 GHz	Performance: C>A>B RADHAZ: B.A>C	This indicator could influence two variables in opposite sense - high power should be negatively correlated with safety but positively correlated with capability/performance (other parameters of the radar being equal).
Environmental resistance standards	Military standards	Commercial standards - IEC 60945	Commercial standards - IEC 60945	A>B=C	Compliance with military standards normally implies greater resistance to environmental condition variation and extremes.

Technical parameter	Technical parameter's value in survey example systems			Implied ranking of systems	Comments
	System A	System B	System C		
Hotel services					
Electrical power required	1100 W	1000 W	2000 W	B≈A>C	Lower power required from platform is generally desired all else being equal.
Cooling load/waste heat	100 W	Not specified	1800 W	C < A or B	Waste heat is generally desired to be minimised to lessen impact on host platform and other systems. Although not specified for System B, it is assumed the judge would realise it was better than System C given input power.
Required data inputs	Synchro	Digital - NMEA 0183	Synchro and/or digital	C>B>A (but not strong difference)	System B - implies little impact to other systems for integration. System B - open standard implies little impact to other modern systems for integration. System C - ease of integration for legacy systems, newer systems may be impacted by need to retrofit with digital-synchro converters.
Weight	*531 kg*	*150 kg*	*400 kg*	B>C>A	Lower weight of components generates less impact to host platform. Impact of antenna assembly weight (top weight in a ship installation) would be expected to dominate considerations.
Display	140 kg	20 kg	200 kg	B>A>C	
Processor	260 kg	40 kg	20 kg	C>B>A	
Antenna assembly	131 kg	90 kg	180 kg	B>A>C	

Technical parameter	Technical parameter's value in survey example systems			Implied ranking of systems	Comments
	System A	System B	System C		
Power efficiency	0.15	0.1	0.14	A≈C>B	Calculated as peak power output over power input (simplistic for this study).
Compatibility					
EMC compliance	Partial compliance with MIL-STD-461D	Not compliant with MIL-STD-461D; compliant with lesser EMC requirements of IEC 60945	Compliant with MIL-STD-461D	C>A>B	
Electrical power standard	Compliant with MIL-STD-1399-300	Partially compliant with MIL-STD-1399-300	Compliant with MIL-STD-1399-300	A=C>B	
Technology					
Radio frequency design	Old	Current	Current	B=C>A	
Signal processing	Old	Current	Current	B=C>A	
Human machine interface	Old (based on CRT monitor and analogue radar video)	New (high resolution LCD display and full digital graphics)	Previous generation (LCD display, analogue radar video and digital overlays)	B>C>A	

Technical parameter	Technical parameter's value in survey example systems			Implied ranking of systems	Comments
	System A	System B	System C		
Multi-mission/multi-function features	Yes (navigational detection; low altitude air search)	No (navigational detection)	Yes (navigational detection; surface search; low altitude air search)	C>A>B	

8.2.2 Indicator models from technical parameters

Table 46 maps the technical parameter from Table 45 against the indicators (used as judgment dimensions) with which the theory developed in Chapter 5 expects them to correlate with (refer to Table 12). This was constructed *a priori* the judgment results. The technical parameters have been mapped into two columns:

- Primary; where it is expected there should be a strong correlation between the technical parameter and the indicator (judgment dimension), and
- Secondary; where it is anticipated that the judge may have relied to some extent on that technical parameter and a correlation may be observed.

Table 46. Hypothesised correlation of technical parameters with indicators of technical integrity.

Indicator (Judgment dimension)	**Technical parameter theorised to correlate with indicator**	
	Primary	**Secondary**
Safety	Hazardous materials (HAZMAT) Antenna safety switch Electrical power interlocks	
Environment Impact		Hazardous materials (HAZMAT) Output Radiated Power
Impact to Systems	EMC compliance Electrical power standard	Output Radiated Power
Fit for Service	Compliance (25 requirements)	Environmental resistance standards
Capability	Capability and Performance Range Range accuracy Bearing accuracy Data rate Automatic tracking capacity	
Effective		Multi-mission/multi-function features
Efficient	Power efficiency	Electrical power required
Quality	Manufacturer's quality system	Numbers in service
Flexible	Multi-mission/multi-function features	

Indicator (Judgment dimension)	Technical parameter theorised to correlate with indicator	
	Primary	**Secondary**
Durable	Shock qualification Environmental resistance standards	Manufacturer's quality system
Survivable	Electronic protection Shock qualification	Environmental resistance standards
Dependable	MTBF Environmental resistance standards	EMC compliance Physical security Computer/information security
Reliable	MTBF	
Supportable	Built in test (BIT) MTTR Maintenance concept	Numbers in service
Maintainable	Built in test (BIT) MTTR Maintenance concept	
Compliance	Compliant Not Compliant	Partially Compliant
Usability		Human machine interface
Security	Physical security Computer/information security	
Technical Complexity	Block diagram and technology descriptions	
High Technology	Radio frequency design Signal processing Human machine interface	

8.2.2.1 Ratings for each indicator based on technical parameters

Based on the expected correlation of technical parameters shown in Table 46 and the actual values for the example systems in Table 45 an indicator model that produced a rating of each of the three example systems was developed in the software tool Criterium Decision Plus (CDP). For each indicator, the model was based on a combination of one or more technical parameters. These technical parameters may be in the form of “quantity variables” that have numerical values or “quality variables” that

are characterised by qualitative information that allows a ranking of the systems on the parameter. CDP allows alternate systems to be compared across user defined decision criteria and provides a rating for each alternate system. The weightings employed within the tool may either be established by AHP pair-wise comparison or by direct estimation. While the weighting of two or more technical parameters (either quantity variables or quality variables in this study) is subjective, in-built sensitivity analysis allows the researcher to clearly see if the rankings of the systems are robust to changes in the weighting and scoring criteria.

In the indicator models, the technical parameters identified as primary in Table 46 were given substantially higher weighting than the secondary technical parameters. The ratings produced, given the inherent subjectivity, are only used for establishing the ranking of the systems. However the subjectivity has been mitigated somewhat by the researcher eliciting and validating weightings and scoring with professional colleagues during the development of the models.

The indicator models and produced ratings are provided in the results section 8.3.4 where the rankings are compared with the corresponding rankings produced for the judgment dimensions.

8.2.3 Sampling approach

Part of the allure of the CAT is that it does not appear that the level of the judge's expertise matters very much nor do they require training in the technique (Amabile, 1996, p. 60, 72). Appropriate judges need only to be persons who are familiar with the domain, but not necessarily experts in the domain of the products/systems being considered. The identification of target respondents is best described as 'purposive sampling' (Tashakkori and Teddlie, 1998, p. 76). Survey respondents were requested to participate based on their familiarity with the definition, design, integration, verification, operation and/or support of complex engineered systems. Demographic data collected allowed examination of variations in judgments against criteria that may be proxies for expertise.

The judgments were obtained via an online survey ensuring independence of the judges (under the CAT the judges are not meant to collaborate) and allowing them to make their judgments as their time commitments allowed. The survey responses were

collected in September through October 2012. 128 requests were made, with 54 of those commencing the survey and 43 completing the survey, representing a 34% completion rate and over a hundred hours of total effort by the respondents.

8.2.4 Demographics of respondents

This section presents the raw demographic information and broad observations about the respondents as a group only. The results sections will use this data to identify if there are any significant differences in judgments and other responses that may be attributable to or explained by demographic groupings.

The charts in this section summarise the demographic data provided by the respondents. It is noted in Figure 75 that there is a preponderance of male respondents in the group. This is not particularly unrepresentative in the engineering of defence systems.

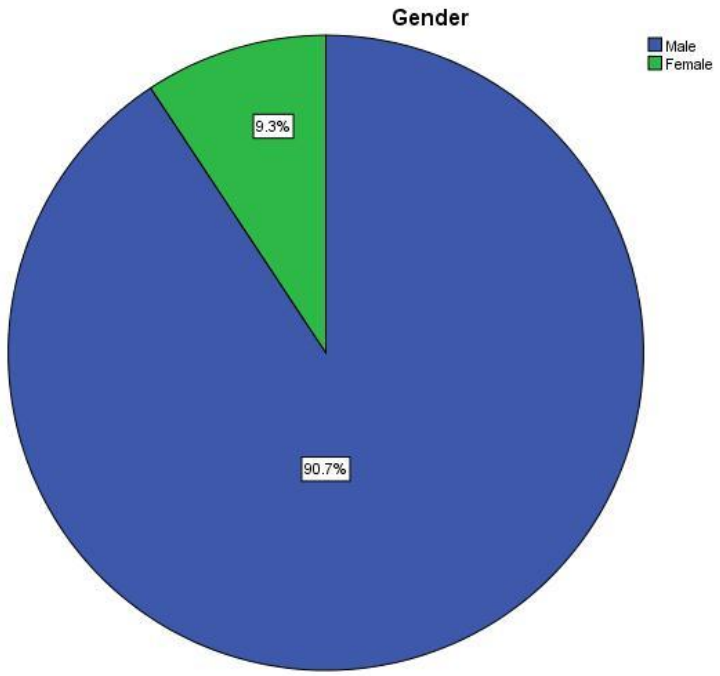


Figure 75. Pie chart of the gender of the respondents (n=43).

The distribution of ages of the respondents is shown in Figure 76. Along with Figure 81 which shows the years of professional experience, it indicates that most of the survey respondents are mature and experienced.

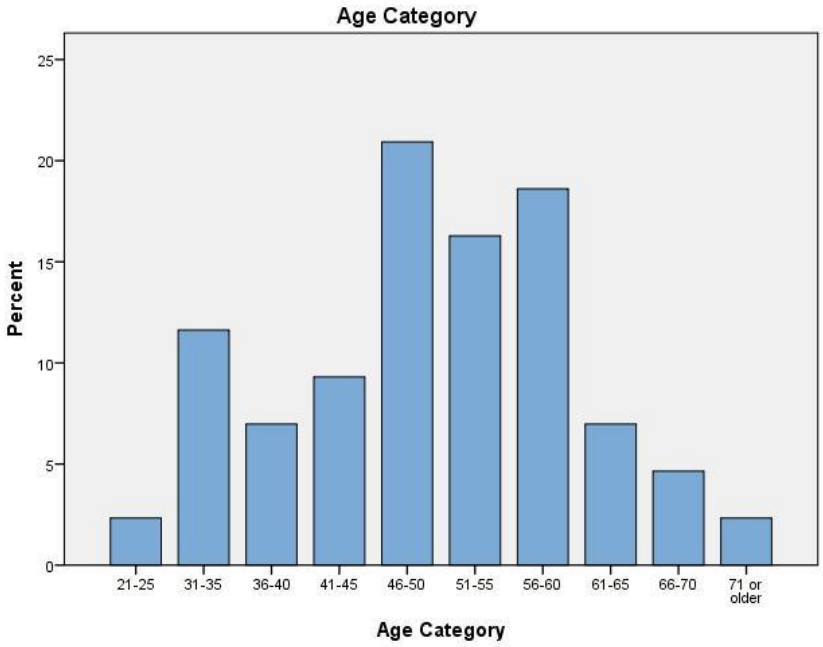


Figure 76. Histogram of the age of the respondents (n=43).

Figure 77 shows the highest level of technical education achieved by the respondents. 95.3% of the respondents have a university level education. This may be compared to the general Australian population where 24% people aged between 15 and 64 years hold a bachelors level degree or higher (Australian Bureau of Statistics, 2011) and demonstrates that the respondents as a group are significantly more educated than the general population as expected in the engineering and usage of complex systems.

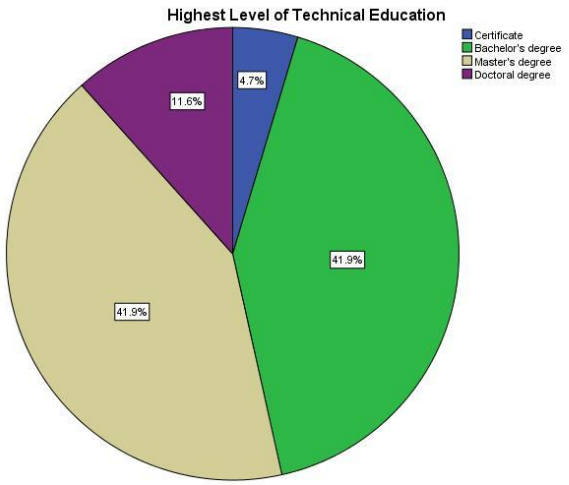


Figure 77. Pie chart of the highest level of technical education of the respondents (n=43).

Figure 78 shows there is a majority of respondents who have identified their main technical education as being in electrical/electronic engineering. This is somewhat unexpected since most technical disciplines are needed in the realisation of complex engineered systems. It may be partly explained in that most, if not all, of these systems embody electrical systems whereas, for example, aircraft based systems do not use the services of naval architects.

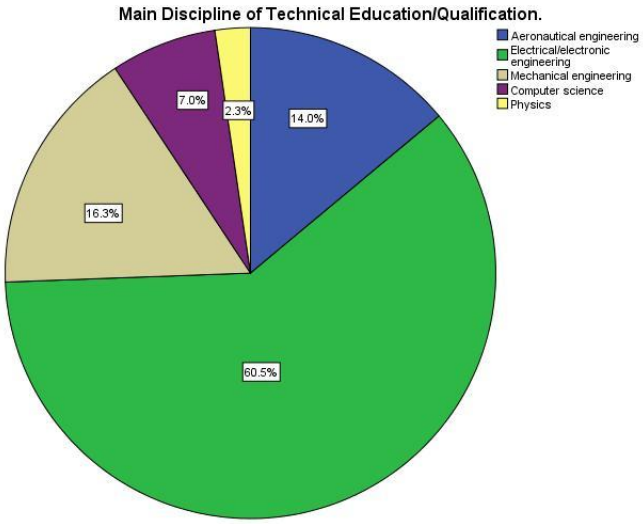


Figure 78. Pie chart of the main discipline of technical education of the respondents (n=43).

Figure 79 and Figure 80 on employment sector reflect a good spread among the defined sectors. Note the larger number of respondents who are currently employed as consultants and academics compared to where most experience had been gained. This could be explained by persons having long careers in the military or government and then taking on consulting work or academic roles later in their careers.

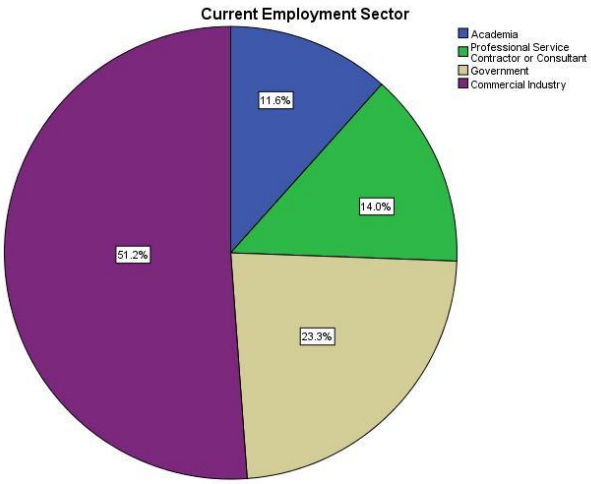


Figure 79. Pie chart of the current employment sector of the respondents (n=43).

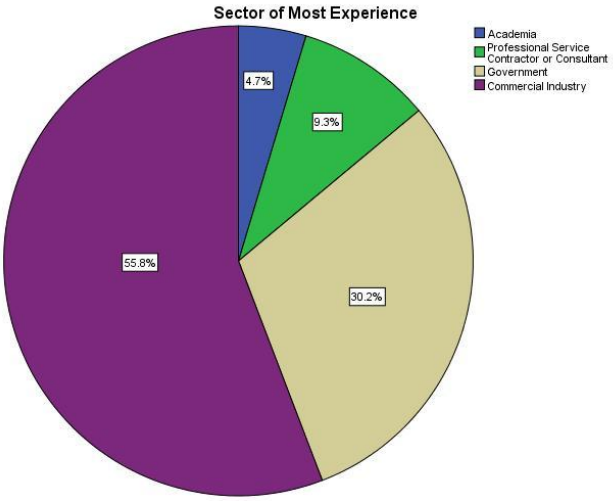


Figure 80. Pie chart of the employment sector of most experience gained by the respondents (n=43).

The group exhibits significant professional experience (most with greater than 10 years as shown in Figure 81) and 55.8% with formal engineering delegations in ADF regulatory frameworks (refer Figure 82 and Figure 84); an experienced and responsible group of professionals.

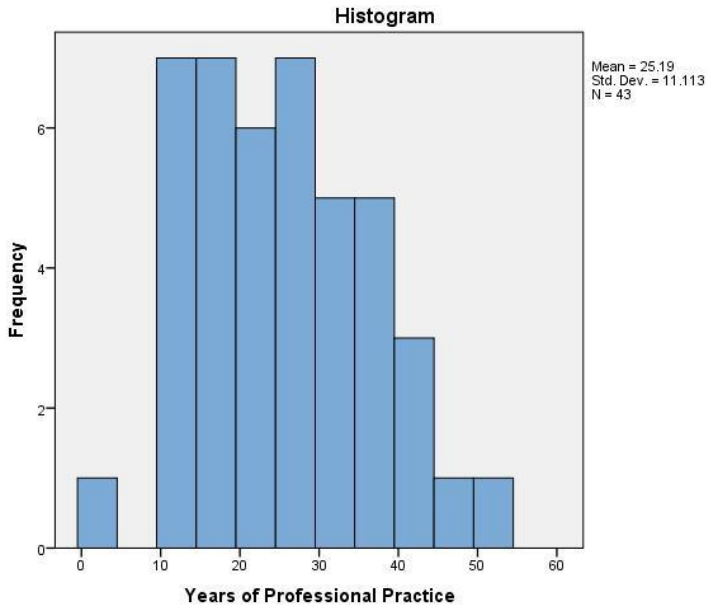


Figure 81. Histogram of the years of professional practice of the respondents (n=43).

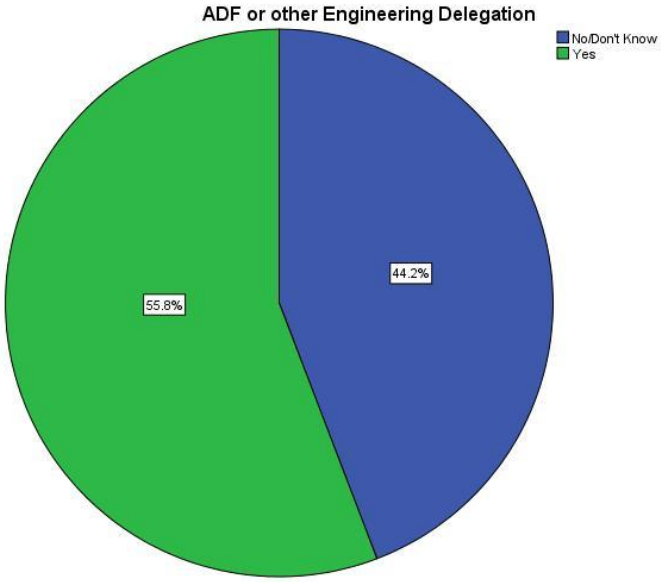


Figure 82. Pie chart of the ADF or other engineering authority delegation held by the respondents (n=43).

Figure 83 shows the system application domain where the respondents have their main experience. A majority of respondents have their main experience in the application domains of Maritime Defence Systems and Aerospace Systems. It is possible that this reflects the researcher's own professional network from which many respondents were invited and also that the area of technical regulation and technical integrity is of high

importance and is currently in active debate amid regulatory reform initiatives within the Australian Air Force and Navy particularly. The relatively small number of respondents, about 4.7%, in Commercial ICT may limit any generalisations that can be made across non-military application domains.

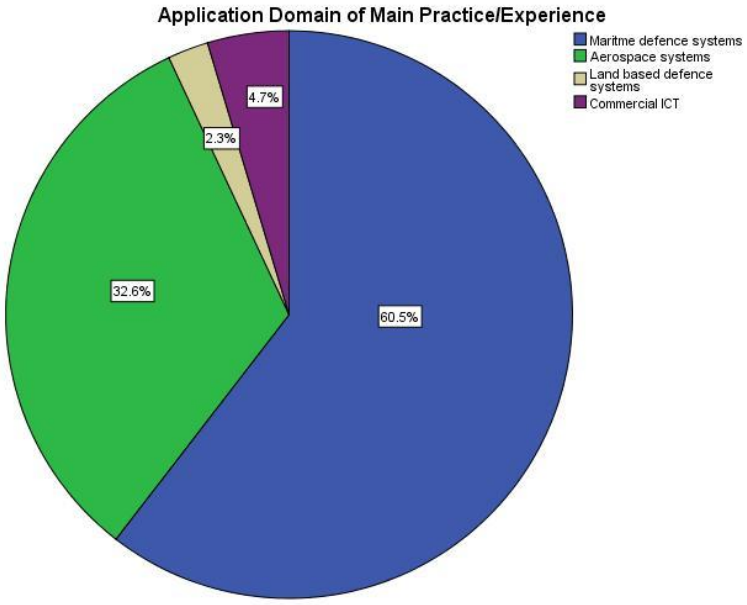


Figure 83. Pie chart of the system application domain where the respondent has most experience (n=43)..

Twenty four (55.8%) of the respondents indicated that they hold or have held formal delegations of Engineering Authority in organisations governed by ADF technical regulations or other regulatory systems. The delegations held are shown in Figure 84. This demonstrates that many of the respondents represent a group of very experienced engineers and that they are, or have previously been, entrusted to make significant engineering decisions for the ADF. There is a bias in representation toward Naval Technical Regulatory System (NTRS) delegations (62.5% of all indicated delegations) that is consistent with the application domain of most experience (60.5% of respondents with main experience in maritime defence systems).

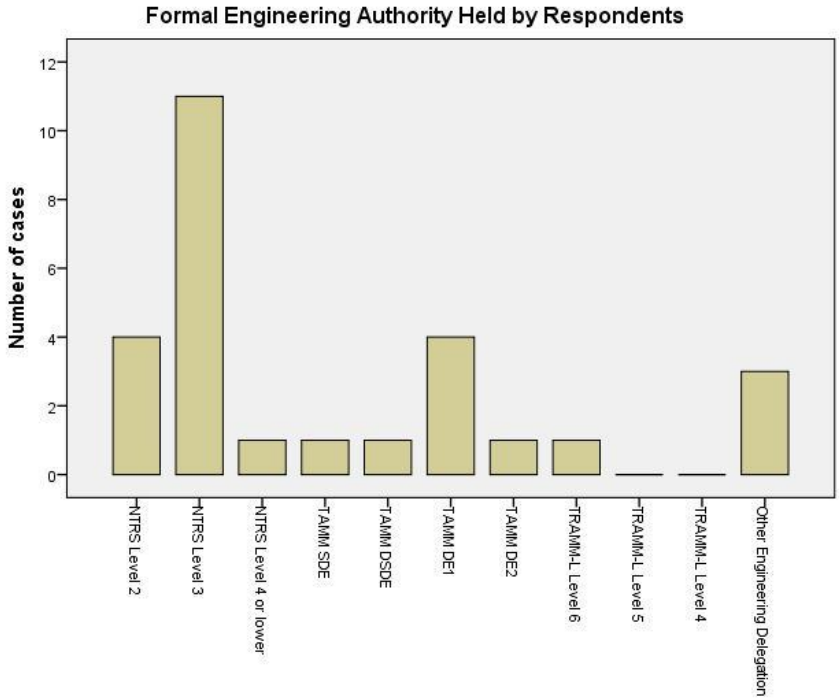


Figure 84. Bar chart of the engineering delegation held by respondents (n=43).

Figure 85 demonstrates that many of the group's experience with complex engineered systems is at what would be considered to be the higher levels of system abstraction: capability definition, system specification and design, system integration and test and project management. The relatively low numbers of respondents who cite experience with system operation, support and sustainment may cause a bias toward judgments being made with a development life-cycle perspective away from an in-service perspective.

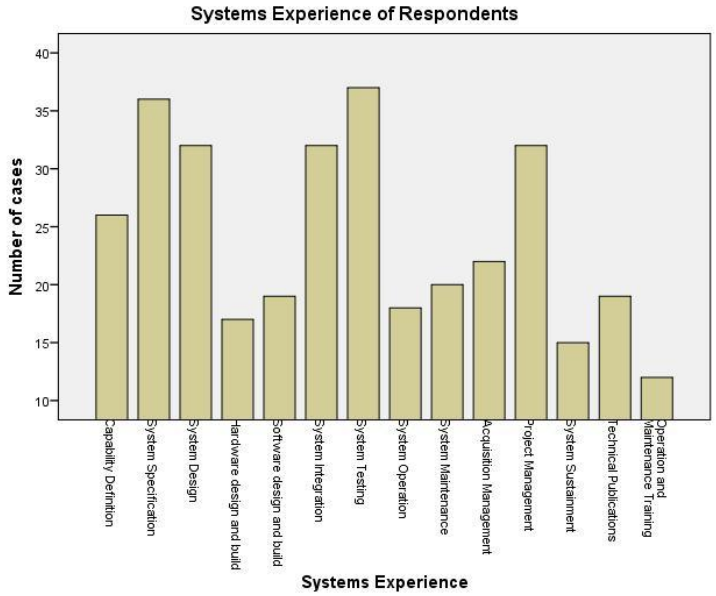


Figure 85. Bar chart of the system systems experience of the respondents (n=43).

8.2.5 Analysis of survey responses

8.2.5.1 Quantitative analysis procedure

The quantitative analysis procedure consists of two parts. The first is the statistical analysis of the ratings returned by the survey respondents. The second part uses quantitative data about the example systems to produce ratings that are compared to the ratings from the respondents to identify correspondence, or not, between the data sets.

8.2.5.1.1 Statistical analysis of ratings data

8.2.5.1.1.1 Non-parametric statistics

The rating scales in the survey instrument were intended to be interval and not to be taken to be a ratio scale. Therefore, while the calculated mean, standard deviation and normal distribution curves were useful to give a 'feel' for the data they could not be validly used for statistical inferences.

The assessment of the results therefore defaulted to the use of non-parametric statistics. The median was used for central tendency and range and interquartile range were used to ascertain the variability in the ratings. The Wilcoxon signed rank test was used to establish the significance in difference of median ratings.

Spearman's rho (r_S) was the appropriate non-parametric statistic for correlation. Note that Spearman's rho does not presume or predict a linear correlation, rather it predicts the extent of a monotonic relationship between two variables.

8.2.5.1.1.2 Interjudge reliability

To ensure comparability of the results of this research with other published results of the use of the CAT, two precedented approaches to interjudge reliability were adopted. (Amabile, 1996, p. 68) utilised several techniques to calculate interjudge reliability of subjective assessments. The second used the Spearman-Brown prediction formula:

$$reliability = \frac{n.r}{1 + (n-1).r}$$

Where *n* is the number of judges and *r* is the mean interjudge correlation.

This is consistent with the method described by (Wuensch, 2007) in using the SPSS intraclass correlation coefficient for *r*. SPSS provides the 'single measure intraclass correlation' which is 'an index of the reliability of the ratings for a typical, single judge' and an 'average measure intraclass correlation' which is 'the reliability for all the judges averaged together'.

Amabile also reported having used the Cronbach's alpha from SPSS for a measure of interjudge reliability and claimed it gave very similar results. Note that Cronbach's alpha gives an estimate of the internal consistency of the rating data. (George and Mallery, 2003, p. 231) provided the following rules of thumb for interpreting Cronbach's alpha:

- $\alpha > 0.9$ – Excellent,
- $\alpha > 0.8$ – Good,
- $\alpha > 0.7$ – Acceptable,
- $\alpha > 0.6$ – Questionable,
- $\alpha > 0.5$ – Poor, and
- $\alpha < 0.5$ – Unacceptable.

'While increasing the value of alpha is partially dependent upon the number of items in the scale, it should be noted that this has diminishing returns. It should also be noted that an alpha of 0.8 is probably a reasonable goal.' These interpretative labels and

Cronbach's alpha were adopted in the main for the results and discussion sections of this thesis.

8.2.5.1.1.3 Scale reliability

Assessing the reliability of a scale was conducted in accordance with the method of (Pallant, 2010, p. Ch. 9). This makes the assumption that a scale is constructed as a sum of the ratings of each of the judgment dimensions (or *items* as they are referred to in scale development).

The scale analysed was formed from the combined ratings for 20 items (the judgment dimension of technical integrity item was not included) across each of the three example system. This provided 125 valid cases for analysis. Cronbach's alpha was again the main measure of reliability produced by this procedure.

8.2.5.1.2 Correspondence of indicators with expert judgments

The use of "correspondence" was in accordance with its common definition: a close similarity, connection, or equivalence. No statistical significance is established.

The ranking of the three systems on each judgment dimension, based on significant differences in the median rating of the judges' rating data was established. Indicator models, based on the technical parameters of the system as introduced in section 8.2.2.1, produced numerical ratings from which rankings were evident. These two sources of rankings were compared and the correspondence between them was established and interpreted.

No precedented criterion for this type of correspondence analysis has been identified. For consistency within this thesis, the following levels of correspondence were defined and used:

- Good correspondence: where ranking of the systems established by statistically significant differences in the medians of judgment data corresponded with the rankings from the indicator model. If medians were not significantly different, good correspondence may still have been ascribed if the indicator model indicated rating differences were small (nominally less than 10%) and the ranking from the observed median corresponded with the indicator model.
- Reasonable correspondence: where the rankings provided by the observed medians corresponded with rankings from the indicator model. Equal ranks may have

occurred in the judgment based ranking, as long as no contradictions in the ranks existed. If the indicator model was sensitive to weighting changes of less than 10% then no higher correspondence was ascribed.

- Marginal correspondence: where at least the rank of the best or worst system corresponded and there was no correspondence in the rank of the other two systems.
- Poor correspondence: where there was no correspondence between the rankings from judgment data and the indicator model.

The comparison of the rankings was made visual by:

- Standardising the three observed median values from the judgment ratings,
- Standardising the indicator model ratings for the systems, and
- Plotting these together on a bar chart.

8.2.5.2 Qualitative analysis procedure

The qualitative analysis procedure used on the open ended survey responses was the same as that used on previous research phases. See section 7.2.2.2 for a description of the procedure.

8.3 Quantitative Results

8.3.1 Reliability of using the CAT for technical integrity measurement

This section presents results related to the research question posed in section 8.1.2: *Using the CAT, how reliable can direct expert ratings of technical integrity be?*

8.3.1.1 Results of technical integrity rating

The frequency of ratings for the judgment of the technical integrity of each of the three example systems was as illustrated in Figure 86, Figure 87 and Figure 88. Normal distribution curves are included for visualisation only. The descriptive statistics associated with these ratings were calculated and are at Table 47.

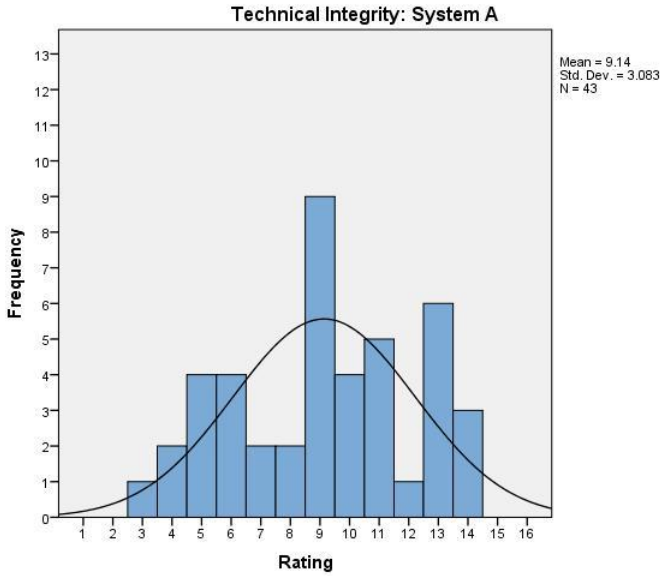


Figure 86. Frequency of technical integrity ratings for System A.

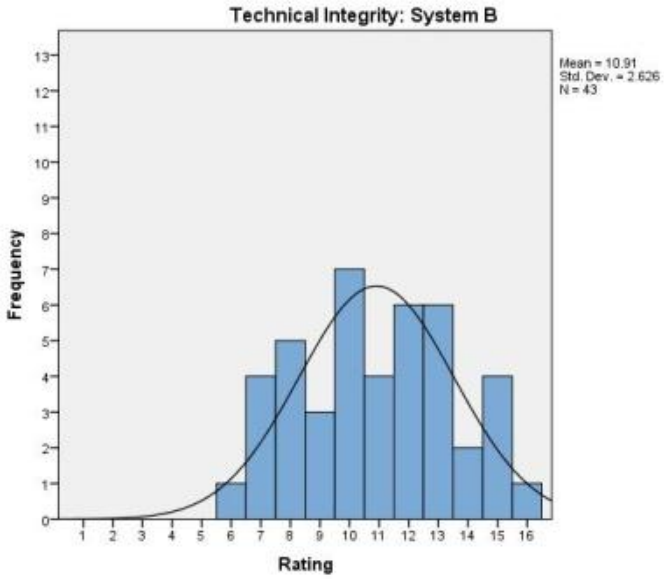


Figure 87. Frequency of technical integrity ratings for System B.

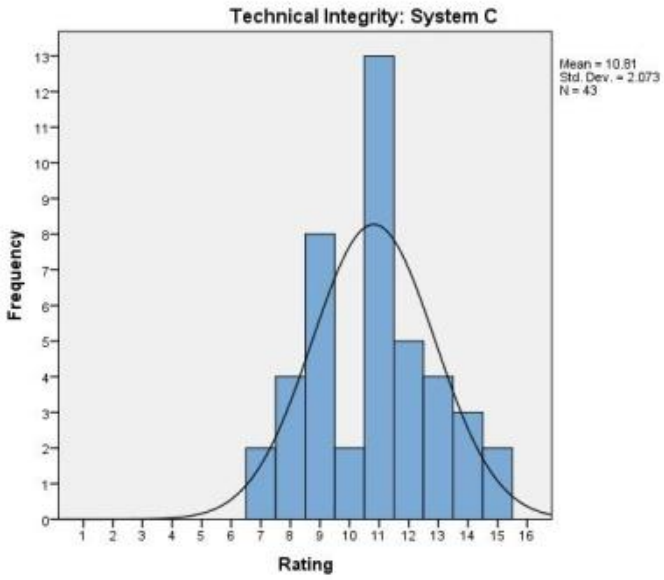


Figure 88. Frequency of technical integrity ratings for System C.

Table 47. Descriptive statistics for technical integrity ratings.

Statistics

		Technical Integrity: System A	Technical Integrity: System B	Technical Integrity: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		9.14	10.91	10.81
Median		9.00	11.00	11.00
Std. Deviation		3.083	2.626	2.073
Range		11	10	8
Minimum		3	6	7
Maximum		14	16	15
Percentiles	25	6.00	9.00	9.00
	50	9.00	11.00	11.00
	75	11.00	13.00	12.00

Figure 89 shows the median, range and inter-quartile range in a box-plot format for visualisation of the results.

Wilcoxon signed rank tests conducted at a significance level of 0.05 established that the median ratings for technical integrity of System A and System B are significantly different, the median ratings of System A and System C are different and the median ratings of System B and System C are not significantly different (the result for the latter is shown in Figure 90). It is concluded from this that the consensus of judges is that the technical integrity of System B and System C are equivalent and exhibit higher

technical integrity than that of System A. It is also noted that the variability as estimated by interquartile range is least for System C and most for System A.

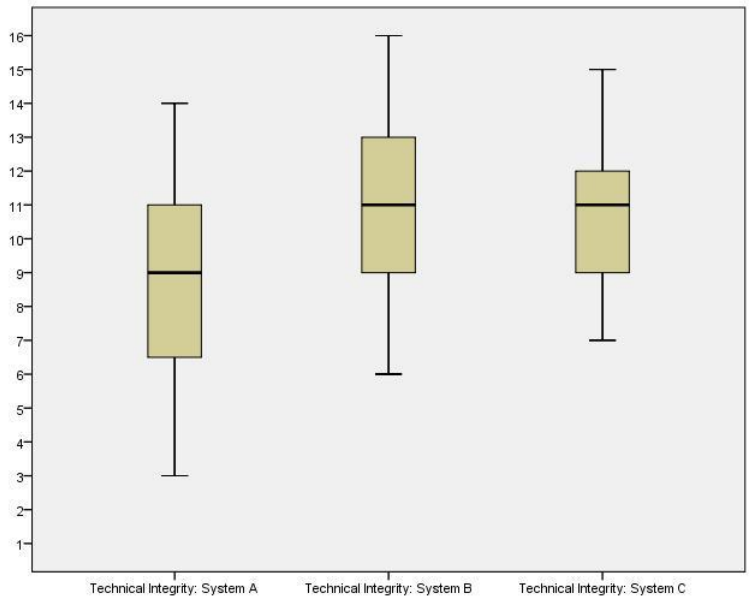
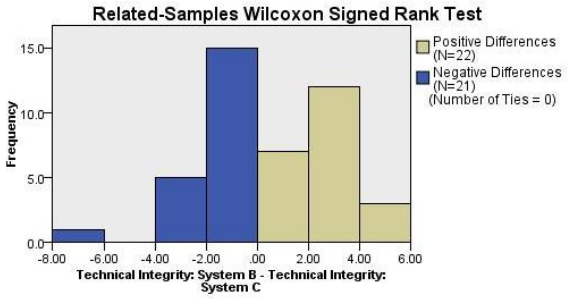


Figure 89. Median, interquartile range and range of technical integrity ratings.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The median of differences between Technical Integrity: System C and Technical Integrity: System B equals 0.	Related-Samples Wilcoxon Signed Rank Test	.692	Retain the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.



Total N	43
Test Statistic	505.500
Standard Error	81.952
Standardized Test Statistic	.397
Asymptotic Sig. (2-sided test)	.692

Figure 90. Result of Wilcoxon signed rank test: technical integrity of System B minus System C.

8.3.1.2 Reliability of technical integrity expert rating

8.3.1.2.1 Interjudge reliability

The intraclass correlation coefficients produced from SPSS for the ratings of technical integrity are provided at Figure 91:

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.113[a]	.021	.850	8.683	2	84	.000
Average Measures	.845[c]	.476	.996	8.683	2	84	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Figure 91. Intraclass correlation coefficients for technical integrity ratings.

Using these intraclass correlation coefficients and *n*=43, the reliability equation yields:

- *reliability* = 0.996 (for an average of judges)
- *reliability* = 0.846 (for a typical judge)

The SPSS provided value for Cronbach's alpha is shown in Figure 92.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.885	.855	43

Figure 92. Cronbach's alpha for technical integrity ratings.

8.3.1.2.2 Characteristics of ratings

8.3.1.2.2.1 Pattern of judge's ratings

Figure 93 shows a scattergram of the mean value of each of the judge's three technical integrity ratings (for Systems A, B and C) versus the range of the same ratings. Each point on the scattergram represents the mean and range of the each judge's 3 ratings of technical integrity. The range of average values was from 7.0 to 14.7 indicating some judges rated the systems much more highly than others. It was also observed that the more optimistic raters (higher mean rating) tended to utilise a lower range of ratings (i.e. all 3 Systems are relatively highly rated). This may just have been an artefact of range compression at the top end of the scale.

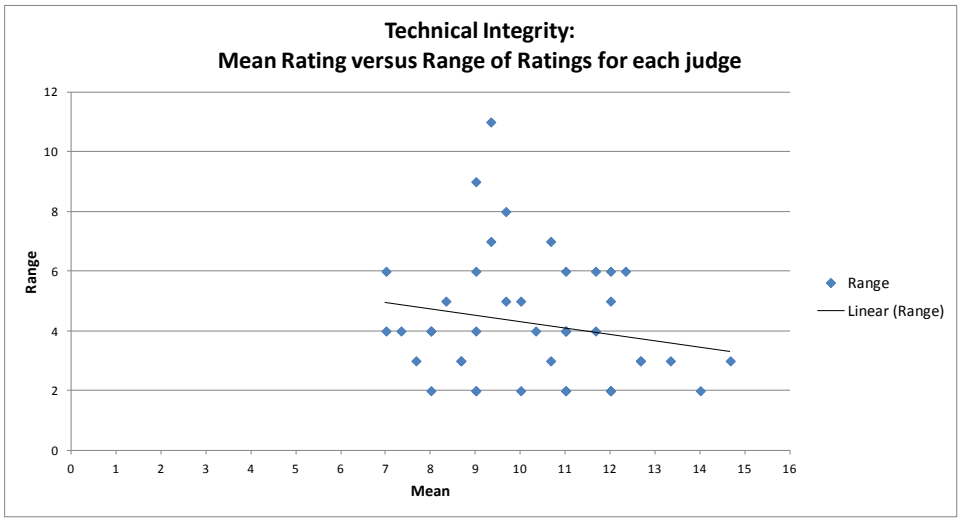


Figure 93. Scattergram of average of 3 ratings versus range of 3 ratings.

8.3.1.2.2.2 Judgment variation across demographic categories

In order to see if there was any correlation with any of the demographic data to explain observed ratings, a nonparametric correlation using Spearman's rho was conducted.

The mean of the three technical integrity ratings per judge was compared via nonparametric correlation to the demographic data comprising: highest level of technical education, gender, age category, years of professional practice, application domain of main practice/experience, current employment sector, sector of most experience, ADF or other engineering delegation and the 14 experience categories. No statistically significant correlations existed. This is consistent with other CAT studies that indicate that judgments from experts and non-experts can be equally reliable (Amabile, 1996, p. 72-73).

8.3.1.3 Correlation of technical integrity rating with rating of other indicators

The judges' ratings for all 3 systems across all 20 dimensions were analysed for correlation to the ratings of technical integrity. Using Spearman's rho (r_S) as the appropriate non-parametric statistic for correlation, the results are as shown in Figure 94. All correlations were statistically significant at the 0.01 level (2-tailed) with the single exception of Environmental Impact which was significant at the 0.05 level (2-tailed).

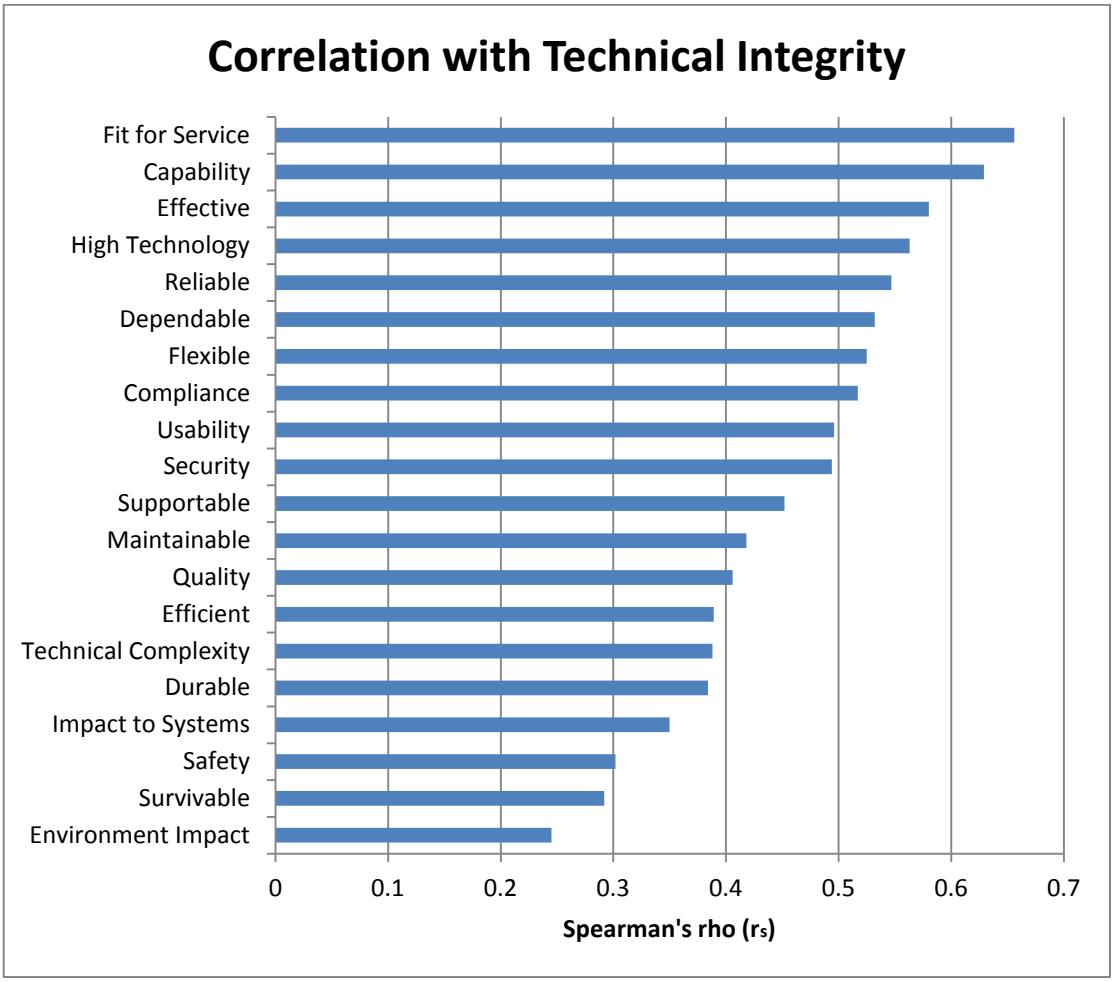


Figure 94. Correlation of judgment dimensions with technical integrity.

The strongest correlation was for Fit for Service at $r_S = 0.656$. A scatter plot with a linear regression line to assist visualisation of the relationship is shown in Figure 95 (i). The weakest correlation was for Impact to the Natural Environment at $r_S = 0.245$. The correlation of Safety was also, surprisingly, weak at $r_S = 0.302$ as shown at in Figure 95 (ii).

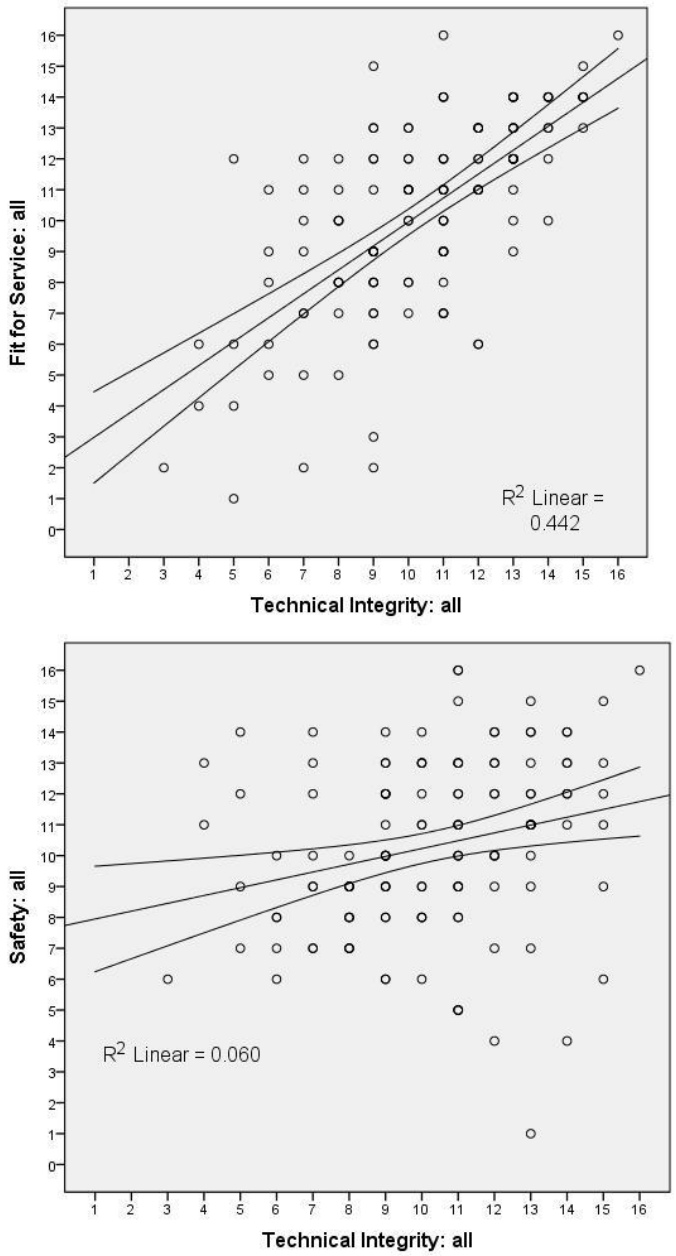


Figure 95. Scattergram of ratings for (i) Fitness for service versus technical integrity ratings; (ii) Safety versus technical integrity ratings.

8.3.2 Reliability of assessing indicators of technical integrity by the CAT

Following the detailed examination of the reliability of CAT judgments for technical integrity, this section presents results related to the corollary research question posed in section 8.1.2: *How reliable can expert ratings for other system quality attributes be?*

Reliability was calculated for each judgment dimension (as listed in section 8.2.1.2), each dimension comprising 3 system cases across all 43 judges, except for the

dimension of technical complexity for which there were 42 complete judgments. Cronbach's alpha was used as the statistic for reliability and was calculated and interpreted as per section 8.2.5.1.1.2.

Table 48 presents the results of this calculation with the judgment dimensions rated from highest to lowest reliability. It was noted that the reliability of Technical Integrity was "good" and the reliability of twelve related judgment dimensions were "excellent". This indicated that a more reliable result for technical integrity may arise from a scale formed from these judgment dimensions.

Table 48. Reliability of ratings for system quality attributes/judgment dimensions.

#	Dimension	alpha	Interpretation
13	Reliable	0.986	Excellent
20	High Technology	0.983	Excellent
16	Compliance	0.976	Excellent
15	Maintainable	0.973	Excellent
17	Usability	0.973	Excellent
2	Environment Impact	0.963	Excellent
12	Dependable	0.959	Excellent
19	Technical Complexity	0.947	Excellent
6	Effective	0.945	Excellent
14	Supportable	0.938	Excellent
11	Survivable	0.934	Excellent
7	Efficient	0.920	Excellent
18	Security	0.898	Good
	Technical Integrity	0.885	Good
8	Quality	0.847	Good
5	Capability	0.823	Good

#	Dimension	alpha	Interpretation
3	Impact to Systems	0.801	Good
4	Fit for Service	0.790	Acceptable
1	Safety	0.685	Questionable
9	Flexible	0.519	Poor
10	Durable	0.226	Unacceptable

8.3.3 Technical integrity scale reliability

Following the evaluation of the reliability of CAT judgments for indicators of technical integrity, this section presents results related to forming a scale from these judgment dimensions. In particular, the research question posed in section 8.1.2: *Can the judgments of the quality attributes form a reliable scale for technical integrity?*

Reliability analysis of the scale was conducted using SPSS v20. The scale analysed was formed from the combined ratings for 20 items (the judgment dimension of technical integrity was not included) across each of the three example system. This provided 125 valid cases for analysis. A scale item in the reliability analysis is synonymous with the judgment dimensions. The following interpretation of results was conducted in accordance with the method of (Pallant, 2010, p. Ch. 9).

The values of the Inter-Item Correlation Matrix were positive which indicated that all items were measuring the same underlying characteristic. Cronbach's alpha, which provides an indication as to how well each item correlates with the total score, was calculated to be 0.917 which suggested very good internal consistency for the scale across these three systems.

The other statistics produced for the scale are shown in Table 49. The 'Corrected Item-Total Correlation' shown in the fourth column indicates the degree to which each item correlates with the total score. All are greater than 0.3, which indicated they are measuring the same underlying property as the rest of the scale. The last column, 'Cronbach's Alpha if Item Deleted', provides an indication of the impact of removing an item from the scale. Removal of 'Environment Impact' would result in improved scale reliability (α would increase to 0.920) and removal of 'Survivable' would not affect scale reliability (α would remain at 0.917).

Table 49. Technical integrity scale: item-total statistics.

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
Safety: all	191.38	1465.011	.415	.340	.916
Environment Impact: all	192.42	1469.795	.309	.289	.920
Impact to Systems: all	192.58	1428.262	.471	.342	.916
Fit for Service: all	191.55	1405.653	.612	.598	.912
Capability: all	191.38	1395.704	.718	.676	.910
Effective: all	191.42	1391.391	.776	.697	.909
Efficient: all	191.80	1420.935	.531	.505	.914
Quality: all	191.33	1446.819	.505	.537	.915
Flexible: all	191.55	1422.185	.627	.532	.912
Durable: all	191.29	1427.578	.555	.560	.914
Survivable: all	191.42	1451.455	.410	.490	.917
Dependable: all	191.62	1397.237	.731	.657	.910
Reliable: all	191.33	1391.674	.696	.780	.910
Supportable: all	191.61	1408.547	.566	.609	.913
Maintainable: all	191.78	1409.477	.557	.650	.914
Compliance: all	191.93	1428.825	.512	.616	.915
Usability: all	191.30	1395.955	.709	.647	.910
Security: all	191.85	1396.630	.554	.496	.914
Technical Complexity: all	192.02	1425.314	.568	.481	.913
High Technology: all	191.58	1362.827	.712	.726	.910

The scale reliability analysis was repeated for the available ratings for each of the three systems individually. Cronbach's Alpha ranged from 0.906 to 0.931which suggested very good internal consistency for the scale across each of these three systems.

Table 50 presents, 'Cronbach's Alpha if Item Deleted' for the reliability analysis for all systems taken together and for each system individually. Highlighted in red and underlined are the values for an item whose removal would increase the Cronbach's Alpha. Removal of 'Environment Impact' (0.920) would result in improved scale reliability in all cases and 'Impact to Systems' would improve reliability considered for Systems B and C. 'Compliance' and 'Security' if removed would increase reliability for one of the systems. Highlighted in yellow and dotted underlined are the values for an item whose removal would not affect the Cronbach's Alpha. This allows a consideration of removal of 'Survivable' from the scale.

Some of these results are not as expected. In particular impact to the natural environment is one of the ADF's explicitly identified aspects of technical integrity, but in the 3 example systems at least, it does not provide any significant contribution to the assessment of technical integrity. Similarly 'Impact to Systems', 'Compliance' and 'Security' would be expected to contribute significantly but do not in this study.

As indicated for System B at least, it would be expected that 'Technical Complexity' could be left from the scale since it is not theorised to contribute to technical integrity, but it would also be expected that 'High Technology' could be similarly left from the scale.

Table 50. Technical integrity scale: scale alpha if items deleted.

	All Systems	System A	System B	System C
Cronbach's Alpha	**.917**	**.931**	**.919**	**.906**
ITEMS:	**Cronbach's alpha if item deleted**			
Safety	.916	.930	.914	.905
Environment Impact	.920	.934	.921	.908
Impact to Systems	.916	.927	.920	.909
Fit for Service	.912	.926	.916	.900
Capability	.910	.924	.914	.895
Effective	.909	.924	.912	.897
Efficient	.914	.928	.912	.907
Quality	.915	.928	.909	.899
Flexible	.912	.927	.915	.898
Durable	.914	.930	.913	.901
Survivable	.917	.929	.913	.898
Dependable	.910	.925	.915	.898
Reliable	.910	.925	.916	.899
Supportable	.913	.925	.916	.901
Maintainable	.914	.925	.913	.900
Compliance	.915	.928	.920	.903
Usability	.910	.927	.913	.899
Security	.914	.935	.912	.900
Technical Complexity	.913	.926	.919	.904
High Technology	.910	.924	.914	.900

Note the previous result for directly rating technical integrity was $\alpha = 0.885$ (refer Figure 92). The values for Cronbach's alpha shown in Table 50 implied a better (more reliable) estimate is obtained by combining the item ratings to form a scale for technical integrity. But the theoretical concept developed in Chapter 5 did not predict a direct additive combination of these items. Indeed the result in Chapter 6, confirmed construct validity for a multiplicative conjoint aggregation of the concepts of capability and dependability as a measurement construct for technical integrity.

8.3.4 Correspondence of indicators with expert judgments

This section presents results related to the research questions posed in section 8.1.2: *Where indicator models of the system quality attributes are available, how well do the models and the corresponding expert ratings of judgment dimensions correspond?* and *What is the best measurement instrument to be used for each indicators (via objective variables or via expert judgment)?*

Indicator models were developed for each indicator based upon both objective variables and values and subjective ranking data for the three example systems. The method of generating indicator models and achieving a ranking for the systems based on the technical parameters of the system was introduced in section 8.2.2.1. In this section, the indicator models and the resultant ratings of the systems from these models is presented. The ranking of the three systems on each CAT judgment dimension, based on significant differences in the median rating of the judges' rating data was established and these are presented for each dimension. Results from establishing the extent of correspondence of the indicator model rankings with the rankings established from expert judgments gained through the CAT are presented and discussed in this section.

8.3.4.1 Safety

As indicated in Table 46, it was expected that the "Safety" indicator would correlate primarily with quality variables for hazardous materials (HAZMAT), the antenna safety switch, and electrical power interlocks. The indicator model and resultant ratings are shown in Figure 96. Inspection of the sensitivity of the model to changes of the weightings shows that the most sensitive parameter is Electrical power interlock which would need to change by 17% to alter the ranking of the systems. The minimum difference in ratings is 7% between System A and System C.

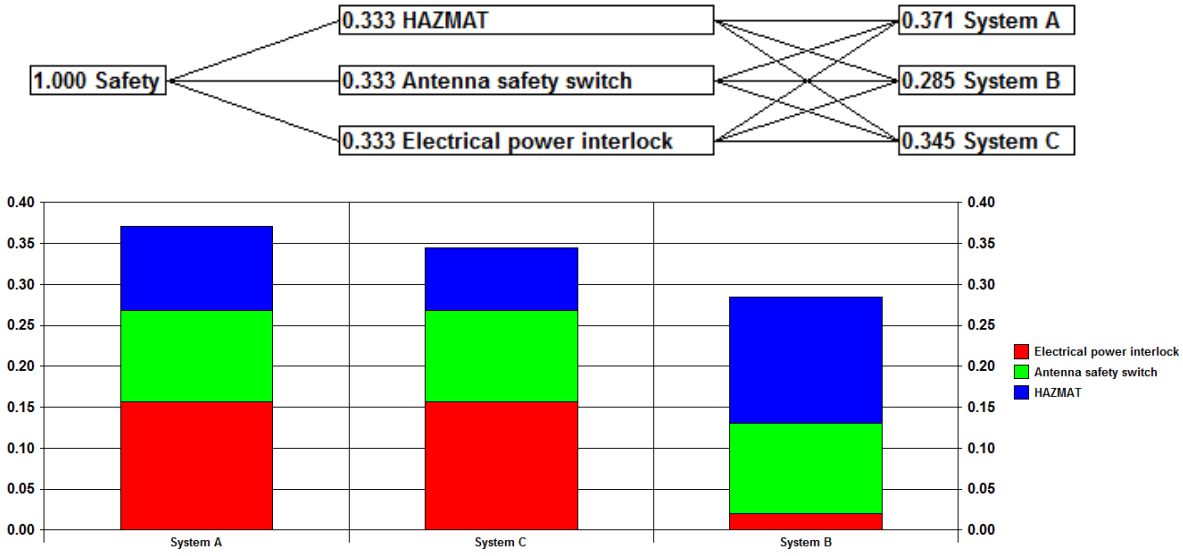


Figure 96. Indicator model and ratings for 'Safety'.

Results from the judgment rating data for "Safety" are presented in the statistics summary at Table 51 and graphically in the boxplot at Figure 97. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test[12] established that the medians of the judges' ratings are significantly different for each pair of systems, except for System A and System C where the difference is not significant.

[12] Wilcoxon signed rank test conducted using the SPSS v20 Non-parametric tests: two or more related samples. All hypothesis testing results for medians in this section are quoted at a significance level of 0.05 unless otherwise stated.

Table 51. Descriptive statistics for judgments of ‘Safety’.

Statistics

		Safety: System A	Safety: System B	Safety: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		10.65	9.70	10.56
Median		11.00	9.00	10.00
Std. Deviation		2.776	2.816	2.864
Range		15	12	11
Minimum		1	4	4
Maximum		16	16	15
Percentiles	25	9.00	7.00	9.00
	50	11.00	9.00	10.00
	75	12.00	12.00	13.00

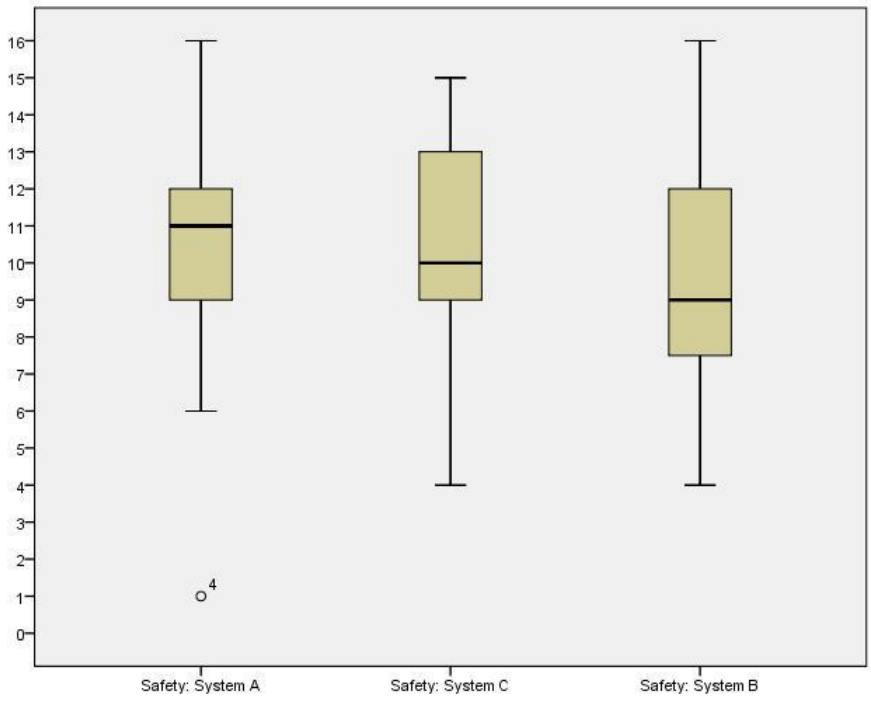


Figure 97. Boxplots for judgments of ‘Safety’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 98. Since the difference in the median of judgment ratings between System A and System C was not significant, and the indicator model rating difference was 7%, there is a **good** correspondence in the rankings of the three systems between the indicator model of safety and the consensus view of the judges.

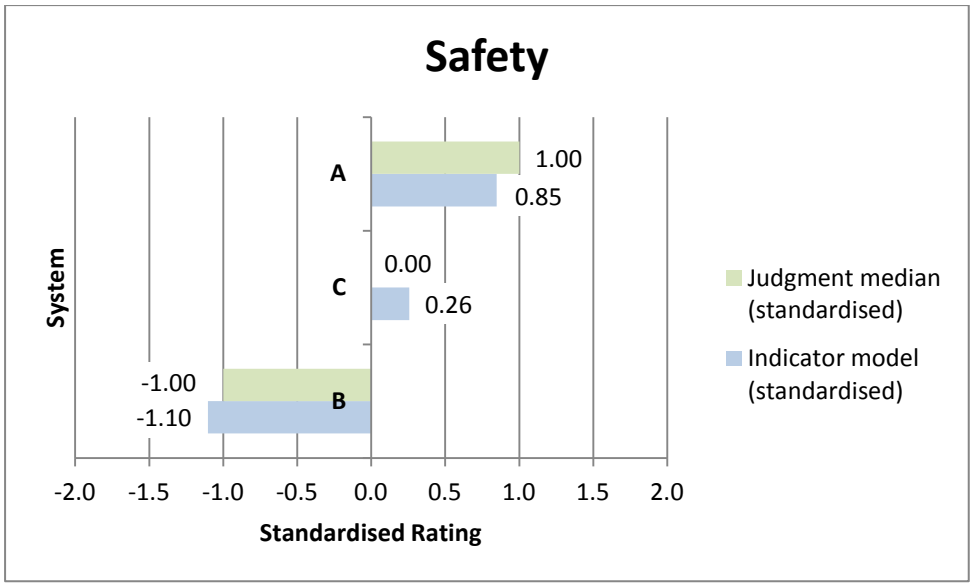


Figure 98. Correspondence of ratings for ‘Safety’.

8.3.4.2 Impact to the natural environment

As indicated in Table 46 it was expected that the variable 'Environment impact' would not primarily correlate with any technical parameters. It was considered that HAZMAT and output radiated power may have some influence on the judgments of this indicator. The indicator model and resultant ratings are shown in Figure 99. Inspection of the sensitivity of the model to changes of the weightings shows any parameter would need to change by more than 82% to alter the ranking of the systems. The minimum difference in ratings is 34% between System A and System C.

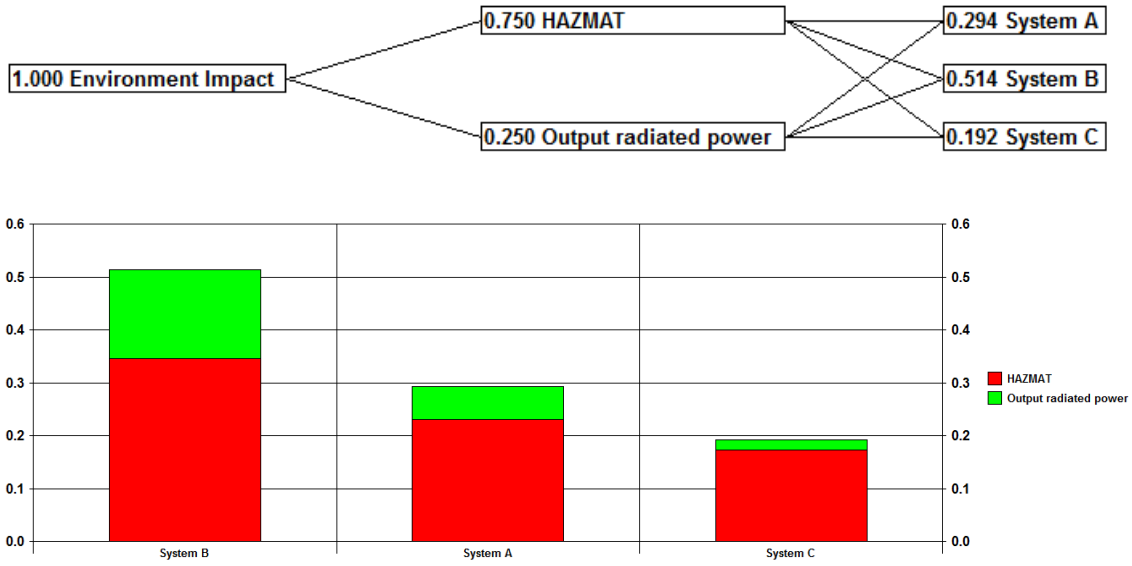


Figure 99. Indicator model and ratings for 'Impact to the natural environment'.

Results from the judgment rating data for 'Environment impact' are presented in the statistics summary at Table 52 and graphically in the boxplot at Figure 100. The variables on the boxplot have been presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are significantly different for each pair of systems, except for System A and System C where there is no observed difference in the medians.

Table 52. Descriptive statistics for ratings of ‘Environment impact’.

Statistics

		Environment Impact: System A	Environment Impact: System B	Environment Impact: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		8.79	10.98	8.09
Median		8.00	11.00	8.00
Std. Deviation		3.196	3.211	3.243
Range		12	14	13
Minimum		3	2	2
Maximum		15	16	15
Percentiles	25	7.00	9.00	6.00
	50	8.00	11.00	8.00
	75	12.00	13.00	10.00

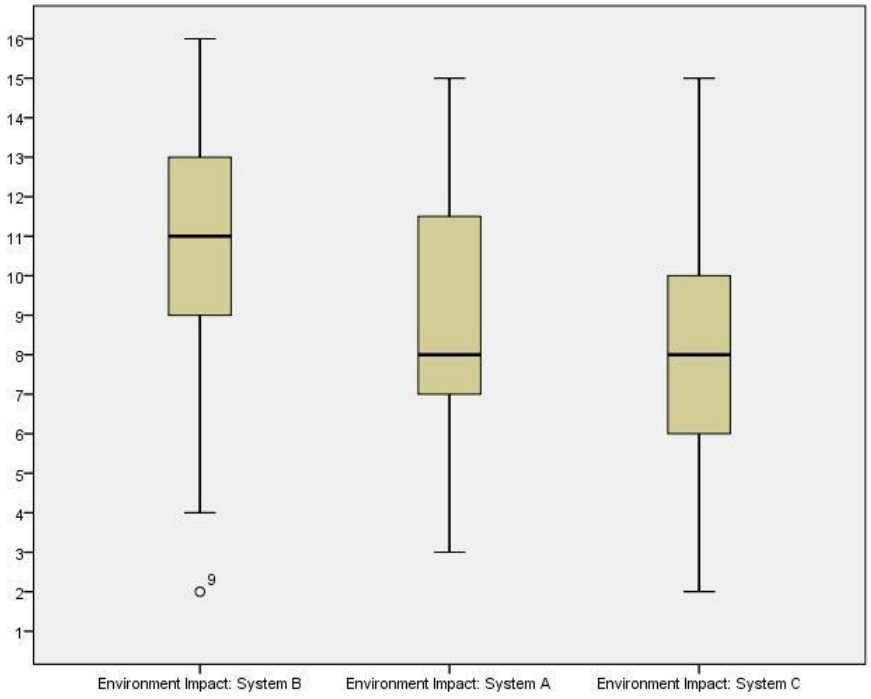


Figure 100. Boxplots for ratings of ‘Environment impact’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 101. Given there was no observed difference in the median of judgment ratings between System A and System C, but there was a significant difference in the indicator model ratings, and that the technical parameters modelled were only considered to be of secondary importance to the indicator, there is a **marginal** correspondence apparent in the rankings of the three systems between the

indicator model of impact to the natural environment and the consensus view of the judges.

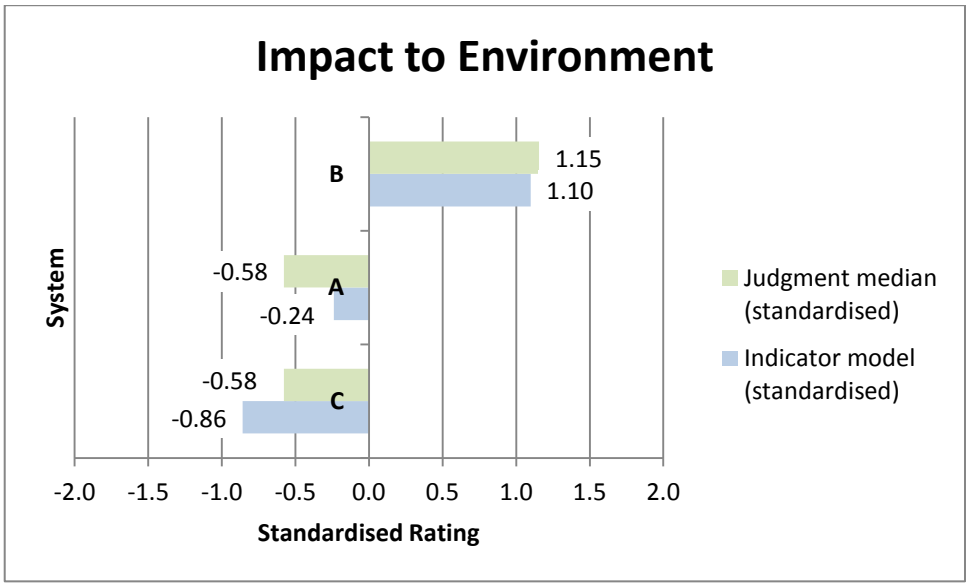


Figure 101. Correspondence of ratings for 'Impact to the natural environment'.

8.3.4.3 Impact to other systems

As indicated in Table 46 it was expected that the 'Impact to systems' indicator would correlate primarily with the quality variables EMC compliance, electrical power standard and secondarily with the quantity variable output radiated power. The indicator model and resultant ratings are shown in Figure 102. Inspection of the sensitivity of the model to changes of the weightings shows that it is sensitive to the weighting of parameters, EMC compliance and electrical power standard, either of which would need to change by 5.4% or more to alter the ranking of the systems. The minimum difference in ratings is 7.9% between System A and System C.

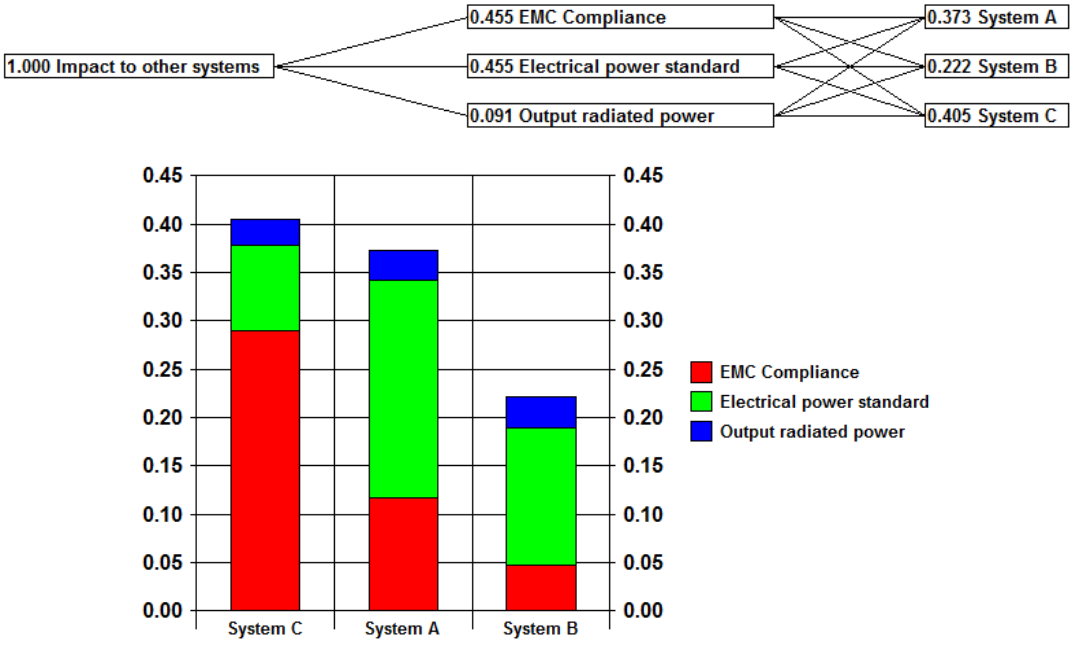


Figure 102. Indicator model and ratings for 'Impact to other systems'.

Results from the judgment rating data for "Impact to other Systems" are presented in the statistics summary at Table 53 and graphically in the boxplot at Figure 103. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are only significantly different for System A and System B.

Table 53. Descriptive statistics for ratings of ‘Impact to systems’.

Statistics

		Impact to Systems: System A	Impact to Systems: System B	Impact to Systems: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		8.28	9.91	9.26
Median		8.00	11.00	9.00
Std. Deviation		3.172	3.435	3.600
Range		11	12	13
Minimum		3	4	2
Maximum		14	16	15
Percentiles	25	5.00	7.00	6.00
	50	8.00	11.00	9.00
	75	10.00	12.00	12.00

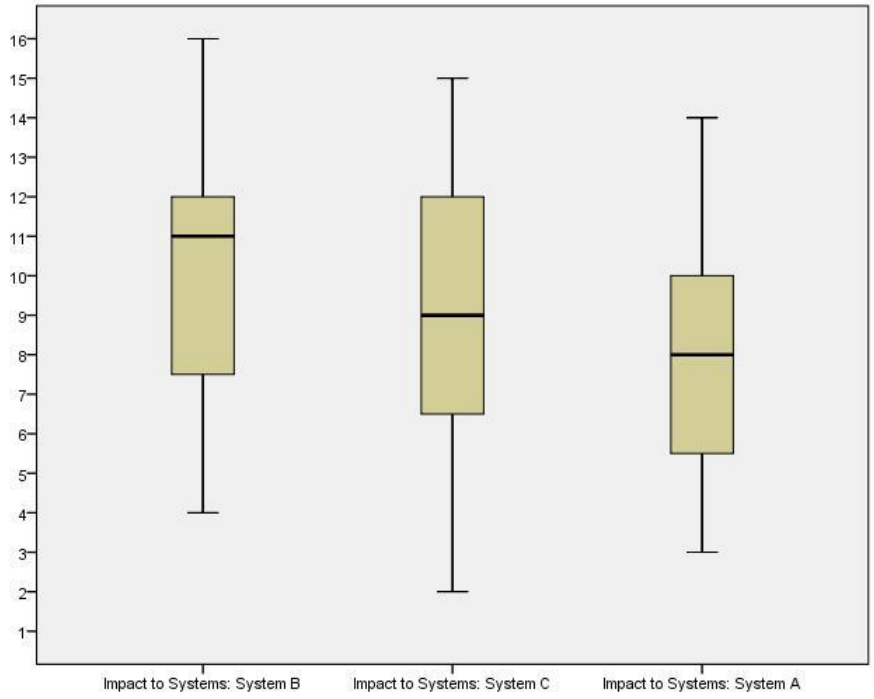


Figure 103. Boxplots for ratings of ‘Impact to Systems’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 98. There is a **poor** correspondence in the rankings of the three systems between the indicator model of ‘impact to other systems’ and the consensus view of the judges.

The judges’ ratings establish that the consensus view is the impact of other systems of System B is less than that of System A. System C is indeterminate between the other

systems. This is not consistent with the indicator model which itself was sensitive to the parameters available. This result implies that the judges were not primarily influenced by the indicators of EMC compliance and Electrical power standard as theorised.

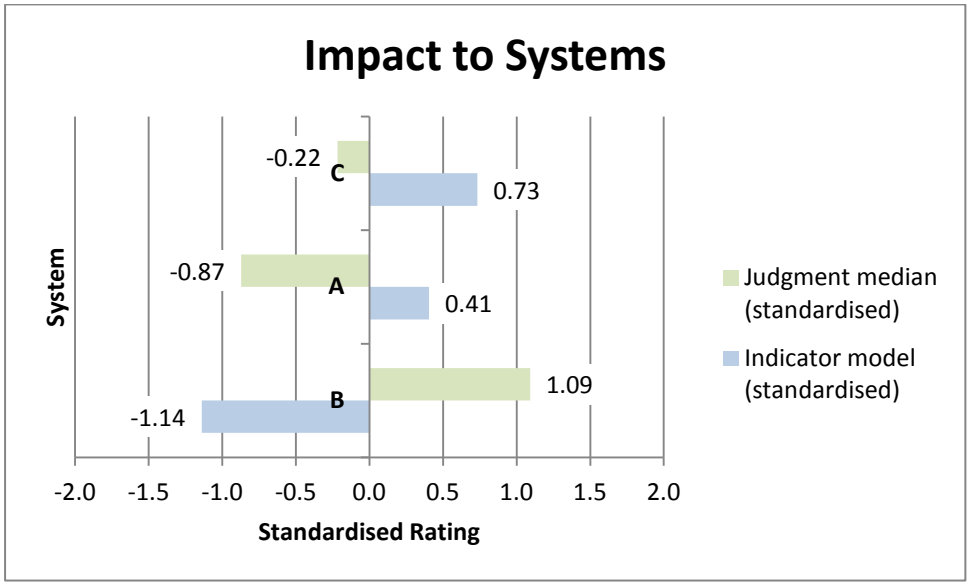


Figure 104. Correspondence of ratings for 'Impact to other systems'.

8.3.4.4 Fit for Service

As indicated Table 46, it was expected that the "Fit for Service" indicator would correlate primarily with quantity variable compliance, and secondarily with environmental resistance standards. The indicator model and resultant ratings are shown in Figure 105. Inspection of the sensitivity of the model shows that it is sensitive to the weighting of both parameters, either of which would need to change by 3.2% or more to alter the ranking of the systems. The minimum difference in ratings is 1.8% between System A and System C.

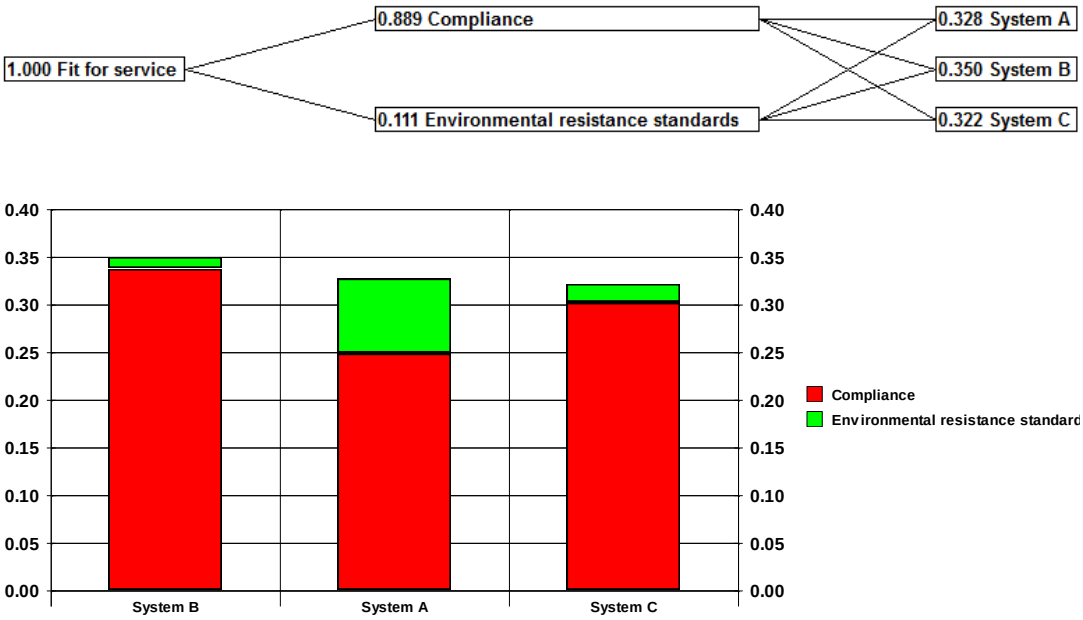


Figure 105. Indicator model and ratings for 'Fit for service'.

Results from the judgment rating data for "Fit for Service" are presented in the statistics summary at Table 54 and graphically in the boxplot at Figure 106. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are only significantly different for System A and System B.

Table 54. Descriptive statistics for ratings of ‘Fit for service’.

Statistics

		Fit for Service: System A	Fit for Service: System B	Fit for Service: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		9.28	11.00	10.26
Median		10.00	12.00	11.00
Std. Deviation		3.239	3.288	2.829
Range		14	14	13
Minimum		1	2	2
Maximum		15	16	15
Percentiles	25	8.00	9.00	8.00
	50	10.00	12.00	11.00
	75	12.00	14.00	12.00

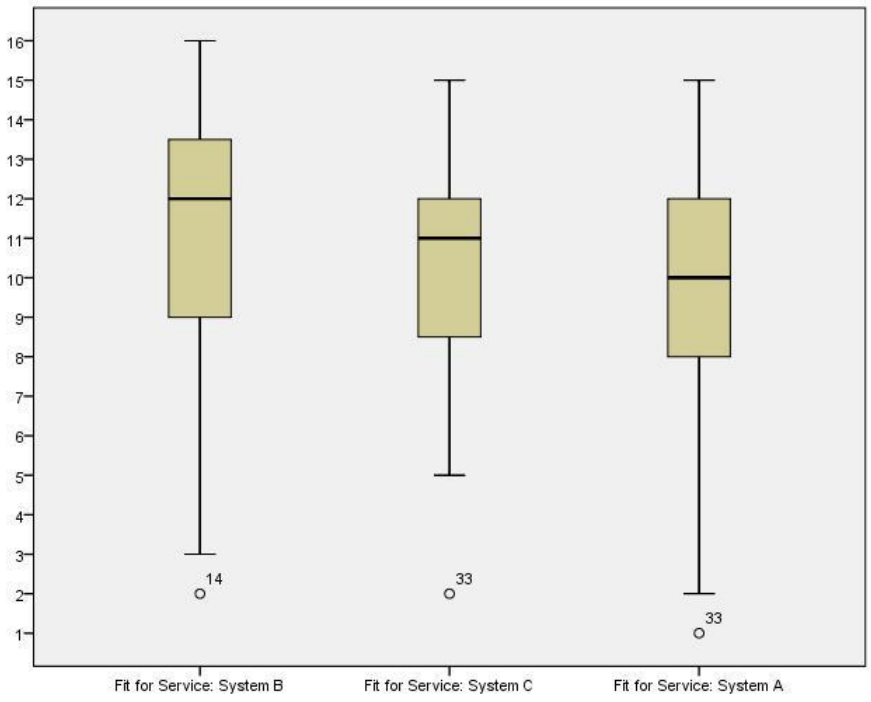


Figure 106. Boxplots for ratings of ‘Fit for Service’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 107. Only the rank of System B as the highest showed correspondence. This result rated as **marginal** correspondence.

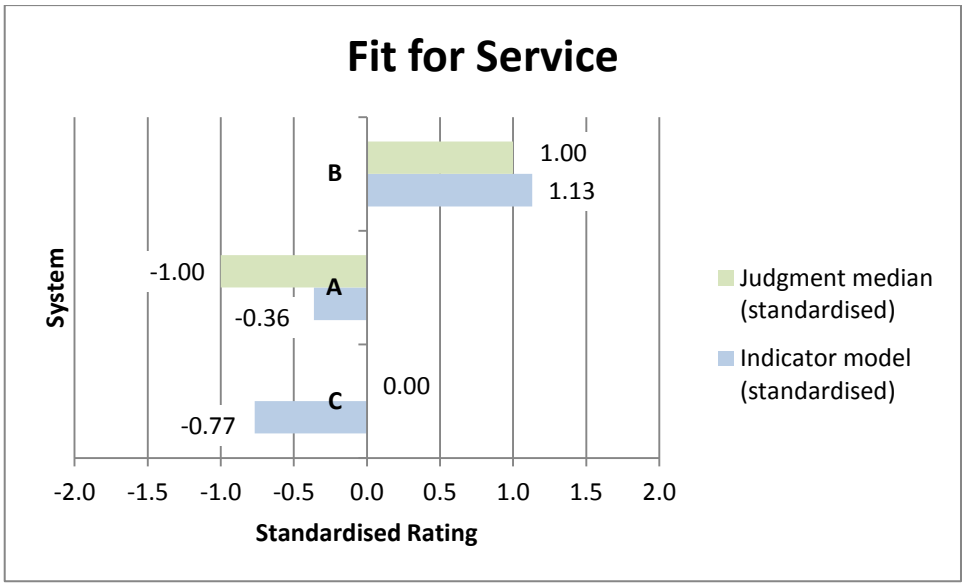


Figure 107. Correspondence of ratings for 'Fit for Service'.

Inspection of Figure 105 indicated that the inclusion of environmental resistance standards, theorised to be a secondary technical parameter, led to the sensitivity of the indicator model and also to the relative ranking of System A and System C. Elimination of environmental resistance standards from the indicator model would improve the correspondence. The revised indicator model is shown at Figure 108. The minimum difference in ratings is 11% between System B and System C. The revised correspondence plot is at Figure **109**. This revised model and the judgments resulted in **good** correspondence. This implies the judges relied upon measures of requirements compliance and perhaps did not include a consideration of environmental resistance standards.

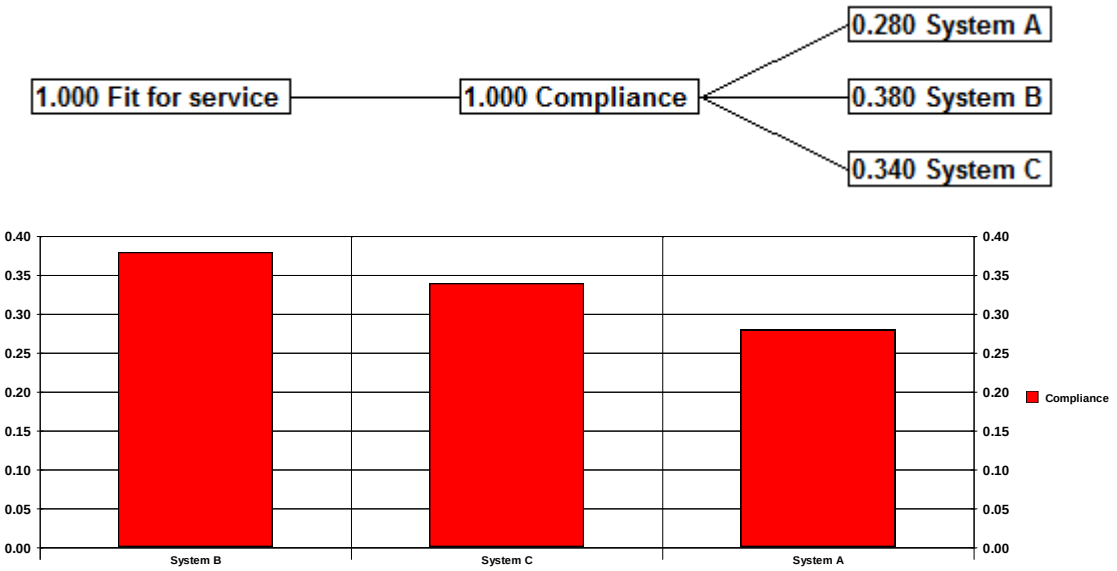


Figure 108. Post facto indicator model and ratings for 'Fit for service'.

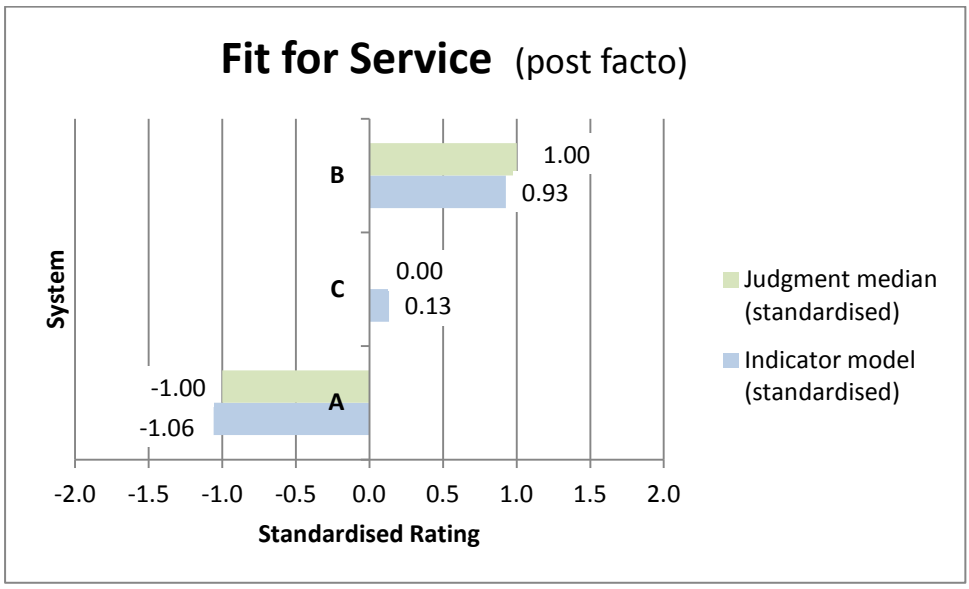


Figure 109. Correspondence of ratings for 'Fit for Service' with post facto indicator model

8.3.4.5 Capability

As indicated in Table 46, it was expected that the "Capability" indicator would correlate primarily with quantity variables for range, range accuracy, bearing accuracy, data rate, and automatic tracking capacity. The indicator model and resultant ratings are shown in Figure 110. Inspection of the sensitivity of the model to changes of the weightings shows that the most sensitive parameter is automatic tracking capacity which would need to change by 6% to alter the ranking of the systems. The minimum difference in ratings is 12.8% between System B and System C.

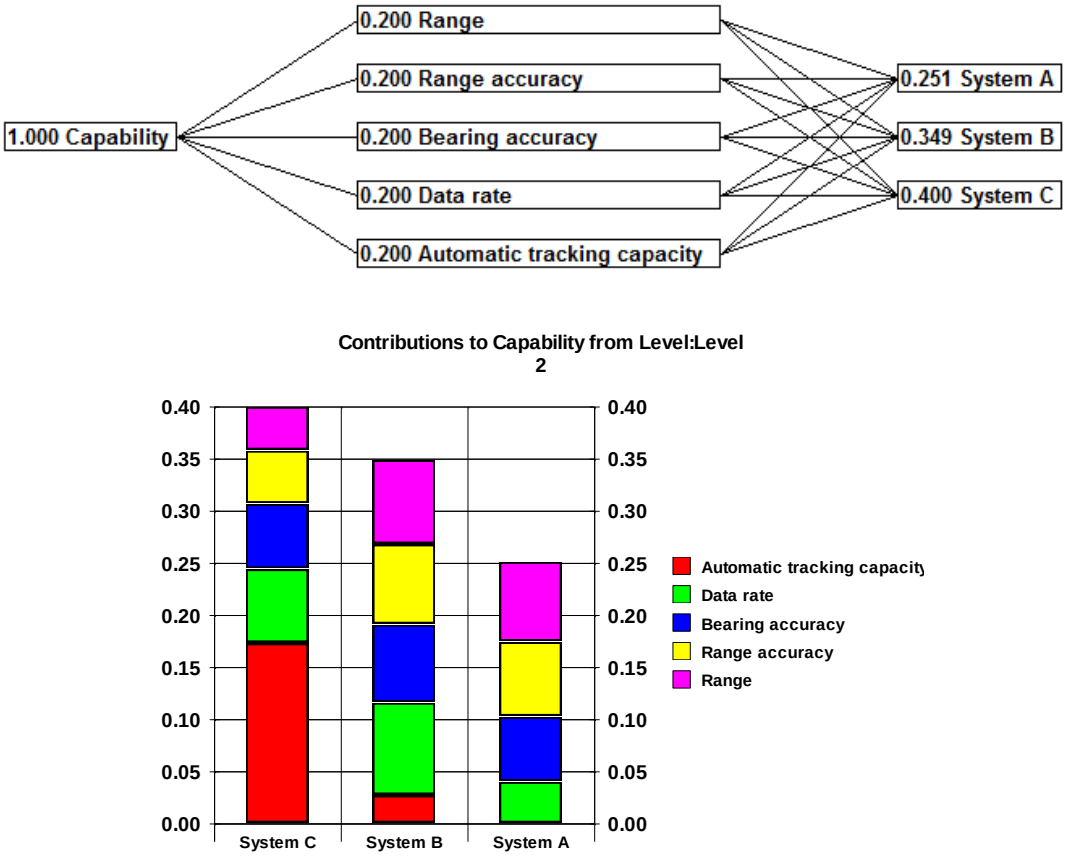


Figure 110. Indicator model and ratings for 'Capability'.

Results from the judgment rating data for "Capability" are presented in the statistics summary at Table 55 and graphically in the boxplot at Figure 111. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are only significantly different for System A and System B.

Table 55. Descriptive statistics for ratings of ‘Capability’.

Statistics

		Capability: System A	Capability: System B	Capability: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		9.40	11.23	10.33
Median		10.00	12.00	11.00
Std. Deviation		2.700	2.409	3.407
Range		12	10	13
Minimum		2	6	2
Maximum		14	16	15
Percentiles	25	8.00	9.00	8.00
	50	10.00	12.00	11.00
	75	11.00	13.00	13.00

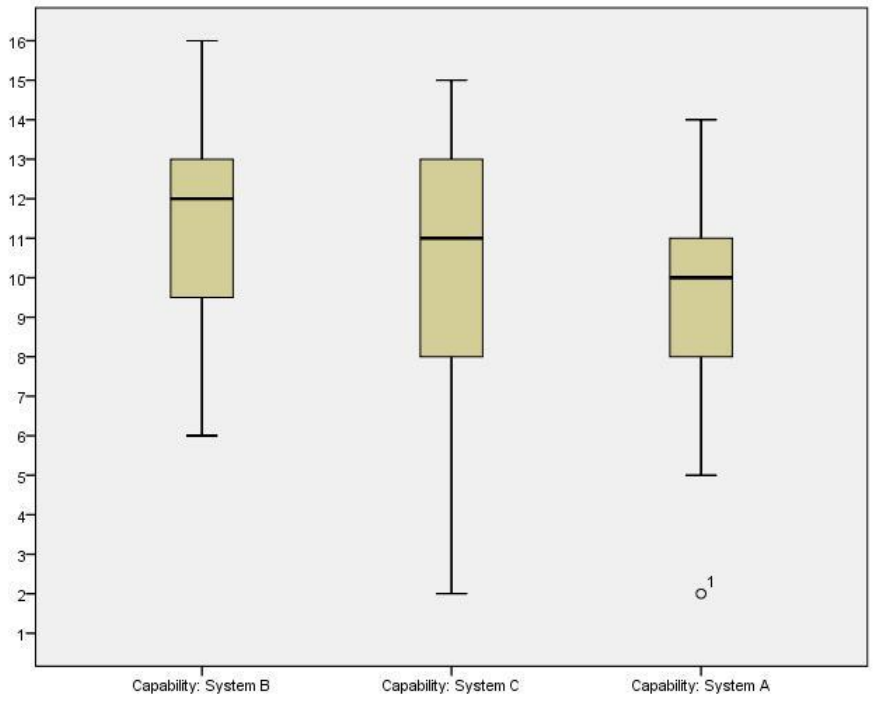


Figure 111. Boxplots for ratings of ‘Capability’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 112. While the judgment difference between System B and System C is not significant, the indicator model gave a 13% difference and inverted the rankings. Noting the sensitivity of the model at 6% weighting change to alter ranking results, the result was that the indicator model and the consensus view of the judges only correspond for System A. This result rates as **marginal** correspondence.

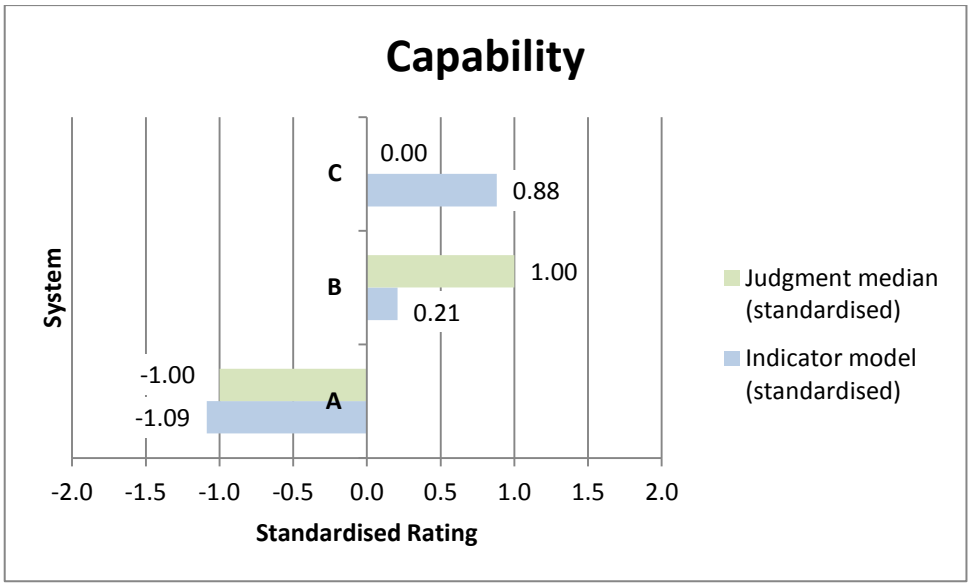


Figure 112. Correspondence of ratings for 'Capability'.

While the rankings of System B and System C do not correspond with the *a priori* indicator model, it can be seen in Figure 110 that the ranking is heavily dependent on the track capacity differences of the systems. It may be that this technical parameter was not considered to be equal in weight to the other technical parameters by the judges. Post facto, the weightings were re-evaluated in the indicator model to decrease the relative importance of automatic track capacity. The results are shown in Figure 113 and Figure 114 and it can be seen that the re-weighted indicator model now produces rankings that accord with the judgment rankings. This confirms the effect of sensitivity of the indicator model and the result therefore still rates as **marginal** correspondence.

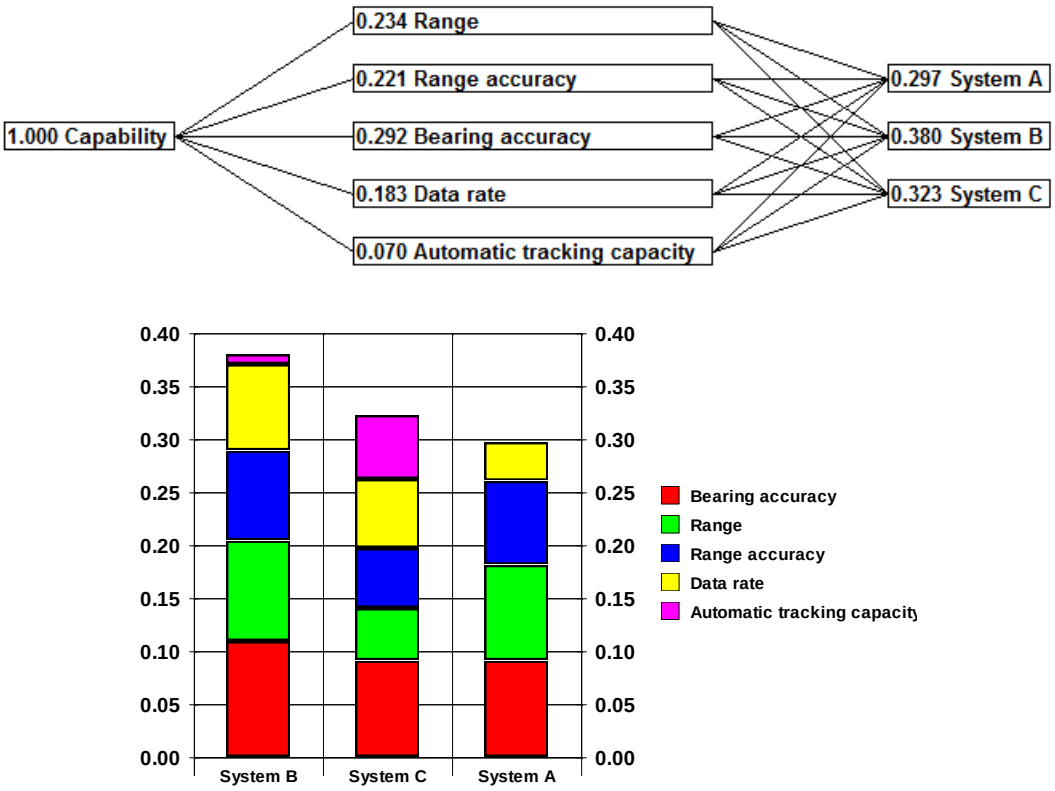


Figure 113. Post-facto Indicator model and ratings for ‘Capability’.

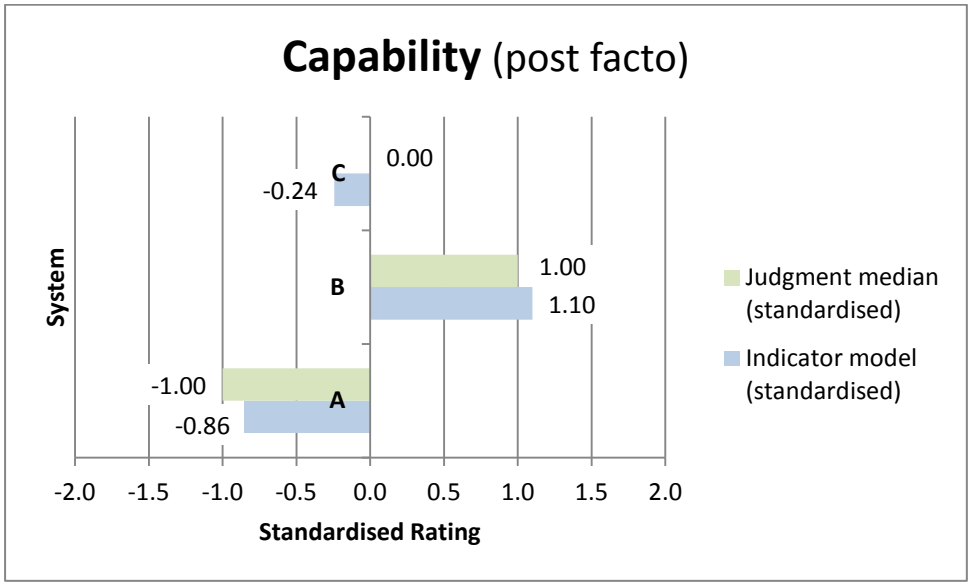


Figure 114. Correspondence of ratings for ‘Capability’ with post facto indicator model.

8.3.4.6 Effectiveness

As indicated in Table 46, it was expected that the "Effective" indicator would not primarily correlate with any technical parameters. It was considered that quality variable multi-mission/multi-function features may have some influence on the judgments of this indicator. This indicated a ranking of the systems from highest to lowest of System C to System A to System B. The indicator model and resultant ratings are shown in Figure 115. The minimum difference in ratings is 5.4% between System A and System C.

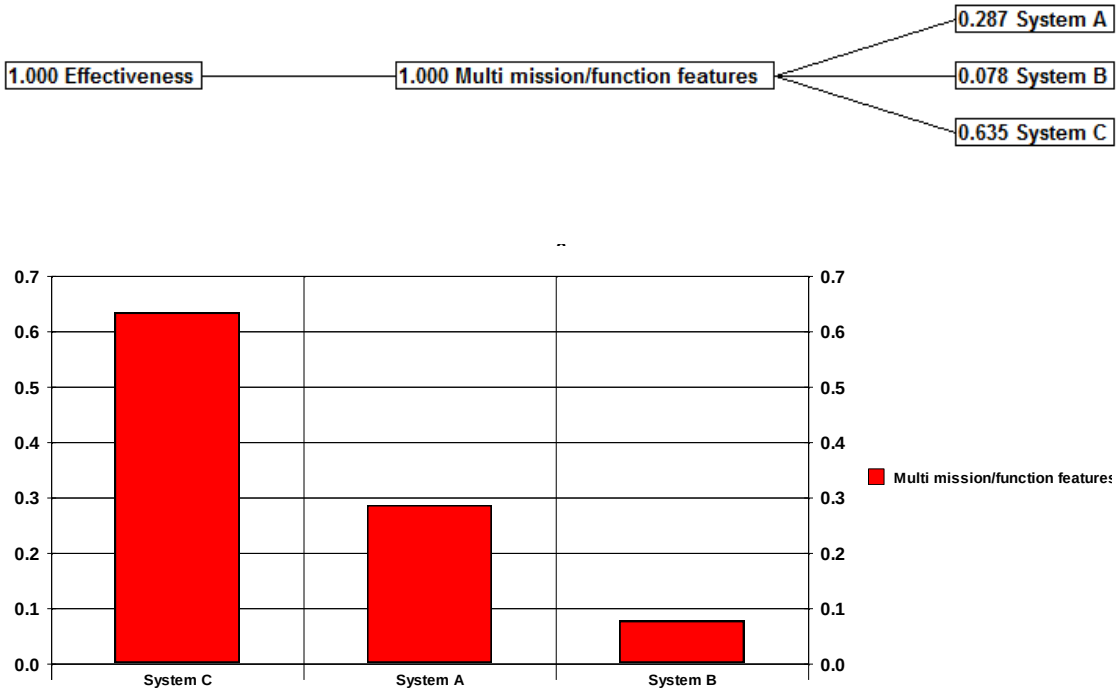


Figure 115. Indicator model and ratings for 'Effectiveness'.

Results from the judgment rating data for "Effective" are presented in the statistics summary at Table 56 and graphically in the boxplot at Figure 116. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the differences in medians of the judges' ratings are significantly different for each pair of systems.

Table 56. Descriptive statistics for ratings of ‘Effective’.

Statistics

		Effective: System A	Effective: System B	Effective: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		8.98	11.65	10.23
Median		9.00	12.00	11.00
Std. Deviation		2.849	2.389	2.553
Range		12	9	11
Minimum		2	7	4
Maximum		14	16	15
Percentiles	25	7.00	10.00	8.00
	50	9.00	12.00	11.00
	75	11.00	14.00	12.00

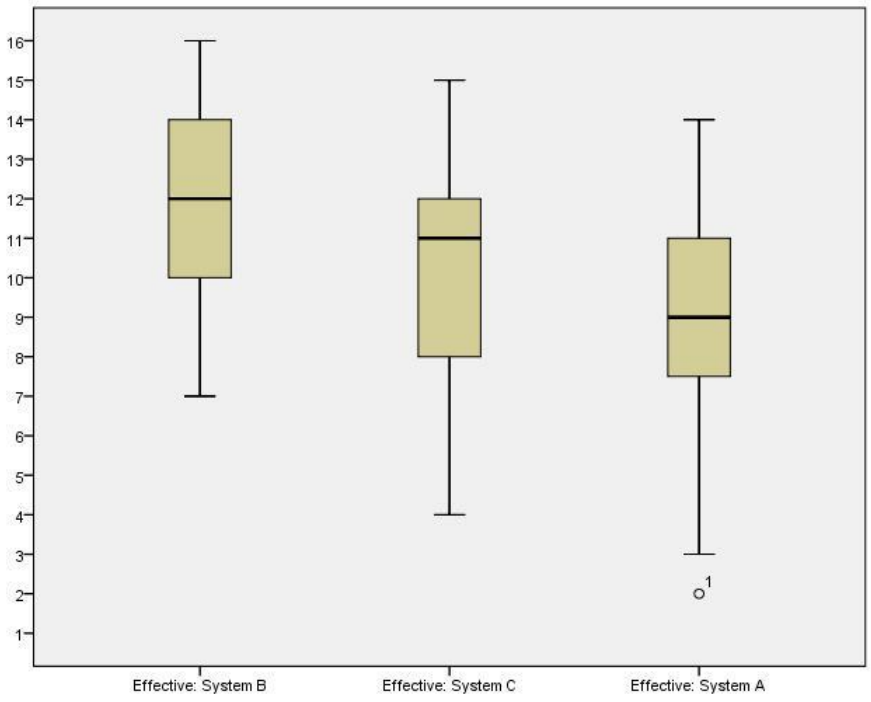


Figure 116. Boxplots for ratings of ‘Effective’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 117. While there is a clear ranking of systems both from judgment data and from the indicator model, the correspondence is **poor**. Not unexpectedly, this implies that the quality variable multi-mission/multi-function features was not a primary determinant in the judges’ ratings.

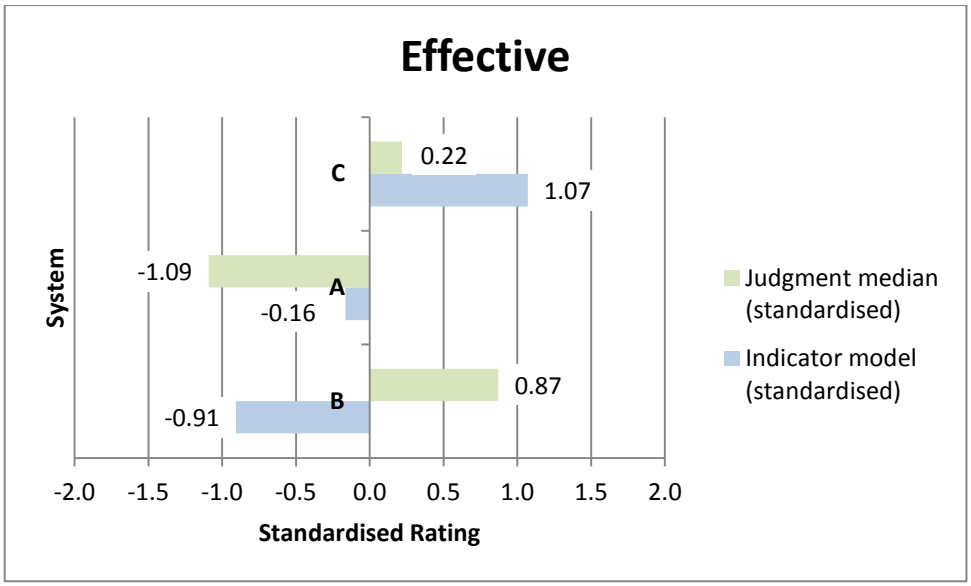


Figure 117. Correspondence of ratings for ‘Effective’.

8.3.4.7 Efficiency

As indicated in Table 46, it was expected that the "Efficient" indicator would correlate primarily with quantity variable Power efficiency and secondarily with Electrical power required. The indicator model and resultant ratings are shown in Figure 118. Inspection of the sensitivity of the model to changes of the weightings shows that either would need to change by 27% to alter the ranking of the systems. The minimum difference in ratings is 11% between System A and System C.

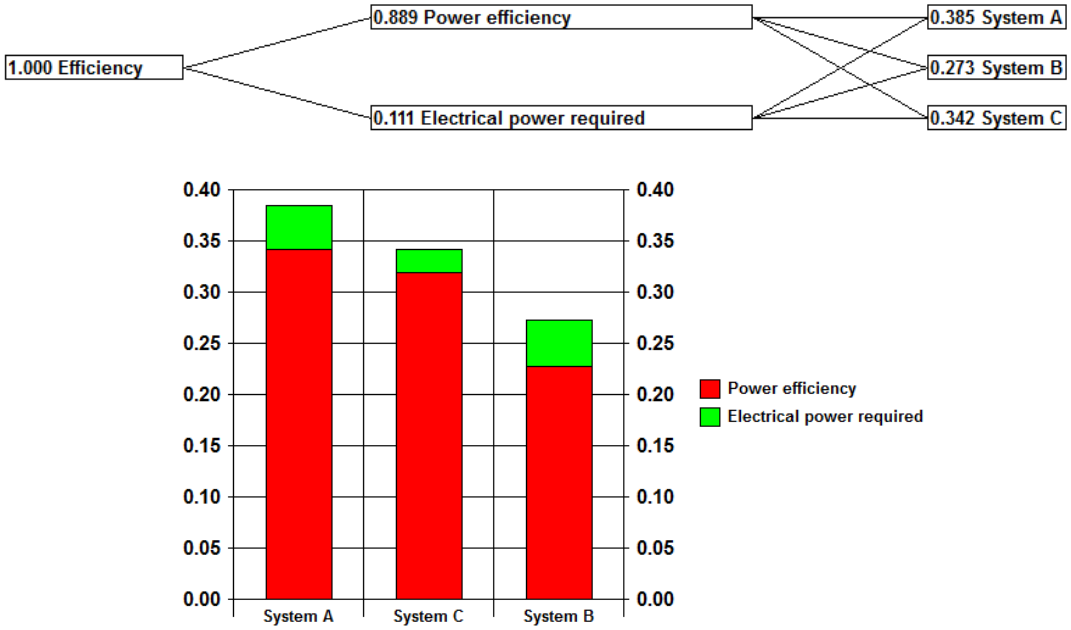


Figure 118. Indicator model and ratings for 'Efficiency'.

Results from the judgment rating data for "Efficient" are presented in the statistics summary at Table 57 and graphically in the boxplot at Figure 119. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are significantly different for each system, except for the System A and System C where the difference is not significant.

Table 57. Descriptive statistics for ratings of ‘Efficient’.

Statistics

		Efficient: System A	Efficient: System B	Efficient: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		9.02	11.63	9.12
Median		9.00	12.00	10.00
Std. Deviation		2.866	2.944	3.289
Range		11	11	14
Minimum		3	5	1
Maximum		14	16	15
Percentiles	25	7.00	9.00	7.00
	50	9.00	12.00	10.00
	75	11.00	14.00	11.00

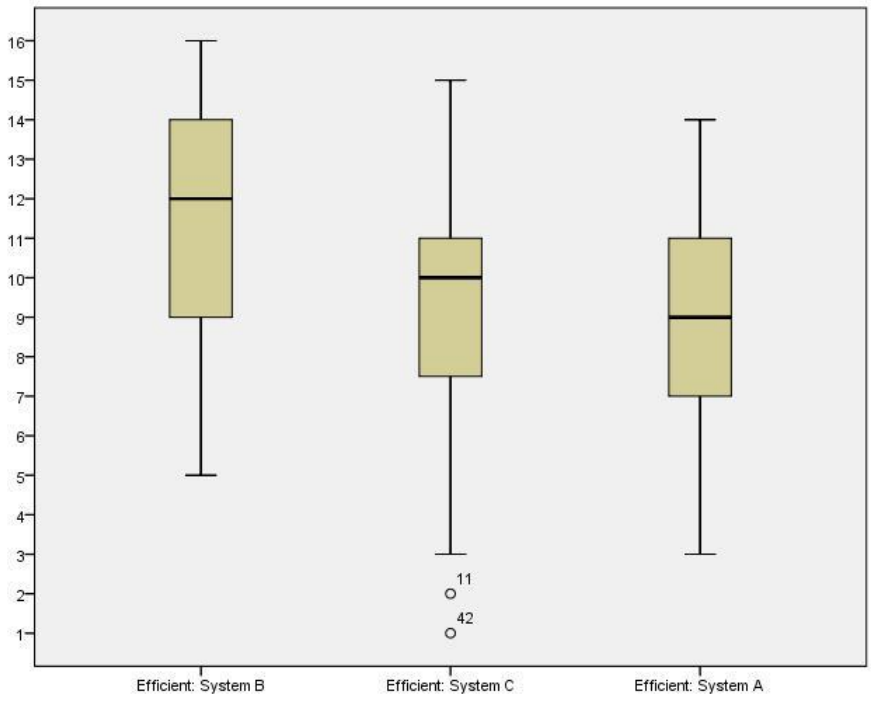


Figure 119. Boxplots for ratings of ‘Efficient’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 120. There is **poor** correspondence apparent in the rankings of the three systems between the indicator model and the consensus view of the judges. This implies the judges were not primarily influenced by the quantity variable of power efficiency as theorised.



Figure 120. Correspondence of ratings for ‘Efficient’.

8.3.4.8 Quality

As indicated in Table 46, it was expected that the “Quality” indicator would correlate primarily with quality variables manufacturer's quality system and secondarily with numbers in service. The indicator model and resultant ratings are shown in Figure 121. Inspection of the sensitivity of the model to changes of the weightings shows that either parameter would need to change by 11% to alter the ranking of the systems. The minimum difference in ratings is 3.4% between System B and System C.

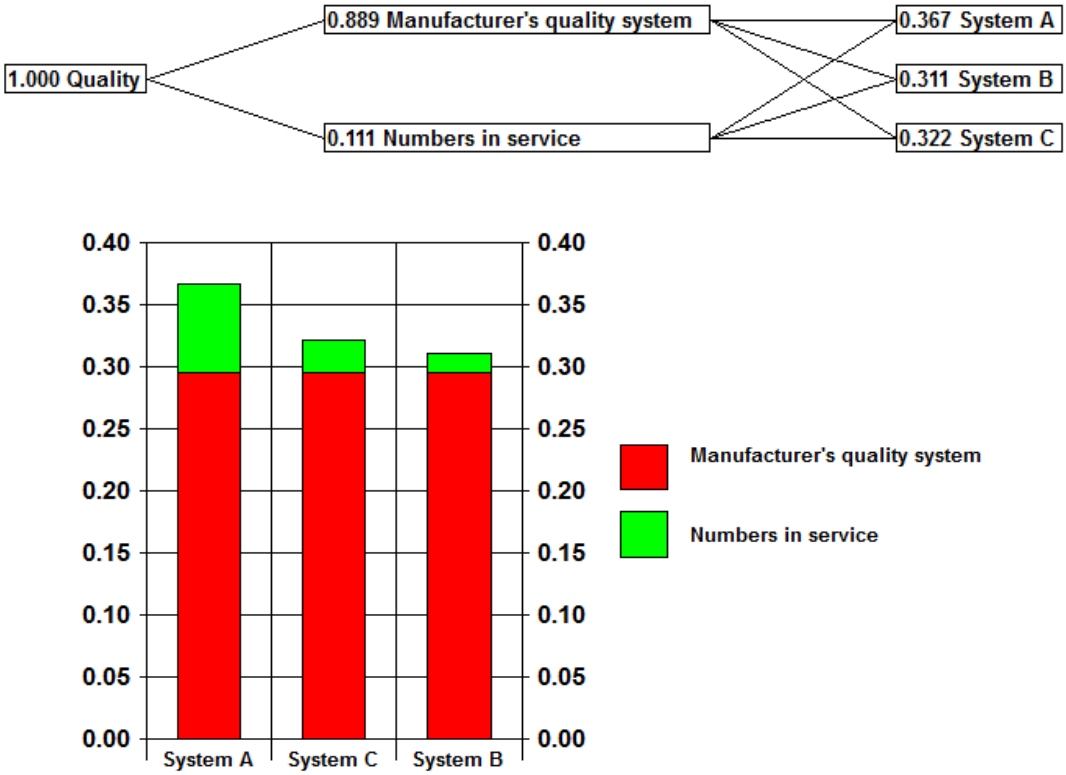


Figure 121. Indicator model and ratings for ‘Quality’.

Results from the judgment rating data for “Quality” are presented in the statistics summary at Table 58 and graphically in the boxplot at Figure 122. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges’ ratings are significantly different for each system, except for the System A and System C where the difference is not significant at the 0.05 level (significance is 0.068).

Table 58. Descriptive statistics for ratings of ‘Quality’.

Statistics

		Quality: System A	Quality: System B	Quality: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		11.16	9.47	10.44
Median		12.00	9.00	11.00
Std. Deviation		2.600	3.165	2.538
Range		11	13	9
Minimum		5	3	6
Maximum		16	16	15
Percentiles	25	9.00	7.00	8.00
	50	12.00	9.00	11.00
	75	13.00	12.00	12.00

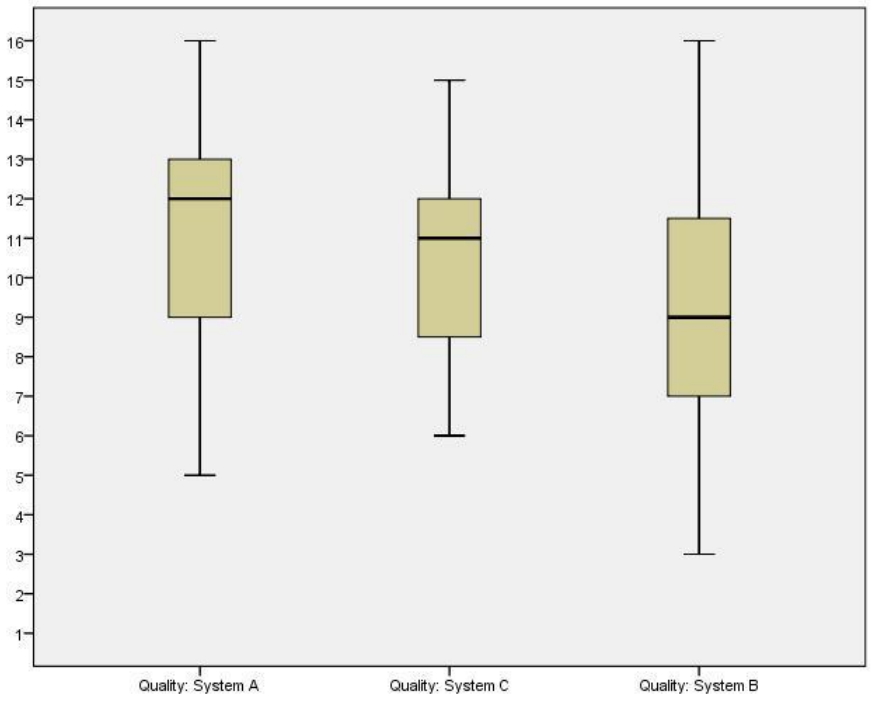


Figure 122. Boxplots for ratings of ‘Quality’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 123. It is noted that there was no difference for the example systems in the primary technical parameter, the only differentiation being in the secondary technical parameter ‘number in service’. The indicator model has only small, 3.4% difference between System B and System C, but the judgment difference was significant. System A and System C were not significantly different in the judges’ ratings but 13% different in the indicator model. Therefore, there is a **marginal**

correspondence apparent in the rankings of the three systems between the indicator model and the consensus view of the judges. This result implies “numbers in service” may not have been the variable upon which the judges relied.

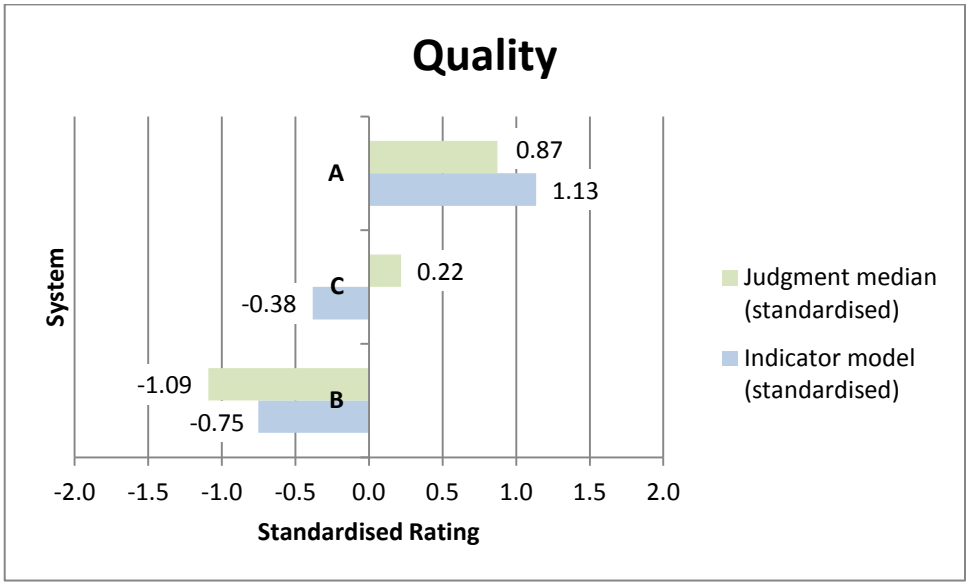


Figure 123. Correspondence of ratings for ‘Quality’.

8.3.4.9 Flexibility

As indicated in Table 46, it was expected that the "Flexible" indicator would correlate primarily with quality variable multi-mission/multi-function features. The indicator model is identical to that for "Effective" as shown at Figure 115.

Results from the judgment rating data for "Flexible" are presented in the statistics summary at Table 59 and graphically in the boxplot at Figure 124. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that only the medians for the System A and System C are significantly different for from each other.

Table 59. Descriptive statistics for ratings of 'Flexible'.

Statistics

		Flexible: System A	Flexible: System B	Flexible: System C
N	Valid	42	43	43
	Missing	1	0	0
Mean		9.69	10.02	10.79
Median		10.00	10.00	11.00
Std. Deviation		2.967	2.858	2.512
Range		12	12	12
Minimum		2	4	3
Maximum		14	16	15
Percentiles	25	7.75	8.00	9.00
	50	10.00	10.00	11.00
	75	12.00	12.00	13.00

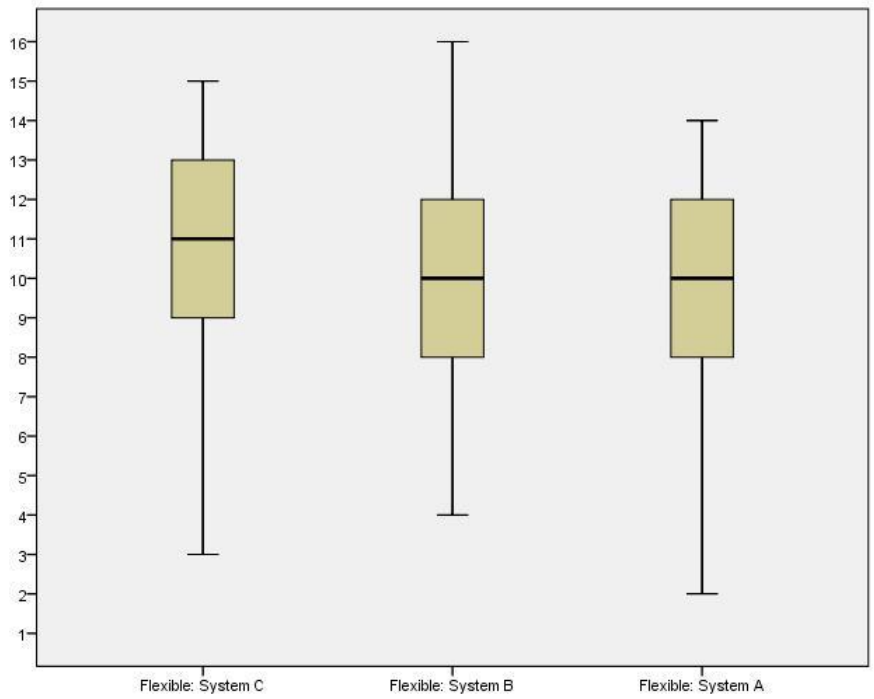


Figure 124. Boxplots for ratings of ‘Flexible’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 125. Notwithstanding that the difference in the median of judgment ratings between system C and B and between systems B and A was not significant, there is a correspondence apparent in the rankings of the three systems between the indicator model and the consensus view of the judges. This is rated as a **reasonable** correspondence.

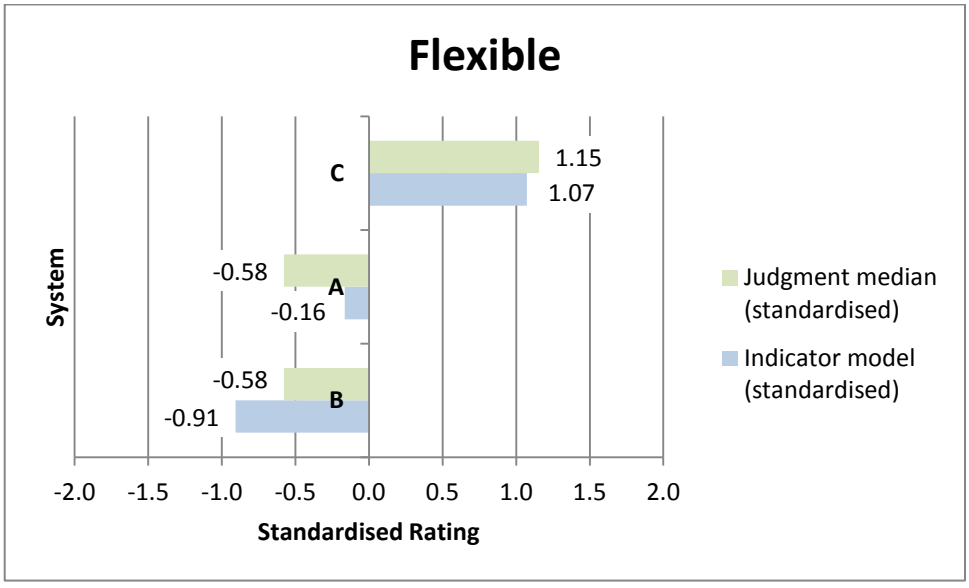


Figure 125. Correspondence of ratings for ‘Flexible’.

8.3.4.10 Durability

As indicated in Table 46, it was expected that the “Durable” indicator would correlate primarily with quality variables shock qualification and environmental resistance standards and secondarily with manufacturer's quality system. The indicator model and resultant ratings are shown in Figure 126. Inspection of the sensitivity of the model to changes of the weightings shows that the most sensitive parameter is Environmental resistance standards which would need to change by 45% to alter the ranking of the systems. The minimum difference in ratings is 58% between System B and System C.

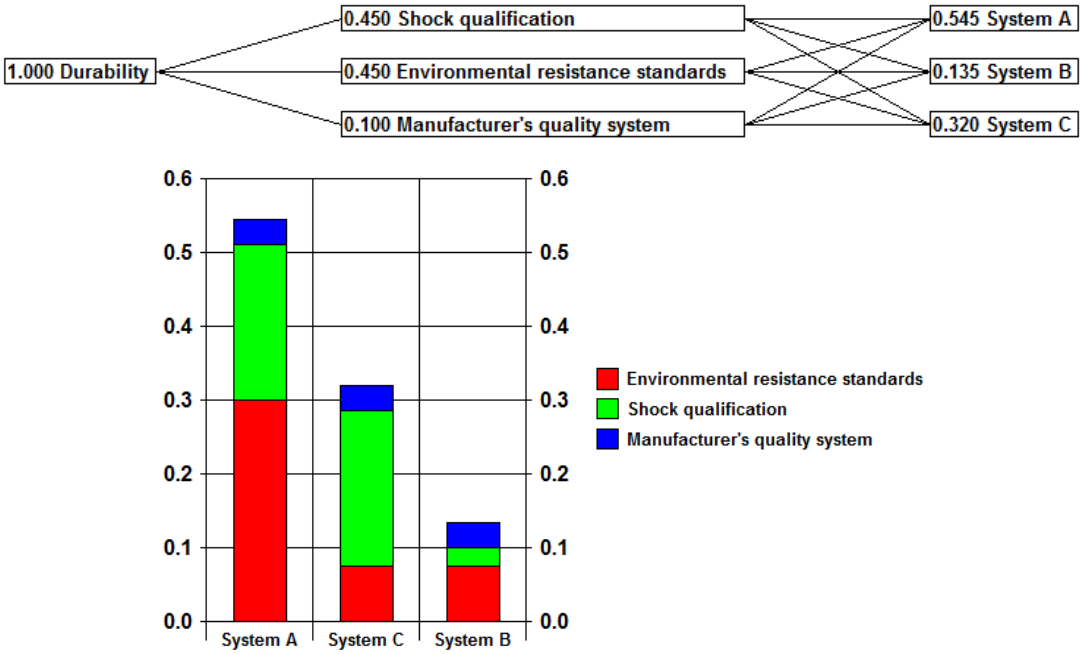


Figure 126. Indicator model and ratings for ‘Durability’.

Results from the judgment rating data for “Durable” are presented in the statistics summary at Table 60 and graphically in the boxplot at Figure 127. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges’ ratings are significantly different only between System A and System C.

Table 60. Descriptive statistics for ratings of ‘Durable’.

Statistics

		Durable: System A	Durable: System B	Durable: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		9.93	10.49	10.84
Median		10.00	11.00	11.00
Std. Deviation		3.334	3.073	2.554
Range		14	14	9
Minimum		2	2	6
Maximum		16	16	15
Percentiles	25	8.00	8.00	9.00
	50	10.00	11.00	11.00
	75	12.00	13.00	13.00

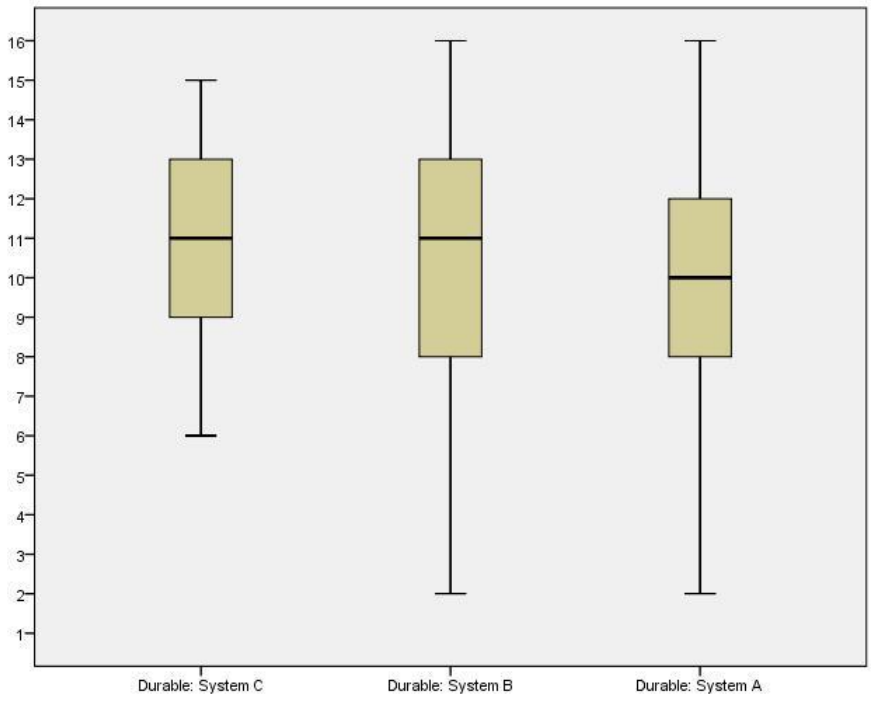


Figure 127. Boxplots for ratings of ‘Durable’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 128. The correspondence is **poor**. This implies that the judges must have used other technical parameters or weighed up different information in rating the durability of the systems. It is also possible that the judges had different interpretations of the meaning of “durable” as a system quality attribute.

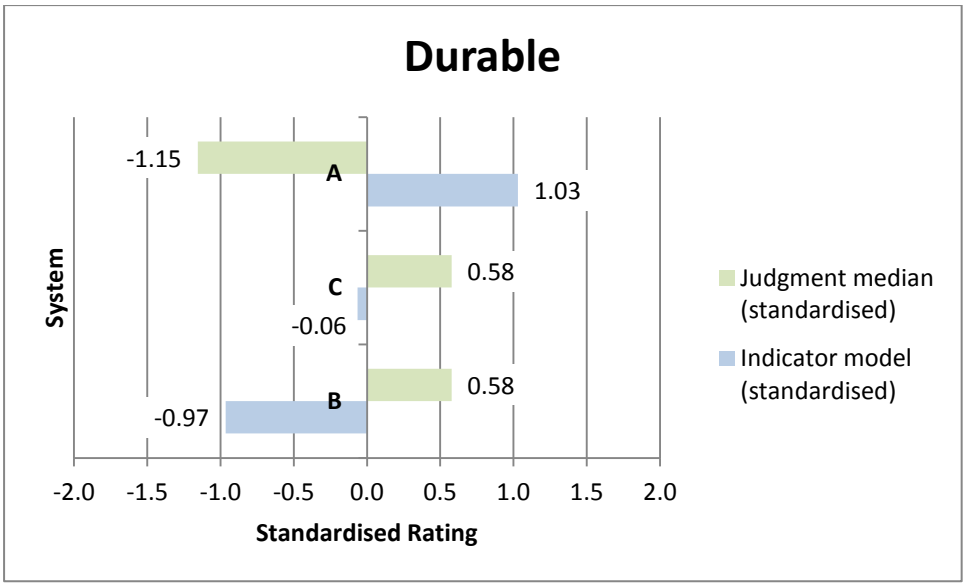


Figure 128. Correspondence of ratings for ‘Durability’.

8.3.4.11 Survivability

As indicated in Table 46, it was expected that the "Survivable" indicator would correlate primarily with quality variables for shock qualification and electronic protection and secondarily with environmental resistance standards. The indicator model and resultant ratings are shown in Figure 129. Inspection of the sensitivity of the model to changes of the weightings shows that the most sensitive parameter is Environmental resistance standards which would need to change by 10% to alter the ranking of the systems. The minimum difference in ratings is 10% between System A and System C.

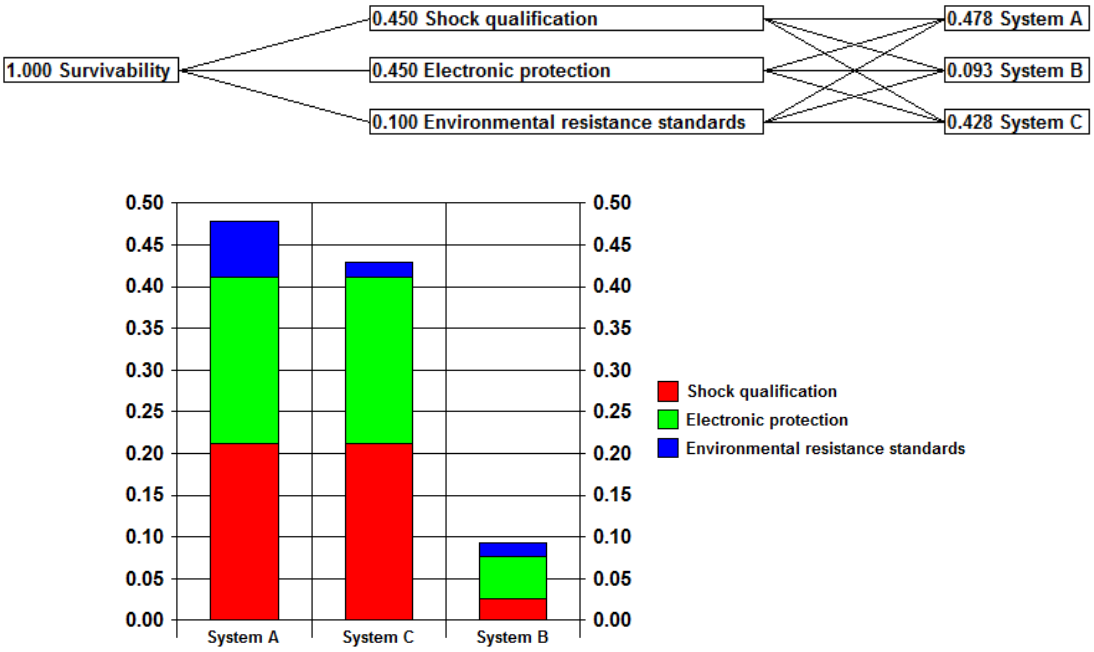


Figure 129. Indicator model and ratings for 'Survivability'.

Results from the judgment rating data for "Survivable" are presented in the statistics summary at Table 61 and graphically in the boxplot at Figure 130. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are significantly different for each system, except for the System A and System C where the difference is not significant.

Table 61. Descriptive statistics for ratings of ‘Survivable’.

Statistics

		Survivable: System A	Survivable: System B	Survivable: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		11.05	8.44	11.35
Median		12.00	8.00	11.00
Std. Deviation		3.207	3.312	2.487
Range		13	14	9
Minimum		3	2	6
Maximum		16	16	15
Percentiles	25	10.00	6.00	9.00
	50	12.00	8.00	11.00
	75	14.00	10.00	14.00

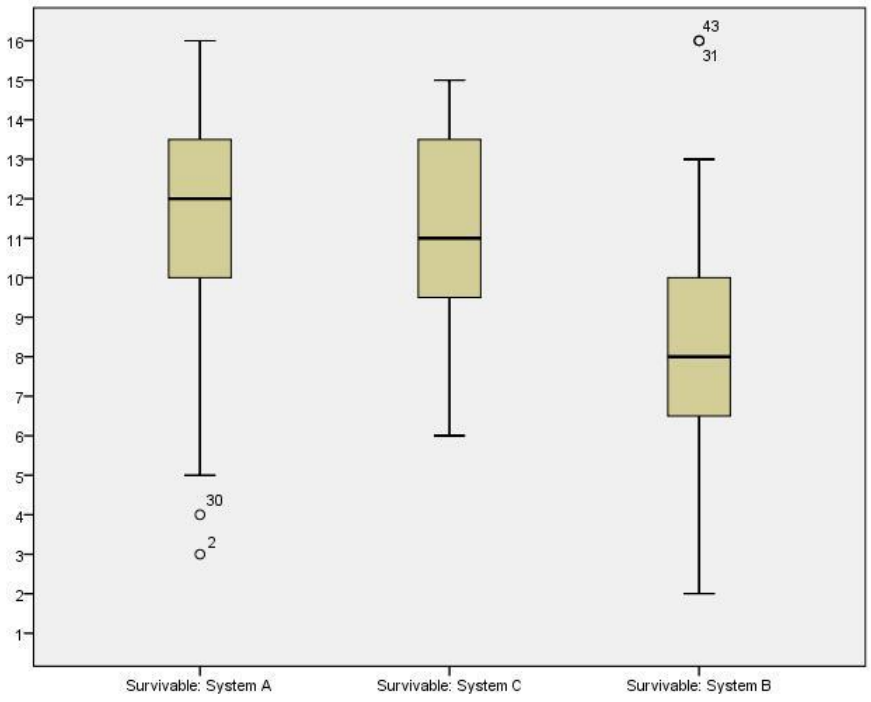


Figure 130. Boxplots for ratings of ‘Survivable’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 131. Noting that the difference in median judgment of System A and System C is not significant and the indicator model produced ratings of these systems within 10%, there is **good** correspondence apparent in the rankings of the three systems between the indicator model of survivable and the consensus view of the judges.

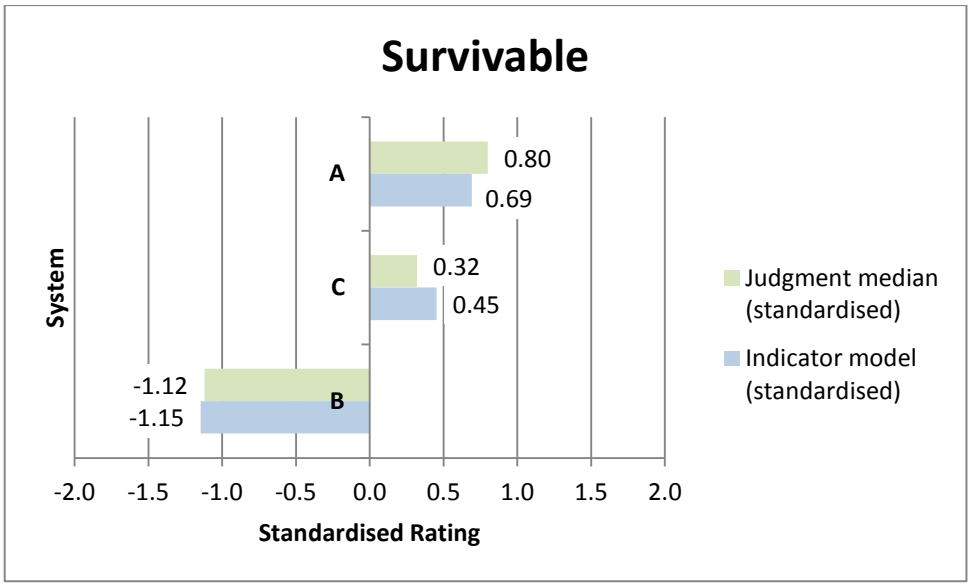


Figure 131. Correspondence of ratings for ‘Survivable’.

8.3.4.12 Dependability

As indicated inTable 46, it was expected that the "Dependable" indicator would correlate primarily with quantity variable MTBF and the quality variable environmental resistance standards and secondarily with quality variables EMC compliance, physical security and computer/information security. The indicator model and resultant ratings are shown in Figure 132. Inspection of the sensitivity of the model to changes of the weightings shows that the most sensitive parameter is Environmental resistance standards which would need to change by 10% to alter the ranking of the systems. The minimum difference in ratings is 1.2% between System B and System C; effectively the same rating.

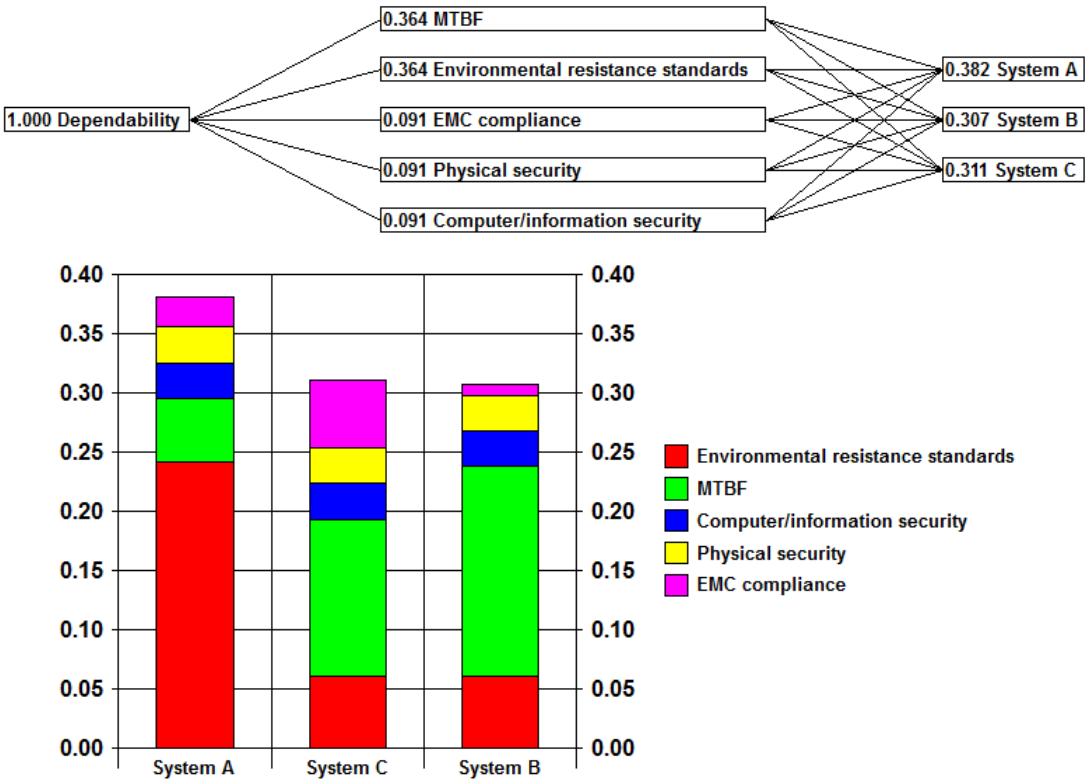


Figure 132. Indicator model and ratings for 'Dependability'.

Results from the judgment rating data for "Dependable" are presented in the statistics summary at Table 60 and graphically in the boxplot at Figure 133. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are significantly different for each system, except for the System B and System C where the difference is not significant.

Table 62. Descriptive statistics for ratings of ‘Dependable’.

Statistics

		Dependable: System A	Dependable: System B	Dependable: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		8.28	11.33	10.58
Median		8.00	11.00	11.00
Std. Deviation		2.906	2.179	2.719
Range		11	9	11
Minimum		2	7	4
Maximum		13	16	15
Percentiles	25	6.00	10.00	8.00
	50	8.00	11.00	11.00
	75	11.00	13.00	13.00

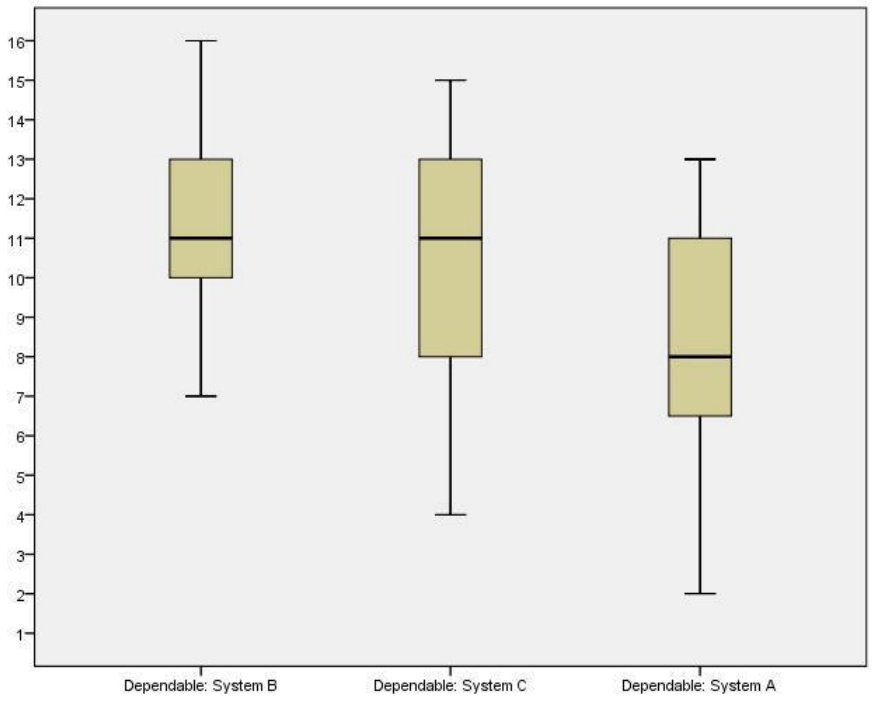


Figure 133. Boxplots for ratings of ‘Dependable’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 98. There is **poor** correspondence apparent in the rankings of the three systems between the indicator model and the consensus view of the judges.

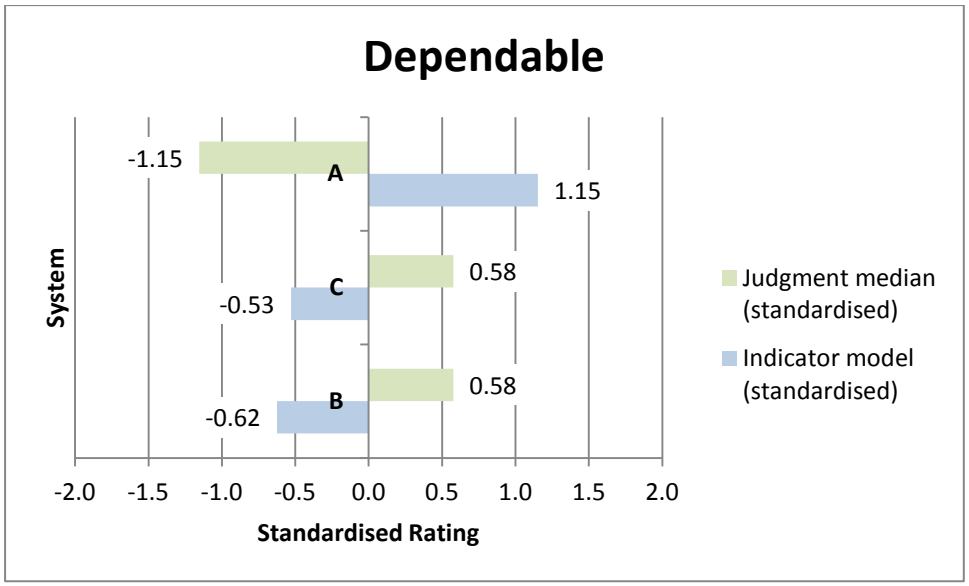


Figure 134. Correspondence of ratings for 'Dependable'.

While the rankings of the systems did not accord with the *a priori* indicator model, it is clear in Figure 126 that the ranking was heavily dependent on the environmental resistance standard differences of the systems. This may not have been an indicator considered to be equal in weight to the MTBF indicator for the judges.

Post facto, the weightings were re-evaluated in the indicator model to decrease the relative importance of environmental resistance standard. The model and results are shown in Figure 135 and Figure 136. This model produced rankings that corresponded with the judges rankings. However, the sensitivity to changing weights for two parameters is less than 4% and it is noted that many other changes to the relative importance of the objective indicators could be made to effect a change in the indicator model rankings to match the judgment rankings. This illustrated why the correspondence, even with a revised indicator model, can be regarded as no higher than **reasonable.**

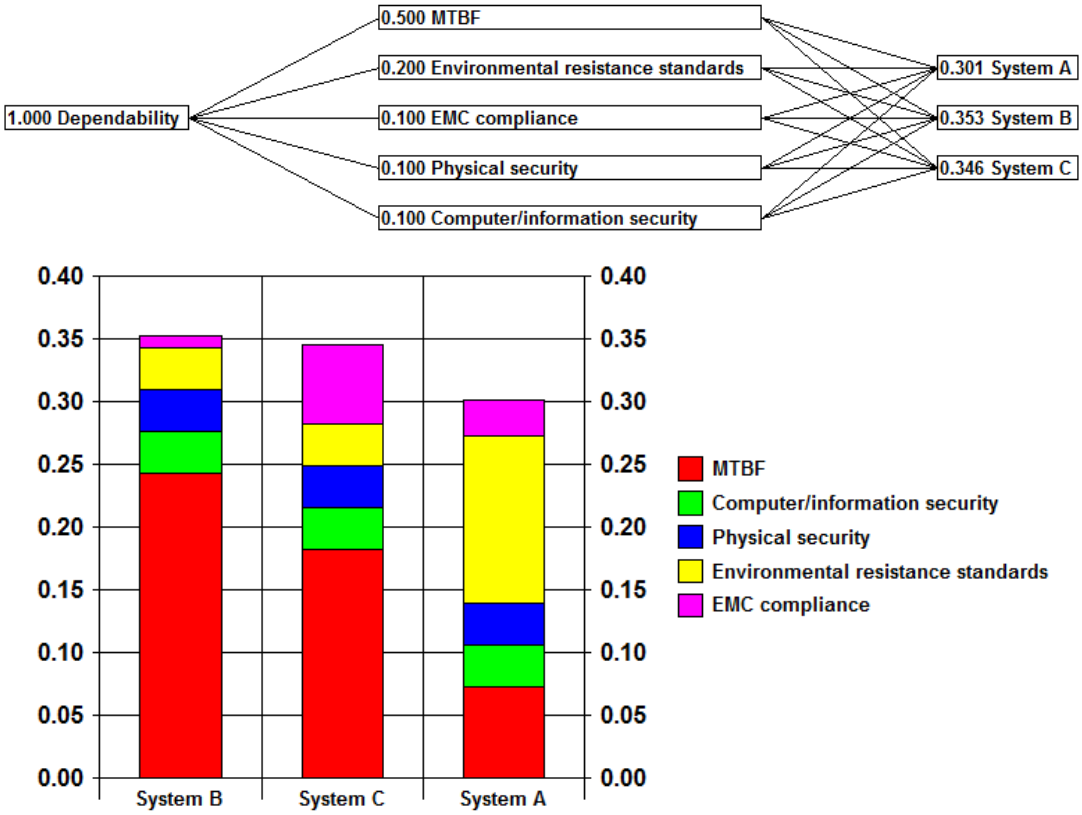


Figure 135. Post facto indicator model and ratings for 'Dependability'.

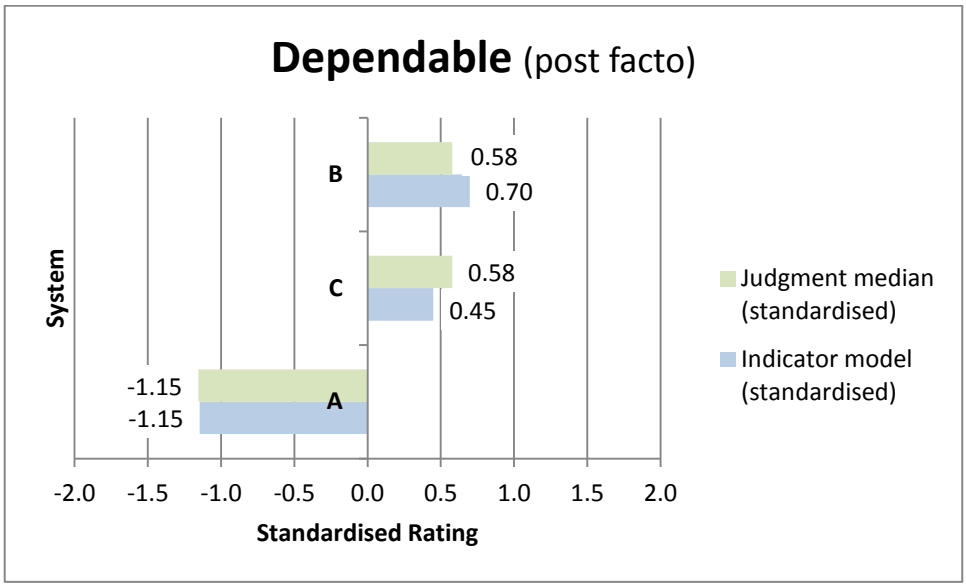


Figure 136. Correspondence of ratings for 'Dependable' with post facto indicator model.

8.3.4.13 Reliability

As indicated in Table 46, it was expected that the "Reliability" indicator would correlate primarily with the quantity variable MTBF. The indicator model and resultant ratings are shown in Figure 137. The minimum difference in ratings is 25% between System B and System C.

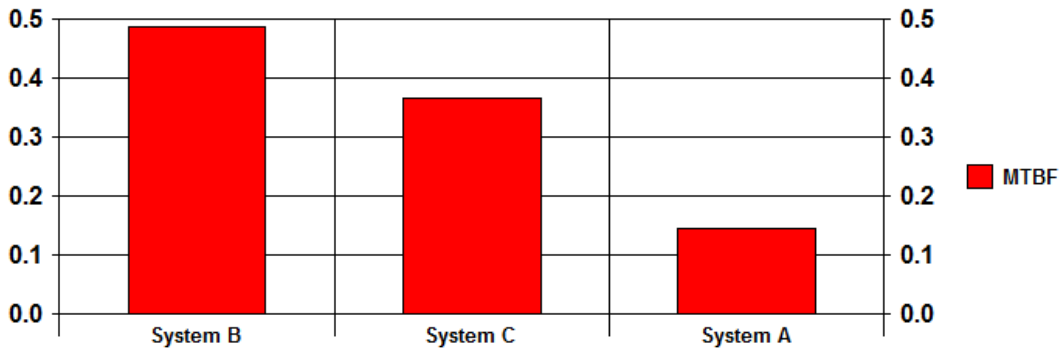


Figure 137. Indicator model and ratings for 'Reliability'.

Results from the judgment rating data for "Reliability" are presented in the statistics summary at Table 63 and graphically in the boxplot at Figure 138. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are significantly different for all systems.

Table 63. Descriptive statistics for ratings of 'Reliable'.

Statistics

		Reliable: System A	Reliable: System B	Reliable: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		7.65	12.21	11.16
Median		8.00	13.00	11.00
Std. Deviation		3.116	2.042	2.104
Range		12	9	8
Minimum		3	7	7
Maximum		15	16	15
Percentiles	25	5.00	11.00	10.00
	50	8.00	13.00	11.00
	75	10.00	14.00	13.00

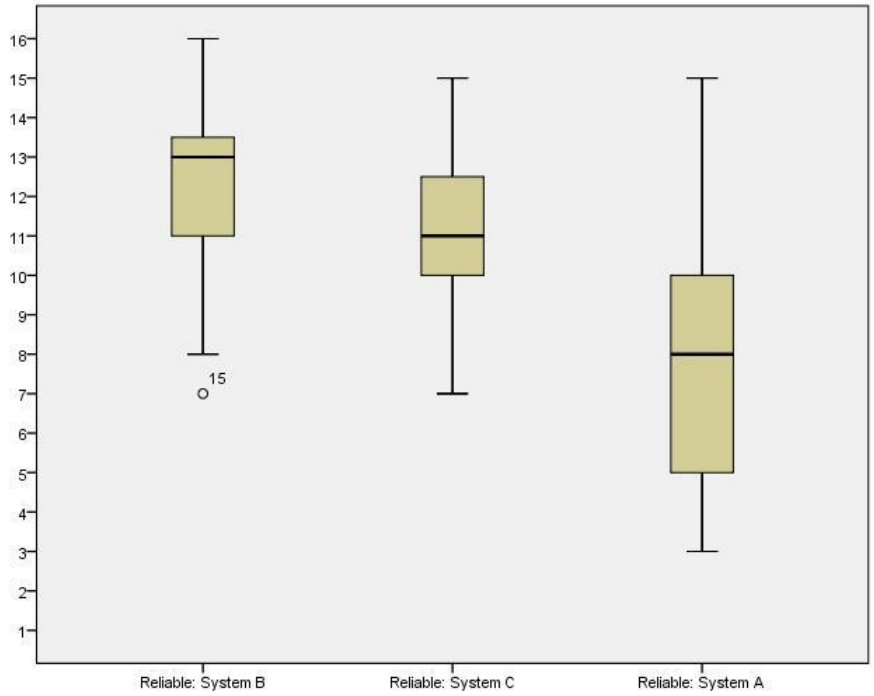


Figure 138. Boxplots for ratings of ‘Reliable’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 139. The indicator model (really just the quantity variable MTBF) was not sensitive, the judges’ ratings are significantly different and the rankings of the systems agreed. Therefore it is concluded there was **good** correspondence in the rankings of the three systems between the indicator model of reliability and the consensus view of the judges.

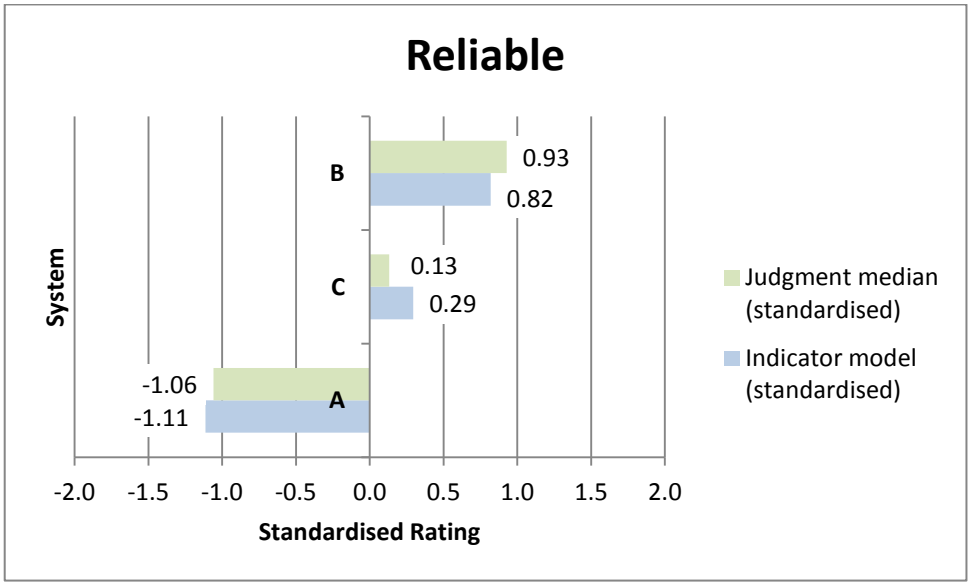


Figure 139. Correspondence of ratings for ‘Reliability’.

Given the good correspondence, and the firm expectation that judges would be likely to have formed their judgment for reliability on MTBF a further question arose that justified analysis: do the ratings of the judges equate linearly with the actual values of MTBF, or do the judges' apply some other value function to MTBF when judging reliability? To investigate this question the median of the judges' ratings is plotted against the numerical value of the MTBF in hours. Figure 140 presents the result with a line of best fit. While the data set was insufficient to establish statistical significance, it appeared that the ratings may be linearly related to the MTBF. However, the intercept was not at the origin as would be expected by the definition of the rating scale and the implicit understanding that a MTBF of zero corresponds to complete absence of reliability. Figure 141 shows that a power law equation may be able to fit the data and better represent the value function implicitly applied by the judges.

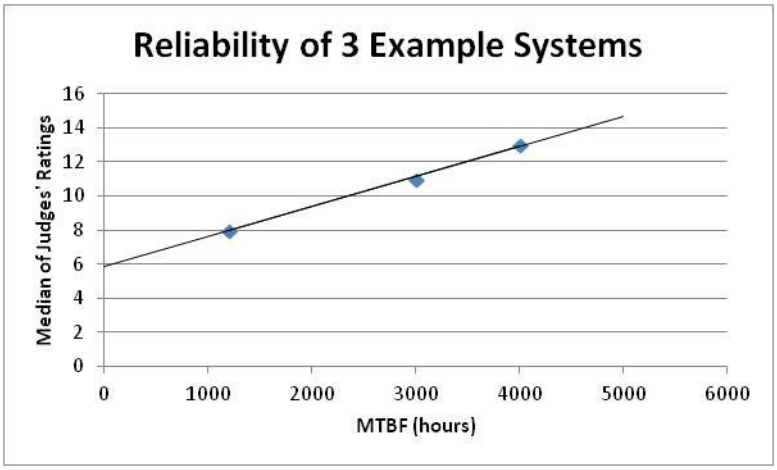


Figure 140. Reliability rating vs MTBF – linear relationship.

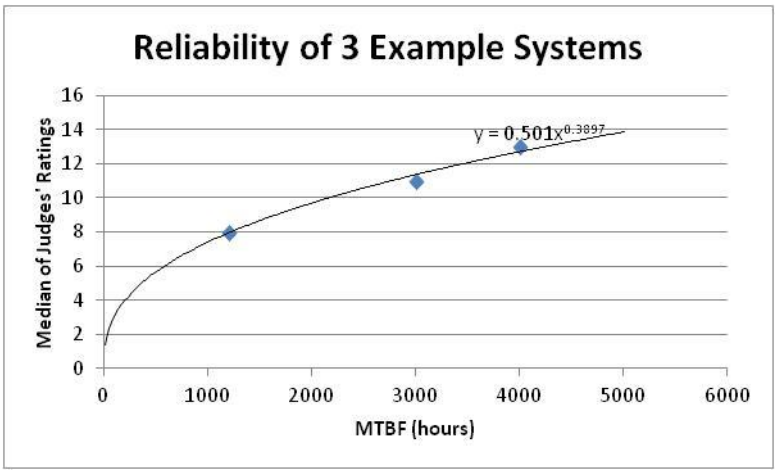


Figure 141. Reliability rating vs MTBF – power law relationship.

8.3.4.14 Supportability

As indicated in Table 46, it was expected that the "Supportable" indicator would correlate primarily with variables for built in test (BIT), mean time to repair (MTTR) and maintenance concept and secondarily to numbers in service. The indicator model and resultant ratings are shown in Figure 142. Inspection of the sensitivity of the model to changes of the weightings shows that the model is sensitive to inverting the ranking of System A and System C with as little as 2% change in numbers in service weighting or 4% change in BIT weighting. The minimum difference in ratings is 2.4% between System A and System C.

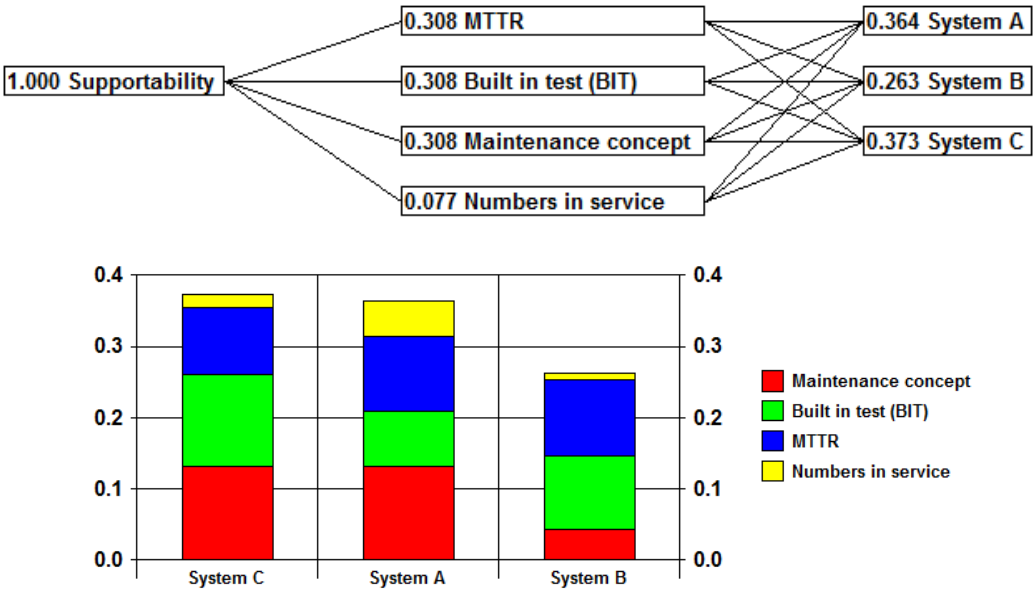


Figure 142. Indicator model and ratings for 'Supportability'.

Results from the judgment rating data for "Supportable" are presented in the statistics summary at Table 64 and graphically in the boxplot at Figure 143. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are significantly different for each system, except for the System A and System B where the difference is not significant.

Table 64. Descriptive statistics for ratings of ‘Supportable’.

Statistics

		Supportable: System A	Supportable: System B	Supportable: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		9.19	8.98	12.05
Median		10.00	9.00	12.00
Std. Deviation		3.375	3.687	2.011
Range		13	14	7
Minimum		2	2	8
Maximum		15	16	15
Percentiles	25	6.00	6.00	11.00
	50	10.00	9.00	12.00
	75	12.00	12.00	14.00

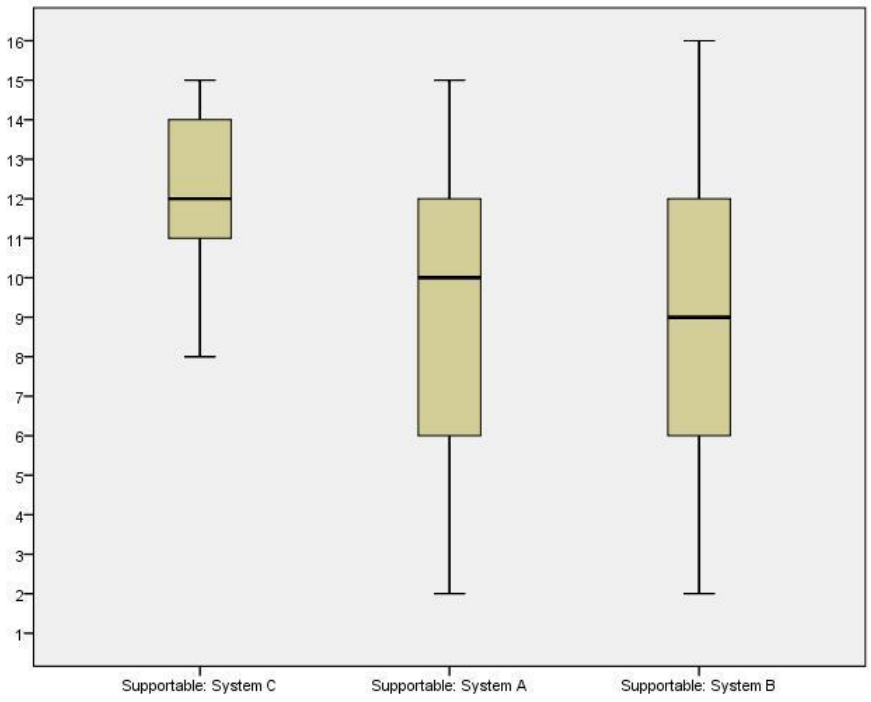


Figure 143. Boxplots for ratings of ‘Supportable’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 144. While the observed rankings agrees, there was sensitivity in the indicator model of 2-4% that could reverse the ranking of System A and System C and the judgments indicated there was no significant difference between the rating of System A and System B. This leads to a rating of **marginal** correspondence apparent in the rankings of the three systems between the indicator model of supportability and the consensus view of the judges.

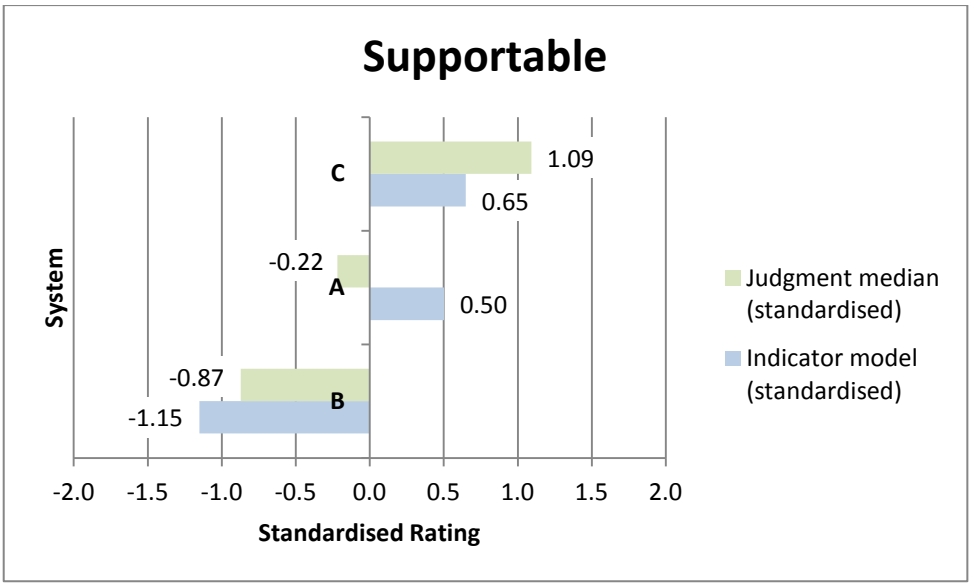


Figure 144. Correspondence of ratings for ‘Supportability’.

8.3.4.15 Maintainability

As indicated in Table 46, it was expected that the "Maintainable" indicator would correlate primarily with variables for with BIT, MTTR and maintenance concept. The indicator model and resultant ratings are shown in Figure 145. Inspection of the sensitivity of the model to changes of the weightings shows that the most sensitive parameter is BIT which would need to change by 24% to alter the ranking of the systems. The minimum difference in ratings is 11% between System A and System C.

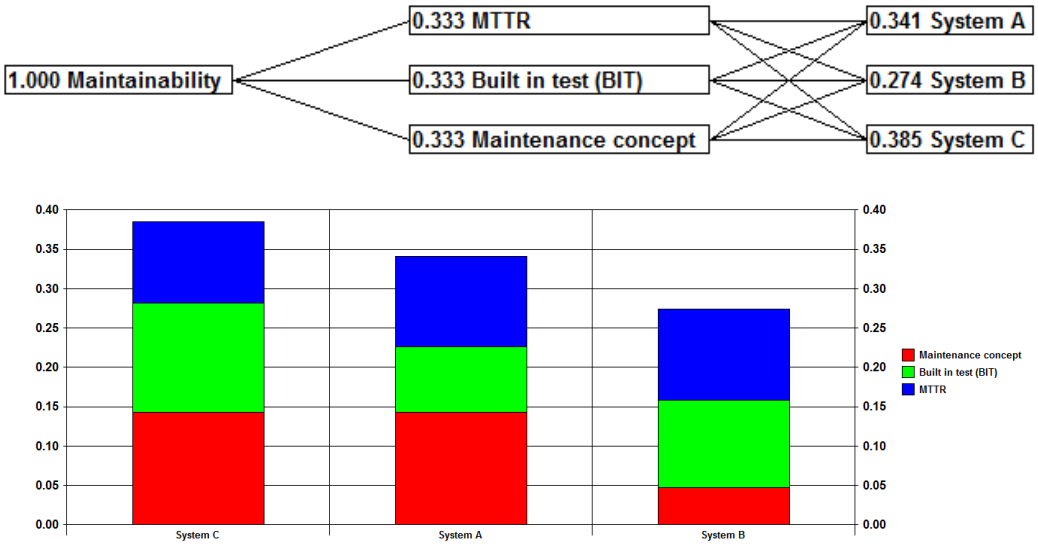


Figure 145. Indicator model and ratings for 'Maintainability'.

Results from the judgment rating data for "Maintainable" are presented in the statistics summary at Table 65 and graphically in the boxplot at Figure 146. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are significantly different for each system, except for the System A and System B where the difference is not significant at the 0.05 level (significance at 0.137).

Table 65. Descriptive statistics for ratings of ‘Maintainable’.

Statistics

		Maintainable: System A	Maintainable: System B	Maintainable: System C
N	Valid	42	43	43
	Missing	1	0	0
Mean		9.29	8.26	12.19
Median		10.00	8.00	12.00
Std. Deviation		2.874	3.619	2.413
Range		11	13	9
Minimum		4	2	7
Maximum		15	15	16
Percentiles	25	7.75	6.00	10.00
	50	10.00	8.00	12.00
	75	11.25	12.00	14.00

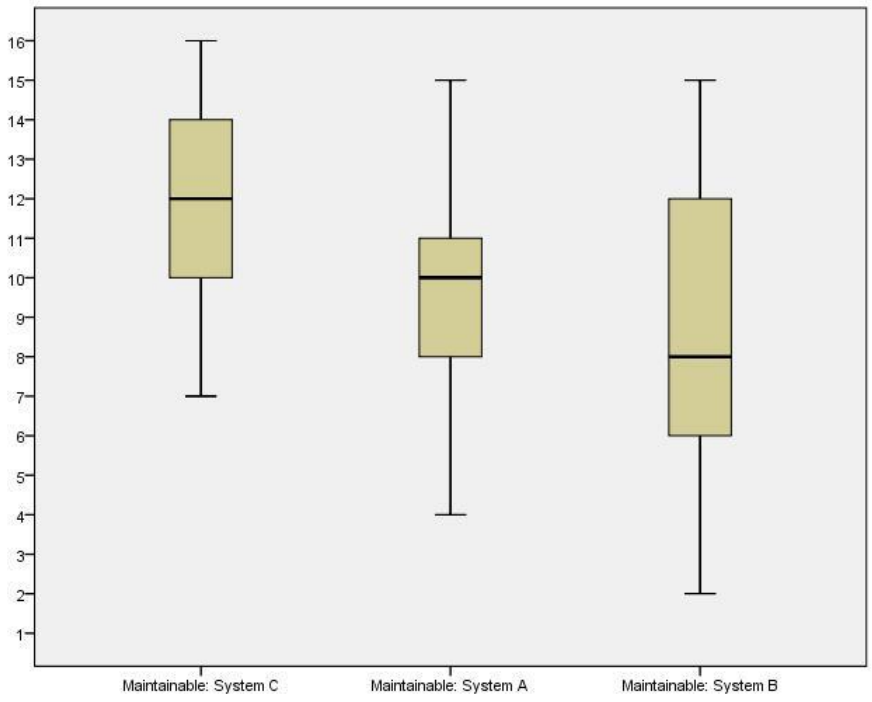


Figure 146. Boxplots for ratings of ‘Maintainable’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 147. While the observed rankings agreed and the indicator model was not sensitive the judgments indicated there was no significant difference between the judges’ rating of System A and System B. This leads to a rating of **reasonable** correspondence apparent in the rankings of the three systems between the indicator model of maintainability and the consensus view of the judges.

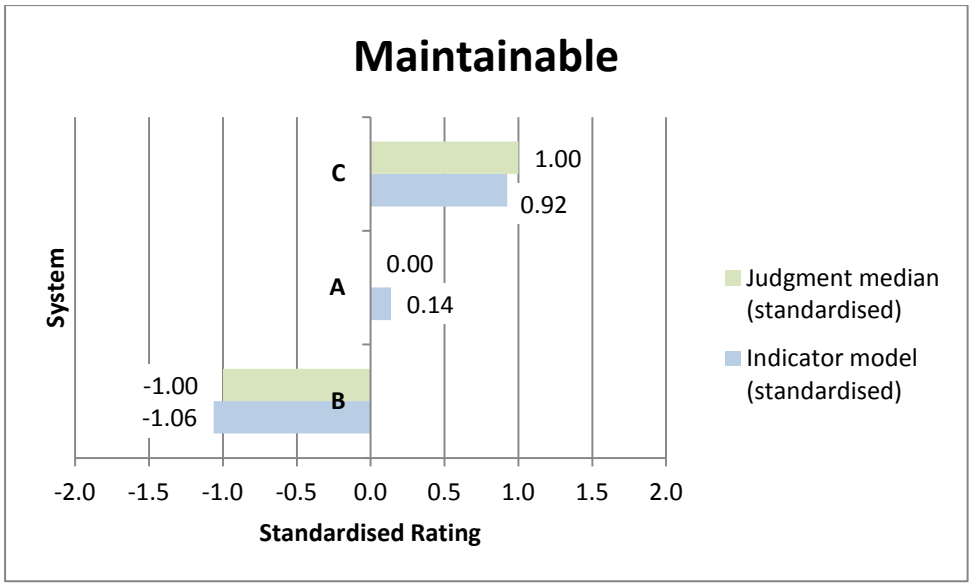


Figure 147. Correspondence of ratings for ‘Maintainability’.

8.3.4.16 Compliance

As indicated in Table 46, it was expected that the "Compliance" indicator would correlate primarily with quantity variables for compliant requirements and non-compliant requirements and secondarily the combination of full and partially compliant requirements. The indicator model and resultant ratings are shown in Figure 148. Inspection of the sensitivity of the model to changes of the weightings shows that the most sensitive parameter is compliant requirements which would need to change by 44% to alter the ranking of the systems. The minimum difference in ratings is 4.6% between System B and System C.

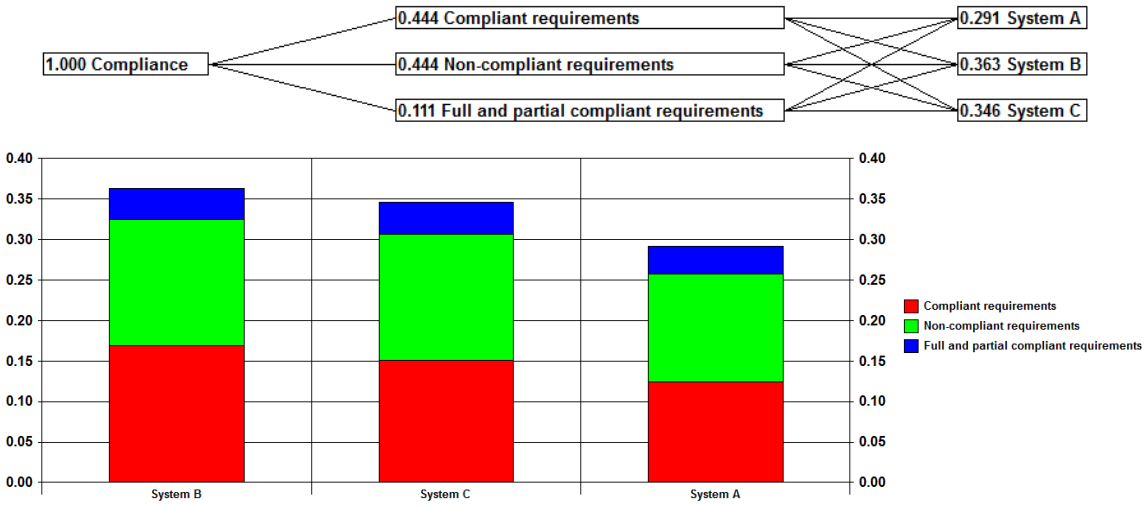


Figure 148. Indicator model and ratings for 'Compliance'.

Results from the judgment rating data for "Compliance" are presented in the statistics summary at Table 66 and graphically in the boxplot at Figure 149. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are all significantly different for each system.

Table 66. Descriptive statistics for ratings of ‘Compliance’.

Statistics

		Compliance: System A	Compliance: System B	Compliance: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		8.09	11.79	9.51
Median		8.00	12.00	10.00
Std. Deviation		3.176	2.386	2.772
Range		14	13	13
Minimum		1	3	2
Maximum		15	16	15
Percentiles	25	6.00	11.00	8.00
	50	8.00	12.00	10.00
	75	11.00	13.00	11.00

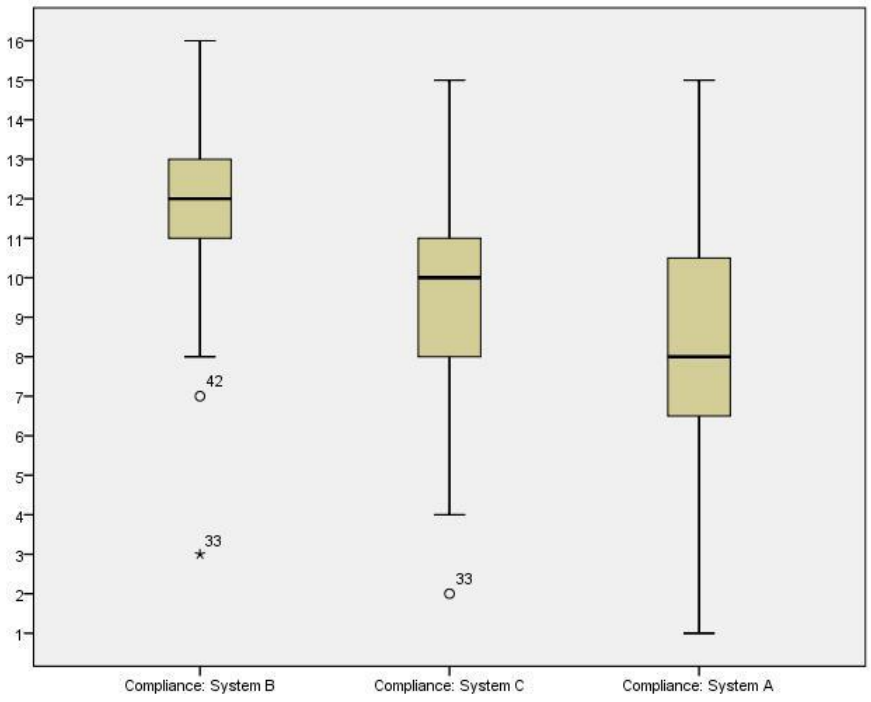


Figure 149. Boxplots for ratings of ‘Compliance’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted inFigure 150. There is a **good** correspondence apparent in the rankings of the three systems between the indicator model of compliance (which was not sensitive) and the consensus view of the judges (with differences in media ratings of systems being significant).

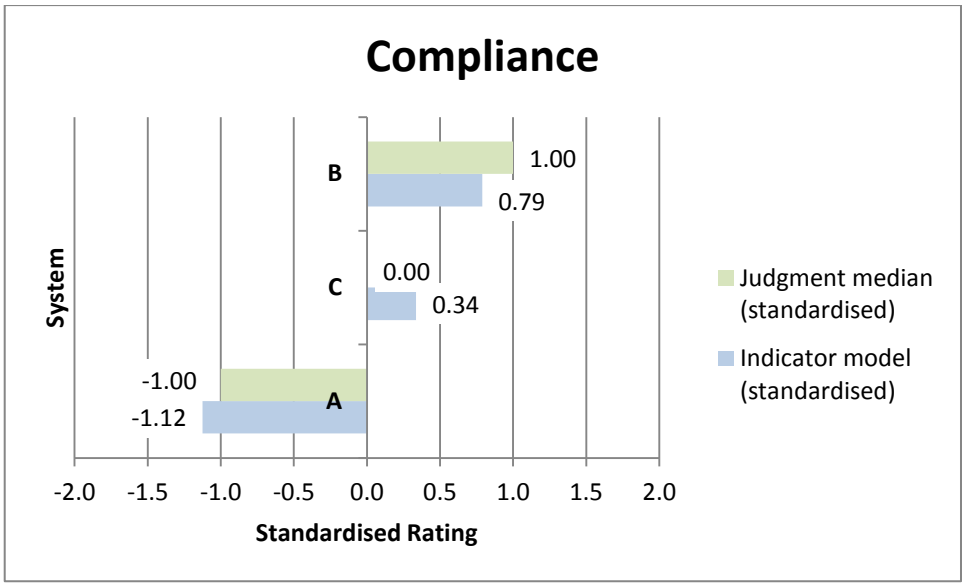


Figure 150. Correspondence of ratings for 'Compliance'.

Given the good correspondence, and the expectation that judges would be likely to have formed their judgment for compliance on the numbers of requirements and their compliance status, a further question arose: do the ratings of the judges equate linearly with the actual numbers of requirements fully-compliant, partially-compliant or non-compliant, or do the judges' apply some other value function to requirement numbers when judging compliance?

To investigate this problem the median of the judges' ratings was plotted against the percentage of requirements in the state of compliant, non-compliant and full and partial compliant as shown in Figure 151, Figure 152 and Figure 153 respectively. The data set is insufficient to establish statistical significance, but assuming linear relationships it was observed that:

- For compliant requirements, if a linear relationship is assumed with an intercept at (0,0), then the data mapped well to a linear relationship that effectively goes to the maximum of the judgment scale when the requirements are fully compliant.

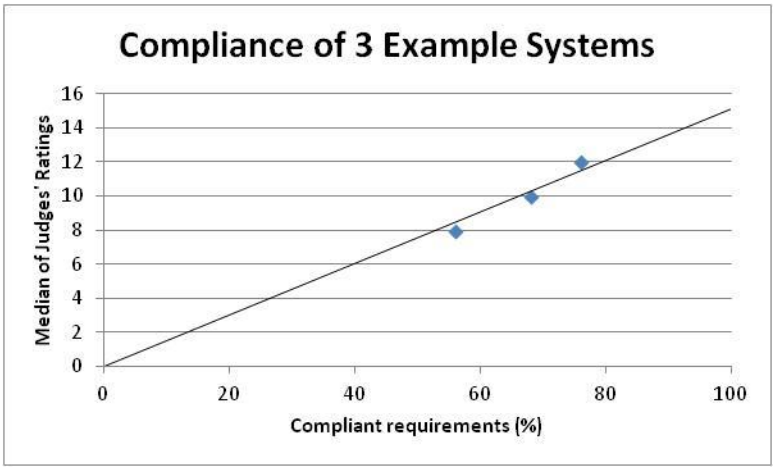


Figure 151. Judges' compliance rating vs compliant requirements.

- For non-compliant requirements, if a linear relationship is assumed with an intercept at (0,16), it was found that when the system compliance would be judged as Not Compliant (judge rating = 0) the number of noncompliant requirements would actually be around 50%.

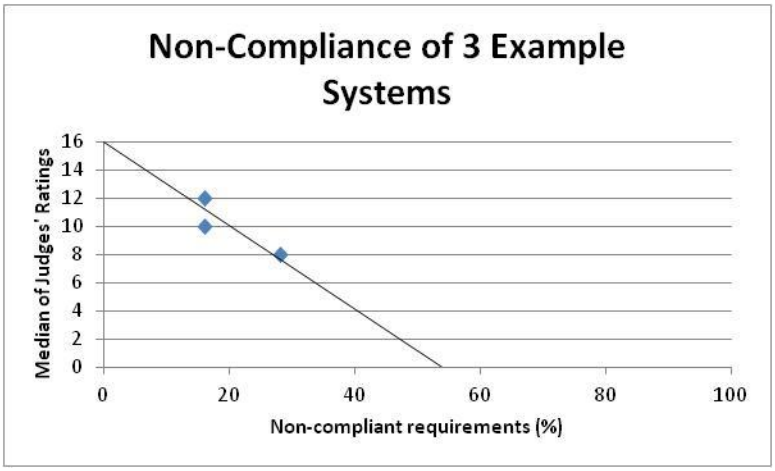


Figure 152. Judges' compliance rating vs number of compliant requirements.

- For a combination of fully and partially complaint requirements, if a linear relationship is assumed with an intercept at (0,0), it was found that a maximum rating of about 13 out of 16 would be expected if there were no non-complaint requirements.

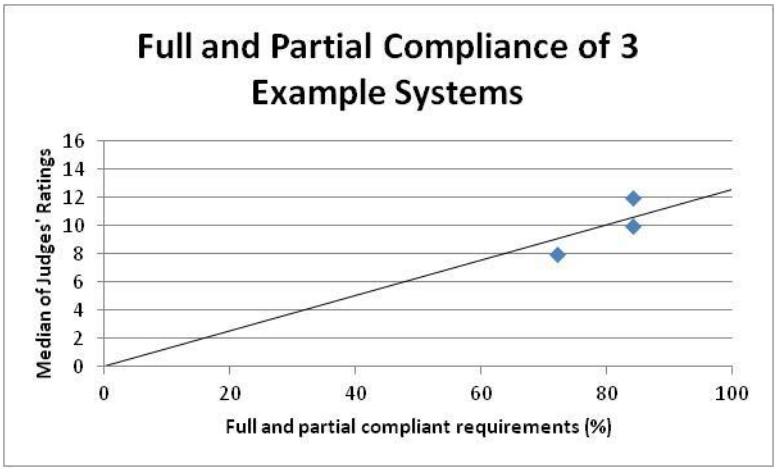


Figure 153. Judges' compliance rating vs full and partial compliant requirements.

These results imply that the percentage of requirements fully-compliant has the best correspondence with the observed ratings of the judges.

8.3.4.17 Usability

As indicated in Table 46, it was expected that the "Usability" indicator would not correlate directly with any technical parameter of the system. It was considered that the technology description of the human machine interface may have some influence on this indicator so a simple indicator model was constructed with resultant ratings shown in Figure 154. The minimum difference in ratings is 49% between System B and System C.

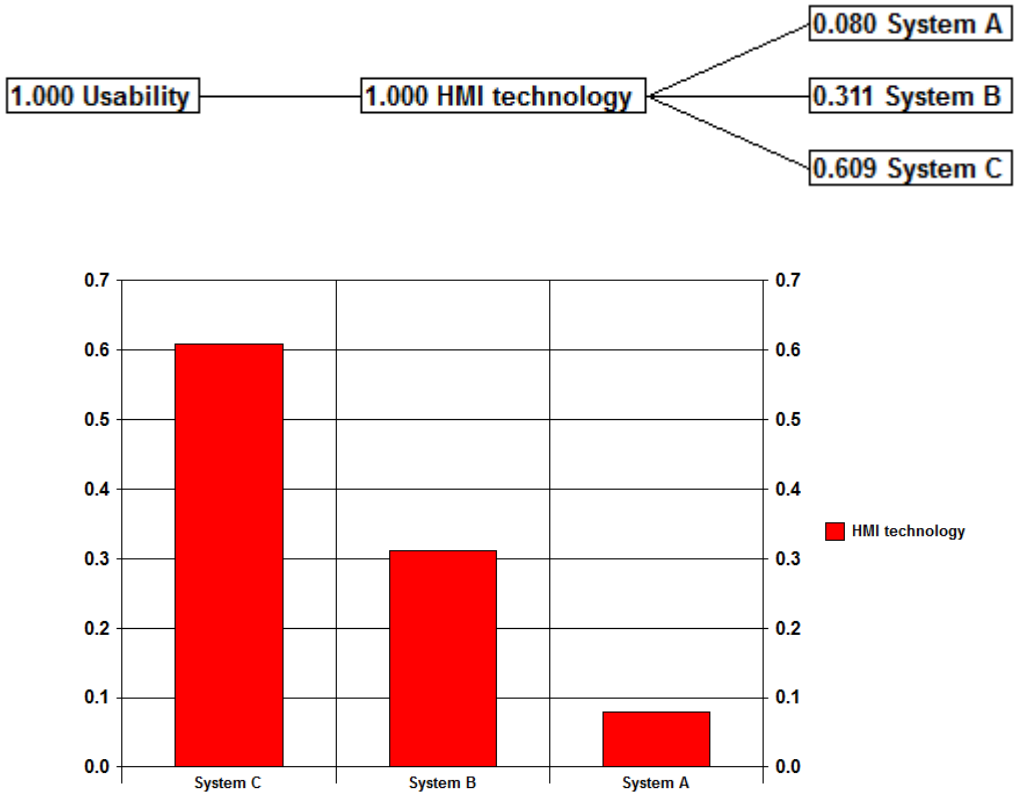


Figure 154. Indicator model and ratings for 'Usability'.

Results from the judgment rating data for "Usable" are presented in the statistics summary at Table 67 and graphically in the boxplot at Figure 155. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the medians of the judges' ratings are significantly different for each system, except for the System B and System C where the difference is not significant at the 0.05 level (significance is at 0.084).

Table 67. Descriptive statistics for ratings of ‘Usability’.

Statistics

		Usability: System A	Usability: System B	Usability: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		8.26	11.84	11.16
Median		9.00	12.00	11.00
Std. Deviation		3.117	2.319	2.181
Range		12	10	8
Minimum		3	6	7
Maximum		15	16	15
Percentiles	25	6.00	11.00	10.00
	50	9.00	12.00	11.00
	75	10.00	14.00	13.00

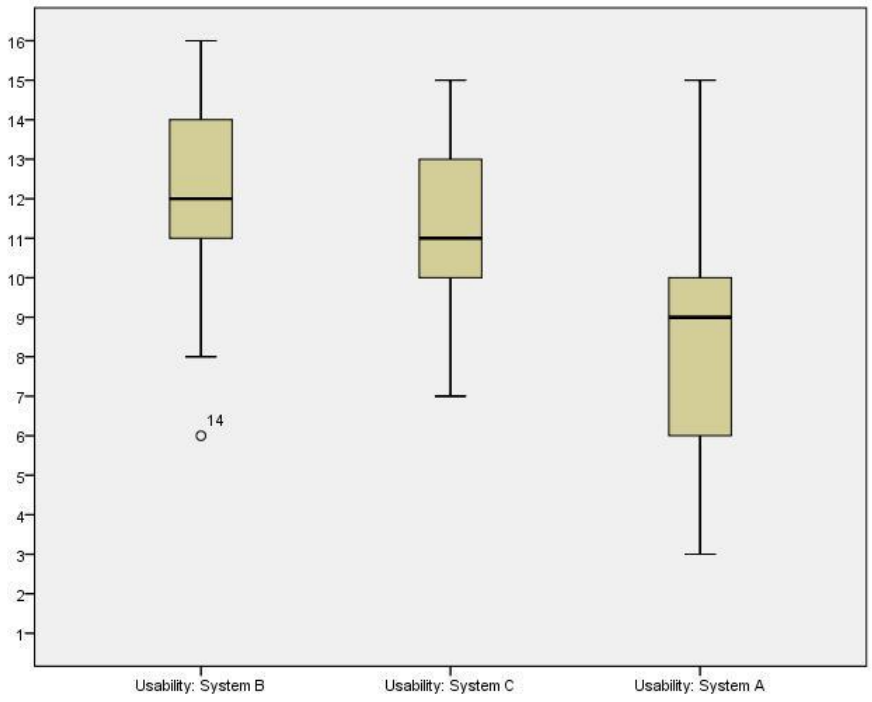


Figure 155. Boxplots for ratings of ‘Usability’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 98. There is a **marginal** correspondence apparent in the rankings of the three systems between the indicator model of usability and the consensus view of the judges. There is agreement in ranking System C as the lowest and the model is not sensitive. While the judgment medians are not significantly different for System B and System C, the indicator model produced widely different ratings. This indicates the

judges were using data other than the currency of the HMI technology to assess the indicator usability.

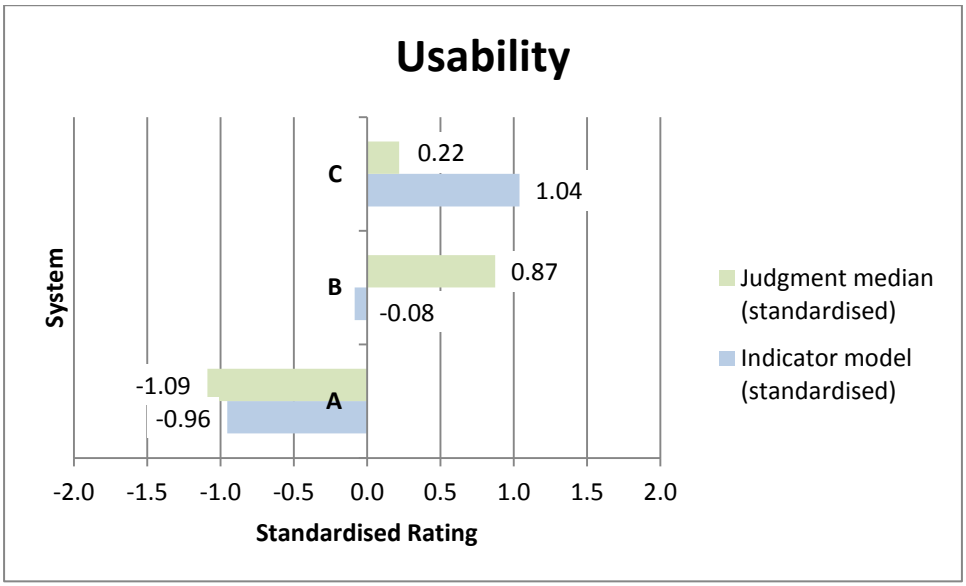


Figure 156. Correspondence of ratings for ‘Usability’.

8.3.4.18 Security

As indicated inTable 46, it was expected that the "Security" indicator would correlate primarily with quality variables physical security and computer/information. However no discrimination in the values of the 3 systems was provided to the judges implying an expected rating of System A= System B = System C.

Results from the judgment rating data for "Security" are presented in the statistics summary at Table 68 and graphically in the boxplot at Figure 157. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the median of the judges' rating for System A was significantly different from System B and System C, but there is no difference between System B and System C.

Table 68. Descriptive statistics for ratings of 'Security'.

Statistics

		Security: System A	Security: System B	Security: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		8.47	10.44	10.42
Median		10.00	11.00	11.00
Std. Deviation		4.306	3.411	3.088
Range		14	13	12
Minimum		1	3	3
Maximum		15	16	15
Percentiles	25	4.00	9.00	9.00
	50	10.00	11.00	11.00
	75	12.00	13.00	13.00

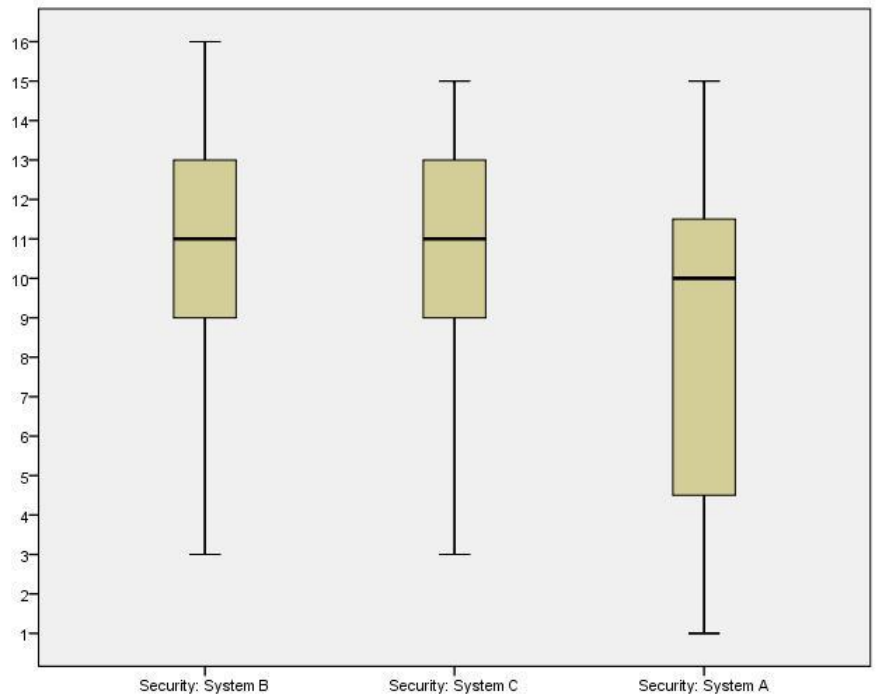


Figure 157. Boxplots for ratings of ‘Security’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 158. The indicator model anticipated no difference in ratings while the judgments indicate System A is to be ranked lowest. Given two of the systems were ranked together by the median of judgments, this result is to be considered a **marginal** correspondence to the indicator model.



Figure 158. Correspondence of ratings for ‘Security’.

8.3.4.19 Technical Complexity

As indicated in Table 46, the indicator "Technical Complexity" had no correlating variables identified. The only point of discrimination that the judges may have used is a subjective assessment of the block diagrams and technology descriptions. No *a priori* ranking was theorised for this variable.

Results from the judgment rating data for "Technical Complexity" are presented in the statistics summary at Table 69 and graphically in the boxplot at Figure 159. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the median of the judges' rating for System A was significantly different from System B and System C, but there is no difference between System B and System C.

Table 69. Descriptive statistics for ratings of 'Technical complexity'.

Statistics

		Technical Complexity: System A	Technical Complexity: System B	Technical Complexity: System C
N	Valid	43	42	42
	Missing	0	1	1
Mean		7.95	10.60	10.71
Median		8.00	11.00	11.00
Std. Deviation		3.109	2.651	2.371
Range		12	11	9
Minimum		2	5	6
Maximum		14	16	15
Percentiles	25	5.00	8.75	8.00
	50	8.00	11.00	11.00
	75	10.00	13.00	12.25

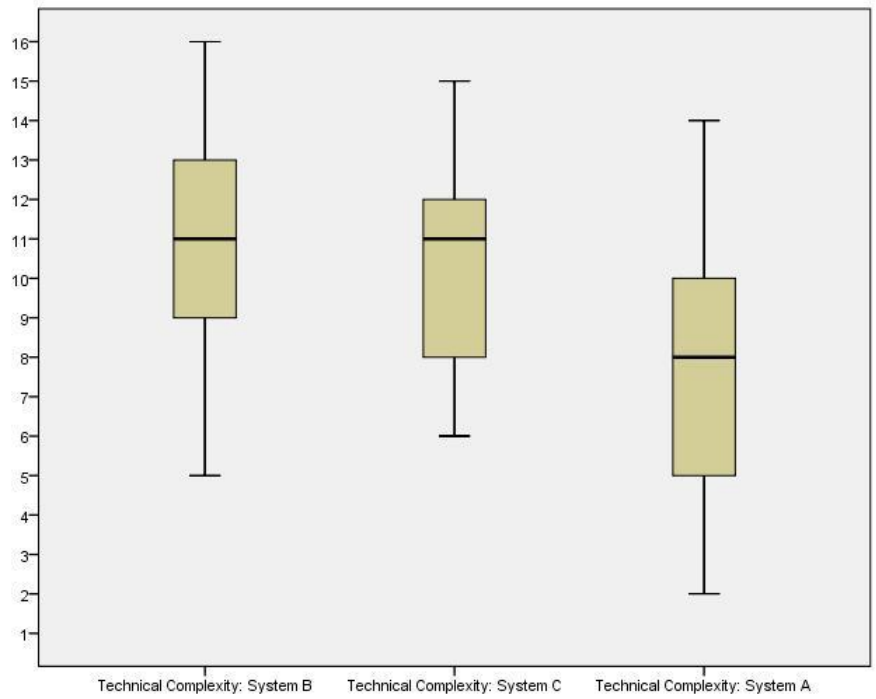


Figure 159. Boxplots for ratings of ‘Technical Complexity’.

The judges’ ratings established that the consensus view is the technical complexity of System B and System C are greater than that of System A, the former two being comparably judged. It is not clear that a subjective assessment of the block diagrams alone would lead to these ratings.

8.3.4.20 High Technology

As indicated in Table 46, it was expected that the "High Technology" indicator would correlate primarily with quality variables for radio frequency design, signal processing, and HMI. The indicator model and resultant ratings are shown in Figure 160. Inspection of the sensitivity of the model to changes of the weightings shows that the model is sensitive to inverting the ranking of System B and System C with as little as 3.5% change in HMI weighting or 4.5% change in signal processing weighting. The minimum difference in ratings is 3.0% between System B and System C.

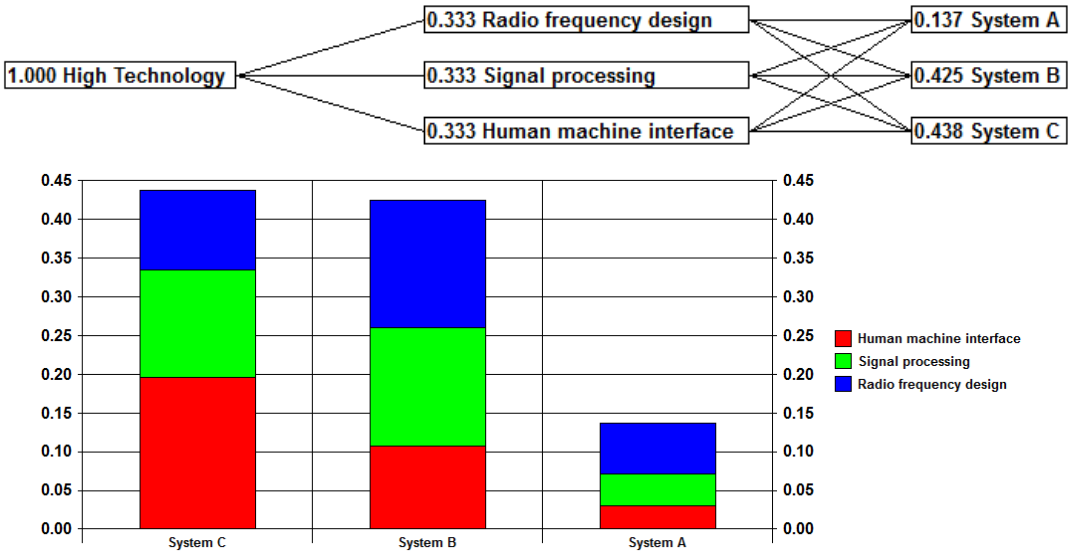


Figure 160. Indicator model and ratings for 'High Technology'.

Results from the judgment rating data for "High Technology" are presented in the statistics summary at Table 51 and graphically in the boxplot at Figure 161. The variables on the boxplot are presented in order of decreasing median left to right. A Wilcoxon signed rank test established that the median of the judges' rating for System A was significantly different from System B and System C, but there is no difference between System B and System C.

Table 70. Descriptive statistics for ratings of ‘High Technology’.

Statistics

		High Technology: System A	High Technology: System B	High Technology: System C
N	Valid	43	43	43
	Missing	0	0	0
Mean		6.86	12.02	11.44
Median		8.00	12.00	12.00
Std. Deviation		3.277	2.474	2.363
Range		13	9	9
Minimum		1	7	6
Maximum		14	16	15
Percentiles	25	4.00	10.00	10.00
	50	8.00	12.00	12.00
	75	9.00	14.00	14.00

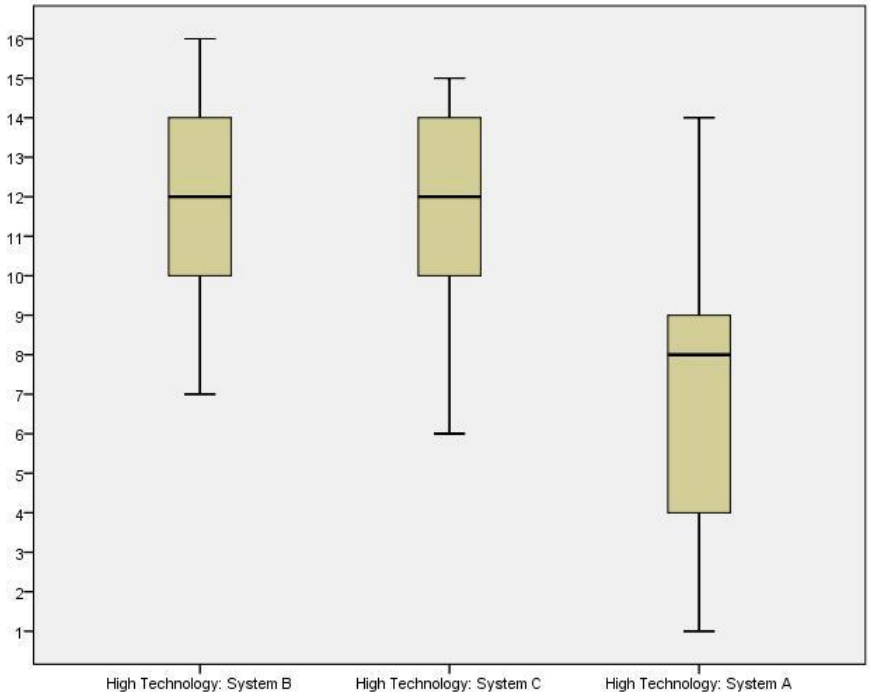


Figure 161. Boxplots for ratings of ‘High Technology’.

The indicator model ratings and the median of the judgment ratings were standardised and plotted in Figure 162. It is observed that the same judgment media of System A and System B corresponds significantly with the indicator model. Therefore it is found that there is a **good** correspondence apparent in the rankings of the three systems between the indicator model of high technology and the consensus view of the judges.

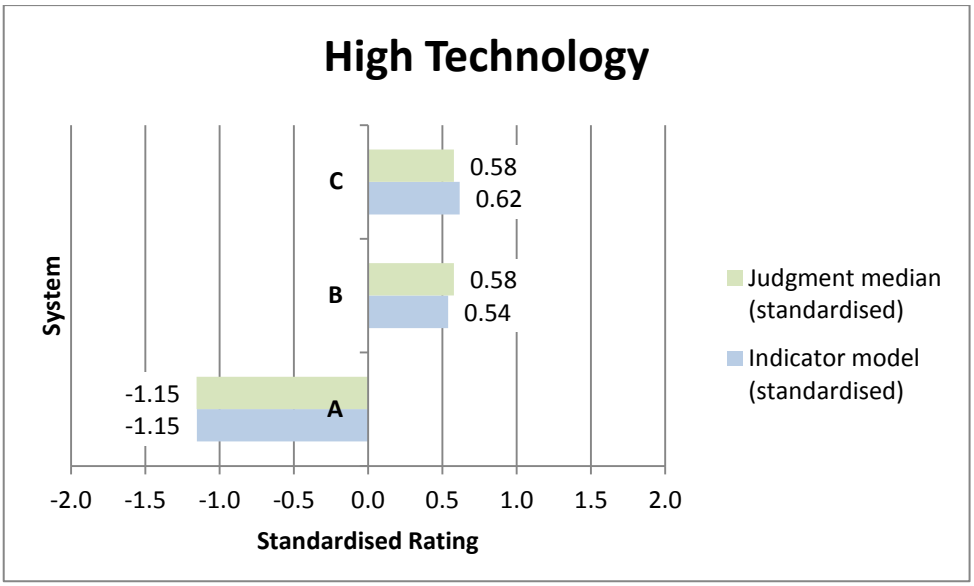


Figure 162. Correspondence of ratings for ‘High Technology’.

8.3.4.21 Summary of correspondence of judgments with objective indicators

Table 71 summarises the correspondence results from the previous sections.

Table 71. Summary of correspondence of judgments with indicator models.

#	Indicator/Judgment Dimension	Correspondence
1	Safety	Good
2	Environment Impact	Marginal
3	Impact to Systems	Poor
4	Fit for Service[13]	Good
5	Capability	Marginal
6	Effective	Poor
7	Efficient	Poor
8	Quality	Marginal
9	Flexible	Reasonable
10	Durable	Poor
11	Survivable	Good
12	Dependable	Reasonable
13	Reliable	Good
14	Supportable	Marginal
15	Maintainable	Reasonable
16	Compliance	Good
17	Usability	Marginal
18	Security	Marginal
19	Technical Complexity	N/A
20	High Technology	Good
	Technical Integrity	N/A

The correspondence is combined with the judgment reliability, first presented in Table 48, in a matrix at Table 72 where the correspondence of judgments and the indicator model are the columns of the grid and the reliability of the CAT judgments makes up the rows of the grid. Each indicator is then shown in the appropriate cell of the grid. The purpose of Table 72 is not to infer a case that something may or may not be a good indicator of technical integrity. Rather it is to show those indicators that can be reliably

[13] Correspondence of the post fact indicator model. A priori indicator model was marginal correspondence.

estimated using the CAT judgment against if that estimate can agree with the indicator model that was constructed from objective data about the system.

Table 72. Indicator correspondence versus judgment reliability.

Reliability of CAT judgment					
	Excellent	Effective Efficient	Environment Impact Supportable Usability	Dependable Maintainable	Survivable Reliable Compliance High Technology
	Good	Impact to Systems	Capability Quality Security		
	Acceptable				Fit for Service
	Questionable				Safety
	Poor			Flexible	
	Unacceptable	Durable			
		Poor	**Marginal**	**Reasonable**	**Good**
		Correspondence of judgment and indicator model			

Consideration of Table 72 provided the following insights:

- The lower rows are where there was large variability in the judges' ratings. It is observed that these rows contain the indicators that are more subjective or open to interpretation, e.g., durable and flexible.
- The upper rows are where there was relatively low variability in the judges' ratings. This implied that there was a mutual understanding between the judges of what the indicator entailed and how to estimate it.
- The indicators in the upper right grid cells had both excellent reliability of the judgments under the CAT and have reasonable to good correspondence with the indicator models for the example systems. It is likely then that the judges used much the same technical parameters in arriving at their rating as was used in the indicator model. E.g., for reliability, the judges almost certainly rated the systems on MTBF as did the indicator model.

- The indicators in the upper left grid cells had excellent reliability, but only poor to marginal correspondence with the indicator model ratings. This implies that the judges did not use the same data, or used other data in addition to or instead of, or applied significantly different criteria weightings in their mental model compared to what had been used in the indicator model.

8.3.4.22 Selection of practical measurement approach for each indicator

Table 73 sorts the indicators in order of decreasing correspondence and adds the judgment reliability as first presented in Table 48 in order to contrast these and to select the most appropriate measurement approach. The measurement approach best suited to practical measurement of each indicator is shown in the last column. In general, the indicated measurement approach is 'Judgment' if the reliability obtained under the CAT was Excellent, or at least Good, and the correspondence with the indicator model was less than Good or if the indicator model was not able to be generalised across disparate system types or if the indicator model was a readily obtainable technical parameter of a engineered system. The rationale for each selected measurement approach is given in the column titled 'Rationale'.

Table 73. Indicator correspondence, judgment reliability and selection of measurement approach.

#	Indicator/Judgment Dimension	Correspondence	Alpha	Reliability	Rationale	Measurement approach
1	Safety	Good	0.685	Questionable	Good correspondence but indicator model not general enough for different systems. Judgment reliability not sufficient.	Indeterminate
11	Survivable	Good	0.934	Excellent	Good correspondence but indicator model not general enough for different systems. Judgment reliability excellent so adopted.	Judgment
13	Reliable	Good	0.986	Excellent	Good correspondence and excellent judgment reliability. Given indicator model is MTBF and the judgments most likely relied upon this technical parameter and that it is readily available for systems in development it is selected as the basis for measurement.	Based upon MTBF.
16	Compliance	Good	0.976	Excellent	Good correspondence and excellent judgment reliability. Given the correspondence observed between judgments and the percentage of requirements fully compliant, and it is unlikely that the judgment would achieve this reliability without the objective data, the objective measure is adopted.	Based upon percentage of requirements fully compliant.
20	High Technology	Good	0.983	Excellent	Good correspondence but indicator model not general enough for different systems. Judgment reliability excellent so adopted.	Judgment

#	Indicator/Judgment Dimension	Correspondence	Alpha	Reliability	Rationale	Measurement approach
4	Fit for Service	Good	0.790	Acceptable	Marginal correspondence with initial model. Judgment reliability only just acceptable. If indicator model is reduced to compliance then correspondence is good. Use compliance measure. Effectively the same measure as the indicator compliance.	Based upon percentage of requirements fully compliant.
9	Flexible	Reasonable	0.519	Poor	Reasonable correspondence but only poor judgment reliability. The only technical parameter theorised to correlate was ‘multi mission/function features’. This quality variable can provide an ordinal rank. This can be adopted pending any better approach.	Based upon ordinal rank established from multi mission/function features of the system.
12	Dependable	Reasonable	0.959	Excellent	Reasonable correspondence but indicator model not general enough for different systems. Judgment reliability excellent so adopted.	Judgment
15	Maintainable	Reasonable	0.973	Excellent	Reasonable correspondence but indicator model not general enough for different systems. Judgment reliability excellent so adopted.	Judgment
2	Environment Impact	Marginal	0.963	Excellent	Marginal correspondence result but excellent reliability with judgments.	Judgment

#	Indicator/Judgment Dimension	Correspondence	Alpha	Reliability	Rationale	Measurement approach
5	Capability	Marginal	0.823	Good	Marginal correspondence but indicator model not general enough for different systems. Judgment reliability is good so adopted.	Judgment
8	Quality	Marginal	0.847	Good	Marginal correspondence but indicator model not general enough for different systems. Judgment reliability is good so adopted.	Judgment
14	Supportable	Marginal	0.938	Excellent	Marginal correspondence result but excellent reliability with judgments.	Judgment
17	Usability	Marginal	0.973	Excellent	Marginal correspondence result but excellent reliability with judgments.	Judgment
18	Security	Marginal	0.898	Good	Marginal correspondence result but good (near excellent) reliability with judgments.	Judgment
3	Impact to Systems	Poor	0.801	Good	Poor correspondence result but good reliability with judgments.	Judgment
6	Effective	Poor	0.945	Excellent	Poor correspondence result but excellent reliability with judgments.	Judgment
7	Efficient	Poor	0.920	Excellent	Poor correspondence result but excellent reliability with judgments.	Judgment

#	Indicator/Judgment Dimension	Correspondence	Alpha	Reliability	Rationale	Measurement approach
10	Durable	Poor	0.226	Unacceptable	Poor correspondence result and unacceptable reliability with judgments. Conclude that the indicator is not well enough defined to measure either via consensus of judgments or via some combination of measureable technical parameters.	None
19	Technical Complexity	N/A	0.947	Excellent	No indicator model theorised in this research. CAT result is excellent and that approach can be adopted.	Judgment
	Technical Integrity	N/A	0.885	Good	No indicator model established. Judgment reliability is good (close to excellent) so this would be a reasonable measurement approach.	Judgment

8.4 Qualitative Results

This section presents the qualitative analysis of the open-ended survey responses. Generally the analysis proceeded by importing the responses into Dedoose. Codes were defined to pursue particular understandings that underlie the quantitative results. Mixed method visualisation and analysis tools within Dedoose were used to further explore for underlying relationships and meanings within the response data.

The open-ended response questions were posed immediately after the rating of judgment dimension questions. The questions were:

- Has your opinion on the nature of "Technical Integrity" changed as a result of completing this survey?
- Can you comment on the extent to which you relied on objective data provided for the systems to make your judgments?
- Can you reflect and comment on how you made judgments when you felt there was little or no objective data provided?
- Any other comments you would like to make?

8.4.1 The nature of technical integrity measurement

40 of 43 respondents provided a response to the first question. Each response was considered, themes identified and text corresponding with those themes extracted and coded in accordance with Table 36. 19 extracts from these responses were coded and are presented in full in Appendix E. Quantitative results of the coding is shown in Figure 42 with discussion following.

Table 74. Code table – nature of technical integrity measurement

Code Name	Criteria for Coding
NATURE OF TI MEASUREMENT	Overall set of extracts from responses to 'What is the nature of technical integrity and its measurement?'
Subjective	Coded when comments that relate to an agreement that the nature of technical integrity and/or its measurement is to be subjectively assessed.
Ambiguous	Coded when a comment reflects the respondent's concern that technical integrity is ambiguous and therefore difficult to assess.
Quality Factor Correspondence	Coded when a comment indicates that the respondent believes there is a high correspondence of one or more quality factors with technical integrity.

Code Name	Criteria for Coding
Aggregation	Coded when the respondent indicates technical integrity should be an aggregation of other variables.

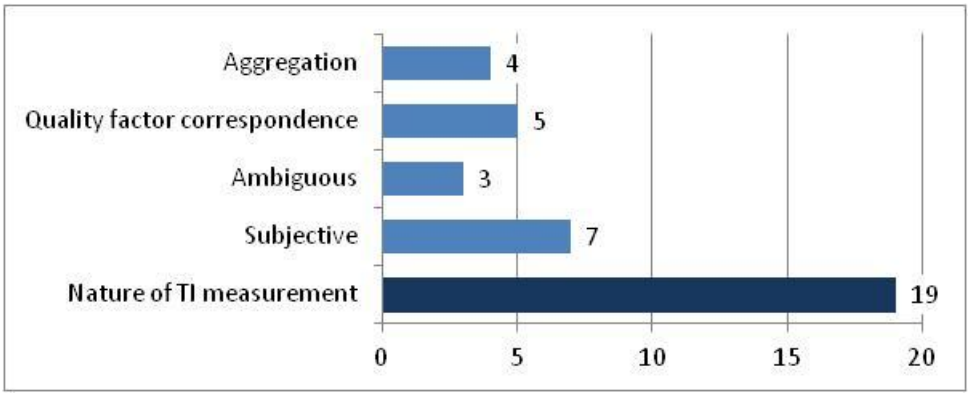


Figure 163. Number of coded extracts – nature of technical integrity measurement.

7 responses clearly indicating technical integrity is considered to be subjective, for example: ‘TI is very subjective and is largely based on the experiences of the "judge" more than simple compliance’ and ‘when it comes to the three arms of technical integrity - safety, fitness for services, and environmental compliance. I think … because of these many potential trade-offs, an overall 'score' for technical integrity may be somewhat subjective.’

5 responses making comment related to the quality factors. One directly equating technical integrity to other judgment dimensions: ‘I changed technical integrity to mean the same as capability, Fitness for Service and Effectiveness. They all scored the same’. Another indicated several judgment dimensions may be assessed similarly: ‘words with slightly different meanings may be assessed similarly’. One respondent identified that the judgment dimensions in the survey were not all expected to correlate with technical integrity: ‘high technical complexity does not necessarily contribute to high technical integrity’. This was intentional in the design of the survey instrument in order to provide data to establish discriminant validity.

4 responses indicated the theme that technical integrity should be regarded as an aggregation of attributes of the system. 3 of the 4 comments indicated that weighting of these attributes by some sort of algorithm would be required, for example: ‘many characteristics of the product contribute to the assessment of technical integrity’. They

do however have different weightings of importance that would somehow have to be taken into account'.

3 respondents indicated they 'feel it [technical integrity] is an ambiguous term'.

8.4.2 Use of objective data in judgments

42 of 43 respondents provided a response to the second question. Each response was considered, themes identified and text corresponding with those themes extracted and coded in accordance with Table 75. 49 extracts from these responses were coded and are presented in full in Appendix E. Quantitative results of the coding is shown in Figure 164 with discussion following.

Table 75. Code table – use of objective data.

Code Name	Criteria for Coding
USE OF OBJECTIVE DATA	Overall set of extracts from responses to 'Can you comment on the extent to which you relied on objective data provided for the systems to make your judgments?' In particular the coding was focused on how is objective data used in rating technical integrity and other system quality factors?
Strongly reliant	Coded when the respondent indicates a primary or strong reliance on the use of the provided objective data to make a judgment.
Reliant but modulated:	Coded when the respondent indicates a reliance on the objective data but modulates that usage with other considerations to make a judgment. Sub-coded to indicated other influences as:
Subjective data	The respondent indicated the use of the qualitative or descriptive data about the system to reach a judgment.
Experience	The respondent indicated the use of experience, or their knowledge of similar systems to reach a judgment.
Context	The respondent indicated the use of knowledge of the context of use of the system was supplemental to reaching a judgment.
Other	Not corresponding to one of the above sub-codes.
Combines data	Coded when the respondent indicated that they used more than one element of objective data and combined that data to form a judgment.

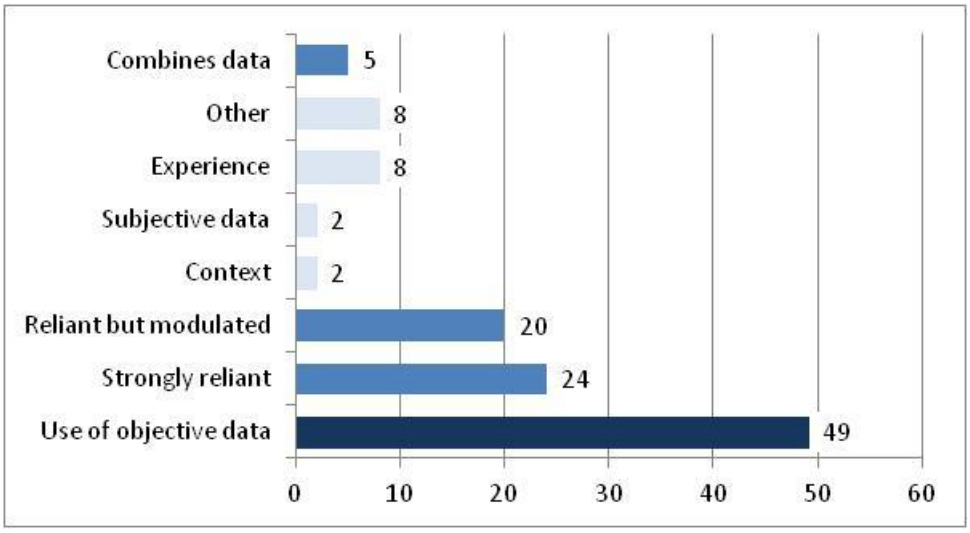


Figure 164. Number of coded extracts – nature of technical integrity measurement.

8.4.2.1 Strongly reliant

The responses strongly indicated that the judges use objective data, when available, to form their judgments. 24 coded responses show a strong reliance on the objective data, as exemplified by:

> I relied very heavily on objective data where appropriate data was provided.
>
> Tried to predominantly use objective data as opposed to overall interpretation and assumptions of the radars to make judgments.
>
> I used the objective data as much as possible.
>
> Where the data was provided I tried to judge based on that data - even though my own "instincts" may have been different.
>
> Where available, objective data used as primary indicator.

This indicates the respondents, attempted at least, to be quite objective in reaching their judgments. Indeed one respondent indicated that given time he would be even more analytically objective:

> If I was completing this survey as a proper engineering task I would spend the time to ensure I was using objective evidence and a complete understanding of each technical area to make my comparisons.

8.4.2.2 Reliant but modulated

20 extracts from the responses were coded as the respondent being reliant on the objective data but their judgments were modulated based on their understanding of the

system's context, their own experience with similar systems, and descriptive data provided about the systems or other subjective considerations.

8.4.2.2.1 Context

The two extracts coded as 'Context' reflect a modulation of the objective data by considering the operational context of the system:

> Where data was conflicting or not conclusive, I used my general experience and knowledge in the context of likely operational uses to form an opinion.
>
> I also tried to go beyond purely the objective data to try and put assessment criteria into context/perspective.

8.4.2.2.2 Subjective data

Two extracts were coded as 'Subjective data' indicating a conscious use of the qualitative or descriptive data about the system to reach a judgment. For example:

> The objective data is used to set the initial ratings, which are then moderated by considering the subjective data.

8.4.2.2.3 Experience

Eight extracts were coded as 'Experience' where the respondent indicated the use of experience, or their knowledge of similar systems to form a judgment. Some themes within these extracts include:

- Use of knowledge, experience and analogy to similar systems when the objective data may be insufficient, inconsistent or inconclusive:

  > where data was conflicting or not conclusive, I used my general experience and knowledge
  >
  > Where no data was available I relied on experience with similar systems.
  >
  > I resorted to my prior experience and knowledge only when the data provided was insufficient

- Another aspect of experience identified by one respondent was the modulation of the objective data itself to reflect experience:

  > In some cases I did use some objectivity - my experience is that MTBFs for commercial equipment (used in a Naval environment) are often far more optimistic (and usually from OEM calcs) than MIL.

- One respondent indicated that experience was brought to bear in order to know what objective data to apply to the judgment dimension:

> The decision to use specific data groups to respond to a question was born out of experience.

8.4.2.2.4 Other

Of the eight extracts that were coded as 'Other' there is no common theme for the influences to modulating the judgment of objective data. There are, however, some insights worth drawing out:

- One respondent indicated the subjective element 'was the value judgment weightings' that made it difficult to trade-off the values of objective elements. This was therefore done subjectively.
- One respondent indicated they 'found it difficult to justify using the same objective data to justify the rankings for different terms'. This is to be expected in the design of the example and the judgment dimensions and indeed the research is intended to identify where the judgment dimensions are redundant and where they may be effectively replaced by objective data proxies. Another respondent indicated he was 'often influenced by MTBF and MTTR' in assessing the judgment dimensions indicating that these may well be significant proxies for some judgment dimensions or elements of a technical integrity scale.
- Some judgment dimensions may require consideration of data across the full-lifecycle of the system, data that is often not available in the design or acquisition phases of a systems lifecycle. For example, one respondent considered that assessment of the impact to the environment of the system would need to include consideration of the disposal phase of the lifecycle:

> Environmental impact is difficult without a full lifecycle analysis of the development of the kit so I put them all close together but simply relied on the Hazmat declaration to separate them (but in reality I suspect construction techniques would make A easier to dismantle and re-use, and given its weight would have a more substantial scrap value.

8.4.2.3 Combines data

Five extracts were coded as 'Combines data' where the respondent indicated that they used more than one element of objective data and combined that data to form a judgment. The comments indicate that two or more elements of objective data were weighted or 'factored' into the judgment. One respondent indicated the judgments were based upon both compliance data and the system technical data: 'Relied on a combination of the specified requirements compliance and the system characteristics'.

8.4.3 Judgment approach with no objective data

40 of 43 respondents provided a response to the third question. Each response was considered, themes identified and text corresponding with those themes extracted and coded in accordance with Table 76. 55 extracts from these responses were coded and are presented in full in Appendix E. Quantitative results of the coding is shown in Figure 165 with discussion following.

Table 76. Code table – judgment approach with no objective data.

Code Name	Criteria for Coding
NO OBJECTIVE DATA APPROACH	Overall set of extracts from responses to 'Can you reflect and comment on how you made judgments when you felt there was little or no objective data provided?' In particular the coding was focused on what judges do when there is no objective data available to assist their judgments.
Rate all equivalently	Coded when the judge indicated that in the absence of objective data they would tend to rate all systems more or less the same.
Subjective evaluation	Coded when the judge indicated he/she would still provided a subjective rating for the dimension. Sub-coded to indicated other influences as:
Context and use	Coded when a response indicated subjective evaluation based on knowledge of the context the system is to be operated within.
Experience	Coded when a response indicated subjective evaluation based on experience of similar systems; or analogy to similar systems.
Assumptions	Coded when a response indicated subjective evaluation based upon assumptions made by the judge.
Technology	Coded when a response indicated subjective evaluation made on the basis of age and type of technology used in the system.
Risk	Coded when a response indicated subjective evaluation of risk associated with lack of objective data being available to characterise the system.
Use related data	Coded when the judge indicated he/she attempted to use the most related objective data to make the judgment.
Would not rate	Coded when judge indicated he/she did not want to or is reticent to rate a dimension when no objective data is available.

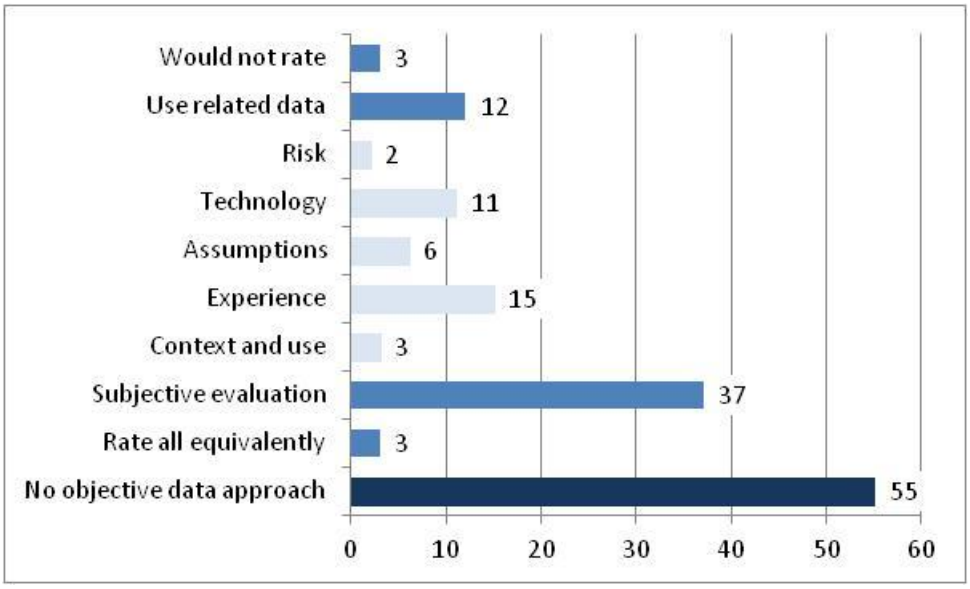


Figure 165. Number of coded extracts – judgment approach with no objective data.

8.4.3.1 Rate all equivalently

Three respondents indicated that, in the absence of objective data their approach was to rate the systems closely together noting that the survey instrument, intentionally by its design, did not allow for the same rating to be given to two or more systems on a judgment dimension:

> Very often where there was little to go on the rankings were clustered together, because we were not allowed to not differentiate.

8.4.3.2 Subjective evaluation

37 respondents indicated they made subjective judgments in the absence of objective data. Some respondents indicated that this was a difficult aspect for them; 'With difficulty!', while at the other end of the spectrum there were indications that 'I had no problems making subjective judgments'. There was clear acknowledgement from several respondents that subjective judgments were considered inferior to the use of objective data:

> Such assessments could not be justified; the rankings are opinions, and unlikely to be coherent. Given that the survey prohibited equal rankings, "I don't know" was excluded from the response possibilities. I had to make a decision, no matter how indefensible it was.

> Where objective data was not available, ranking is made by experience and trust in previously utilised systems encountered in my career, and experiences, results, and trust of these items and their performance. Again, this is entirely subjective, and not coherent.

While others acknowledged that making engineering judgments without all desired data this was often the reality in practice: ‘Somewhat frustratingly, which I suppose can be reflective of reality’.

8.4.3.2.1 Context and use

Three respondents indicated they used the understanding of the system’s likely operational use in forming a subjective judgment. For example:

> Judgments were made based on my understanding of the operational environment and the way the system is used operationally, taking into consideration the different attributes that may confer tactical and operational flexibility.

8.4.3.2.2 Experience

15 of the extracts indicate a reliance on experience to make subjective judgments. This reliance ranges from ‘Guesswork based on experience’ and ‘Gut feel’ to thoughtful reliance ‘on experience of similar systems capabilities and limitation’.

8.4.3.2.3 Assumptions

Six extracts indicated the respondents resorted to assumptions or guesswork in the absence of objective data. For the most part these assumptions were based upon extrapolations from the system descriptions. For example:

> I used my best guess about what characteristics products of that type have.

> Previous experience with each of the different genres of radar ("old" MIL, commercial and "current" MIL) which provided assumptions (correct or otherwise) on likely performance/reliability.

8.4.3.2.4 Technology

Eleven extracts indicated that a subjective evaluation was made on the basis of age and type of technology used in the system. Three distinct themes were evident in the comments: age of technology, civil versus military technology, and the quantity of the systems in service.

Some comments indicated they were influenced by the age of technology reflecting the quality of the system: ‘recent technology and implementation suggested higher quality (given modern quality systems)’. Others indicate the age of technology could also influence the ratings on some judgment dimensions in a more balanced way:

> The lower technology level (but more mature technology) along with less “bells and whistles” suggested a system that could and would be well understood.

Similarly consideration of military versus civil construction standards and ‘on number of units in service (even though no evidence on how well those units were performing)’ influenced the ratings.

8.4.3.2.5 Risk

Two respondents indicated risk considerations in reaching a judgment without objective data.

8.4.3.3 Use of related data

Twelve respondents indicated the attempt to use related data as best they could to rate judgment dimensions when objective data was not available. One respondent indicated a deferral to the general level of requirement compliance. Others selected one or more objective data elements as closely reflecting the judgment dimension as possible and/or augmented that with assessment of the descriptive data of the systems:

> My judgments were based on aligning what I considered the relevant characteristics of the systems were that related to the judgment category.
>
> The pictures actually provided me with a "feeling" about the system and that may have influenced the judging. Generally I tried to find at least one objective data item that allowed a justification of the judgment. If one system did not provide the objective data, then the distance from the others was not large as I assumed that they would be similar.

8.4.3.4 Would not rate

Three respondents clearly indicate their preference to not rate the systems at all or had no confidence in their own judgments in the absence of adequate objective data.

> I felt a bit uncomfortable about areas where there was little of no objective data, particularly areas that are not my expertise.
>
> Making judgments against subjective criteria was difficult. There was little objective data provided and this raised a lot of questions for me about context and interpretation of the assessment criteria.

8.4.4 Views on obtaining better ratings

27 of 43 respondents provided a response to the fourth question. Each response was considered, themes identified and text corresponding with those themes extracted and coded in accordance with Table 77. 15 extracts from these responses were coded and are presented in full in Appendix E. Quantitative results of the coding is shown in Figure 166 with discussion following.

Table 77. Code table – views on obtaining better ratings.

Code Name	Criteria for Coding
BETTER RATINGS	Overall set of extracts from responses: 'Any other comments you would like to make?' In particular, the coding was focused on what else the judges identified that may allow for better rating of technical integrity and other judgment dimensions/quality attributes of the system.
Definitions	Coded where the comments indicated that the provision of definitions of the judgment dimensions would make for better ratings.
Group judgments	Coded where the comments indicate an approach of group based or consensus judgments may make for better ratings.
System data	Coded where comments indicated another type of data about the system would improve rating ability.
Rating scales	Coded where comments indicated changes to ratings scales may be considered. Number of rating points, allowing ties, etc.

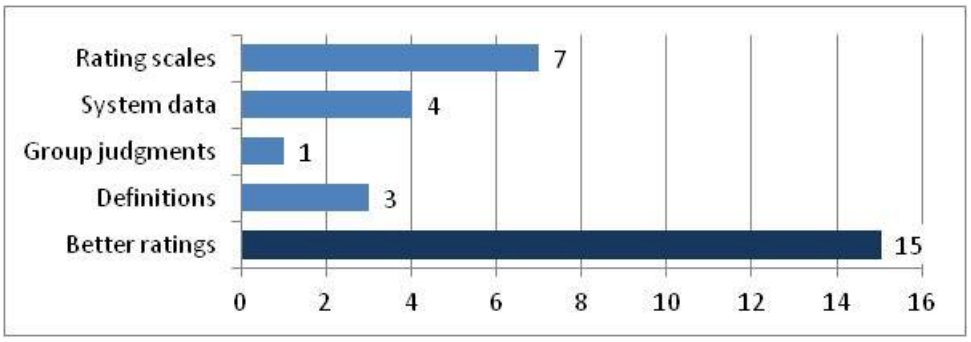


Figure 166. Number of coded extracts – views on obtaining better ratings.

8.4.4.1 Definitions

Three respondents provide comments that indicated they struggled to fully discern the meaning of some judgment dimensions and about the difference between the judgment dimensions:

> I did struggle in thinking about the differences between: Effectiveness vs Capability, Dependability vs Reliability vs Durability, and Technical Complexity vs High Technology.

'Durable' was cited in all three extracts as being difficult to define and assess. The implication from these comments is that definitions be provided for each of the judgment dimensions. Although the survey instrument intentionally did not provide definition for the judgment dimensions in accordance with the CAT, it may be a valid consideration for a scale-based measurement instrument.

8.4.4.2 Group judgments

One comment indicated a group approach to judgments should be considered:

> It would benefit from a group approach so that domain knowledge shortfalls could be compensated for. At the end of group discussion each member could then marked independently to allow for divergences of view to be catered for.

This would be contrary to the CAT approach upon which the survey instrument was based, but may be a valid direction for future research where recognised techniques like Delphi method (Skulmoski et al., 2007) may be evaluated for technical integrity measurement.

8.4.4.3 System data

Four comments were coded that requested additional system data be provided to support judgments. Two of these suggested an explicit statement of the intended use of the system would allow for better judgments;

> Very difficult to make subjective decisions without knowing the purpose of the system. Some idea can be assumed from the requirements but still better to know its intended purpose and environment.

This is a good suggestion that should be adopted in future studies. Intended use is of course key to the assessment of fitness for purpose and similar judgment dimensions. Fit for *intended* purpose must be considered, not fitness for *any* purpose.

One comment suggests the price of the system be provided to assist in judgments:

> I think the missing piece of the puzzle is the price (i.e. market value, in dollars). Knowing the relative price of the three radars would have helped answer a number of the questions, especially some of the ones on capability and effectiveness. Extending that line of logic, if I wanted to check the level of technical integrity of something I would look to the market price. How do I compare the value of an iPhone to a pencil? A house to an education? A horse to a bomb? Any one thing to something else?.

While the technical integrity, and other judgment dimensions, may well correlate with price (i.e., once could reasonably expect high integrity systems to cost more to design and produce than lower integrity systems) it is dismissed by *reductio ad absurdum* that price or cost not be considered as a reasonable basis for assessing technical integrity since spending more money is no causal to increased technical integrity.

8.4.4.4 Rating scales

Seven responses were coded that indicated some change to the rating scales used in the survey instrument is considered beneficial. Three of the comments reflected that the 15 points of the scale were too many and another comment reflecting that only relative scoring is considered important. Two comments suggested that equal ratings be allowed.

These aspects the survey instrument had been considered in its design as discussed previously in the method section.

8.5 Discussion

This section takes the quantitative results from section 8.3 and discusses the implications and outcomes afforded by these results. Where possible the qualitative results from section 8.4 are used in support of the discussion.

8.5.1 Reliability of using the CAT for technical integrity measurement

The reliability of a measurement instrument is an assessment as to the degree to which measured values are correlated with the true value of the property. In many cases, including technical integrity, where the value of the property cannot be directly observed or estimated, reliability assessment focuses on the correlation of observed scores on parallel applications of the instrument. That is, does the instrument objectively provide the same value when used by different people and/or at different times[14]?

(Amabile, 1996, p. 69-71) reports the reliability results of studies using the CAT across three domains with a range of Cronbach's alpha values as follows:

- Verbal creativity (e.g., stories and prose) – 0.67 to 0.91.
- Artistic creativity (e.g., drawings, paintings, sketches) – 0.53 to 0.92.
- Problem-solving creativity (e.g., computer programs, high-technology services, structures) – 0.61 to 0.87.

14 The reader is referred to ALLEN, M. J. & YEN, W. M. 1979. *Introduction to measurement theory*, Brooks/Cole Monterey, CA. and DEVELLIS, R. F. 2011. *Scale development: Theory and applications*, Sage Publications, Incorporated. for a detailed description of measurement reliability.

In this thesis, a value of Cronbach's alpha of 0.865 was obtained for ratings of technical integrity across the three example systems. This compares well with the higher end of the range of values reported by these other studies. An alpha of greater than 0.70 is generally considered to be an acceptable reliability for a measurement instrument for most social science research. The indirect measurement of technical integrity via expert judgment is reliable for the type of system used in this survey research at least. Although this study was limited in scope to a particular type of engineered system (maritime radars), this is substantive support for the assessment of technical integrity via the CAT.

Most of the coded responses in section 8.4.1 indicated that technical integrity is a subjective concept and therefore apt for estimation by judgment. However, several respondents did indicate that some type of weighted aggregation of technical parameters would be more appropriate but they were not specific as to how this may be practically achieved.

8.5.1.1 Correlation of judgment dimensions

The application of the CAT also involves the judgment of related indicators. In this research, 18 other expected indicators of technical integrity were gathered. The correlation of technical integrity ratings with ratings of these indicators provided some unexpected results. Most notably, the indicators of safety and environmental impact (both of which are explicitly in the ADF definition of technical integrity) were weakly correlated with technical integrity (r_s = 0.302 and 0.245 respectively). As was expected, the indicator Fit for Service, explicit in the ADF definition of technical integrity showed the highest correlation with technical integrity (r_s = 0.656).

The technical integrity judgments are positively correlated with all other judgment dimensions, however, there are no strong correlations[15]. The largest correlations are with Fitness for Service (r_S = 0.656) and Capability (r_S = 0.629) with other moderate correlations[16] being effectiveness, reliability, dependability, flexibility and compliance. The qualitative results at section 8.4.1 showed that at least one respondent found technical integrity, capability, fitness for service and effectiveness to be the same

15 A correlation is generally interpreted to be 'strong' when Spearman's rho is greater than 0.7.

16 A correlation is generally interpreted to be 'moderate' when Spearman's rho is in the range 0.5 to 0.7.

phenomena. Correlation of these judgment dimensions with technical integrity is as expected. 'High technology' also shows moderate correlation with technical integrity in this study. It is not expected that the level of technology employed in a system should be directly related to the technical integrity of the system and it would not have been surprising if this showed little or no correlation. This was also noticed by one respondent as indicated in section 8.4.1. Equally however, it may be readily construed that newer technology systems may often be considered to be technically "better" than older technology systems.

The judgment dimensions of safety, impact to the natural environment, and survivability all show a weak correlation[17] with technical integrity ratings. The results for safety and impact to the natural environment are not as expected given that these judgment dimensions directly correspond to the elements of technical integrity explicitly defined in ADF technical regulations.

The other judgment dimensions exhibit a low correlation with technical integrity. The result that 12 of the 20 judgment dimensions have weak to low correlation with technical integrity suggests that inclusion of these dimensions as variables in a scale for technical integrity may not be warranted.

8.5.2 Technical integrity scale reliability

The 18 indicators of technical integrity and 2 other indicators, not expected to correlate with technical integrity, were combined into a scale for technical integrity. Analysis using the expert ratings indicated that a more reliable estimate of technical integrity may be obtained by use of this scale (Cronbach's alpha = 0.917) than by indirect estimation of technical integrity alone (Cronbach's alpha = 0.885). However, the improvement in reliability comes at significantly more effort, i.e., judges rating 20 indicators vice a single rating of technical integrity.

The identified example of using a scale to estimate a system characteristic with some similarities to technical integrity was SUS for usability (refer section 3.4.8) claimed to be 'robust and reliable' but without any statistical results that may be compared.

17 A correlation is generally interpreted to be 'weak' when Spearman's rho is in the range 0.16 to 0.3.

8.5.3 Correspondence of indicators with expert judgments

Indicator models were developed for each indicator based upon both objective variables and values and subjective ranking data for the three example systems. Consideration of the correspondence of these models with the judgments gained through the CAT provided the following conclusions:

- Where a system quality attribute has clearly identifiable quantity variables associated with it, the reliability of judgments also tends to be excellent and the correspondence between the two is good. The quality attributes of reliability and technical compliance are examples. This is readily explained by understanding that the judge can readily access, understand and use the same numerical values in support of their judgments. The number of respondents who indicated they are reliant on objective data in section 8.4.2 supports this explanation. The rating of systems on this quality attribute should be accomplished by obtaining the actual numerical values for the systems (since the judges would require this to provide such reliable ratings in any case).
- Excellent judgment reliability is still evident for many system quality attributes even when the results of the indicator model do not completely correspond. The quality attributes of supportability, maintainability and usability are examples. These quality attributes are well defined and agreed upon terms but the corresponding indicator models generally consist of several quantity and quality variables. The result may be understood as the judge being able to better combine a system's values into a reliable ranking than is achieved with a weighted additive model of variables. Or, as indicated by the results of section 8.4.2.2 where many respondents supplement the provided system data with an understanding of the system's context of use, their own experience with similar systems and descriptive data about the system, the rating of systems on these system quality attributes should be accomplished by utilising competent judgments.
- Further, excellent judgment reliability is still evident for some system quality attributes even when the results of the indicator model do not correspond. The quality attributes of effectiveness and efficiency are examples. In these cases there may be none or only weakly related variables that could be used in an indicator model or little difference in the values of the variables with which to differentiate

the systems. The rating of systems on this quality attribute should be accomplished by utilising expert judgment.

- Some system quality attributes exhibit poor to unacceptable judgment reliability and poor correspondence with the results of the indicator model. The quality attribute durability is the primary example. In the qualitative results of section 8.4.4.1, three respondents cited durable as being difficult to define and assess. Durability is more subjective than most of the other system quality attributes and is open to relatively wide interpretation explaining the low reliability of judgment ratings. This is supported by the qualitative results at section 8.4.3 that indicated purely subjective estimates are frequently made when objective data is not available to assist in the judgments, but other approaches like rating all systems the same are also used by some respondents. Also there is little in the way of clearly associated quality variables upon which to base a judgment or an indicator model upon. Consequently this type of system quality attributes may not be able to be reliably estimated.
- The system quality attribute safety had only questionable level of reliability in judges' ratings but there was good correspondence with an indicator model for safety. However the indicator model was specific to the example systems and not readily generalised to other types of systems. Safety, while being an often used term, is not often objectively assessed (at least not in a total system sense) and there is no clear and general indicator model from readily obtainable quantity or quality variables. Safety may not be able to be reliably estimated, but given the oft lack of numerical data, it may be that expert judgment be the nominal mechanism for practical estimation.

8.5.4 Safety

Previous phases of research highlighted responses that indicated that the indicator safety may dominate the estimation of the dependability component of technical integrity or that it may be better to consider it as an additional third component of technical integrity (refer to sections 6.6.3 and 7.6.3). The results of section 8.3.1.3 indicated that the judgments of the indicator safety were not strongly correlated with the judgments of technical integrity for these example systems. This suggests that it is not a dominant contributor to technical integrity or to the concept of dependability and it is further implied that it will not for a third component of technical integrity. This latter implication will be further explored in Chapter 9.

8.6 Conclusion

This section begins by restating each of the research questions developed in section 8.1.2 and providing a summary answer. The overall conclusion and the intent for the next chapter are then presented.

How reliable can direct expert ratings of technical integrity be? Rating of technical integrity using the CAT was conducted on three example systems and good reliability was obtained (Cronbach's alpha of 0.865).

How reliable can expert ratings for other system quality attributes be? Twenty other system quality attributes were subject to ratings using the CAT on the same three example systems. Twelve of these showed excellent reliability (Cronbach's alpha greater than 0.9).

Can the judgments of the quality attributes form a reliable scale for technical integrity? The 18 system quality attributes expected to be indicators of technical integrity were combined into a scale. The scale reliability was excellent (Cronbach's alpha of 0.917) and was improved compared to the direct rating of technical integrity (Cronbach's alpha of 0.865).

Where values and information are available for technical parameters of the system, how well do these values and expert ratings correspond? Indicator models were developed for 19 of the 20 system quality attributes based on technical parameters of the example systems. The ratings of these models were compared to the rankings established from the combined ratings of the survey respondents. Results were mixed with 9 of 19 models showing reasonable to good correspondence with the judges' ratings with the other 10 of 19 showing poor to marginal correspondence.

What is the best measurement instrument to be used for each indicator (via empirical estimates of technical parameters or via expert judgment)? In comparing the results of correspondence of the indicator models of technical parameters with judges ratings and the reliability of the judges' ratings a decision as to the most practical approach to estimate each of the 20 indicators was made. 14 of the indicators were found to be best estimated by way of judges' ratings and 4 are best measured by way of technical parameters of the system. For one indicator, safety, the best method is indeterminate

from the data available. For one other indicator, durable, there appears to be no practical or reliable measurement approach.

The judgments of the indicator safety were not strongly correlated with the judgments of technical integrity for these example systems, despite being an explicitly defined element of the ADF definition of technical integrity. This suggests, counter to respondent comments, that it is not a dominant contributor to technical integrity or to the concept of dependability

While it has been found that the CAT is a reliable measurement approach for the estimation of technical integrity, analysis of a scale constructed from the expected indicators of technical integrity indicated that higher reliability was possible. However, the measurement scale alone does not account for or take advantage of the theory of technical integrity developed in Chapter 5. Chapter 9 seeks to confirm the theory of technical integrity from the survey ratings data and by synthesising and evaluating four candidate measurement instruments for technical integrity using the results obtained in this chapter.

9 Technical integrity theory and measurement instrument

9.1 Introduction

In Chapter 6 it was established that it was valid to form a conjoint aggregation of the concepts of capability and dependability into a measure of technical integrity. Chapter 7 extended this to look at the relative importance of capability and dependability to arrive at weightings for the aggregation. Chapter 8 assessed the consensual assessment technique for estimation of technical integrity and for a large number of the expected indicators of technical integrity. It found that the reliability of directly assessing technical integrity could be good and that a scale formed from indicators could provide excellent reliability. It also compared the reliability of assessing the indicators via the CAT to ratings produced by indicator models constructed from technical parameters of the example systems thereby nominating a preferred approach for the practical measurement of each indicator.

The influence diagram, first introduced in Chapter 8, Figure 71 is extended to include the conjoint measurement construct of technical integrity in Figure 167. In this diagram the estimated values of quality attributes may be combined into scales or other aggregation constructs for one or both of the concepts capability and dependability.

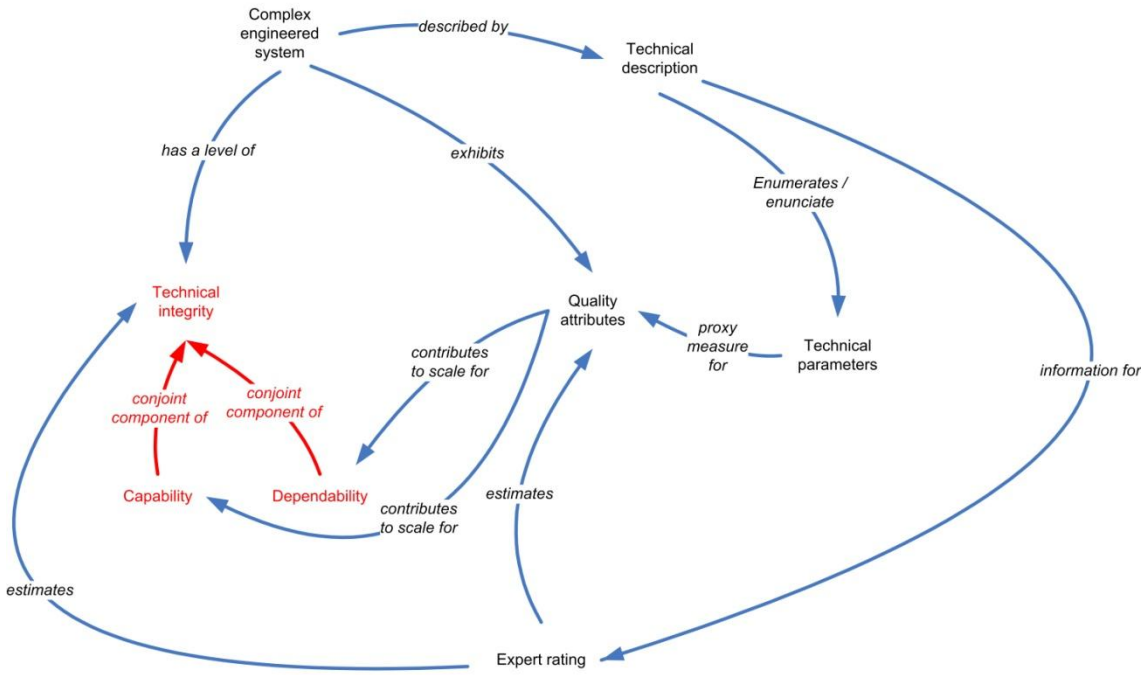


Figure 167. Influence diagram of the aggregate technical integrity and quality attributes of a complex engineered system.

This model assisted in the identification of additional research questions that are the subject of this chapter. They are:

- Does the judgment rating data confirm the measurement construct of technical integrity as being composed of capability and dependability?
- Using expert ratings of quality attributes, can it be shown that there are two components that contribute to technical integrity?
  - What does factor analysis indicate about the number and nature of components of technical integrity from the empirical judgments?
  - How well do these factors correspond to what was theorised as the orthogonal concepts of capability and dependability?
  - Can practical and valid scales be formed for these components?
  - How can these scales be combined to form a practical measurement instrument for technical integrity?

9.2 Method

This phase of the research was based upon the same survey data obtained as previously described in section 8.2. The analytical methods used within this phase of the research

focused on using that data to confirm and amend the theory of technical integrity and its measurement construct. The main two techniques used were confirmatory factory analysis and exploratory factor analysis.

9.2.1 Confirmatory factor analysis

Confirmatory factor analysis is a set of techniques to test specific hypotheses or theories concerning the structure underlying a set of variables (Pallant, 2010, p. 181). Using judgment data for the indicators across the three example systems confirmatory factor analysis aimed to confirm the theory of technical integrity developed previously in Chapter 5 and amendments to that theory to establish a best fit model for use as a measurement construct. Confirmatory factor analysis is conducted by way of Structural Equation Modelling (SEM) enabled by the software tool IBM SPSS AMOS. The method follows the steps of (Moss, 2008):

- A diagram of the theorised model is input into AMOS and error terms incorporated.
- Analysis is executed by IBM SPSS AMOS.
- Regression weights and correlations obtained from AMOS were interpreted.
- The fit of the data to the hypothesised model was assessed.
- The model is assessed in order to determine if it is supported or refuted by the data set. If the model is either supported or refuted, a means to refine the model is investigated.

9.2.2 Exploratory factor analysis

Exploratory factor analysis is used to gather information about the interrelationships among a set of variables (Pallant, 2010, p. 181).

Exploratory factor analysis was conducted to gather information about the interrelationships of the judgment dimensions using rating data obtained for the three example systems. SPSS v20 was used for the analysis. Principal Component Analysis (PCA) was used since an empirical summary of the data set is desired and PCA is considered the best choice for this (Pallant, 2010, p. 182).

The 18 variables theorised to contribute to technical integrity were included in the PCA. ‘Technical Complexity’ and ‘High Technology’ were not included in the analysis since

they were not expected to contribute to technical integrity, having been included in the judgment dimensions of the CAT as a means of establishing discriminant validity.

9.3 Results

9.3.1 Confirmatory factor analysis of theorised model of technical integrity

Recall the research question: *can it be shown that there are two components that contribute to technical integrity and how well do these components correspond to what was theorised?* The hypothesised factors of technical integrity are capability and dependability and the indicators expected to cluster with these concepts was identified in Table 12. A SEM model consistent with that hypothesized earlier was constructed in AMOS and is shown in Figure 168.

Firstly, the estimates of normality of each indicator (or variable in AMOS terms) were reviewed. All variables showed skewness less than 2 and kurtosis less than 5 indicating the data was suitable for using Generalised Least Square Estimates as recommended for sample sizes of less than 500. Then, the indication of outliers was inspected. One case showed a p value less than 0.001 indicating it was an outlier and was removed from subsequent analysis. This left 128 samples upon which the analysis was conducted.

The calculated regression weights showed a critical ratio (CR) that was much greater than 2 for all variables. This indicated that the factors seemed to influence each of the indicators as expected. The standardised regression weights are shown in Table 78. No weights are smaller than 0.5 indicating all the variables are at least reasonably aligned with the factors.

The covariance between the factors dependability and capability showed a CR of 5.87, well above the value of 2, which indicates a significant correlation. The correlation coefficient between the factors dependability and capability was estimated at 0.930 indicating that these factors are highly correlated. This implies that the theorised orthogonality of capability and dependability is not supported, by the data for the three example systems at least.

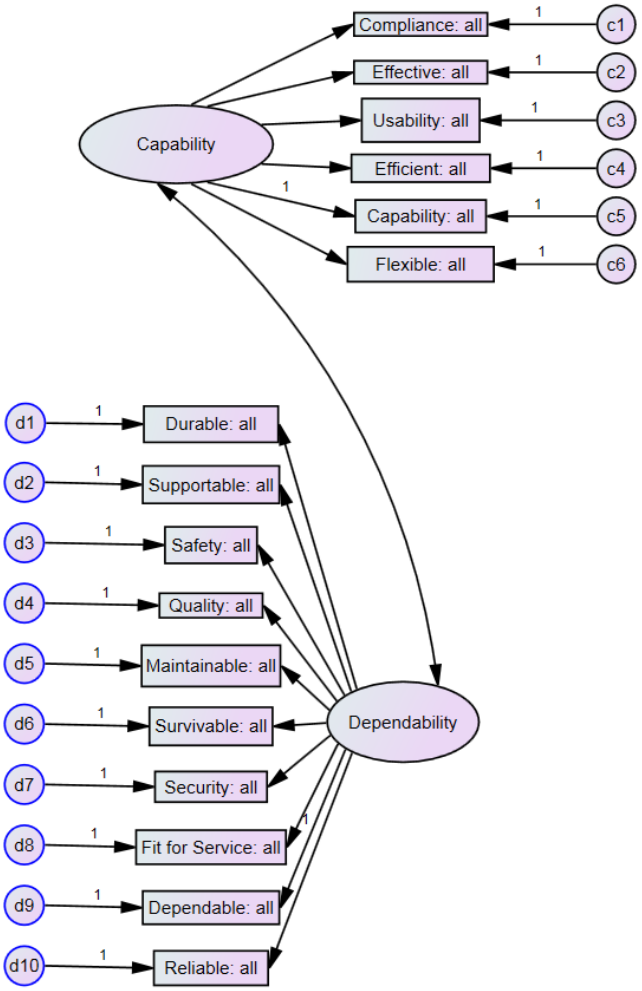


Figure 168. SEM model[18] of technical integrity concepts and indicators.

The fit of the model is examined in accordance with the recommendations of (Moss, 2008) since the conclusions drawn from the CR values may not be accurate, particularly for a relatively small sample size. Indices to be assessed for model fit, their value, recommended value for fit and a result for each are presented in Table 79. To declare a good model fit, it is usual to have all values complying with the recommended value. In this case, none do. This suggests the model is inaccurate and the hypothesised model of technical integrity is unsuitable for explaining the judgment ratings for the indicators of technical integrity.

[18] Graphical notation of the SEM model is: rectangles represent observed variables; ellipses represent latent variables that are assumed to give rise to the observed variables; circles are error terms, single headed arrows show an assumed predictive relationship; double-headed arrows indicate an association or correlation.

Table 78. Standardised regression weights - theorised technical integrity construct.

Variable		Factor	Estimate
Durable_all	<---	Dependability	.717
Support_all	<---	Dependability	.796
Safety_all	<---	Dependability	.572
Quality_all	<---	Dependability	.709
Maint_all	<---	Dependability	.811
Survive_all	<---	Dependability	.597
Secure_all	<---	Dependability	.679
FitServ_all	<---	Dependability	.789
Depend_all	<---	Dependability	.796
Reliable_all	<---	Dependability	.895
Comply_all	<---	Capability	.785
Effective_all	<---	Capability	.854
Useable_all	<---	Capability	.806
Efficient_all	<---	Capability	.648
Capability_all	<---	Capability	.859
Flexible_all	<---	Capability	.726

Table 79. Model-fit parameters – theorised technical integrity construct.

Parameter	Value from AMOS	Recommended value for fit	Result
CMIN [1]	207.777	-	-
p [2]	0.000	> 0.05	Fail
CMIN/DF [3]	2.017	< 2	Fail (just)
RMR [4]	1.391	<0.05.	Fail
GFI [4]	0.795	> 0.9	Fail
AGFI [5]	0.730	> 0.9	Fail
NFI [6]	0.242	> 0.9	Fail
CFI [7]	0.320	> 0.9	Fail

(1) Minimum value of the discrepancy (chi-square distributed). (2) p value associated with CMIN. (3) CMIN divided by degrees of freedom. (3) Root Mean square Residual, 0 indicates perfect fit. (4) Goodness of Fit Index. (5) Adjusted Goodness of Fit Index. (6) Bentler-Bonett Normed Fit Indicator. (7) Comparative Fit Index.

The AMOS analysis output also provided modification indices. Consideration of these indices and reiterations of the analysis after allowing correlations and or deletions of some variables from the model, did not appreciably improve the model fit.

The researcher therefore reverted to exploratory factor analysis to identify the number of factors and their association with the judgment dimensions.

9.3.2 Exploratory factor analysis

The failure of confirmatory factor analysis to find support for the previously hypothesized model of technical integrity led the researcher to utilise exploratory factor analysis to identify the number of factors and their association with the judgment dimensions. The nature of the analysis is iterative and the results in this section reflect the iterations undertaken.

9.3.2.1 First iteration of PCA

Prior to performing the PCA, the suitability of the data for factor analysis was assessed. Inspection of the correlation matrix revealed the presence of many coefficients of 0.3 and above as highlighted in green in Table 80.

Table 80. Correlation matrix – 18 variables.

Correlation Matrix

		Safety: all	Environment Impact: all	Impact to Systems: all	Fit for Service: all	Capability: all	Effective: all	Efficient: all	Quality: all	Flexible: all	Durable: all	Survivable: all	Dependable: all	Reliable: all	Supportable: all	Maintainable: all	Compliance: all	Usability: all	Security: all
Correlation	Safety: all	1.000	0.157	0.247	0.224	0.261	0.259	0.212	0.417	0.322	0.321	0.450	0.228	0.133	0.371	0.386	0.133	0.253	0.292
	Environment Impact: all	0.157	1.000	0.310	0.143	0.241	0.310	0.400	0.096	0.163	0.109	0.043	0.281	0.170	0.104	0.012	0.173	0.223	0.337
	Impact to Systems: all	0.247	0.310	1.000	0.346	0.299	0.373	0.388	0.221	0.293	0.174	0.196	0.401	0.363	0.293	0.228	0.250	0.345	0.283
	Fit for Service: all	0.224	0.143	0.346	1.000	0.454	0.493	0.314	0.373	0.379	0.357	0.236	0.441	0.408	0.392	0.355	0.632	0.388	0.419
	Capability: all	0.261	0.241	0.299	0.454	1.000	0.723	0.452	0.467	0.608	0.390	0.321	0.511	0.588	0.365	0.412	0.526	0.546	0.362
	Effective: all	0.259	0.310	0.373	0.493	0.723	1.000	0.533	0.406	0.501	0.404	0.242	0.626	0.642	0.423	0.400	0.526	0.621	0.408
	Efficient: all	0.212	0.400	0.388	0.314	0.452	0.533	1.000	0.320	0.276	0.219	0.154	0.452	0.362	0.130	0.124	0.373	0.469	0.416
	Quality: all	0.417	0.096	0.221	0.373	0.467	0.406	0.320	1.000	0.460	0.485	0.508	0.239	0.217	0.447	0.485	0.170	0.225	0.265
	Flexible: all	0.322	0.163	0.293	0.379	0.608	0.501	0.276	0.460	1.000	0.436	0.438	0.366	0.430	0.375	0.404	0.298	0.397	0.440
	Durable: all	0.321	0.109	0.174	0.357	0.390	0.404	0.219	0.485	0.436	1.000	0.521	0.463	0.482	0.404	0.493	0.233	0.425	0.242
	Survivable: all	0.450	0.043	0.196	0.236	0.321	0.242	0.154	0.508	0.438	0.521	1.000	0.194	0.154	0.386	0.474	0.007	0.240	0.216
	Dependable: all	0.228	0.281	0.401	0.441	0.511	0.626	0.452	0.239	0.366	0.463	0.194	1.000	0.682	0.375	0.409	0.412	0.603	0.489
	Reliable: all	0.133	0.170	0.363	0.408	0.588	0.642	0.362	0.217	0.430	0.482	0.154	0.682	1.000	0.418	0.332	0.489	0.726	0.390
	Supportable: all	0.371	0.104	0.293	0.392	0.365	0.423	0.130	0.447	0.375	0.404	0.386	0.375	0.418	1.000	0.734	0.234	0.392	0.295
	Maintainable: all	0.386	0.012	0.228	0.355	0.412	0.400	0.124	0.485	0.404	0.493	0.474	0.409	0.332	0.734	1.000	0.157	0.379	0.303
	Compliance: all	0.133	0.173	0.250	0.632	0.526	0.526	0.373	0.170	0.298	0.233	0.007	0.412	0.489	0.234	0.157	1.000	0.471	0.176
	Usability: all	0.253	0.223	0.345	0.388	0.546	0.621	0.469	0.225	0.397	0.425	0.240	0.603	0.726	0.392	0.379	0.471	1.000	0.369
	Security: all	0.292	0.337	0.283	0.419	0.362	0.408	0.416	0.265	0.440	0.242	0.216	0.489	0.390	0.295	0.303	0.176	0.369	1.000

The Kaiser-Meyer-Olkin value was 0.863 exceeding the recommended value of 0.6 and Bartlett's Test of Sphericity reached statistical significance, supporting the factorability of the correlation matrix.

PCA revealed the presence of three components with eigenvalues exceeding 1, explaining 40.1%, 11.4% and 7.1% of the variance respectively as shown in Table 81.

Table 81. Components and explained variance: 18 variables.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.211	40.059	40.059	7.211	40.059	40.059	4.479	24.881	24.881
2	2.047	11.371	51.431	2.047	11.371	51.431	3.749	20.828	45.710
3	1.283	7.125	58.556	1.283	7.125	58.556	2.312	12.846	58.556
4	.971	5.392	63.948						
5	.891	4.951	68.899						
6	.756	4.200	73.099						
7	.674	3.745	76.844						
8	.665	3.694	80.538						
9	.631	3.503	84.041						
10	.597	3.316	87.357						
11	.438	2.432	89.789						
12	.382	2.122	91.911						
13	.332	1.847	93.758						
14	.268	1.491	95.248						
15	.259	1.442	96.690						
16	.248	1.376	98.066						
17	.194	1.076	99.142						
18	.154	.858	100.000						

Extraction Method: Principal Component Analysis.

The corresponding scree plot at Figure 169 revealed a reasonably clear break after the second component suggesting that only two components be retained for further investigation.

A Parallel Analysis was conducted with results as shown in Table 82. This revealed that only two components have eigenvalues exceeding the corresponding criterion values for a randomly generated data matrix of the same size (18 variables by 129 respondents). This analysis confirmed the decision to retain two components.

The two component solution explained a total of 51.43% of the variance, with Component 1 contributing 40.06% and Component 2 contributing 11.37%.

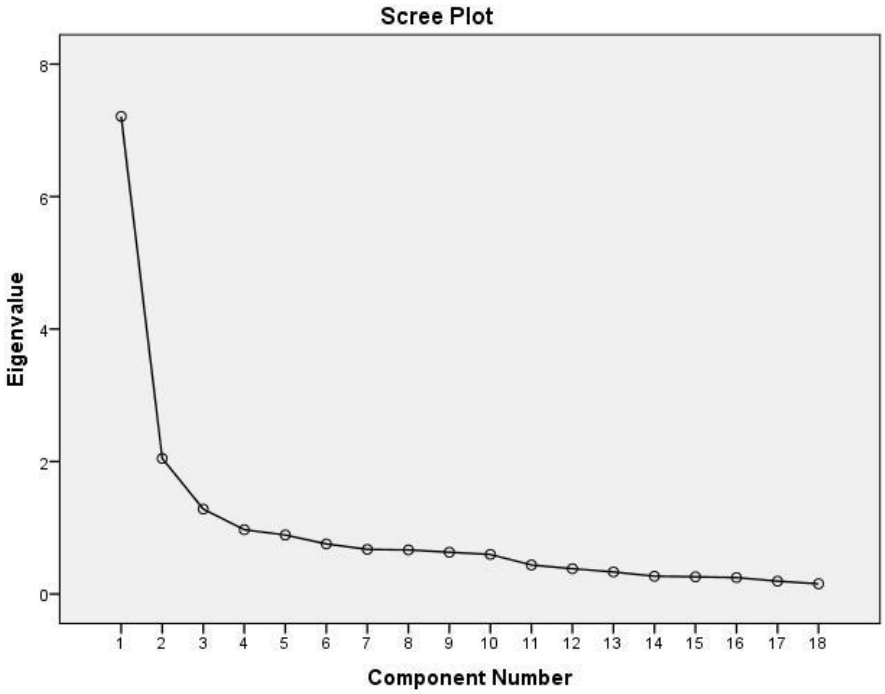


Figure 169. Scree plot of 18 variables.

Table 82. Parallel analysis results.

Component	Eigenvalues from PCA			Parallel Analysis Criterion Value	Decision
	Total	% of Variance	Cumulative %		
1	7.211	40.059	40.059	1.7242	Retain
2	2.047	11.371	51.431	1.5635	Retain
3	1.283	7.125	58.556	1.4501	Reject
4	0.971	5.392	63.948	1.3603	Reject
5	0.891	4.951	68.899	1.2777	Reject
6	0.756	4.200	73.099	1.1957	Reject
7	0.674	3.745	76.844	1.1283	Reject
8	0.665	3.694	80.538	1.0607	Reject
9	0.631	3.503	84.041	0.9933	Reject
10	0.597	3.316	87.357	0.9258	Reject
11	0.438	2.432	89.789	0.8709	Reject
12	0.382	2.122	91.911	0.8052	Reject
13	0.332	1.847	93.758	0.7534	Reject
14	0.268	1.491	95.248	0.6979	Reject
15	0.259	1.442	96.690	0.6388	Reject
16	0.248	1.376	98.066	0.5851	Reject
17	0.194	1.076	99.142	0.5206	Reject
18	0.154	.858	100.000	0.4483	Reject

9.3.2.2 Two factor solution

The PCA was repeated but with a fixed number of factors to extract set at 2. Note that the theorised construct for technical integrity consisted of two components (capability and dependability) and these were thought to be orthogonal. Notwithstanding this consideration, the recommendation of (Pallant, 2010, p. 185) to commence with an oblique rotation, using the Oblimin method, to check the degree of correlation between the components was accepted to assist in the interpretation of the two components.

The resulting rotated solution revealed a structure where most of the variables loaded substantially on only one component. However, 4 of the 18 components loaded moderately to strongly on both components complicating the structure. The Pattern Matrix shows the factor loadings in each of the variables while the Structure matrix shows the correlations between the variables and the factors (refer to Table 83). The component plot demonstrates this graphically with two clusters of variables broadly aligning with the extracted components. The resulting Component Correlation Matrix shows a correlation of 0.408 which is above 0.3 indicating that an oblique rotation was the appropriate method to be used. The component plot at Figure 170 showed a structure around the two components with only the variable ‘Flexible’ loading moderately on both extracted components.

Table 83. Pattern and structure matrices: 2 components,18 variables, Oblimin rotation.

Pattern Matrix[a]

	Component	
	1	2
Effective: all	.768	
Reliable: all	.768	
Compliance: all	.754	
Dependable: all	.740	
Usability: all	.727	
Efficient: all	.719	
Capability: all	.643	
Fit for Service: all	.553	
Environment Impact: all	.527	
Impact to Systems: all	.508	
Security: all	.502	
Survivable: all		.842
Maintainable: all		.795
Quality: all		.734
Supportable: all		.678
Safety: all		.629
Durable: all		.624
Flexible: all	.337	.492

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Structure Matrix

	Component	
	1	2
Effective: all	.831	.468
Reliable: all	.793	.376
Dependable: all	.781	.402
Usability: all	.771	.404
Capability: all	.749	.521
Compliance: all	.695	
Efficient: all	.681	
Fit for Service: all	.638	.434
Security: all	.574	.381
Impact to Systems: all	.539	
Environment Impact: all	.456	
Maintainable: all	.352	.806
Survivable: all		.774
Quality: all	.336	.749
Supportable: all	.406	.731
Durable: all	.435	.697
Flexible: all	.538	.629
Safety: all		.626

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

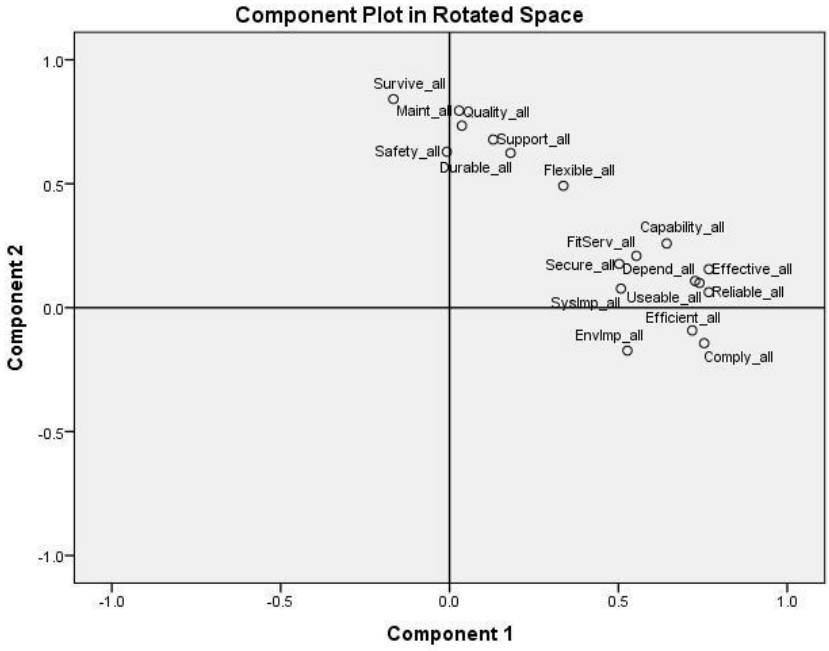


Figure 170. Component plot: 2 components,18 variables, Oblimin rotation.

The communalities matrix at Table 84 shows how much the variance in each item is explained. Values below 0.3 could indicate the item does not fit well with other items in

the component and are therefore candidates for removal from the scale. This implied that ‘Environment Impact’ and ‘Impact to Systems’ were candidates for removal.

Table 84. Communalities matrix: 2 components,18 variables, Oblimin rotation.

Communalities

	Initial	Extraction
Safety: all	1.000	.392
Environment Impact: all	1.000	.233
Impact to Systems: all	1.000	.295
Fit for Service: all	1.000	.444
Capability: all	1.000	.616
Effective: all	1.000	.710
Efficient: all	1.000	.471
Quality: all	1.000	.562
Flexible: all	1.000	.491
Durable: all	1.000	.514
Survivable: all	1.000	.622
Dependable: all	1.000	.618
Reliable: all	1.000	.632
Supportable: all	1.000	.548
Maintainable: all	1.000	.650
Compliance: all	1.000	.501
Usability: all	1.000	.604
Security: all	1.000	.356

Extraction Method: Principal Component Analysis.

9.3.2.3 Removal of poor-fit variables

The results from both the communalities matrix (Table 84) and the results at Table 50 suggested that Impact to the natural environment [Environment Impact] and Impact to other systems [Impact to Systems] may be able to be left out of a scale for technical integrity. The PCA was re-run without these two variables included. The 16 remaining variables were included in the PCA. The suitability of the data for factor analysis was re-assessed. Inspection of the correlation matrix again revealed the presence of many coefficients of 0.3 and above as highlighted in green in Table 85.

Table 85. Correlation matrix: 16 variables.

Correlation Matrix

		Safety: all	Fit for Service: all	Capability: all	Effective: all	Efficient: all	Quality: all	Flexible: all	Durable: all	Survivable: all	Dependable: all	Reliable: all	Supportable: all	Maintainable: all	Compliance: all	Usability: all	Security: all
Correlation	Safety: all	1.000	.224	.261	.259	.212	.417	.322	.321	.450	.228	.133	.371	.386	.133	.253	.292
	Fit for Service: all	.224	1.000	.454	.493	.314	.373	.379	.357	.236	.441	.408	.392	.355	.632	.388	.419
	Capability: all	.261	.454	1.000	.723	.452	.467	.608	.390	.321	.511	.588	.365	.412	.526	.546	.362
	Effective: all	.259	.493	.723	1.000	.533	.406	.501	.404	.242	.626	.642	.423	.400	.526	.621	.408
	Efficient: all	.212	.314	.452	.533	1.000	.320	.276	.219	.154	.452	.362	.130	.124	.373	.469	.416
	Quality: all	.417	.373	.467	.406	.320	1.000	.460	.485	.508	.239	.217	.447	.485	.170	.225	.265
	Flexible: all	.322	.379	.608	.501	.276	.460	1.000	.436	.438	.366	.430	.375	.404	.298	.397	.440
	Durable: all	.321	.357	.390	.404	.219	.485	.436	1.000	.521	.463	.482	.404	.493	.233	.425	.242
	Survivable: all	.450	.236	.321	.242	.154	.508	.438	.521	1.000	.194	.154	.386	.474	.007	.240	.216
	Dependable: all	.228	.441	.511	.626	.452	.239	.366	.463	.194	1.000	.682	.375	.409	.412	.603	.489
	Reliable: all	.133	.408	.588	.642	.362	.217	.430	.482	.154	.682	1.000	.418	.332	.489	.726	.390
	Supportable: all	.371	.392	.365	.423	.130	.447	.375	.404	.386	.375	.418	1.000	.734	.234	.392	.295
	Maintainable: all	.386	.355	.412	.400	.124	.485	.404	.493	.474	.409	.332	.734	1.000	.157	.379	.303
	Compliance: all	.133	.632	.526	.526	.373	.170	.298	.233	.007	.412	.489	.234	.157	1.000	.471	.176
	Usability: all	.253	.388	.546	.621	.469	.225	.397	.425	.240	.603	.726	.392	.379	.471	1.000	.369
	Security: all	.292	.419	.362	.408	.416	.265	.440	.242	.216	.489	.390	.295	.303	.176	.369	1.000

The Kaiser-Meyer-Olkin value was 0.868 (compared with the value of 0.863 for the 18 variable analysis) exceeding the recommended value of 0.6 and Bartlett's Test of Sphericity reached statistical significance, supporting the factorability of the correlation matrix.

The analysis was again constrained to a two component solution which explained a total of 55.32% (c.f. 51.43% for the 18 variable analysis) of the variance, with Component 1 contributing 43.00% (c.f. 40.06%) and Component 2 contributing 12.32% (c.f.11.37%) as shown in Table 86. Review of the scree plot and parallel analysis further justified the retention of a 2 component solution.

Table 86. Components and explained variance: 16 variables.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	6.879	42.996	42.996	6.879	42.996	42.996	6.040
2	1.971	12.319	55.315	1.971	12.319	55.315	4.899
3	1.040	6.501	61.816				
4	.940	5.874	67.690				
5	.839	5.241	72.930				
6	.718	4.490	77.420				
7	.663	4.144	81.565				
8	.616	3.848	85.413				
9	.444	2.772	88.185				
10	.382	2.390	90.575				
11	.335	2.093	92.668				
12	.283	1.770	94.438				
13	.268	1.676	96.114				
14	.257	1.605	97.719				
15	.201	1.254	98.972				
16	.164	1.028	100.000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

The rotated solution, using the Oblimin method, produced a component correlation matrix that showed a correlation of 0.440 which is above 0.3 indicating that an oblique rotation was the appropriate method to be used.

The pattern and structure matrices are shown at Table 87. The revealed structure is simplified compared to the 18 variable PCA. All but one of the variables, that being ‘Flexible’, loaded substantially on only one component. The component plot at Figure 171 demonstrated this graphically with two clusters of variables broadly aligning with the extracted components.

Table 87. Pattern and structure matrices: 2 components,16 variables, Oblimin rotation.

Pattern Matrix[a]

	Component	
	1	2
Reliable: all	.833	
Compliance: all	.818	
Effective: all	.797	
Usability: all	.772	
Dependable: all	.761	
Capability: all	.686	
Efficient: all	.672	
Fit for Service: all	.593	
Security: all	.469	
Survivable: all		.860
Maintainable: all		.763
Quality: all		.737
Safety: all		.670
Supportable: all		.655
Durable: all		.588
Flexible: all	.355	.469

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 6 iterations.

Structure Matrix

	Component	
	1	2
Effective: all	.843	.457
Reliable: all	.823	.345
Usability: all	.792	.385
Dependable: all	.786	.390
Capability: all	.776	.507
Compliance: all	.718	
Fit for Service: all	.661	.415
Efficient: all	.639	
Security: all	.555	.402
Maintainable: all	.404	.793
Survivable: all		.783
Quality: all	.364	.755
Supportable: all	.439	.721
Durable: all	.481	.686
Safety: all		.648
Flexible: all	.562	.625

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

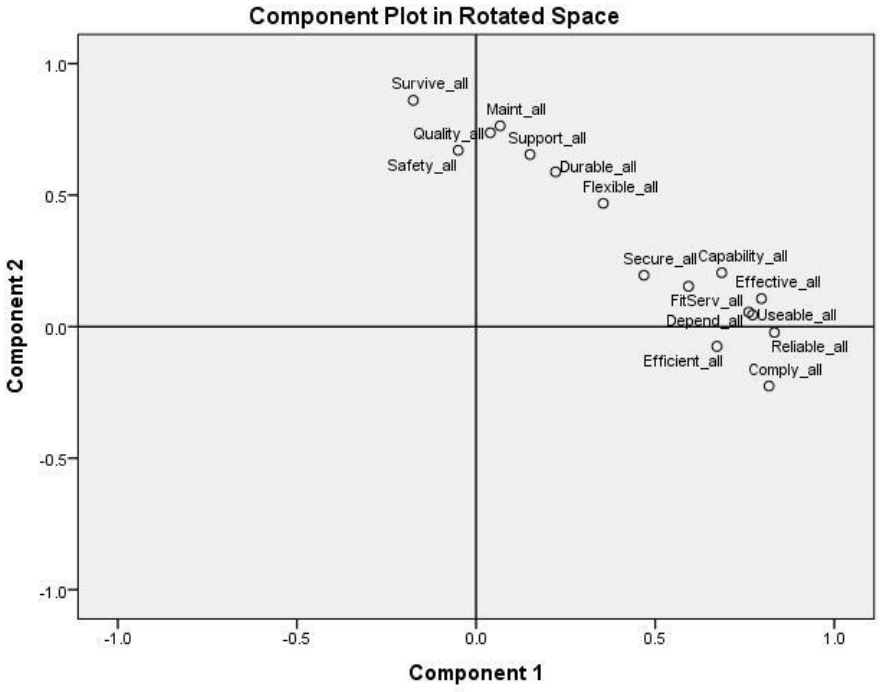


Figure 171. Component plot: 2 components,16 variables, Oblimin rotation.

9.3.2.4 Refinement by removal of 'Flexibility'

Given the moderate loading of Operational Flexibility (the variable 'Flexible') to both extracted components in previous PCAs and the indications that it may not be a discriminator for differentiating the technical integrity of different systems (at least for the systems in this experiment) another iteration of the PCA was conducted without 'Flexible' leaving 15 variables. The suitability of the reduced data set for factor analysis was re-assessed. Inspection of the correlation matrix revealed the presence of many coefficients of 0.3 and above as highlighted in green in Table 88.

Table 88. Correlation matrix: 15 variables.

Correlation Matrix

		Safety: all	Fit for Service: all	Capability: all	Effective: all	Efficient: all	Quality: all	Flexible: all	Durable: all	Survivable: all	Dependable: all	Reliable: all	Supportable: all	Maintainable: all	Compliance: all	Usability: all	Security: all
Correlation	Safety: all	1.000	.224	.261	.259	.212	.417	.322	.321	.450	.228	.133	.371	.386	.133	.253	.292
	Fit for Service: all	.224	1.000	.454	.493	.314	.373	.379	.357	.236	.441	.408	.392	.355	.632	.388	.419
	Capability: all	.261	.454	1.000	.723	.452	.467	.608	.390	.321	.511	.588	.365	.412	.526	.546	.362
	Effective: all	.259	.493	.723	1.000	.533	.406	.501	.404	.242	.626	.642	.423	.400	.526	.621	.408
	Efficient: all	.212	.314	.452	.533	1.000	.320	.276	.219	.154	.452	.362	.130	.124	.373	.469	.416
	Quality: all	.417	.373	.467	.406	.320	1.000	.460	.485	.508	.239	.217	.447	.485	.170	.225	.265
	Flexible: all	.322	.379	.608	.501	.276	.460	1.000	.436	.438	.366	.430	.375	.404	.298	.397	.440
	Durable: all	.321	.357	.390	.404	.219	.485	.436	1.000	.521	.463	.482	.404	.493	.233	.425	.242
	Survivable: all	.450	.236	.321	.242	.154	.508	.438	.521	1.000	.194	.154	.386	.474	.007	.240	.216
	Dependable: all	.228	.441	.511	.626	.452	.239	.366	.463	.194	1.000	.682	.375	.409	.412	.603	.489
	Reliable: all	.133	.408	.588	.642	.362	.217	.430	.482	.154	.682	1.000	.418	.332	.489	.726	.390
	Supportable: all	.371	.392	.365	.423	.130	.447	.375	.404	.386	.375	.418	1.000	.734	.234	.392	.295
	Maintainable: all	.386	.355	.412	.400	.124	.485	.404	.493	.474	.409	.332	.734	1.000	.157	.379	.303
	Compliance: all	.133	.632	.526	.526	.373	.170	.298	.233	.007	.412	.489	.234	.157	1.000	.471	.176
	Usability: all	.253	.388	.546	.621	.469	.225	.397	.425	.240	.603	.726	.392	.379	.471	1.000	.369
	Security: all	.292	.419	.362	.408	.416	.265	.440	.242	.216	.489	.390	.295	.303	.176	.369	1.000

The Kaiser-Meyer-Olkin value was 0.857 (compared with the value of 0.863 obtained for the 18 variable analysis) exceeding the recommended value of 0.6 and Bartlett's Test of Sphericity reached statistical significance, supporting the factorability of the correlation matrix.

The analysis was again constrained to a two component solution which explained a total of 56.02% (c.f. 51.43% for the 18 variable analysis) of the variance, with Component 1 contributing 43.00% (c.f. 40.06%) and Component 2 contributing 13.02% (c.f.11.37%) as shown in Table 89. Review of the scree plot and parallel analysis confirmed the retention of a 2 component solution.

Table 89. Components and explained variance: 15 variables.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	6.449	42.996	42.996	6.449	42.996	42.996	5.756
2	1.953	13.023	56.019	1.953	13.023	56.019	4.473
3	1.023	6.817	62.837				
4	.939	6.262	69.099				
5	.830	5.531	74.630				
6	.664	4.424	79.053				
7	.649	4.329	83.382				
8	.476	3.176	86.559				
9	.444	2.957	89.516				
10	.374	2.497	92.013				
11	.295	1.967	93.980				
12	.276	1.838	95.818				
13	.262	1.746	97.563				
14	.201	1.337	98.901				
15	.165	1.099	100.000				

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings cannot be added to obtain a total variance.

The rotated solution, using the Oblimin method, produced a component correlation matrix that showed a correlation of 0.428 which is above 0.3 indicating that an oblique rotation was the appropriate method to be used.

The pattern and structure matrices are shown at Table 90. The revealed structure was further simplified compared to the previous PCAs, where all but one of the variables loaded substantially on only one component. The component plot at Figure 172 demonstrated this graphically with two clusters of variables clearly aligning with the extracted components.

Table 90. Pattern and structure matrices: 2 components,15 variables, Oblimin rotation.

Pattern Matrix[a]

	Component	
	1	2
Reliable: all	.833	
Compliance: all	.815	
Effective: all	.801	
Usability: all	.773	
Dependable: all	.763	
Capability: all	.693	
Efficient: all	.674	
Fit for Service: all	.597	
Security: all	.475	
Survivable: all		.847
Maintainable: all		.771
Quality: all		.727
Safety: all		.671
Supportable: all		.666
Durable: all		.587

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 5 iterations.

Structure Matrix

	Component	
	1	2
Effective: all	.844	.443
Reliable: all	.823	.334
Usability: all	.795	.381
Dependable: all	.791	.391
Capability: all	.770	.477
Compliance: all	.718	
Fit for Service: all	.663	.410
Efficient: all	.643	
Security: all	.551	.383
Maintainable: all	.411	.806
Survivable: all		.780
Quality: all	.367	.751
Supportable: all	.447	.735
Durable: all	.485	.688
Safety: all		.656

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

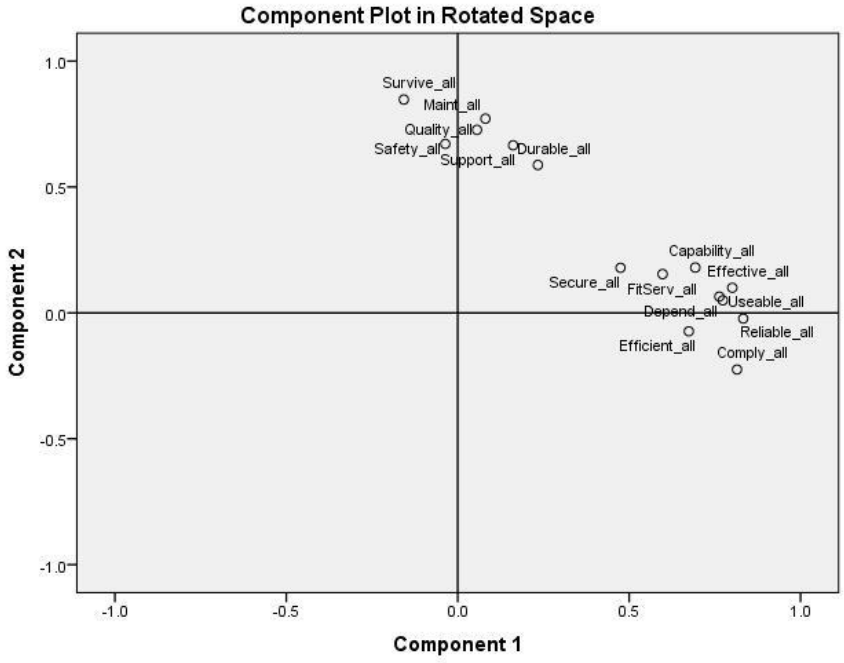


Figure 172. Component plot: 2 components,15 variables, Oblimin rotation.

9.3.2.5 Interpretation of components

The pattern matrix at Table 90 presented the simplest structure for interpretation[19]. Table 10 presents data from the pattern matrix and arranges it in order of variable loading on each component and highlights (bolds) those with the highest loadings (loading > 0.7).

Table 91. Variable loading on extracted components.

Extracted Factor/Component	Component 1	Component 2
Variable Loading on Component (Highest to lowest; >0.7 bolded)	**Reliability [Reliable]** **Technical Compliance [Compliance]** **Effectiveness [Effective]** **Usability [Usability]** **Dependability [Dependable]** Capability [Capability] Efficiency [Efficient] Fitness for Service [Fit for Service] Security [Security]	**Survivability [Survivable]** **Maintainability [Maintainable]** **Construction Quality [Quality]** Safety [Safety] Supportability [Supportable] Durability [Durable]

9.3.2.5.1 Correspondence of components with concepts of capability and dependability

Recall that the model developed from theory predicted two components of technical integrity: *capability* and *dependability*. Table 92 presents the variables against the components upon which they primarily load in the second and third columns. The hypothesised correspondence of those same variables to the technical integrity model components is shown in the fourth and fifth columns. Hypothesising that Component 1 corresponds to *capability* and that Component 2 corresponds to *dependability*, the sixth column shows if the variable loading and theorised correspondence align and therefore if the judgment results support the theorised model.

[19] It is noted that there is significant debate in the literature on if the Pattern matrix or the Structure matrix should be used as the basis for interpretation. In keeping with the review of this debate by BROWN, J. D. 2009. Choosing the Right Type of Rotation in PCA and EFA. *SHIKEN - JALT Testing & Evaluation SIG Newsletter,* 13**,** 20-25., the approach taken herein has been to use the matrix that provides the simplest structure for interpretation and that the use of the structure matrix in this study yields the same association of components with variables for interpretation.

Table 92. Variable loading on extracted components compared to theorised model of technical integrity.

	Judgment Results		**Model of TI Components**		**Theorised Model Supported**
	Component 1	**Component 2**	***Capability***	***Dependability***	
Reliability [Reliable]	x			x	No
Technical Compliance [Compliance]	x		x		Yes
Effectiveness [Effective]	x		x		Yes
Usability [Usability]	x		x		Yes
Dependability [Dependable]	x			x	No
Capability [Capability]	x		x		Yes
Efficiency [Efficient]	x		x		Yes
Fitness for Service [Fit for Service]	x			x	No
Security [Security]	x			x	No
Survivability [Survivable]		x		x	Yes
Maintainability [Maintainable]		x		x	Yes
Construction Quality [Quality]		x		x	Yes
Safety [Safety]		x		x	Yes
Supportability [Supportable]		x		x	Yes
Durability [Durable]		x		x	Yes
Flexibility [Flexible)	-	-	x		No (1)
Impact to the natural environment [Environment Impact]	-	-		x	No (2)
Impact to other systems [Impact to Systems]	-	-		x	No (2)

(1) Fairly evenly loaded on both components prior to removal from scale.

(2) Loaded moderately to component 1 prior to removal from scale.

The results show that the correspondence of Component 2 to *dependability* component of the technical integrity model is reasonably good:

- All of the variables that load onto Component 2 were presumed to be associated with dependability.
- Impact to the Natural Environment and Impact to Other Systems were expected to associate with *dependability*, but did not associate well with the scale in this experiment (to be discussed later, but this may well be a limitation of this experiment in that there was little to differentiate the example systems in respect of these variables).
- However, reliability, dependability, fitness for service and security loaded primarily on Component 1 but were expected to associate more strongly with *dependability* of the model.

The results show that the correspondence of Component 1 to *capability* component of the technical integrity model is reasonably good:

  - 5 of the 6 variables theorised to associate with capability loaded primarily on Component 1.
  - The other variable, Flexibility, was expected to be associated primarily to capability, but loading in the PCAs that included it showed it moderately loaded to both components.
  - However, reliability, dependability, fitness for service and security were not theorised to be primarily associated with *capability*, but do load most heavily upon Component 1.

In summary, there was not complete correspondence with the theorised components of technical integrity.

9.3.2.5.2 Correspondence of components with quality characteristics

The quality model of (Firesmith et al., 2009a) was reviewed in section 3.3.3. To see if that model of internal and external quality characteristics could assist in interpreting the result of factor analysis, the variables with the highest loadings identified in Table 91

were compared to the quality models represented in Figure 6 and Figure 7[20]. With the exception of the variable 'maintainability' which is an internal quality characteristic in the quality model, the remainder of the variables correspond to external quality characteristics. Noting the software focus of the quality model, it is fair to consider that Maintainability is synonymous with Serviceability in the quality model, which would also make it an external quality characteristic. The variables were highlighted in Figure 173 in order to see if any assistance to interpretation was provided. Variables loading on Component 1 are highlighted in red and those loading on component 2 in blue. Solid outlines are used where there is a one-to-one correspondence between the terms and dashed lines where near synonyms to the variables that are not otherwise represented in the quality model have been identified.

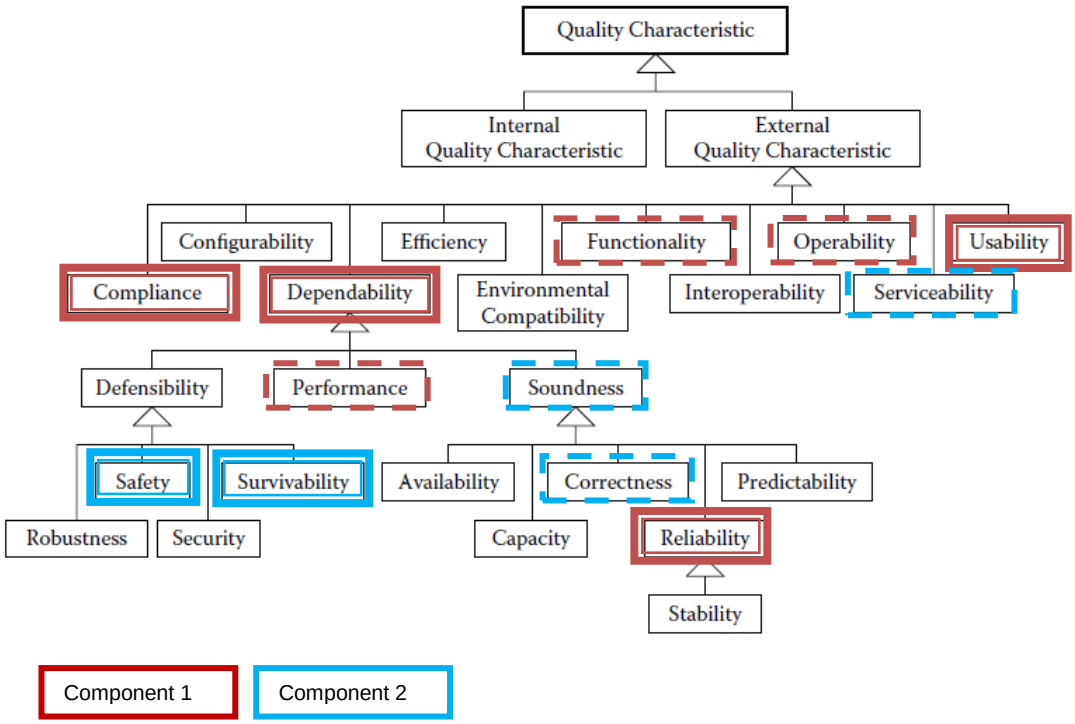


Figure 173. Correspondence of variables to Firesmith's external quality characteristics.

Irrespective of the component upon which the variables loaded, it was observed that effectively all the components correspond with external quality characteristics of the

[20] As a cautionary note, the model of Firesmith et.al., is focused on the characteristics of software intensive systems rather than the broader view taken for engineered systems in this thesis, however it does provide a useful model for comparison of results.

system. As would be expected, technical integrity can be seen to be comprised of ‘black box quality characteristics’ that is, externally visible qualities of the system as per the quality model of (Firesmith et al., 2009a). However, the mapping presented in Figure 173 provided little assistance in the interpretation of components suggest by the factor analysis.

9.3.2.5.3 Correspondence of components with benefit and consequence

The concepts of benefit and sacrifice in order to establish value of a system to a stakeholder was identified in section 3.3.2. A possible interpretation of the components identified for technical integrity may be as some acceptable balance between benefit and sacrifice. To examine this interpretation further, Table 93 was constructed. For this purpose, the view that is taken is that any variable for which an infinite value would be of most value to a stakeholder, e.g., an infinite value of reliability would mean that the system would never fail, any finite value must therefore represent a compromise and can be considered a sacrifice. Variables that relate to the desired characteristics of the system, e.g., compliance with stated requirements and effectiveness, are considered a benefit. While this is a subjective process it is useful in searching for an interpretation. The variable dependability has a case that could be argued for a benefit or a sacrifice and therefore has not been indicated as either in this analysis.

Table 93. Correspondence of variables associated with benefit and sacrifice.

Component	Variable	Benefit	Sacrifice
Component 1	Reliability		✓
	Technical Compliance	✓	
	Effectiveness	✓	
	Usability	✓	
	Dependability	?	?
Component 2	Survivability		✓
	Maintainability		✓
	Construction Quality		✓
	Safety		✓

With the exception of reliability and the possible exception of dependability, Table 93 reveals that it may be reasonable to associate component 1 with benefit and component 2 with sacrifice, but this is not a completely satisficing interpretation.

9.3.2.5.4 New interpretations

Since there was not complete correspondence with the theorised components of technical integrity or with other concepts identified in the literature, the researcher consulted professional colleagues to informally provide their interpretation of the primary variable loading. The colleagues were presented with Table 94 and asked to consider what the underlying characteristics of a system might be that would be correlated with these. Their ideas were captured, discussed in detail with the researcher and subsequently documented by the researcher in Table 95.

Table 94. Variables associated with components of technical integrity.

Component	Component 1	Component 2
Variables strongly correlated with component	Reliability Technical Compliance Effectiveness Usability Dependability	Survivability Maintainability Construction Quality Safety

Table 95. Interpretations of the nature of components.

Component 1	Component 2	Discussion
Objective	Subjective	Concept was that most of the items in Component 1 could be readily objectively evaluated. Whereas those items in Component 2 tended to not be so readily assessed objectively – e.g. Safety is often dealt with subjectively via safety case argument and qualitative hazard risk assessments.
Measurable	Experience	Similar to above, most of the items in Component 1 could be readily measured, even by analysis during design and development. Whereas those items in Component 2 tended to be assessed from experience with the actually built system.
Explicit expectation	Implicit expectation	The items in Component 1 align with attributes of the system you would expect to see explicitly manifest in the design – the intended attributes of the system. The items in Component 2 are implicitly expected but often not specified. E.g. It is explicitly assumed that they system should be survivable over the range of reasonably foreseeable conditions.
Normal operation	Extreme circumstances Abnormal circumstances	The items in Component 1 are manifest in the system in its normal operating condition and environment. The items in Component 2 are only experienced in extreme or abnormal circumstances. E.g. survivability is only manifest when a threat is extant. E.g. maintainability is only extant when failures are present. E.g. safety is only manifest when causal factors and/or hazards are present.
Demonstrable	Documentable/ Argued	The items in Component 1 are generally readily demonstrated over all conditions of interest (similar to normal operation). The items in Component 2 may not be demonstrable or in practice only practically demonstrable for a small number of cases (e.g. can practically demonstrate maintainability for several defined fault scenarios, but cannot practically demonstrate maintainability for all possible faults of a complex engineered system). Because of this the level of the item in a system is often argued qualitatively – e.g. safety is argued via a safety case rather than demonstrated.
Testable	Untestable / Recorded anomalous behaviour	The items in Component 1 are generally readily tested over all conditions of interest (as for Demonstrable above). The items in Component 2 cannot be practically tested for all circumstances. Evaluation may rely on the observation and recording of anomalous behaviour over time.

9.3.3 Confirmatory factor analysis of amended model of technical integrity

The results of exploratory factor analysis provided the basis for an amended model of technical integrity. After further consideration by the researcher, the concepts of explicit and implicit expectations of technical integrity in third row of Table 95 were accepted for further analysis and refinement. An amended SEM model that was constructed in AMOS is shown in Figure 174. The amended concepts of explicit technical integrity and implicit technical integrity are shown as "ExpTI" and "ImpTI" respectively.

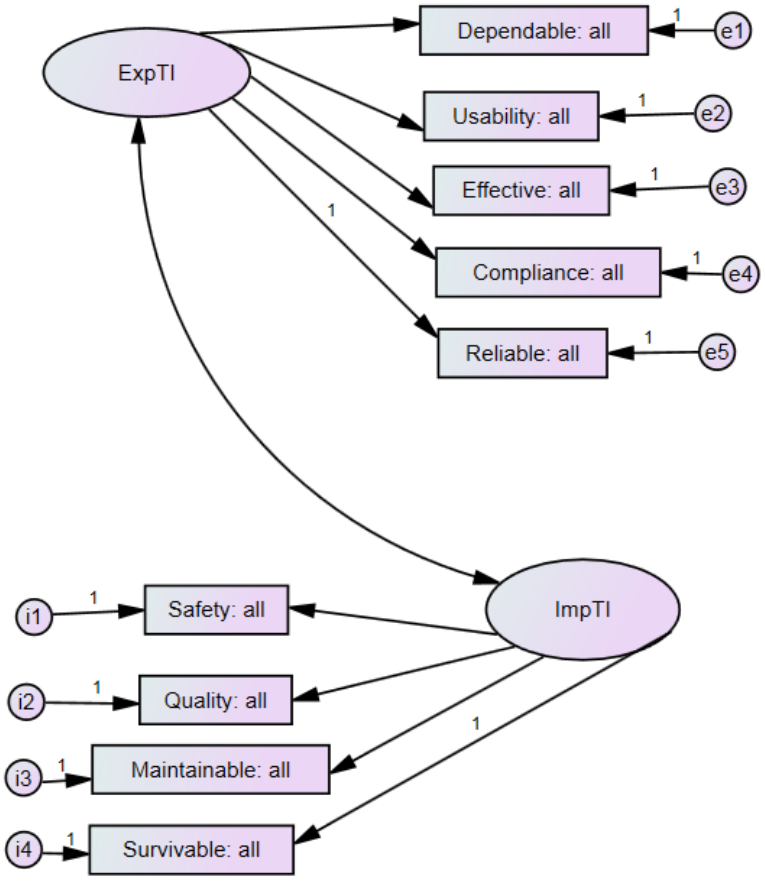


Figure 174. SEM model of amended technical integrity concepts and indicators.

The confirmation of sufficient normality of the data and the elimination of an outlier in order to use Generalised Least Square Estimates was described in section 9.3.1 and holds for this subsequent analysis.

The calculated regression weights showed a critical ration (CR) that was much greater than 2 (all greater than 5 in fact) for all variables. This indicated that the factors seem to influence each of the indicators as expected. The standardised regression weights are in

Table 96. No weights are smaller than 0.5 indicating all the variables are at least reasonably aligned with the factors.

Table 96. Standardised regression weights - amended technical integrity construct.

Variable		Factor	Estimate
Comply_all	<---	ExpTI	.610
Effective_all	<---	ExpTI	.828
Useable_all	<---	ExpTI	.827
Depend_all	<---	ExpTI	.771
Reliable_all	<---	ExpTI	.863
Safety_all	<---	ImpTI	.680
Quality_all	<---	ImpTI	.717
Survive_all	<---	ImpTI	.682
Maint_all	<---	ImpTI	.698

The covariance between the factors ExpTI and ImpTI showed a CR of 4.177, well above the value of 2 which indicates a significant correlation. The correlation coefficient between the factors ExpTI and ImpTI was estimated at 0.555 indicating that these factors are moderately correlated. This implies that the concepts of implicit and explicit technical integrity are not completely orthogonal but significantly more so that was the case with the concepts of capability and dependability that were highly correlated (0.930) when modelled in section 9.3.1.

The fit of the model is examined in accordance with the recommendations of (Moss, 2008) since the conclusions drawn from the CR values may not be accurate, particularly for a relatively small sample size. Indices to be assessed for model fit, their value, recommended value for fit and a result for each are presented in Table 97. To declare a good fit, it is usual to have all values complying with the recommended value. In this case, they do not all pass the test. This suggests the model is not completely accurate, but by comparison to Table 79, the model is definitely a better fit to this data set (the three example systems) than was the model of technical integrity based on capability and dependability.

Table 97. Model-fit parameters – amended technical integrity construct.

Parameter	Value from AMOS	Recommended value for fit	Result
CMIN [1]	40.027	-	-
p [2]	0.039	> 0.05	Fail (just)
CMIN/DF [3]	1.540	< 2	Pass
RMR [4]	0.803	< 0.05.	Fail
GFI [4]	0.930	> 0.9	Pass
AGFI [5]	0.879	> 0.9	Fail (just)
NFI [6]	0.648	> 0.9	Fail
CFI [7]	0.820	> 0.9	Fail (just)

9.3.3.1 Refinement of the model

The AMOS output provided modification indices. The highest modification index value of 4.548 indicated a correlation between e3 and i2, the error terms associated with effectiveness and quality respectively. Allowing correlation between these error terms would compromise the clarity of the model. Therefore, removal of the associated variables was trialed. Removal of the quality variable from the model resulted in worse model fit indicators than with it included. The effectiveness variable was then removed from the model and produced a better fit model as demonstrated by the results of Table 98.

Table 98. Model-fit parameters – amended technical integrity construct (effectiveness removed).

Parameter	Value from AMOS	Recommended value for fit	Result
CMIN [1]	26.111	-	-
p [2]	0.127	> 0.05	Pass
CMIN/DF [3]	1.374	< 2	Pass
RMR [4]	0.784	< 0.05.	Fail (but better)
GFI [4]	0.949	> 0.9	Pass
AGFI [5]	0.903	> 0.9	Pass
NFI [6]	0.741	> 0.9	Fail (but better)
CFI [7]	0.903	> 0.9	Pass

The modification index identified that the next terms correlating strongly that could lead to an improved model were e4 and i4, the error terms associated with the variables compliance and survivability respectively. Technical compliance is thought to be pivotal to the concept of technical integrity so this was not considered for removal. Survivability was removed from the model and the model fit was observed to be further enhanced as shown in Table 99.

Table 99. Model-fit parameters – amended technical integrity construct (effectiveness and survivability removed).

Parameter	**Value from AMOS**	**Recommended value for fit**	**Result**
CMIN [1]	16.792	-	-
p [2]	0.209	> 0.05	Pass
CMIN/DF [3]	1.292	< 2	Pass
RMR [4]	0.596	< 0.05.	Fail (but better)
GFI [4]	0.962	> 0.9	Pass
AGFI [5]	0.919	> 0.9	Pass
NFI [6]	0.808	> 0.9	Fail (better and close)
CFI [7]	0.943	> 0.9	Pass

No further refinement of the model was feasible. The remaining indexes that indicated a less than satisfactory model fit were not significantly changed by deleting other variables. It is likely that no better model fit is possible with this data set noting it is relatively small (128 samples) for SEM analysis. Consequently, the reduced model shown at Figure 175 is considered the best construct available, grounded in the available ratings data.

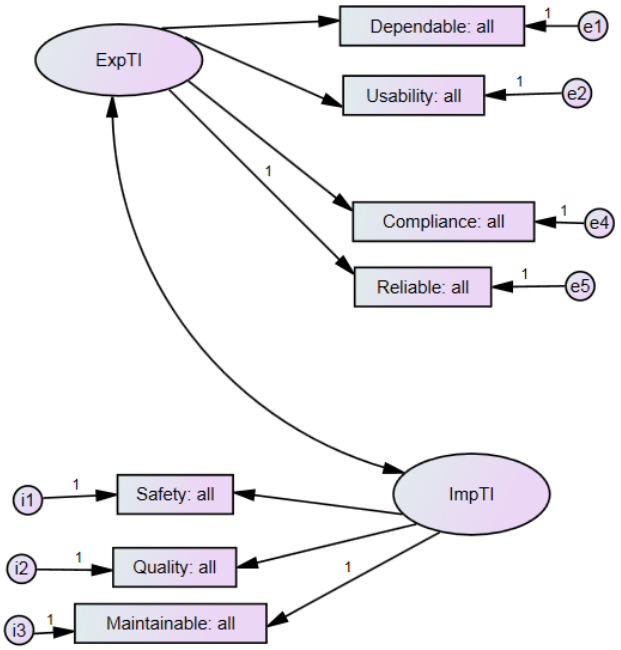


Figure 175. Reduced SEM model of technical integrity concepts and indicators.

The standardised regression weights are as shown in Table 100. No weights were smaller than 0.5 indicating all the variables remained at least reasonably aligned with the factors.

Table 100. Standardised regression weights - reduced technical integrity construct.

Variable		Factor	Estimate
Comply_all	<---	ExpTI	.574
Useable_all	<---	ExpTI	.821
Depend_all	<---	ExpTI	.764
Reliable_all	<---	ExpTI	.891
Safety_all	<---	ImpTI	.625
Quality_all	<---	ImpTI	.714
Maint_all	<---	ImpTI	.714

9.4 Discussion

9.4.1 Revision to the theory of technical integrity

The confirmatory factor analysis of section 9.3.3 resulted in a refined measurement construct, grounded in ratings data from the survey research, for technical integrity

based around the factors termed ‘explicit technical integrity’ and ‘implicit technical integrity’. The relationship of these factors to technical integrity and the indicators correlated to them are illustrated in Figure 176.

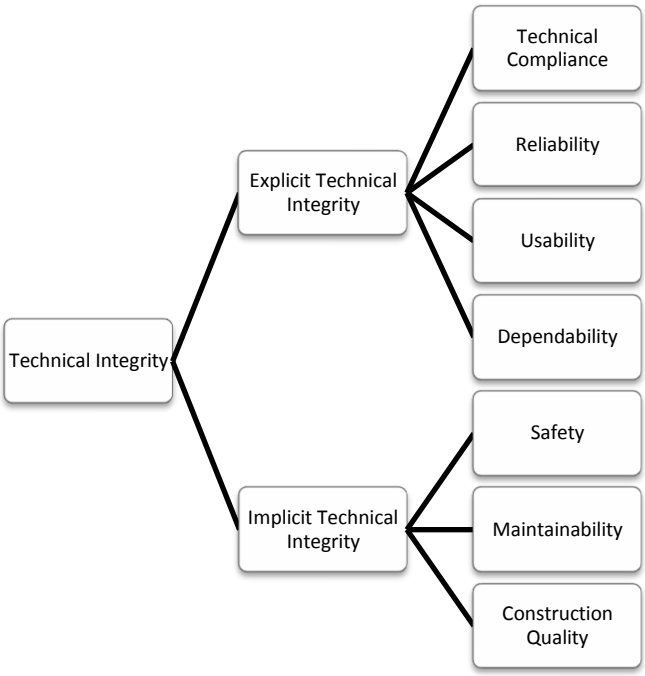


Figure 176. Theory of technical integrity comprising explicit and implicit components.

This in effect is an alternate theory of technical integrity. Compared to the theory proposed in Chapter 5, the primary indicators in this theory mostly align with the concept of dependability. The exception is the indicator technical compliance which was theorised to correlate with the concept of capability, but the other indicators of explicit technical integrity were not expected to correlate with the concept of capability. A different measurement construct to that proposed for technical integrity comprising capability and dependability concepts is inferred. A conjoint measurement approach to aggregate estimates of explicit technical integrity and implicit technical integrity is relevant for assessment of an instrument.

Safety is an element of this model, but the factor analysis did not reveal it, or any other indicator of group of indicators, as being a third component of technical integrity. Nor does it load onto a component that is solely associated with the risk of unintended, undesirable consequences. The qualitative data extracted from the survey respondents that indicated that safety would dominate the concept of dependability and that it should

form a third element of technical integrity is not supported by the ratings data from those same respondents.

9.4.2 Assessment of measurement instruments

Four different measurement constructs were identified by previous phases of this research. In this section, an instrument is synthesised for each of the constructs and, using that instrument, an estimate of technical integrity was made for each of the three systems that were the subject of the survey. The goodness of each of the measurement instruments is then summarised and compared. Determination of the "best" measurement instrument is then determined and discussed.

9.4.2.1 Direct measurement of technical integrity using the CAT

The method and results of the direct measurement of technical integrity were presented in detail in section 8.3.1. In summary, the values obtained for the technical integrity each system are shown in Table 101. The mean and median ratings from the judgment ratings are presented. The normalised rating takes the mean rating and divides by 16 in order to compress to a [0..1] numeric scale.

Table 101. Technical integrity measures using the CAT.

System	Rating		Normalised Rating
	Median	Mean	
System A	9	9.14	0.571
System B	11	10.91	0.682
System C	11	10.81	0.676

Recall that Cronbach's alpha for direct measurement of technical integrity was 0.885, regarded as good reliability.

9.4.2.2 Technical integrity scale of indicators

Section 8.3.3 investigated the reliability of forming a scale from the twenty indicators from which judgments were obtained for the three example systems. It found that reliability of such a scale was excellent ($\alpha > 0.9$). It was also clear that inclusion of the indicator 'Environment Impact' detracted from the reliability. The indicators 'Technical Complexity' and 'High Technology' are also not theorised to be indicators of technical integrity. Removal of these three indicators resulted in a seventeen item scale from which ratings for each of the three systems was established.

The scale was calculated in SPSS using the ratings data for the three systems. This provided 43 valid ratings for each of the 17 indicators for each of the three systems. Table 102 shows the resultant reliability of the scale obtained for each system, the scale mean and scale standard deviation for each system and the scale mean normalised to a [0..1] interval. The latter is the rating of technical integrity produced by this measurement instrument.

Table 102. Technical integrity ratings from scale of indicators.

System	Scale Rating		Normalised Rating
	Reliability (Cronbach's α)	**Mean / Standard Deviation**	
System A	Excellent (α = 0.92)	157.39 / 35.524	**0.579**
System B	Excellent (α = 0.92)	178.37 / 33.184	**0.656**
System C	Excellent (α = 0.90)	180.23 / 28.743	**0.663**

9.4.2.3 Conjoint measurement from aggregation of capability and dependability

The theory of technical integrity established in Chapter 5 proposed a conjoint measurement construct of capability and dependability. While this construct has not been fully supported by the results of factor analysis of the survey data, it is still reasonable, particularly considering the content validity established for the construct in Chapter 6, to synthesise a measurement instrument as a candidate measurement approach. Recall that a multiplicative conjoint measurement had been proposed and for defence systems the aggregation coefficients could be set to provide the same relative importance to capability and dependability based on the results of Chapter 7. The approach of using the percentage of technical requirements compliant had also been validated as a proxy measurement for the capability component.

Measurement of dependability on a [0..1] numerical scale had been found to be valid in section 6.5.5 but no measurement instrument with significant support arose from the survey data or from other precedents in the literature. In order to realise a measurement instrument for dependability and assess that instrument it was decided to construct a scale for dependability. The scale is constructed from the eleven indicators postulated to correlate with dependability as shown in Table 12. The scale was calculated in SPSS

using the ratings data for the three systems. This provided 43 valid ratings for each of the 11 indicators for each of the three systems.

Table 103 shows the values obtained for capability, dependability and the resultant value from a multiplicative conjoint for technical integrity. The reliability of the dependability scale obtained for each system, the scale mean and scale standard deviation for each system and the scale mean normalised to a [0..1] interval are shown. The last column is the rating of technical integrity produced by this measurement instrument. It is obtained simply by multiplying the value for capability by the normalised value for dependability and taking the square root of the result as per the equation in section 7.6.1.

Table 103. Technical integrity ratings from a conjoint of capability and dependability.

System	Capability (compliance)	Dependability Scale Rating			Technical Integrity Rating
		Reliability (Cronbach's α)	Mean / Standard Deviation	Normalised	
System A	0.56	Good (α = 0.886)	113.61 / 25.612	0.646	**0.601**
System B	0.76	Good (α = 0.893)	120.23 / 25.334	0.683	**0.720**
System C	0.68	Good (α = 0.837)	119.09 / 18.222	0.677	**0.678**

9.4.2.4 Conjoint measurement from aggregation of explicit and implicit technical integrity

Using the amended theory of technical integrity from section 9.4.1 and the findings of practical means of indicator measurement from section 8.3.4, a measurement instrument for technical integrity was synthesised as illustrated in Figure 177. The measurement instrument consists of two scales, the results of which are combined in a multiplicative conjoint. The two scales are:

- Explicit technical integrity comprising elements:
  - Readily obtainable values for the variables of MTBF and Percentage of technical requirements complied with.
  - Expert judgments of indicators: usability and dependability.
- Implicit technical integrity comprising expert judgments of indicators: maintainability, construction quality and safety.

It is noted there is some reservation relating to the method to be used to estimate the indicator safety. Rating of safety by judgments was of questionable reliability (α = 0.685) and while the correspondence of ratings between the indicator model and the judgments was good (see Table 73) the indicator model was not general enough for application across a wide range of systems. Judgment data was used in the instrument in lieu of any other general method being identified (a general measure for safety of a complex engineered system remains a question for future research).

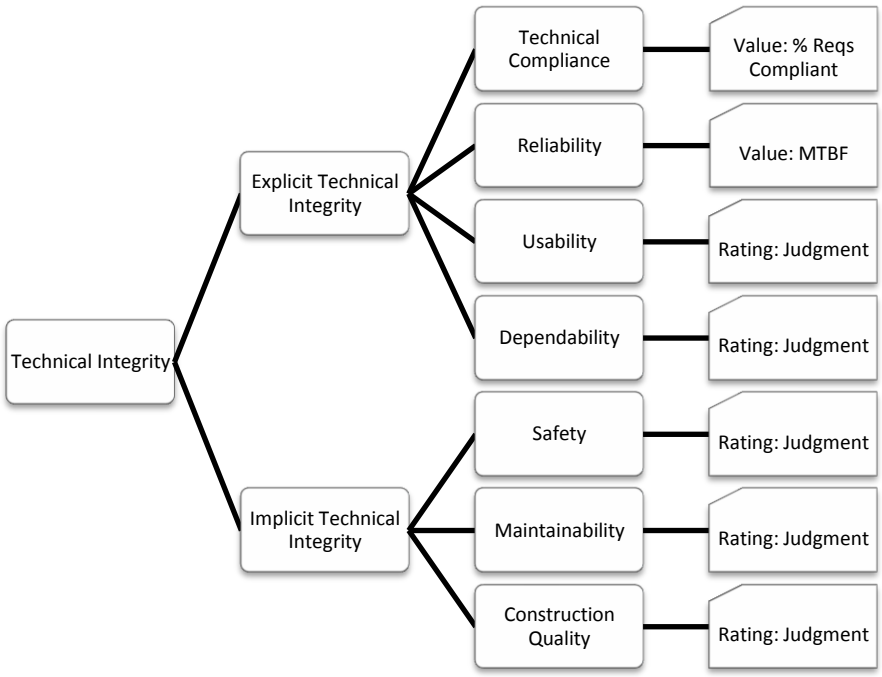


Figure 177. Measurement construct for technical integrity from explicit and implicit components.

The instrument requires the combination of indicator values into two scales:

- Explicit technical integrity: Cronbach's alpha for a scale constructed from judgment data of the four indicators was 0.836 representing good reliability. This was considered to be a lower bound for reliability since the indicators of reliability and technical compliance will be measured from objective values and therefore contribute more reliably than the judgment data used in this calculation. Removal of any one of the indicators from the scale does not improve reliability.
- Implicit technical integrity: Cronbach's alpha for a scale constructed from judgment data of the three indicators was 0.689 representing questionable reliability. This is a realistic estimation since all of the indicators in the construct are to be assessed via

judgments. Removal of any one of the indicators from the scale does not improve reliability.

Chapter 7 established the validity of combining the concepts of capability and dependability into a multiplicative conjoint equation with the coefficients set to the same value for defence systems. While this result cannot be read directly across to this instrument, the considerations that led to the proposed form of conjoint equation in section 5.3.3 are equivalent to those that require consideration for the conjoint aggregation of explicit and implicit technical integrity. For the purpose of generating a practical measurement instrument, the coefficients of the conjoint equation are also both set to 0.5 and this is assumed to be valid at least for the defence systems of importance to the research problem.

Table 104 shows the results of forming a scale from the indicator ratings defined for explicit technical integrity. While not part of the intended measurement instrument it serves to establish a lower bound for the reliability of this part of the measurement instrument. The reliability for the scale ranges from poor to good across the systems. Overall the reliability of this scale is good ($\alpha > 0.8$).

Table 104. Explicit technical integrity ratings from scale of indicators.

System	Scale Rating		Normalised Rating
	Reliability (Cronbach's α)	**Mean / Standard Deviation**	
System A	Good (α = 0.80)	32.28 9.772	**0.504**
System B	Poor (α = 0.67)	47.16 6.343	**0.737**
System C	Acceptable (α = 0.70)	42.42 7.122	**0.663**

Table 105 shows the actual values used for calculating the value explicit technical integrity in the measurement instrument. The values in the compliance column are the ratio of requirements compliant to the total number of requirements. The values for reliability those explicitly defined for each system, normalised to a [0..1] numerical scale as in the indicator model of section 8.3.4.13. Usability is the mean rating value from Table 67 and similarly for dependability from Table 62. The last column shows the normalised rating for explicit technical integrity calculated from the indicators.

Table 105. Explicit technical integrity ratings from additive model of indicators.

System	Compliance	Reliability	Usability	Dependability	Explicit TI Rating
System A	0.56	0.146	8.26	8.28	**0.435**
System B	0.76	0.488	11.84	11.33	**0.674**
System C	0.68	0.366	11.16	10.58	**0.601**

Table 106 shows the results of forming a scale from the indicator ratings defined for implicit technical integrity. The reliability for the scale ranges from poor to good across the systems. Overall the reliability of this scale is questionable ($\alpha > 0.6$).

Table 106. Implicit technical integrity ratings from scale of indicators.

System	Scale Rating		Normalised Rating for Implicit TI
	Reliability (Cronbach's α)	**Mean / Standard Deviation**	
System A	Questionable ($\alpha = 0.67$)	31.17 6.446	**0.649**
System B	Good ($\alpha = 0.81$)	27.42 8.212	**0.571**
System C	Poor ($\alpha = 0.51$)	33.19 5.56	**0.691**

Table 107 presents the results of combining the ratings of explicit and implicit technical integrity. It is obtained simply by multiplying the value for explicit technical integrity by the value for implicit technical integrity and taking the square root of the result.

Table 107. Technical integrity conjoint rating from explicit and implicit components.

	Explicit Technical Integrity	Implicit Technical Integrity	Technical Integrity
System A	0.435	0.649	**0.531**
System B	0.674	0.571	**0.620**
System C	0.601	0.691	**0.645**

9.4.2.5 Goodness of measurement instruments

It is important to ensure that the instrument or instruments selected to measure technical integrity both accurately measure it and are actually measuring the characteristic of a complex engineered system that is the assumed latent variable technical integrity. 'Goodness' is assessed by looking at the reliability and validity achieved by the

measurement constructs. (Sekaran, 2003, p. 202-209). In this research, reliability was assessed primarily by Cronbach's alpha (α) of the ratings data used to establish a measure. Validity was assessed as broadly as possible and was considered in terms of content validity, criterion-related validity and construct validity, the latter in turn assessed through convergent and discriminant validity.

Criterion-related validity (concurrent validity in this research) assessed if the measurement discriminates individuals known to be different. The three example systems were intentionally contrived to be different across a large range of system quality attributes and technical parameters and it was thought unlikely, though possible, that any two would exhibit the same value for technical integrity. The values obtained for technical integrity in sections 9.4.2.1 through 9.4.2.4 are plotted for each of the measurement instruments in Figure 178 and this was used subsequently to assess concurrent validity of the instruments.

The assessment of the reliability and validity of each instrument is summarised in Table 108. The reliability of each instrument is good or excellent except for instrument 4, the conjoint construct of explicit and implicit technical integrity.

The use of an ordinal scale for technical integrity was established through consensus of expert judgments and is a minimum level of content validity that was established and is applicable to each instrument. The content validity of the conjoint construct of capability and dependability was rigorously determined as described in Chapter 6.

Instrument 3 was found to provide the most discrimination between the three example systems while instruments 1 and 2 did not discriminate between the technical integrity of System B and System C (notwithstanding they could have equivalent technical integrity).

Convergent reliability looks for correlation between the results of measurement instruments measuring the same underlying characteristic. Instrument 1 and instrument 2 have provided similar results in ranking and ratings and agree with instrument 4 in rankings. Instrument 3 does not correlate with these other instruments for the ranking of Systems B and C but does correlate with the low ranking of System A.

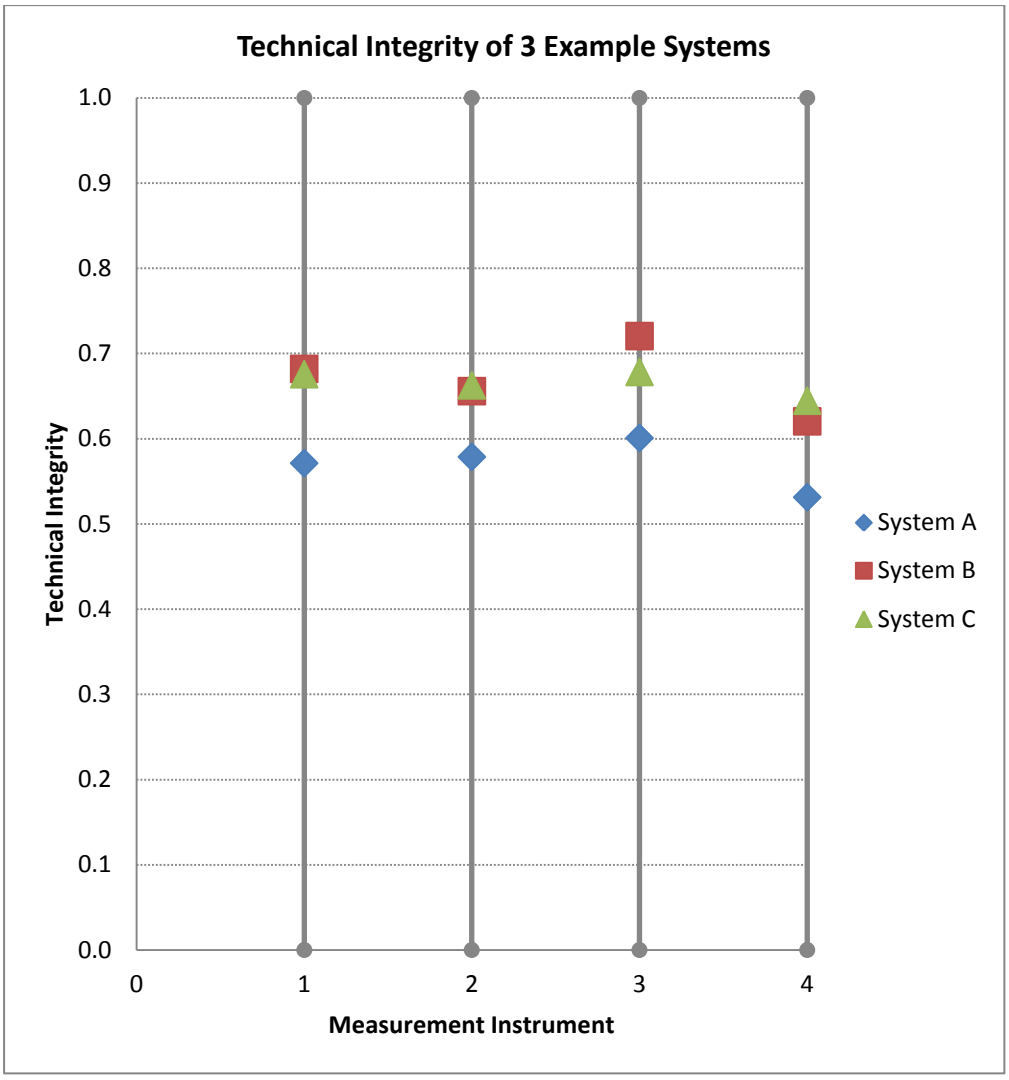


Figure 178. Values of technical integrity for the three example systems using four measurement instruments.

Two indicators were included in the survey research as a means of assessing discriminant validity. Those indicators were technical complexity and high technology. It has been seen that high technology has been reasonably highly correlated with technical integrity in the results for all instruments. It is considered, with hindsight, that this could be correlated with many other indicators of technical integrity, e.g., newer, higher technology systems would generally have improved usability, better reliability and more features compared to lower technology systems. So, discriminant validity should not be discounted because of this. Technical complexity is not strongly correlated with technical integrity which supports discriminant validity of the instruments.

The instruments are practical in the sense that they are readily able to provide an estimate of the technical integrity and are general enough to be applied over a wide range of system types and domains of system application. However, it is not readily apparent which instrument could be considered to have the most ‘goodness’ and be both the most reliable and the most valid.

Table 108. “Goodness” of measurement instruments.

Measurement instrument	**Reliability**	**Content validity**	**Criterion related validity (concurrent validity)**	**Construct validity**	
				Convergent validity	**Discriminant validity**
(1) **Direct CAT** TI Direct Measurement via CAT	Good (α = 0.885)	Ordinal scale for technical integrity is valid by virtue of expert consensus.	Discriminated technical integrity of older radar system (System A) from the two newer radar systems (Systems B and C). Did not discriminate between System B and C.	Yes. Ranking of systems correlates with instrument (2) and reasonably well with instrument (4).	Rating data for the indicator Technical Complexity is not strongly correlated with this measure (r_S = 0.388). Rating data for the indicator High Technology is more strongly correlated (r_S = 0.563) and ranks as one of the highest indicators correlating with the technical integrity ratings.
(2) **Scale of Indicators** Scale from theorized indicators of technical integrity	Excellent (α in range 0.90 to 0.92)	Ordinal scale for technical integrity is valid by virtue of expert consensus. Face validity for indicators of technical integrity (no expert consensus gathered).	Discriminated technical integrity of older radar system by 12% (System A) from the two newer radar systems (Systems B and C). Did not well discriminate between System B and C (~1% difference in rating).	Yes. Ranking of systems correlates with instrument (1) and reasonably well with instrument (4).	As above.

Measurement instrument	Reliability	Content validity	Criterion related validity (concurrent validity)	Construct validity	
				Convergent validity	Discriminant validity
(3) **Capability and Dependability Conjoint** Conjoint construct of capability and dependability concepts	Good – Excellent α = 1 for capability. α in range 0.837 to 0.893 for dependability scale.	Ordinal scale for technical integrity is valid; conjoint construct of capability and dependability is valid; capability measurement on 0..1 scale by way of requirement compliance is valid; dependability measurement on a 0..1 scale is valid (but no method validated) – all by virtue of expert consensus. Face validity for indicators in scale for dependability (no expert consensus gathered).	Discriminated on differences between the systems. Particularly differences in compliance. System B discriminated from system C by 11% of rating, and system C from system A by 22%.	No. Only correlates with other instruments in terms of ranking of System A last. Inverts rankings of Systems B and C compared to instrument 4 and instruments 1 and 2 hold these as equivalent technical integrity.	The compliance term of the construct will be uncorrelated with Technical Complexity and with High Technology. The items in the scale for dependability show correlations in the order of $r_S \approx 0.3$ and $r_S \approx 0.5$ for these same discriminant indicators.

Measurement instrument	Reliability	Content validity	Criterion related validity (concurrent validity)	Construct validity	
				Convergent validity	Discriminant validity
(4) **Explicit and Implicit Conjoint** Conjoint construct of explicit and implicit technical integrity concepts	Poor-Questionable α ranges from 0.67 to 0.80 for scale of explicit technical integrity. α ranges from 0.51 to 0.81 for scale of implicit technical integrity.	Ordinal scale for technical integrity is valid by virtue of expert consensus. Face validity of construct of implicit and explicit technical integrity claimed by virtue of interpretation of FCA results with group of professional colleagues.	Discriminated on differences between the systems. System B discriminated from system C by 8% of rating, and system C from system A by 22%.	Yes. Correlates with other instruments in terms of ranking System A lowest. Close rating of Systems B and C is not hugely inconsistent with measurement instruments 1 and 2 that rank these equivalently. Rankings of systems B and C inverted with respect to instrument 3.	The items of explicit technical integrity that will rely on judgment ratings are usability and dependability. These are both moderately correlated with Technical Complexity (r_S = 0.44) and more strongly correlated with High Technology $r_S \approx$ 0.61). The other items will be directly measured and therefore will be uncorrelated with these discriminant indicators. The items of implicit technical integrity that will rely on judgment ratings are safety, quality and maintainability. These are effectively not correlated with Technical Complexity (all $r_S < 0.27$) or with High Technology (r_S <0.24 except for maintainability where r_S =0.35).

9.4.2.6 Determination of best measurement instrument

It was not clear which of the four measurement instruments evaluated could be considered best in terms of reliability and validity. Nor are these the only characteristics of the measurement instrument that are to be considered in recommending a measurement approach.

Pragmatically, cost[21] is also a significant issue in this measurement problem and needs to be considered in selection of an instrument. Table 109 qualitatively describes the considerations as to the cost associated with each instrument. While calculated estimates for the percent of requirements compliant and for MTBF may be considered an expense, it is usual for these estimates to be readily available during a design and acquisition phase of a system project. This data is required in the system descriptions for the reliability of judgments observed in this research to be obtained. So these are not a differentiator for cost. Cost is therefore driven, assuming the same number of judges in each case, by the number of indicators or judgment dimensions that need to be rated. Clearly then, instrument 1 is the least expensive means to estimate technical integrity. It is to be noted that the relativity in the costs is subjective and in absolute terms, a particular research objective may be able to justify higher costs.

Table 109. Relative cost of measurement instrument.

Measurement instrument	Calculated estimates	Indicator ratings	Relative cost
(1) **Direct CAT** TI Direct Measurement via CAT	Nil	1 rating from multiple judges	Inexpensive
(2) **Scale of Indicators** Scale from theorised indicators of technical integrity	Nil	17 ratings from multiple judges	Very expensive
(3) **Capability and Dependability Conjoint** Conjoint construct of capability and dependability concepts	Percent of requirements compliant	11 ratings from multiple judges	Expensive
(4) **Explicit and Implicit Conjoint** Conjoint construct of explicit and implicit technical integrity concepts	Percent of requirements compliant MTBF	5 ratings from multiple judges	Moderate

[21] i.e., a measure of the difficulty and degree of effort required to obtain the estimate.

In order to identify the best measurement instrument from considerations of practicality (i.e., cost), reliability and validity, a decision model was constructed in CDP as shown in Figure 179. The criterion at each level was simply weighted equally as a default approach. The criterion of reliability was directly rated for each alternative using an estimate, including uncertainty range, of α. For the other criteria, the alternatives were pair-wised compared using the summary of validity in Table 108 and the relative costs of Table 109 as bases for the comparison.

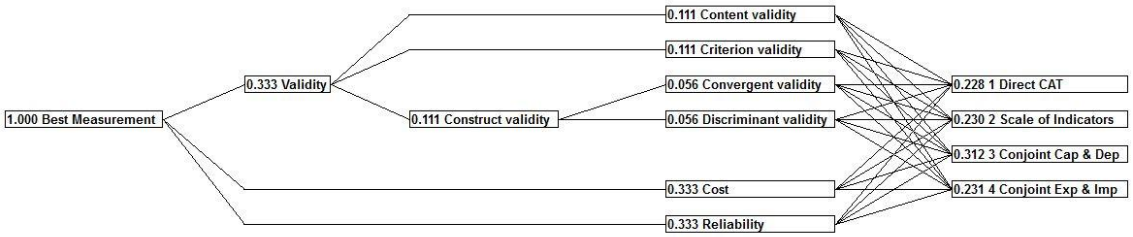


Figure 179. Decision model for best measurement instrument.

Figure 180 shows the results of the decision model. Instrument 1 Direct TI is clearly the best instrument based on this model. Notable is the dominance of the cost element in ranking instrument 1 as the best.

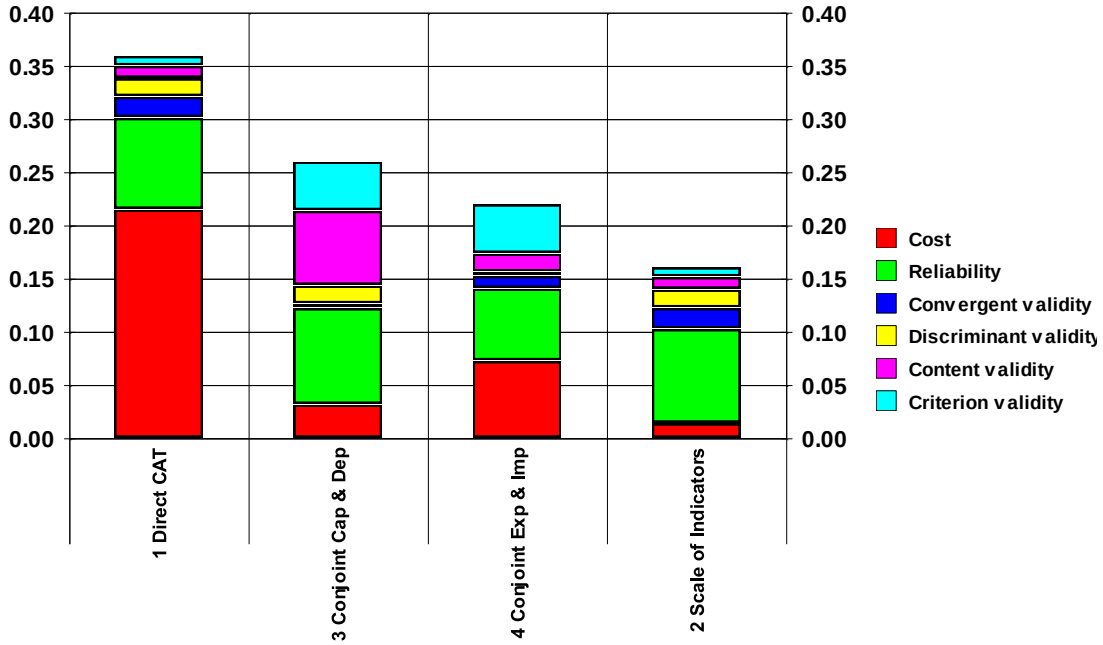


Figure 180. Results of decision model for best measurement instrument.

This result is most sensitive to the weighting of cost. As shown in Figure 181 and Table 110 reducing the weighting of cost by 15% would provide a result that Instrument (3) Capability and Dependability Conjoint is best. Similarly an increase in the weighting of validity by 20% would indicate Instrument (3) Capability and Dependability Conjoint to be best. The decision model is reasonably insensitive to other weighting changes.

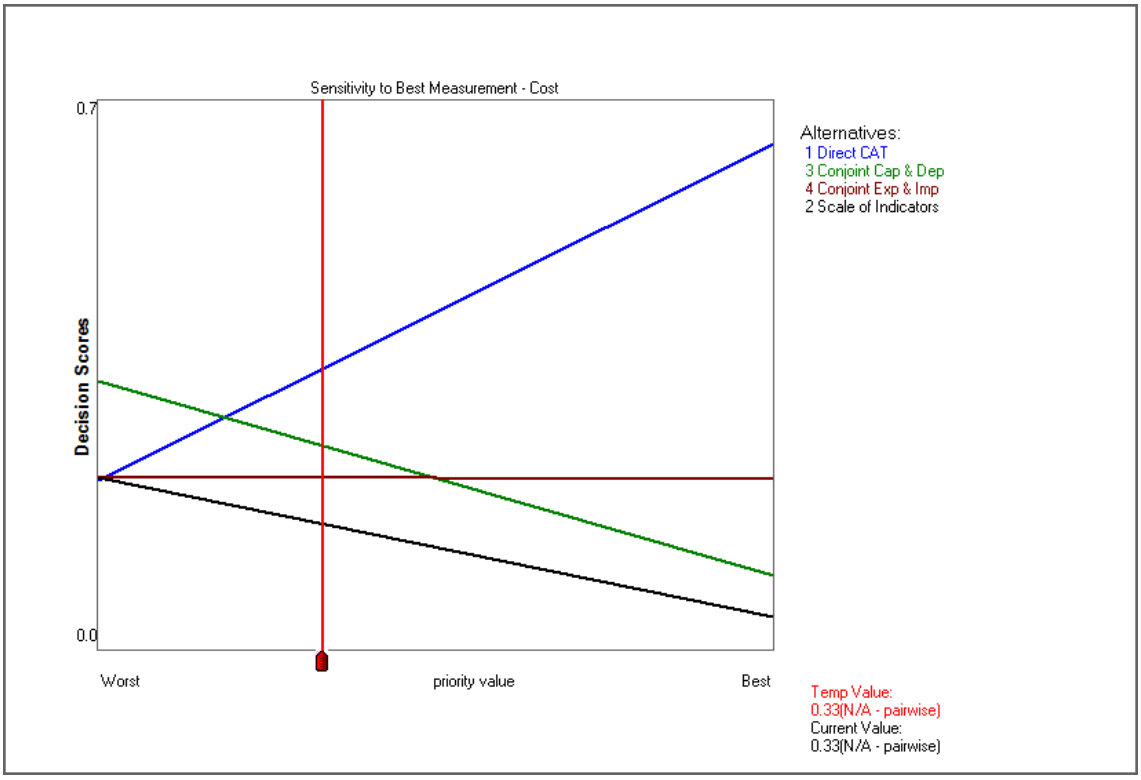


Figure 181. Sensitivity of best measurement instrument result to cost.

Table 110. Sensitivity of leading alternative, Instrument 1, to weights.

Criterion	Sub-criterion	Criticality	Usurper
Best Measurement	Cost	-15%	3 Conjoint Cap & Dep
Best Measurement	Validity	20%	3 Conjoint Cap & Dep
Validity	Construct validity	-52%	3 Conjoint Cap & Dep
Best Measurement	Reliability	59%	3 Conjoint Cap & Dep
Validity	Content validity	69%	3 Conjoint Cap & Dep
Validity	Criterion validity	100%	N/A
Construct validity	Convergent validity	100%	N/A
Construct validity	Discriminant validity	100%	N/A

To examine the determination of the best instrument if cost were not a discriminant, the relative costs of the instruments are set to be equal in the CDP model. The result is shown in Figure 182. The conjoint measurement approach based upon measure of capability and dependability is favoured due in large part to the establishment of its content validity. The relative lack of measurement reliability based upon the indicators used in instrument (4) Explicit and Implicit Conjoint counts against it, even though it reflects the revised theory of technical integrity.

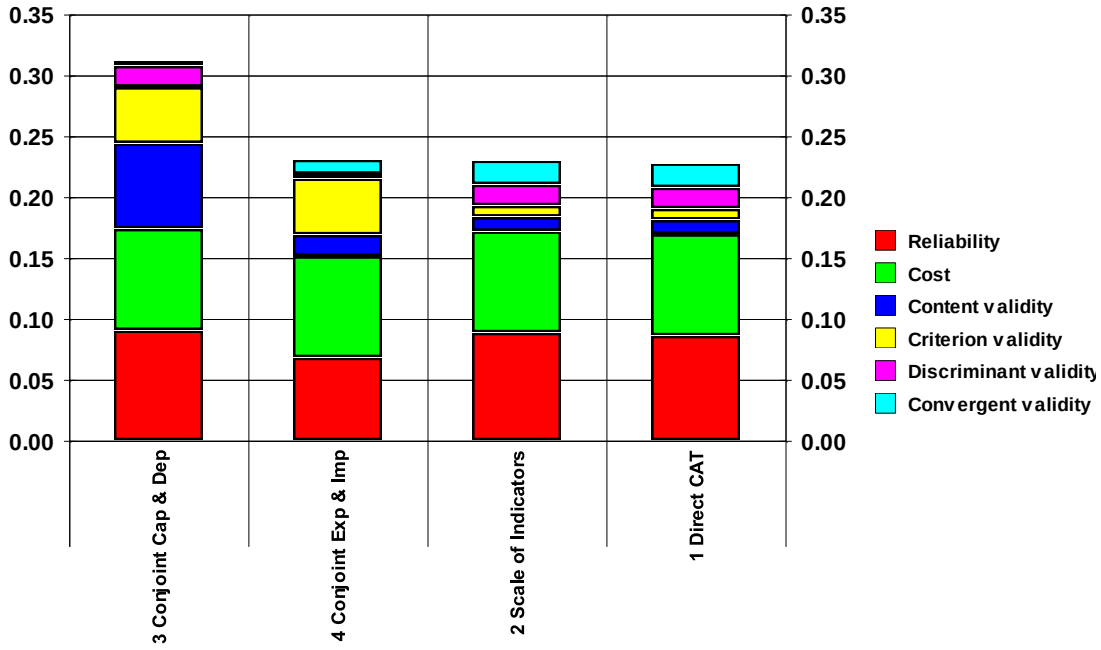


Figure 182. Results of decision model for best measurement instrument, cost not discriminated.

9.4.3 Presentation of technical integrity measurement values

The technical integrity ratings produced from the four measurement instruments were all relatively close numerically and this was perhaps to have been expected for similar systems like those of the examples. However instruments (3) Capability and Dependability Conjoint and (4) Explicit and Implicit Conjoint that combine two factors of technical integrity provide an opportunity to visualise those values to assist in discrimination of the results.

Figure 183 provides a contour chart of technical integrity on the axes of capability and dependability. Isolines of technical integrity at 0.1 intervals are plotted. Overlayed on this chart are the results from Table 103 for systems A, B and C. With this instrument, it

can be seen that the systems had approximately the same rating for dependability and differed mostly on capability. It is the difference in capability ratings that give rise to difference in technical integrity ratings for the systems in this study.

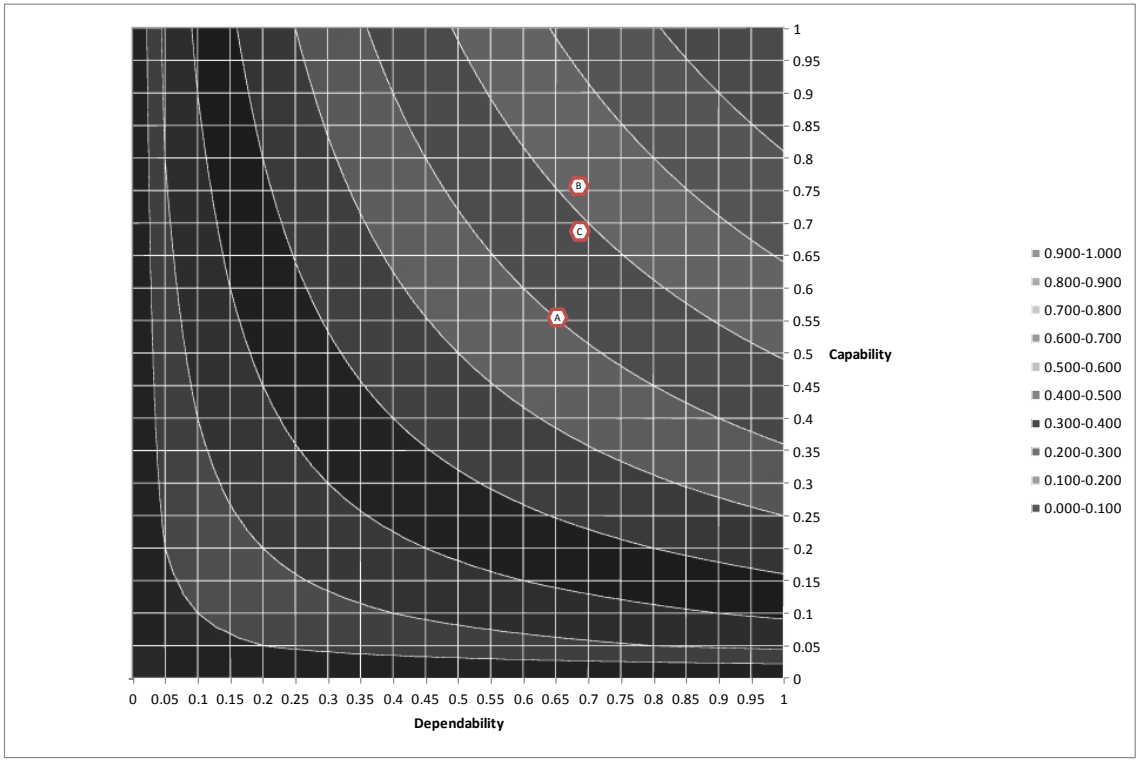


Figure 183. Plot of technical integrity on axes of capability and dependability.

Similarly, Figure 184 provides a contour chart of technical integrity on the axes of explicit technical integrity and implicit technical integrity. Overlayed on this chart are the results from Table 107 for systems A, B and C. With this instrument, it can be seen that the systems differed most in explicit technical integrity. Systems B and C whose overall technical integrity ratings are approximately the same can be clearly differentiated on the contour chart: System B having higher explicit technical integrity and System C having implicit technical integrity.

Using these contour charts as aids to visualising the results of measurement instruments (3) Capability and Dependability Conjoint and (4) Explicit and Implicit Conjoint improves concurrent validity of these instruments relative to instruments (1) Direct TI and (2) Scale of Indicators that do not provide component ratings needed for this type of visualisation.

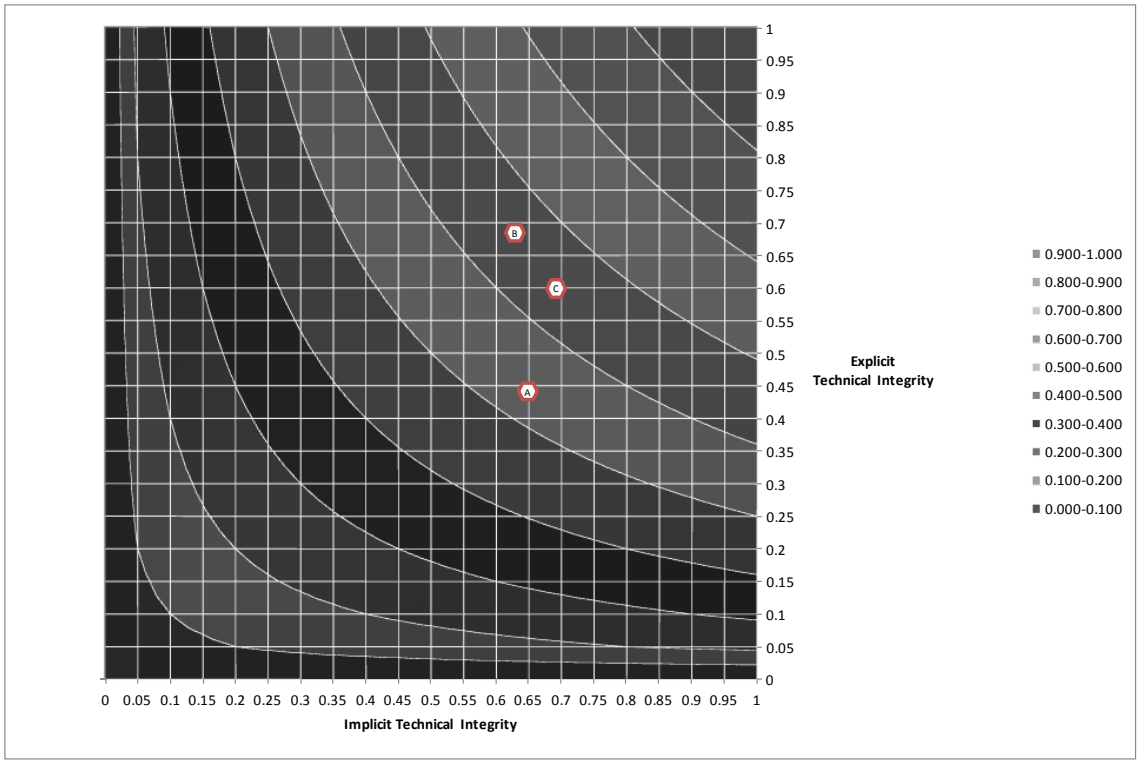


Figure 184. Plot of technical integrity on axes of implicit and explicit technical integrity.

9.5 Conclusions

This section summarises the findings and limitations arising from the research that was reported in this chapter.

9.5.1 Findings

The theory of technical integrity involving two orthogonal concepts of capability and dependability is not supported by the ratings data and may be wrong. The grounded theory of technical integrity derived from the ratings data from judgments of three example systems indicates two distinct factors that have been labelled Explicit Technical Integrity and Implicit Technical Integrity.

Four viable measurement instruments have been identified and assessed. The "best" instrument is the direct measurement of technical integrity using the CAT. Practicality, as estimated by a relative cost, dominated the assessment and if cost is not a discriminator then the conjoint measurement approach based upon the concepts of capability and dependability was favoured due in large part to the establishment of its content validity and the reliability of its constituent measures.

The measurement approach of using the CAT to directly measure technical integrity is available to future research to further solutions to the overall research problem of assessing the effectiveness of technical regulations. Use of the measurement instrument will provide further validation for the measurement approach.

9.5.2 Limitations

The survey research undertaken during this project consumed hundreds of hours from experienced professionals. This has provided a constraint on obtaining additional data that could further support the findings.

While the conclusions drawn have been grounded in data and are significant for the representative systems used in the survey research, further studies are required to extrapolate the findings to systems that are different in application, domain of use and size, scope and complexity.

Both the inductively derived theory of technical integrity (inclusive of the concepts of capability and dependability) and the revised theory, grounded in ratings data from the three representative systems, confirm that technical integrity exhibits two factors. However, the "best" approach to the measurement of technical integrity does not provide any estimate of these factors. This may limit the interpretation of measurements undertaken in future applications. In particular, some provisions of technical regulation could conceivably affect only one of the identified factors, but this would not be readily identified. This consideration may still favour an approach of the conjoint measurement construct particularly if the cost of obtaining measurements is of lesser importance than the reliability and validity of the measurements.

10 Summary, conclusions and implications

10.1 Summary

The primary research problem considered by this thesis was: *How effective are the provisions of technical regulations in assuring the technical integrity of complex engineered systems?*

To begin this investigation a broad background survey of technical regulation was undertaken. In particular, in the ADF where the stated objective of technical regulation is technical integrity, the provisions of technical regulation and the nature of technical integrity were examined. This examination highlighted knowledge gaps including:

- The lack of a model for how technical regulation is meant to achieve its technical aims.
- There has been no cost-effectiveness or cost-benefit analysis undertaken to justify the introduction of, or extent of, regulatory frameworks.
- While the explicit goal of ADF technical regulation is *technical integrity*, there is not a clear, objective definition of technical integrity established and the subjective definitions vary across the air, land, maritime and explosive ordnance regulatory domains.
- Fitness for service is a key element in the subjective definition of technical integrity but the definitions of fitness for service are not objective and differ across the devolved regulatory domains.

- While heuristically supportable, the provisions required by ADF technical regulations have not been objectively shown to be sufficient or effective in assuring the technical integrity of complex engineered systems.

The research was delimited to complex engineered systems, and the main concepts of these were reviewed and summarised including an appreciation of the properties and characteristics of such systems. The contemporary means for designing and implementing these systems was identified in order to compare these to the provisions of technical regulation. Highlighted knowledge gaps included:

- There is no evidence available to support the implicit claim that provisions of ADF technical regulations, including the processes of Systems Engineering, positively affect the technical integrity of materiel.
- The technical regulatory system does not deal with current trends in the development and realisation of complex systems and systems-of-systems which increasingly are constrained by an acquisition paradigm of procuring off-the-shelf components to reduce development risk.
- There is no universally agreed definition of a complex engineered system. This consideration led to derivation of one for use in this research:

> A complex engineered system is defined to be a system that has been intentionally engineered from numerous components: technological (hardware and software) as well as human, each of which has several interconnections and is interdependent, yielding a system with both intended and unintended emergent behaviour.

At this point, it became clear that a focus on establishing a way to measure the technical integrity of a complex engineered system was a research question that must necessarily be addressed first as it is a prerequisite for the objective assessment of the effectiveness of the provisions of technical regulation. The key research question pursued became:

How can the technical integrity of a complex engineered system be measured?

The literature survey sought to identify previous approaches or similar concepts to measure technical integrity and to review relevant measurement theory and methodologies necessary to construct a measurement instrument for technical integrity. The literature survey found that:

- There is no consensus of support for measuring abstract system properties like technical integrity in the literature.
- There is no precedented measurement construct or measurement instrument for technical integrity.
- There is no pragmatic measurement approach for estimation of the risk (unintended, undesired outcomes) of a system or to compare that risk across a class of similar systems or across dissimilar systems.
- There is no precedented measurement approach for technical integrity.

Several measurement approaches were identified for investigation under the subsequent research phases: the Consensual Assessment Technique; conjoint measurement; scales and indices; and the use of proxy measurements for some variables thought to be indicators of technical integrity.

A phased approach that began as *exploratory* research and progressed to *descriptive* research was adopted for this thesis. The phased approach was designed to:

- In the exploratory research phase:
  - Develop theory, concepts and a measurement construct for technical integrity.
- In the descriptive research phases:
  - Assess the content validity of the measurement construct for technical integrity,
  - Assess the feasibility of direct estimation of technical integrity by expert judgement,
  - Establish parameters for the measurement instrument, and
  - Establish the variables to be included in candidate measurement instruments and establish the reliability and validity of those measurement instruments.

The concept development phase of the research yielded a generalised theory of *technical integrity* as consisting of orthogonal concepts of capability and dependability, where:

- Capability is defined to mean that the system meets all specified requirements for performance, function and quality attributes in the defined environment and when

used as intended. Capability is the designed set of emergent properties of a system that are intended and useful (i.e., required); and

- Dependability is defined to mean that the risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. It includes indicators such as robustness (system is tolerant of usage that may exceed specified conditions without catastrophic consequences), reliability (failures do not happen often), and system safety (risk of unidentified hazards and failure modes resulting in undesirable outcomes is minimised).

This theory allowed consideration of appropriate forms of measurement and in particular for a conjoint measurement construct for technical integrity to be developed. Quality attributes of a complex engineered system that could be indicators of the technical integrity concepts and corresponding variables that should contribute to or be full or partial proxies for technical integrity were also identified in order to support development of a practical measurement instrument.

The research progressed to establish the content validity of the technical integrity measurement construct. A survey instrument gathered both rating data to particular questions seeking the level of agreement with various aspects of the conjoint measurement construct for technical integrity and asked open-ended questions through which additional insights into the reasons behind the expert ratings emerged. Both quantitative and qualitative analysis methods were used to analyse the data gathered from 116 respondents. The quantitative analysis procedure tested the following hypotheses:

- Technical integrity can be measured on an ordinal scale.
- Technical integrity can be measured by a single numeric value.
- Technical integrity can be measured on a ratio scale.
- Capability of a complex engineered system can be measured on a scale of range 0 to 1.
- Capability of a system in excess of specified requirements is valued.
- The proposed capability measure is best.
- Dependability of a complex engineered system can be measured on a scale of range 0 to 1.

- Dependability of a complex engineered system better than the risk acceptability threshold is valued.
- The proposed dependability measure is best.
- Technical integrity can be measured by a conjoint measurement from the components capability and dependability.
- The conjoint measurement of technical integrity allows for comparison of similar systems.
- The conjoint measurement of technical integrity allows for comparison of dissimilar systems.

The same survey provided quantitative ratings which allowed determination of an approach to aggregate measures of capability and dependability into the conjoint measurement construct of technical integrity. The quantitative analysis procedure tested the hypotheses that *the ratings of the relative importance of capability versus dependability are the same across the different system domains*. Qualitative analysis of the elicited reasoning behind the judgements was undertaken and triangulated with the quantitative results to achieve a better indication of the validity of the aggregation construct.

A second survey instrument was developed based on the CAT methodology. It elicited expert ratings of technical integrity and 20 other quality attributes of three representative complex engineered systems. 18 of these quality attributes were identified as indicators likely to correlate with or contribute to a measurement of technical integrity. Some of these quality attributes were readily assessable by objective variables while others were intentionally much more subjective. Two of the quality attributes were subjective characteristics of a complex engineered system that were not necessarily expected to correlate with technical integrity. The CAT was used as the underlying methodology for the survey development, but results obtained were able to be assessed for the estimation of technical integrity via instruments that could include scales, conjoint measurement and the direct measurement of proxy variables.

The survey instrument also asked open-ended questions about the formation of the ratings through which additional insights into the judges' reasoning and use of objective data was gleaned. Both quantitative and qualitative analysis methods have been used to

analyse the data gathered from 43 complete responses. The analysis procedure explored the following research questions:

- Using the CAT, how reliable can direct expert ratings of technical integrity be?
- How reliable can expert ratings for other system indicators (system quality attributes) be?
- Can the judgments of the indicators (system quality attributes) form a reliable scale for technical integrity?
- Where objective variables of the indicators (system quality attributes) are available, how well do the variables and expert ratings of the indicators correspond?
- What is the best measurement instrument to be used for each indicator (via direct measurement of a proxy variable or via expert judgment)?
- Using expert ratings of indicators, can it be shown that there are two factors that contribute to technical integrity and how well do these factors correspond to the concepts of capability and dependability?
- Can reliable scales be formed from these indicators and variables to produce a measurement instrument for technical integrity?

The results from these research questions provided the basis for synthesising four candidate measurement instruments which were evaluated for reliability, validity and cost.

10.2 Conclusion about the research questions

The primary research question addressed by this thesis was: how can the technical integrity of a complex engineered system be measured? This section summarises the conclusions from each of the research phases conducted prior to summarising them. There is little previous research directly related to this research question, therefore the findings cannot be comparatively assessed against any previous theory and results.

10.2.1 Conjoint measurement construct for technical integrity

A theory and concepts for technical integrity was developed in Chapter 5. The main contribution in this development was a theory of technical integrity consisting of two orthogonal concepts arsing from the intended, desired emergent properties and the unintended, undesired emergent properties of a complex system. These concepts were termed capability and dependability.

From this theory, and a consideration of measurement theory, a conjoint measurement construct to combine measures of capability and dependability was justified. An approach to measure capability via a calculation of the percentage of specified requirements achieved was identified. For dependability, such a straightforward and objective approach was not apparent, which lead to a proposal of estimating this via expert judgment.

In order to form a basis for the research in this area, the technical goals of a range of technical regulations were identified and a correspondence with quality attributes of a complex engineered system was established. This allowed for an extensive set of indicators of technical integrity to be defined for subsequent research phases. In addition this work has been structured into an ontology for technical regulation that will support future research in this area.

10.2.2 Content validity of the technical integrity measurement construct

The content validity of the technical integrity measurement construct was assessed by survey research that sought expert judgments. Chapter 6 found support that the conjoint measurement construct for technical integrity is considered valid for the ranking of similar systems for the purpose of assessing the impact of technical regulation. Specifically it has been found that:

- Measurement of *technical integrity* on an ordinal scale is valid.
- Constructing a conjoint measurement from measures of *capability* and *dependability* is valid.
- Measurement of *capability* on a 0 to 1 range is valid.
- *Capability* in excess of that specified is valued.
- Measurement of *dependability* on a 0 to 1 range is valid.
- *Dependability* better than the acceptability threshold of the system's stakeholders is valued.

10.2.3 Conjoint measurement aggregation

It was not known if the relative importance of the concepts of capability and dependability differed across different application domains of complex engineered system, an answer being necessary to establish aggregation weights in a conjoint

construct for technical integrity. Analysis of expert survey responses in Chapter 7 identified a new concept: that the relative importance of capability and dependability falls into three categories:

- 'Military systems' and 'Civil ICT systems' where capability and dependability are of equal importance,
- 'Important Civil systems' where dependability is considered more important than capability, and
- 'Critical Civil systems' where dependability is considered much more important than capability.

Significantly, it is also noted that in no system domain considered does the consensus view indicate that concept of capability outweighs concept of dependability.

For the purpose of using a conjoint measurement construct for military systems developed under the provisions of ADF technical regulations, the conjoint equation for technical integrity is:

$$TI = \sqrt{C.D}$$

10.2.4 Reliability of using the CAT for technical integrity measurement

A novel application of the CAT was tested on a set of three representative radar systems. Results in section 8.3.1 indicate that indirect estimation of the technical integrity of these systems is reliable ($\alpha = 0.885$) and compares favourably with results obtained in other applications of the CAT. This result suggests that indirect estimation of technical integrity of a complex engineered system via a consensus of judges can be a feasible and pragmatic measurement instrument.

The application of the CAT also involves the judgment of related indicators. In this research, 18 other expected indicators of technical integrity were gathered. The correlation of technical integrity ratings with ratings of these indicators provided some unexpected results in section 8.3.1.3. Most notably, the indicators of safety and environmental impact (both of which are explicitly in the ADF definition of technical integrity) were weakly correlated with technical integrity ($r_s = 0.302$ and 0.245 respectively, significant at the 0.01 level). As was expected, the indicator Fit for

Service, explicit in the ADF definition of technical integrity showed the highest correlation with technical integrity (r_s = 0.656, significant at the 0.01 level).

10.2.5 Reliability of assessing indicators of technical integrity by the CAT

While the judgment reliability of technical integrity was good (α = 0.885), judgments of 12 of the other indicators of technical integrity displayed excellent reliability (α > 0.9) in section 8.3.2.

Indicators that are more subjective in definition, and for which objective data is often not readily available, show reliability results that are less than what is normally acceptable for use. In particular, indicators representing definitional elements of technical integrity, safety and fitness for service exhibit α = 0.685 and α =0.790 in this research. It would be desirable for these to exhibit higher reliability since they are, at best, difficult to estimate from objective data.

10.2.6 Technical integrity scale reliability

The 18 indicators of technical integrity and 2 other indicators, not expected to correlate with technical integrity, were combined into a scale for technical integrity in section 8.3.3. Analysis using the expert ratings indicated that a more reliable estimate of technical integrity may be obtained by use of this scale (α = 0.917) than by indirect estimation of technical integrity alone (α = 0.885). However, the improvement in reliability comes at significantly more effort (i.e., expert judging of 20 indicators vice a single judgment of technical integrity).

10.2.7 Correspondence of indicators with expert judgments

Indicator models were developed for each indicator based upon both objective variables and values and subjective ranking data for the three example systems. Consideration of the correspondence of these models with the expert judgments gained through the CAT in section 8.3.4 provided the following conclusions:

- Where a system quality attribute has clearly identifiable objective variables associated with it, the reliability of judgments also tends to be excellent and the correspondence between the two is complete. The quality attributes of reliability and technical compliance are examples. This is readily explained by understanding that

the judge can readily access, understand and use the same objective values in support of their judgments. The rating of systems on this quality attribute should be accomplished by obtaining the actual values for the systems (since the judges would require this to provide such reliable ratings in any case).

- Excellent judgment reliability is still evident for many system quality attributes even when the results of the indicator model do not completely correspond. The quality attributes of supportability, maintainability, survivability and usability are examples. These quality attributes are well defined and agreed upon terms but the corresponding indicator models generally consist of several objective and subjective variables. The result may be understood as the judge being able to better combine a systems' values into a reliable ranking than an weighted additive model of variables. The rating of systems on these indicators (system quality attributes) should be accomplished by utilising expert judgment.
- Further, excellent judgment reliability is still evident for some system quality attributes even when the results of the indicator model do not correspond. The quality attributes of effectiveness, efficiency and dependability[22] are examples. In these cases there may be none or only weakly related variables that could be used in an indicator model or little difference in the values of the variables with which to differentiate the systems. The rating of systems on this quality attribute should be accomplished by utilising expert judgment.
- Two system quality attributes exhibit poor to unacceptable judgment reliability and no correspondence with the results of the indicator model. The quality attributes of durability and flexibility are in this grouping. These attributes are more subjective than most of the others and are open to relatively wide interpretation explaining the low reliability of judgment ratings. Also there is little in the way of clearly associated variables upon which to base a judgment upon. These system quality attributes may not be able to be reliably estimated.
- The system quality attribute safety had only questionable level of reliability in judges' ratings and there was only moderate correspondence with an indicator model for safety. An initial explanation of this result is that safety, while being an often used term, is not often objectively assessed (at least not is a total system sense)

[22] The indicator, not the concept.

and there is no clear indicator model from readily obtainable objective variables. Safety may not be able to be reliably estimated, but given the oft lack of objective data, it may be that expert judgment be the nominal mechanism for a practical assessment.

10.2.8 Components of technical integrity

The judgment data for the indicators across the three example systems was analysed in section 9.3 using both confirmatory factor analysis and exploratory factor analysis to confirm the theory of technical integrity developed previously. Exploratory factor analysis, using PCA, confirmed that a theory of technical integrity comprising two components was supported. However the two components and the indicators that most significantly loaded upon each of them do not correspond with the concepts of capability and dependability or with the predicted correlation of indicators with these concepts. The interpretation of the result was not straightforward since it did not accord with any of the identified groupings of system quality attributes. The interpreted result required an alternate the theory of technical integrity be postulated and confirmed through factor analysis.

This thesis then contributes a modified theory of technical integrity, grounded in data. The theory derived in section 9.4.1 is that technical integrity combines the concepts of:

- Explicit technical integrity. This concept embodies aspects of a complex engineered system that are usually explicitly required and the level of achievement may be objectively verified. The primary indicators associated with this concept are reliability, technical compliance, effectiveness, usability and dependability.
- Implicit technical integrity. This concept embodies aspects of a complex engineered system that are most often not the subject of explicit technical requirements but are implicitly assumed to be necessary for a critical system. These aspects are often not readily assessable except via an experienced assessor, by logical argument or by actual experience of the system fielded in real, operational conditions. The primary indicators associated with this concept are survivability, maintainability, construction quality and safety.

10.2.9 A measurement instrument for technical integrity

Using both the inductively derived and the amended theories of technical integrity and the available ratings data for the three representative systems, four viable measurement instruments have been identified and assessed in section 9.4.2:

(1) Technical integrity direct measurement using the CAT.
(2) Scale from theorised indicators of technical integrity.
(3) Conjoint construct of capability and dependability concepts.
(4) Conjoint construct of explicit and implicit technical integrity concepts.

The "best" instrument from considerations of cost, reliability and validity is the direct measurement of technical integrity using the CAT.

Despite the results suggesting an alternate theory of technical integrity, the conjoint construct of capability and dependability concepts, was the second "best" measurement instrument and could usurp the best instrument with 20% or less change in the weighting of decision criteria of cost or validity. This measurement instrument has, by virtue of its two components, an ability to further differentiate close ratings of systems by allowing visualisation of both components in two dimensions (see section 0).

Both instruments are available to future research that may use them to further solutions to the overall research problem. Use of the instruments will provide further validation data for the measurement techniques.

10.2.10 Summary of conclusions

An exploratory and descriptive research project has been conducted and has contributed a practical measurement instrument for assessing the technical integrity of complex engineered systems – the 'best" measurement instrument is indicated to be the direct measurement of technical integrity using the CAT (i.e., via expert judgment of design data during system development). The conjoint construct of capability and dependability concepts, was the second "best" measurement instrument with advantages including the ability to further discriminate technical integrity values through the underlying theoretical concepts.

This contribution was underpinned by the development of a theory of technical integrity, a measurement construct and establishing its content and construct validity. A

novel application of the CAT to the estimation of technical integrity and other quality attributes of a complex engineered system provided data to ground and amend the theory of technical integrity, to assess the reliability of expert judgments and to decide if the measurement of technical integrity and its concepts and indicators should be measured via expert judgments or via readily available objective data.

10.3 Conclusion about the research problem

Recall that the prime research problem explored by this thesis was *How effective are the provisions of technical regulations in assuring the technical integrity of complex engineered systems?* While it was not possible to directly assess this without a means for measuring technical integrity, the quest for this measurement instrument becoming the main research question pursued in this thesis, the exploratory and descriptive nature of the research allows some general conclusions to be drawn about the research problem. This section summarises findings and insights derived from the qualitative data that was collected and analysed and summarises the main contributions this research has made toward resolving the research problem.

10.3.1 Findings and insights

Qualitative data collected and analysed in this research supported that the quest for a practical measurement instrument for technical integrity was a valid research objective. However, doubts about the tractability and usefulness of a measurement of technical integrity including practical difficulty and concerns about abstraction leading to loss of information were also evident (refer section 6.4.3). It may be that there is limited application of the measurement instrument beyond that intended by this research or that the validity of application for wider purposes must be further established to overcome these doubts. This accorded with results of the literature survey where alternative views about measuring abstract properties of engineered systems were found (refer section 3.2).

Dichotomous results about the nature of technical integrity of a complex engineered system were also evident (refer section 8.4.1). There was a general concurrence that technical integrity is a subjective aspect of a system and not readily reduced to direct measurement and it requires experience and judgment to estimate. However, when asked to rate the technical integrity of example systems, the respondents indicated a

strong reliance on objective data to reach ratings of system quality attributes including technical integrity. When objective data was not available for a system quality attribute, many respondents indicated an ability to make a subjective rating although they considered this to be a more difficult task, reliant on experience and inferior to when objective data was available.

While the proposal to measure the concept of capability (the desired, intended emergent properties of a system) by the extent of compliance with a system's specified requirements was generally accepted, there exists a body of experienced opinion that the technical requirements of a system do not well match the true capability needs of stakeholders of the system (refer section 6.4.1.1). This may indicate first-hand experiences with the breakdown in the system engineering processes of translating user and other stakeholder needs into valid technical requirements for the system (indeed the failure of the requirements engineering process or the failure to conduct an effective requirements engineering process are often cited as being the cause for systems project failures, e.g. (The Standish Group, 1995) identifies 'Incomplete requirements' and 'Lack of user involvement' as being the top 2 factors of why projects are impaired). Consequently, there is a view that the capability concept may only be assessed after the system has been realized and is trialed or used in realistic environments (refer 6.4.1.5). If this view is accepted, and considering the period of system definition, design, development and transition to service is often measured in the order of a decade for a complex engineered system, assessing the effectiveness of technical regulation in a useful timeframe may not be possible.

The approach to measuring the concept of dependability (undesired, unintended emergent properties of a system) by using some type of overall risk assessment measure was not strongly supported (refer 6.4.2.1). However it was found that there was no clear proposal of an alternative approach. A strong view was expressed that system safety should be emphasised as a explicit element of the technical integrity measurement construct since it was expected that system safety would dominate the dependability concept (refer 7.4.3.3). However the quantitative data from the judgments of the three systems, which indicated that judgments of safety were not strongly correlated with judgments of technical integrity, did not support this view (refer 8.5.4).

Qualitative results supported the quantitative findings of the relative importance of the concepts of capability and dependability over different system domains and provided insights into the derived grouping of system domains where the capability and dependability relative importance differs (refer section 7.5). Dependability starts to dominate capability in the judge's considerations when the consequence of failure of the system is high (e.g. in the nuclear power domain). The comparable importance of the concepts of capability and dependability in the military domains is explained by a conscious trade-off of the inherent risk the system poses to its own personnel (i.e., safety risk) against system capability (the lack of which may also endanger own personnel in the face of a military threat).

An interesting insight was identified that capability shortfalls of a system may be more readily accepted by stakeholder than dependability shortfalls (refer section 7.4.3.2). This is explained in terms that workarounds could be put in place for a capability shortfall (e.g., use the system differently, avoid circumstances in which the capability limitation would become apparent) but cannot be readily put in place for dependability shortfalls (e.g. inherent poor reliability of a system makes it extremely difficult to achieve a reasonable availability at a reasonable cost).

10.3.2 Contribution to the research problem and area

Clear contributions to the research problem and area have been made by this research thesis. Paraphrasing Lord Kelvin (see page 54), technical integrity can be expressed in numbers and in doing so we now 'know something about it' and have advanced the understanding of technical integrity 'to the stage of science'.

Specifically, to the research area of technical regulation and its effectiveness, a practical measurement instrument has been developed ready for use and further validation arising from use. This instrument will allow the assessment of the technical integrity of a complex engineered system, thereby allowing the effectiveness of technical regulations to be assessed.

Of broader application, is the contribution that has shown that use of the CAT can obtain reliable estimates of subjective attributes of a complex engineered system. The CAT may therefore be used as an estimation approach in other research areas which

require a straightforward and reliable method to estimate a latent variable of engineered entities.

10.4 Implications for theory

Many of the concepts for the engineering of complex systems have been derived as a result of experience and reflection upon those experiences. Systems engineering and architecture as disciplines may be considered to be in large part a set of heuristics for the development of systems. It is evident that many extant concepts for explaining the characteristics of complex engineered systems are similarly inferred by this approach.

For example, the work of (Honour, 2007) in characterising the types of emergent properties of a system, built upon in Chapter 5 of this thesis, experientially "makes sense" and in turn is built upon other concepts that also "makes sense" but cite nothing in the way of objective evidence to support the theory or concepts. The results of Chapter 9 have contradicted this theory by providing an alternate model, grounded in data, for grouping the emergent properties of a complex system (the characteristics of a complex system are seen to be aligned with the implicit and explicit aspects of technical integrity). While the data is not wholly objective in terms of directly measuring tangible characteristics of a system, rather it is rating data emerging from a consensus of expert opinion, it does represent a rigorous and repeatable approach that generates results that provide support for the validity of the theory, concepts and measurement constructs.

The generalised implication from this, is that development of systems theory, may seek more data from similarly conducted research to provide support or otherwise for systems concepts that today are generally accepted on the basis of inference. This is in contrast to some influential views about system characteristics that 'many of the most important factors are not measurable' (Maier and Rechtin, 2009, p. 2)

The approach of using the CAT, developed for and used mostly for "soft" measurands like creativity embodied in works of art and literature, was novel in application for what would be considered by most to be a "hard" technical measurand of an engineered system. The resulting measurement reliability obtained for estimation of technical integrity, a latent variable not directly measurable, indicates that further application of the CAT in the engineering and management of engineered systems may be warranted.

It also infers that it would be reasonable to consider a broader usage of the CAT in other domains where attributes of a subject may not be directly measurable.

10.5 Implications for policy and practice

10.5.1 Practical implications and usage of the findings

The measurement instrument developed in this thesis provides a practical means to obtain reliable estimates of technical integrity of a complex engineered system during its development by using readily available objective data and system design descriptions and by eliciting the judgment of reasonably well experienced practitioners. This points to the measure being able to be utilised for Technical Performance Measurement (TPM) thereby providing a leading indicator during development for the likelihood of the requisite technical integrity being achieved and may highlight other problems that may impact the in-service system.

10.5.2 Implications for policies

Recall that the ADF defines technical integrity as an item's fitness for service, safety and compliance with regulations for environmental protection. For the representative systems, the subject of this thesis in Chapters 8 and 9, it was found that the rating of the indicators 'Impact to Environment' and 'Safety' did not correlate strongly with the judge's ratings of technical integrity (refer section 8.3.1.3). The judgments of 'Fitness for Service' did correlate strongly with the judge's ratings of technical integrity. Given the majority of judges involved in generating these ratings are well familiar with the ADF technical regulations this was a surprising result and must have some implication for the ADFs technical regulatory policies.

A possible implication for consideration is whether ADF technical regulation is effectively communicating its objective of technical integrity? Or, are the defined elements really better addressed as disparate concepts that may require different regulatory approaches and technical provisions?

Technical integrity cannot be defined as absolutely present or not present. A complex engineered system will always engender some level of safety risk, and there will be shortfalls in capability or fitness for service. Consequently, the technical regulations that seek to assure technical integrity should provide a means for estimating technical

integrity both as an objective basis for justifying the provisions of the technical regulation itself, but also to determine if a requisite level of technical integrity is likely to be or has been achieved for a system prior to its introduction into service.

It was concluded in this thesis that in no system application domain considered does the consensus view indicate that the concept of capability (intended emergent features) outweighs dependability (unintended, undesired emergent features). However, what is seen in practice is significant effort provisioned for and expended in defining and then assuring that the technical requirements of a system development are achieved but provisions for the prevention of unintended and undesired design outcomes are of less a priority in what is expressly required for the acquisition and development of complex engineered systems for the ADF. For instance, often specific, measureable technical goals for system safety are not well established, rather the provision of regulations and of contracts is for a safety program to be executed rather than for particular and measurable safety goals to be achieved. The implication is that development of ADF technical regulations should include provisions to assure that system quality attributes that are related to the dependability concept and to implicit technical integrity are defined and achieved.

10.6 Limitations

This section summarises the main limitations associated with the aforementioned conclusions, detailed limitations having been identified at the conclusion of each of Chapters 6 through 9 inclusive. These limitations became apparent during the course of the research. Delimitations to the scope of the research were defined in Chapter 1.

10.6.1 Cautions to use of findings

The following limitations are identified as cautions to the use of the findings of this thesis and as a seed for future research:

- It is not considered valid to measure *technical integrity* on a ratio scale. This may limit future research goals that will use the measurement instruments (e.g., it may never be able to produce results that indicate x% more spent on technical regulation has increased *technical integrity* of systems by y%).
- It is not considered valid to use the conjoint measure of technical integrity to compare systems across different domains (but this may be accounted for by using

different values in the conjoint equation, these values would need to be determined for other than military systems).

- Other applications of measurement of technical integrity may not be valid (e.g., for the purpose of judging the acceptability of a system). This will not impact continued effort on this research problem, but may detract from the potential value of the instrument for adjacent research problems.
- Doubts and scepticism as to the validity of the system's written requirements as a basis for measurement have been expressed in the survey results. These views may limit the acceptance of percentage compliance in the conjoint measurement instrument by some stakeholders.
- The survey research undertaken during this project consumed hundreds of hours from experienced professionals. This has provided a constraint on obtaining additional data that could further support the findings and this will likely be a practical limitation for obtaining more validation data particularly for larger and more complex systems.
- While the conclusions drawn have been grounded in data and are significant for the example systems used in the experiment, it requires further research to extrapolate the findings to systems that are different in application, domain of use and size, scope and complexity.
- Both the inductively derived theory of technical integrity (inclusive of the concepts of capability and dependability) and the revised theory, grounded in ratings data from the three example systems, confirm technical integrity exhibits two factors. However, the "best" approach to the measurement of technical integrity does not provide any estimate of these factors. This may limit the interpretation of measurements undertaken in future applications. In particular, some provisions of technical regulation could conceivably affect only one of the identified factors, but this would not be readily identified. This consideration may favour using the conjoint measurement construct.

10.6.2 Validity of assumptions

Key assumptions made to delimit the scope and allow the pragmatic execution of the research were identified in Section 1.7.2. The validity of those assumptions post the research and the resulting limitations are considered in this section.

- The first assumption made was that there was indeed a common underlying property, a latent variable, of a complex engineered system which can be assessed as technical integrity. This assumption remains valid and the research has determined valid measurement instruments for technical integrity.
- It was assumed there would be some convergence of opinion between the diverse stakeholders with respect to differentiating a system with higher technical integrity from one with lower technical integrity. While this has been shown to have held true in the two surveys reported herein, it is a limitation of the research that neither survey was repeated at a different time which does not allow us to know if the opinions of one or more stakeholder groups could change with the passage of time.
- The last key assumption was that the imposition of technical regulation has a measurable effect on technical integrity and that this effect can be discerned from other drivers of technical integrity. No findings of this research have either raised or lowered the confidence in this assumption.

10.6.3 Threats to validity of findings

This section addresses threats to the validity of the reported findings in general and what has been done in the research design and analysis to mitigate these threats. 'All research strategies and methods are seriously flawed' and the strength associated with a choice of one method comes with research dilemmas that must be recognized. While it is desirable to simultaneously maximise generalisability, precision and realism it is never possible and the resulting compromise must be recognized as a threat to the validity of the research findings (McGrath, 1981).

The survey research method adopted has achieved good generalisability across the population due to the number and range of respondents. However, the precision and realism have limited the validity of extrapolating the results to real systems and real engineering assessments. These threats were mitigated by using example systems which were realistic, if not actually real, and a level of fidelity that is typical of what might be encountered in a real scenario of forming engineering judgments. The judgments were made is a typical engineering workspace, i.e. on an office computer. Secondly, both quantitative and qualitative data were gathered in each of the surveys and analysed for convergence in order to compensate for vulnerabilities in either method.

The scope of the second survey was limited to one study of realistic maritime radar systems. However, the systems are realistic, albeit not real. A move toward the use of actual systems would have run into problems with access to proprietary data and would have heavily constrained the survey to a small number of respondents who were willing and able to enter Non-Disclosure Agreements with participating radar manufacturers. It is indeed the case that the respondents may not have had sufficient experience with the example systems, or the type of system more broadly, but this is regularly the case in professional practice where these type of judgments are routinely required with no more data than is provided in the product descriptions used in this survey. This further supports the appropriateness of using the CAT for this type of problem. While this study yielded useful findings and insights the validity of these is threatened if they were to be generalised too broadly.

The finding that the relative importance of capability and dependability differ across different classes or domains of systems emerged directly from the survey data. While the finding is statistically supported, the questioning was very high level and broad. The use of both quantitative analysis and qualitative analysis results is an attempt to improve the generalisability of the result. In this case, both quantitative and qualitative results corresponded giving support to the notion that this may be a general result, but of course these results come from a limited sample of judges, not all expert in all the domains, and so there would be a remaining threat to the validity of a broad general claim.

Interpretation of respondents who did not provide an alternative measure for dependability as really supporting the proposed measure is threatened by other credible interpretations of the qualitative data (e.g., they may not have had enough time to come up with an alternative measurement approach).

10.7 Implications for further research

The main conclusions of this thesis have been drawn from one set of example systems: that being, three, representative, naval/maritime navigation radars. Constraining the study to these systems was a necessary delimitation in order to reasonably limit the amount of time respondents would need to voluntarily contribute and thereby ensure a sufficient number of responses were received for significance. Relieving this constraint and repeating the experiment for different types of systems would allow further

validation of findings for technical integrity measurement construct over a wider range of complex engineered systems.

Assuming the approach for measuring technical integrity continues to be valid, it may be used for quantitatively seeking resolution of the research problem, i.e., researching the effectiveness of technical regulations in assuring the technical integrity of complex engineered systems. The research is implied to require a post-facto longitudinal study over the period where ADF technical regulations were being introduced and matured. A longitudinal study of effectiveness of changes to technical regulations is also an area for future research.

Arriving at a means to reliably measure system characteristics which are described in terms of risks has not been adequately resolved in this thesis. The primary example is the quality attribute of safety. The literature survey revealed that no widely accepted method for a generalised measurement of a system's overall safety has been established nor are there any methods to combine a set of estimated safety risks (be they quantitatively or qualitatively assessed risks). Estimation of safety via the CAT was not demonstrated to be particularly reliable in this research. Further investigation of using the CAT on different example systems, perhaps with more differentiated safety variables and values in the system descriptions is suggested. Other approaches to reliably measure system safety need to be identified and tested noting that the objective is that the approach needs to be practical and cost-effective, precluding specific and intensive analytic methods like PRA.

Contemporary research looking at the effectiveness of engineering processes for the development of complex engineered systems (e.g., the work of (Honour, 2010, Elm, 2008)) provide results primarily in the cost and schedule dimensions of project performance. The effectiveness of engineering processes in terms of outcomes for the technical characteristics (e.g., integrity, quality, dependability, capability) of complex engineered systems has not been so clearly determined. The measurement approach established for technical integrity should prove useful for extending the research into the effectiveness of systems engineering on technical characteristics of systems.

References

ADF 2002. Defence Instruction (General) LOG 08-15. *Regulation of Technical Integrity of Australian Defence Force Materiel.* Canberra, ACT, Australia: Department of Defence.

ALLEN, M. J. & YEN, W. M. 1979. *Introduction to measurement theory*, Brooks/Cole Monterey, CA.

AMABILE, T. M. 1996. Consensual Assessment Technique. *Creativity in Context.* Boulder, CO, USA: Westview Press.

ARMY 2003. Defence Instructions (Army) LOG 1/2003. *Regulation of the Technical Integrity of Land Materiel.* AMNDT 1 ed. Canberra, ACT: Department of Defence, Army Headquarters.

ASHBY, W. R. 1955. *An introduction to cybernetics*, Taylor & Francis.

AUSTRALIAN BUREAU OF STATISTICS 2011. 6227.0 - Education and Work, Australia, May 2011 Australian Bureau of Statistics.

AUSTRALIAN DEFENCE FORCE 2001. Australian Air Publication 7001.048(AM1) ADF Airworthiness Manual. Airworthiness Coordination and Policy Agency, Commonwealth of Australia.

AUSTRALIAN DEFENCE FORCE 2004a. Australian Air Publication 7001.054 Airworthiness Design Requirements Manual. Director General Technical Airworthiness, Commonwealth of Australia.

AUSTRALIAN DEFENCE FORCE 2004b. Defence Instruction (General) LOG 08-15 (AM 1). *Regulation of Technical Integrity of Australian Defence Force Materiel.* Canberra, ACT, Australia: Department of Defence.

AUSTRALIAN DEFENCE FORCE 2005. Australian Air Publication 7001.053(AM1) Technical Airworthiness Management Manual. Director General Technical Airworthiness, Commonwealth of Australia.

AUSTRALIAN DEFENCE FORCE 2010. Defence Instruction (General) LOG 4-5-012 Regulation of Technical Integrity of Australian Defence Force Materiel. Canberra, ACT, Australia: Department of Defence.

BAHR, N. J. 1997. System Safety Engineering and Risk Assessment: A Practical Approach, Philadelphia, PA, Taylor & Francis.

BALDWIN, R. & CAVE, M. 1999. *Understanding Regulation. Theory, Strategy and Practice,* Oxford, UK, Oxford University Press.

BALE, E. A. & EDWARDS, D. W. 2000. Technical Integrity - An Engineer's View. *Trans IChemE,* Vol 78, Part B**,** 355-361.

BARBACCI, M., KLEIN, M. H., LONGSTAFF, T. A. & WEINSTOCK, C. B. 1995. Quality Attributes. Pittsburgh, Pennsylvania: Software Engineering Institute, Carnegie Mellon University.

BASS, L., CLEMENTS, P. & KAZMAN, R. 2003. *Software Architecture in Practice,* Boston, Addison-Weley.

BLANCHARD, B. 2008. *System engineering management,* John Wiley & sons.

BOEHM, B. 1978. *Characteristics of Software Quality,* New York, American Elsevier.

BOEHM, B., VALERDI, R. & HONOUR, E. 2008. The ROI of Systems Engineering: Some Quantitative Results for Software-Intensive Systems. *Systems Engineering,* 11**,** 14.

BROOKE, J. 1996. SUS - A Quick and Dirty Usability Scale. *In:* JORDAN, P. W., THOMAS, B., WEERDMEESTER, B. A. & MCCLELLAND, I. L. (eds.) *Usability Evaluation in Industry.* London.: Taylor & Francis.

BROWN, J. D. 2009. Choosing the Right Type of Rotation in PCA and EFA. *SHIKEN - JALT Testing & Evaluation SIG Newsletter,* 13**,** 20-25.

BROWNING, T. R., DEYST, J. J. & EPPINGER, S. D. 2002. Adding Value in Product Development by Creating Information and Reducing Risk. *IEEE Transactions on Engineering Management,* Vol. 49.

BROWNING, T. R. & HONOUR, E. C. 2008. Measuring the Life-Cycle Value of Enduring Systems. *Systems Engineering,* Vol. 11, No. 3**,** 16.

CARIFIO, J. & PERLA, R. J. 2007. Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *J. Soc. Sci.,* 3**,** 106-116.

CENTER FOR CHEMICAL PROCESS SAFETY 2008. Risk-Based Determination of the Adequacy of Safeguards. *Guidelines for Hazard Evaluation Procedures.* 3rd Edition ed.: John Wiley and Sons.

CLEMEN, R. T. 1996. Making Hard Decisions. An Introduction to Decision Analysis., Belmont, Ca., Duxbury Press.

CONANT, R. C. & ASHBY, W. R. 1970. Every good regulator of a system must be a model of that system. *International Journal of Systems Science* 1**,** 8.

COOK, R. I. 2000. How Complex Systems Fail. Cognitive Technologies Laboratory, University of Chicago.

CRESWELL, J. 2003. Research Design Qualitative, Quantitative, and Mixed Methods Approaches. *Handbook of mixed methods in social & behavioral research*.

CROPLEY, A. 2002. Qualitative Research Methods. An Introduction for Students of Psychology and Education, Riga, Zinatne Publishers.

CROPLEY, D. H. & HARRIS, M. B. 2004. Too hard, too soft, just right… Goldilocks and three research paradigms in SE. *Systems Engineering, managing complexity and change–14th Annual International Symposium Proceedings.*

DEFENCE MATERIEL ORGANISATION 2007. DMO Acquisition and Sustainment Manual. Canberra. ACT: Department of Defence, Australian Government.

DEPARTMENT OF DEFENCE 2006. Navy Hazard Risk Calculator. *ABR 6303 NAVSAFE Manual - Naval Safety Management.*

DEPARTMENT OF DEFENCE 2007. Defence Ordnance Publication 102 (AM1) Technical Integrity of Explosive Ordnance. Commonwealth of Australia.

DEPARTMENT OF DEFENCE 2009. Technical Regulation ADF Materiel Manual-Land (TRAMM-L). Iss 4.0 ed. Melbourne, Vic: Director of Technical Regulation - Army, Commonwealth of Australia.

DEPARTMENT OF DEFENCE 2010. *The Strategic Reform Program - Making It Happen,* Canberra, ACT, Commonwealth of Australia.

DEPARTMENT OF DEFENCE 2011. Defence Instructions (General) 4–1–006. *Safety of Explosive Ordnance.* AMDNT NO 2 ed. Canberra ACT: Department of Defence (CJLOG).

DEPARTMENT OF DEFENSE 1993. MIL-STD-882C System Safety Program Requirements. Department of Defense, USA.

DEPARTMENT OF DEFENSE 2000. MIL-STD-882D. *Standard Practice for System Safety.* Department of Defense, USA.

DEPARTMENT OF DEFENSE 2008. Systems Engineering Guide for Systems of Systems. Version 1.0 ed. Washington, DC: Office of the Deputy Under Secretary of Defense for Acquisition and Technology.

DEPARTMENT OF ENERGY 2006. DOE O 413.3A Program and Project Management for the Acquisition of Capital Assets. Washington, D.C.: US Department of Energy.

DEPARTMENT OF ENERGY 2008. DOE G 413.3-1 Managing Design and Construction Using Systems Engineering. Washington, D.C.: US Department of Energy.

DEROSA, J. K., GRISOGONO, A.-M., RYAN, A. J. & NORMAN, D. 2008. A Research Agenda for the Engineering of Complex Systems. *IEEE International Systems Conference.* Montreal, Canada,: IEEE.

DEVELLIS, R. F. 2011. *Scale development: Theory and applications*, Sage Publications, Incorporated.

DOBES, L. 2009. A practical guide to cost-benefit analysis. A practical guide: evidence for policy and decision-making. Sydney, University of New South Wales Press Ltd.

EDWARDS, M. 2010. An Ontology of Technical Regulation. *Systems Engineering Test and Evaluation 2010.* Adelaide, SA, Australia.

EIA 1998a. EIA-632 Processes for engineering a system. Electronic Industries Alliance.

EIA 1998b. EIA-649 National Consensus Standard for Configuration Management. Electronics Industries Alliance.

ELM, J. P. 2008. A Study of Systems Engineering Effectiveness - Initial Results. *Systems Conference, 2008 2nd Annual IEEE.* Montreal, Canada: IEEE.

ELM, J. P., GOLDENSON, D. R., EMAMM, K. E., DONATELLI, N., NEISA, A. & NDIA SE EFFECTIVENESS COMMITTEE 2007. A Survey of Systems Engineering Effectiveness - Initial Results. Software Engineering Institute, Carnegie Mellon University and the National Defense Industrial Association.

FAULKNER, J. 2009. *Minister of Defence's Forward to "Defence Capability Plan 2009"* [Online]. Available: http://www.defence.gov.au/dmo/id/dcp/html/foreword/index.html [Accessed 19 Nov 2011].

FENTON, N., LITTLEWOOD, B., NEIL, M., STRIGINI, L., SUTCLIFFE, A. & WRIGHT, D. 1998. Assessing dependability of safety critical systems using diverse evidence. *IEE Proceedings - Software,* Vol. 145, No. I.**,** 5.

FINKELSTEIN, L. 2003. Widely, strongly and weakly defined measurement. *Measurement,* 34**,** 39-48.

FIRESMITH, D. G., CAPELL, P., FALKENTHAL, D., HAMMONS, C. B., LATIMER, D. & MERENDINO, T. 2009a. Architecture and Quality. *The Method Framework for Engineering System Architectures.* Boca Raton, FL: CRC Press, Taylor & Francis Group, LLC.

FIRESMITH, D. G., CAPELL, P., FALKENTHAL, D., HAMMONS, C. B., LATIMER, D. & MERENDINO, T. 2009b. Architecture and Quality. *The Method Framework for Engineering System Architectures.* Boca Raton, FL: CRC Press, Taylor and Francis Group.

FIRESMITH, D. G., CAPELL, P., FALKENTHAL, D., HAMMONS, C. B., LATIMER, D. & MERENDINO, T. 2009c. *The Method Framework for Engineering System Architectures,* Boca Raton, FL, CRC Press, Taylor and Francis Group.

GARVEY, P. R. 2009. Analytical Methods for Risk Management. A Systems Engineering Perspective., Boca Raton, FL, CRC Press, Taylor & Francis Group.

GEORGE, D. & MALLERY, P. 2003. SPSS for Windows step by step: A simple guide and reference., Boston, Allyn & Bacon.

GILB, T. 2005. Scales of Measure. How to Quantify. Competitive Engineering. A Handbook for Systems Engineering, Requirements Engineering and Software Engineering using Planguage. Butterworth-Heinemann.

GLINZ, M. Year. On non-functional requirements. *In:* Requirements Engineering Conference, 2007. RE'07. 15th IEEE International, 2007. IEEE, 21-26.

GREEN, P. E. & RAO, V. R. 1971. Conjoint Measurement for Quantifying Judgmental Data. *Journal of Marketing Research,* Vol. VIII**,** 355-363.

HACKETT, M. 2008. *Technical Integrity and Process Safety* [Online]. National Offshore Petroleum Safety Authority. Available: http://www.nopsa.gov.au/presentation/ [Accessed 20 October 2010].

HARRIS, D. 2004. Is Aviation Safety a Self-Organising Critical System. *SETE 2004 Focussing on Project Success.* Adelaide, SA, Australia: Systems Engineering Society of Australia.

HITCHINS, D. K. 2003. *Advanced Systems Thinking, Engineering and Management,* Norwood, MA, USA, Artech House.

HITCHINS, D. K. 2007. *Systems Engineering, A 21st Century Methodology,* Chichester. UK, John Wiley & Sons, Ltd.

HONOUR, E. 2007. Connecting an Architecture with its Emergent Properties. *Conference on Systems Engineering Research.* Hoboken, NJ , USA.

HONOUR, E. 2010. Systems Engineering Return on Investment. *INCOSE International Symposium 2010.* Chicago, USA.

HUBBARD, D. W. 2007. How To Measure Anything. Finding the Value of Intangibles in Business., Hoboken, NJ, John Wiley and Sons.

HUBBARD, D. W. 2009. The failure of risk management: Why it's broken and how to fix it, Wiley.

HUMPHREYS, P. 1977. Application of multi-attribute utility theory. *Decision Making and Change in Human Affairs***,** 165-207.

HUNTER, R. D. 2009. *Standards, Conformity Assessment, and Accreditation For Engineers,* Boca Raton, FL, USA, CRC Press, Taylor and Francis Group.

HUREVICH, V. L. & SERENKOV, P. S. 2009. System of Technical Regulation and Standardization as Object for Risk Management. *International Conference on Risk Assessment and Management.* Geneva.

IEEE 1990. IEEE Standard Glossary of Software Engineering Terminology. New York: Institute of Electrical and Electronic Engineers.

INCOSE 2010. Systems Engineering Handbook. A Guide for System Life Cycle Processes and Activities. v 3.2 ed.

INTERNATIONAL MARITIME ORGANIZATION 2007. Formal Safety Assessent. *In:* COMMITTEE, M. S. (ed.). IMO.

IRVING, I. 2008. Emergent Challenges Facing Systems Engineering in the Defence Industry in Australia. *Systems Engineering Society of Australia Newsletter.* Australia.

ISO 2007. ISO/IEC 26702 IEEE Std 1220-2005. *Systems Engineering - Application and Management of the Systems Engineering Process.* First edition 2007-07-15 ed.: ISO/IEC and IEEE.

ISO 2008. ISO/IEC 15288:2008 Systems and software engineering—System life cycle processes. Geneva: International Organization for Standardization.

ISO/IEC 9126 1991. Information technology -- Software product evaluation - Quality characteristics and guidelines for their use. Geneva.

IVERSON, G. & LUCE, R. D. 1998. The representational measurement approach to psychophysical and judgmental problems. *Measurement, judgment, and decision making***,** 1-79.

JOHNSON, C. W. 2006. What are emergent properties and how do they affect the engineering of complex systems? *Reliability Engineering and System Safety,* 91**,** 1475-1481.

KAPLAN, A. 1964. The conduct of inquiry: Methodology for behavioral science.

KASSER, J. 2002. The acquisition of a System of Systems is just a simple multi-phased parallel-processing paradigm. *Engineering Management Conference, 2002. IEMC '02. 2002 IEEE International.*

KASSER, J. 2009. "System" or "System of Systems": Much Ado about Nothing? *INCOSE Insight.* USA.

KEENEY, R. L. & RAIFFA, H. 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. New York: John Wiley & Sons.

KING, L. W. 2008. *The Code of Hammurabi, Translated by L. W. King* [Online]. New Haven, CT: Lillian Goldman Law Library. Available: http://avalon.law.yale.edu/ancient/hamframe.asp [Accessed 2012].

KONTOVAS, C. A. & PSARAFTIS, H. N. 2009. Formal Safety Assessment: A Critical Review. *Marine Technology,* 46**,** 15.

KOSSIAKOFF, A., SWEET, W. N., SEYMOUR, S. J. & BIEMER, S. M. 2011a. Structure of Complex Systems. *Systems Engineering Principles and Practice.* 2nd ed.: John Wiley & Sons.

KOSSIAKOFF, A., SWEET, W. N., SEYMOUR, S. J. & BIEMER, S. M. 2011b. *Systems Engineering Principles and Practice,* John Wiley & Sons.

KRITZINGER, D. 2006. *Aircraft system safety. Military and civil aeronautical applications.,* Cambridge, England, CRC Press, Woodhead Publishing Limited.

LAMACRAFT, R., WALSH, E. G., SCHEDLICH, R. B., FILOR, C. W. & CUNEO, L. 1998. Report of the Board of Inquiry into the fire in HMAS Westralia on 5 May 1998. Commonwealth of Australia.

LAWSHE, C. H. 1975. A Quantitative Approach to Content Validity. *Personnel Psychology,* 28**,** 563-575.

LEE, J. C. & MCCORMICK, N. J. 2012a. Risk-informed regulations and reliability-centered maintenance. *Risk and safety analysis of nuclear systems.* Wiley.

LEE, J. C. & MCCORMICK, N. J. 2012b. Risk and Safety of Enigneered Systems. *Risk and safety analysis of nuclear systems.* Wiley.

LEEDY, P. D. & ORMOND, J. E. 2005. *Practical Research. Planning and Design,* Upper Saddle River, NJ, Pearson Education International.

LEVESON, N. G. 2011. Systems Theory and its Relationship to Safety. *Engineering a Safer World. Systems Thinking Applied to Safety.* Cambridge, MA: MIT Press.

LLOYD'S REGISTER OF SHIPPING Lloyd's register of shipping : rules and regulations for the construction and classification of steel vessels. *Lloyd's register of shipping : rules and regulations for the construction and classification of steel vessels*.

LUCE, R. D. & TUKEY, J. W. 1964. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology,* 1**,** 1-27.

MAGEE, C. L. & DE WECK, O. L. 2002. An Attempt at Complex System Classification. *ESD Internal Symposium.* Cambridge, MA: Massachusetts Institute of Technology, Engineering Systems Division.

MAIER, M. W. & RECHTIN, E. 2009. *The Art of Systems Architecting,* Boca Raton, FL, CRC Press, Taylor & Francis Group.

MARRISON, C. 2002. Operating Risk. *Fundamentals of Risk Measurement.* New York: McGraw Hill.

MCCALL, J. A. & MATSUMOTO, M. T. 1980. Software Quality Measurement Manual. General Electric Company.

MCDERMID, J. A. 2001. Software Safety: Where's the Evidence? *6th Australian Workshop on Industrial Experience with Safety Critical Systems and Software.* Brisbane, Australia: Australian Computer Society, Inc.

MCGRATH, J. E. 1981. Dilemmatics: The Study of Research Choices and Dilemmas. *American Behavioral Scientist,* 25**,** 179-210.

MEYER, M. A. & BOOKER, J. M. 2001. *Eliciting and Analyzing Expert Judgment,* ASA SIAM.

MICHELL, J. 2005. The logic of measurement: A realist overview. *Measurement,* 38**,** 285-294.

MILLER, G. 2010. Engineering Complex Systems using Off-The-Shelf Products. *Systems Engineering Test & Evaluation.* Adelaide, SA.

MODARRES, M. 2006. Risk Analysis in Engineering. Techniques, Tools and Trends., Boca Raton, FL, Taylor & Francis.

MOSS, D. S. 2008. *Confirmatory factor analysis* [Online]. Melbourne, Australia: psych-it.com.au. Available: http://www.psych-it.com.au/Psychlopedia/article.asp?id=163 [Accessed 25 November 2012].

MURAVYOV, S. & SAVOLAINEN, V. 2001. Special interpretation of formal measurement scales for the case of multiple heterogeneous properties. *Measurement,* 29**,** 209-223.

NARDO, M., SAISANA, M., SALTELLI, A., TARANTOLA, S., HOFFMAN, A. & GIOVANNINI, E. 2005. Handbook on Constructing Composite Indicators: Methodology and User Guide. *OECD Statistics Working Papers.*

NASA 2007. Systems Engineering Handbook. Rev 1 ed.: National Aeronautics and Space Administration.

NAVSEA 2005. NAVSEA Instruction 9410.2 Naval Warfare Systems Certification Policy. Washington Navy Yard, DC: Department of the Navy, Naval Sea Systems Command.

PALLANT, J. 2010. SPSS survival manual: A step by step guide to data analysis using SPSS, Open University Press.

PARNELL, G. S., DRISCOLL, P. J. & HENDERSON, D. L. 2008. *Decision Making in Systems Engineering and Management,* Hoboken, New Jersey, John Wiley &Sons Inc.

PERROW, C. 1984. Normal accidents: Living with high risk systems. New York: Basic Books.

PERROW, C. B. 2008. Complexity, catastrophe, and modularity. *Sociological Inquiry,* 78**,** 162-173.

PI ASSET INTEGRITY. 2007. *Technical Integrity Audits* [Online]. PI Asset Integrity Ltd. Available: http://www.piai-ltd.com/Technical_Audits.htm [Accessed 20 October 2010].

PRASAD, D. K. 1998. Dependable Systems Integration using Measurement Theory and Decision Analysis. Doctor of Philospohy, University of York.

RAYTHEON AUSTRALIA. 2012. Depth of Expertise. Systems Engineering Capability to Support Australia's Future Naval Projects. Available: http://www.raytheon.com.au/rtnwcm/groups/public/documents/content/rau_depth_of_expertise_pdf.pdf.

REASON, J. 2000. Human Error: Models and Management. *British Medical Journal,* 320**,** 768-770.

RIZZO, P. J. 2011. Plan to reform Support Ship Repair and Management. Canberra, ACT: Defence, Commonwealth of Australia.

ROBERTS, F. S. 1984. Measurement theory with applications to decision making, utility, and the social sciences. *Encyclopedia of Mathematics and its Applications.* Cambridge Cambridgeshire and New York, NY, USA: Cambridge University Press.

ROSKIES, R. 1965. A measurement axiomatization for an essentially multiplicative representation of two factors. *Journal of Mathematical Psychology,* 2**,** 266-276.

ROYAL ACADEMY OF ENGINEERING 2007. Creating systems that work: Principles of engineering systems for the 21st century. *In:* ELLIOTT, D. C. & DEASLEY, P. P. (eds.) *Working Party on Integrated System Design.* London: The Royal Academy of Engineering.

ROYAL AUSTRALIAN NAVY 2003a. ABR 6492 Navy Technical Regulations Manual (NTRM). Commonwealth of Australia.

ROYAL AUSTRALIAN NAVY 2003b. Defence Instructions (Navy) LOG 47-3. *Regulation of Technical Integrity of Australian Defence Force Maritime Materiel.* AMDNT NO 1 ed. Canberra ACT: Department of Defence (Navy Headquarters).

ROYAL AUSTRALIAN NAVY 2007. Nias Island Sea King Accident Board of Inquiry Report. Royal Australian Navy.

ROYAL AUSTRALIAN NAVY 2009. Defence Instructions (Navy) LOG 47-3. *Regulation of Technical Integrity of Australian Defence Force Maritime Materiel.* AMDNT NO 2 ed. Canberra ACT: Department of Defence (Navy Headquarters).

SAJI, G. 2003. Safety Goals in 'risk-informed, performance-based' regulation. *Reliability Engineering and System Safety,* 80**,** 163-172.

SCHMIDT, N. 2012. Military Airworthiness. *Aviation Business Asia Pacific.* May/June 2012 ed.

SEKARAN, U. 2003. Research methods for business: A skill building approach, New York, John Wiley & Sons.

SHEARD, S. A. & MOSTASHARI, A. 2008. Principles of Complex Systems for Systems Engineering. *Systems Engineering***,** 17.

SKULMOSKI, G. J., HARTMAN, F. T. & KRAHN, J. 2007. The Delphi method for graduate research. *Journal of information technology education,* 6**,** 1.

STANDARDS AUSTRALIA 2004. Risk Management Guidelines. Companion to AS/NZS 4360:2004. Sydney NSW: Standards Australia international Ltd.

STEPHENSON, J. 1991. SYSTEM SAFETY 2000. A Practical guide for Planning, Managing, and Conducting System Safety Programs, New York, John Wiley & Sons, Inc.

SUPPES, P. 2009. Measurement Theory and Engineering In: MEIJERS, A. (ed.) Handbook of the Philosophy of Science, Volume 9: Philosophy of Technology and Engineering Sciences. Elsevier BV.

SYDENHAM, P. H. 1982. *Handbook of Measurement Science,* Chichester, UK, John Wiley and Sons.

TASHAKKORI, A. & TEDDLIE, C. 1998. *Mixed methodology: Combining qualitative and quantitative approaches*, Sage Publications, Incorporated.

TEDDLIE, C. & TASHAKKORI, A. 2009. Foundations of mixed methods research. Integrating quantitative and qualitative approaches in the social and behavioral sciences, Los Angeles, Sage.

THE STANDISH GROUP 1995. The Standish Group Report: Chaos.

THOMSON, W. 1883. Lord Kelvin's Lecture on "Electrical Units of Measurement" *Popular Lectures* Vol 1.

TRAINOR, T. & PARNELL, G. Year. Using stakeholder analysis to define the problem in systems engineering. *In:* Proc 17 th Annual International Symposium INCOSE, San Diego. Utne, IB (2006), Systems engineering principles in fisheries management. In Marine Policy, 2007. 624-634.

TVERSKY, A. 1967. A general theory of polynomial conjoint measurement. *Journal of Mathematical Psychology,* 4**,** 1-20.

UCLA: ACADEMIC TECHNOLOGY SERVICES, S. C. G. *Statistical analyses using SPSS* [Online]. Available: http://www.ats.ucla.edu/stat/spss/whatstat/whatstat.htm [Accessed 5 November 2012 2012].

USAF 1969. MIL-STD-499A. *System Engineering Management.* Department of Defense, USA.

WALLENIUS, J., DYER, J., FISHBURN, J. S., STEUR, P. C., ZIONTS, R. E. & DEB, K. 2008. Multiple criteria decision making, multiattribute utility theory. Recent accomplishments and what lies ahead. *Management Science,* 54**,** 1336-1349.

WALLIMAN, N. 2010. *Research methods: The basics,* London, Taylor & Francis.

WASSON, C. S. 2006. System Analysis, Design, and Development Concepts, Principles, and Practices: A John Wiley & Sons. *Inc, Hoboken, New Jersey*.

WIKIPEDIA. 2010. *List of system quality attributes* [Online]. Available: http://en.wikipedia.org/wiki/List_of_system_quality_attributes [Accessed 27Jan2010].

WILSON, F. R., PAN, W. & SCHUMSKY, D. A. 2012. Recalculation of the Critical Values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development,* 45**,** 197-210.

WUENSCH, K. L. 2007. *Inter-Rater Agreement* [Online]. East Carolina University. Available: core.ecu.edu/psyc/wuenschk/docs30/InterRater.doc [Accessed 12 October 2012 2012].

Bibliography

The literature in this section is not directly referenced within this thesis but contributes to a broader understanding of the research problem and context.

AFIFI, A., CLARK, V. A. & MAY, S. 2004. Computer-Aided Multivariate Analysis, Boca Raton, Chapman & Hall / CRC.

ANON. 2009. Theory of conjoint measurement [Online]. Wikipedia. Available: http://en.wikipedia.org/wiki/Theory_of_conjoint_measurement [Accessed 9 March 2009 2009].

ARMY 2004. Technical Regulation of Army Materiel Manual (TRAMM). Issue 3.1 ed. Melbourne, Victoria: Directorate of Technical Regulation - Army.

ARVIDSSON, F. & FLYCHT-ERIKSSON, A. 2002. Ontologies I. Homepage of PhD Course: Semantic Web, Linkoping University, Sweden.

ASLAKSEN, E. 2009. Designing Complex Systems: Foundations of design in the functional domain, Auerbach Pub.

AUSTRALIAN COMMUNICATIONS AND MEDIA AUTHORITY. 2008. Evidence for Communications and Media Regulation [Online]. Available: http://www.acma.gov.au/webwr/_assets/main/lib310671/atug_regional_conf_2008.pdf [Accessed 25 January 2009].

AVEN, T. 2008. A semi-quantitative approach to risk analysis, as an alternative to QRAs. Reliability engineering & System safety, 93, 790-797.

AVEN, T. 2010. On how to define, understand and describe risk. Reliability engineering & System safety, 95, 623-631.

BARBACCI, M. R., ELLISON, R., LATTANZE, A. J., STAFFORD, J. A., WEINSTOCK, C. B. & WOOD, W. G. 2002. Quality Attribute Workshops. Architecture Tradeoff Analysis Initiative. 2nd ed. Pittsburgh, PA: Carnegie Mellon Software Engineering Institute.

BOEHM, B. & LANE, J. A. 2007. Using the Incremental Commitment Model to Integrate System Acquisition, Systems Engineering, and Software Engineering. CrossTalk, 19, 15.

BOEHM, B. & LANE, J. A. 2009. Better Management of Development Risks: Early Feasibility Evidence. 7th Annual Conference on Systems Engineering Research 2009 (CSER 2009). Loughborough University.

BOSCH-REKVELDT, M., JONGKIND, Y., MOOI, H., BAKKER, H. & VERBRAECK, A. 2011. Grasping project complexity in large engineering projects: The TOE (Technical, Organizational and Environmental) framework. International Journal of Project Management, 29, 728-739.

BRAHA, D., SUH, N., EPPINGER, S., CARAMANIS, M. & FREY, D. 2003. Complex Engineered Systems.

BUCKLEY, J., JAMES, P., CUCKSON, B. & MARSHALL, S. Naval Ship Assurance – A New Approach. BMT Defence Services Ltd / Lloyd's Register.

CALVANO, C. N. & JOHN, P. 2004. Systems Engineering in an Age of Complexity. Systems Engineering, 7, 10.

COLWELL, B. 2004. Design Fragility. Computer. IEEE Computer Society.

COLWELL, B. 2005. Complexity in Design. Computer. IEEE Computer Society.

COOK, S. C. & FERRIS, T. L. J. 2007. Re-evaluating Systems Engineering as a Framework for Tackling Systems Issues. Systems Research and Behavioral Science, 24, 14.

CROLL, P. R. 2008. Quality Attributes - Architecting Systems to Meet Customer Expectations. SysCon 2008 - IEEE International Systems Conference. Montreal, Canada.

DYER, J. S. & SARIN, R. K. 1979. Measurable multiattribute value functions. Operations Research, 27, 810-822.

ENGEL, A. & BROWNING, T. R. 2008. Designing Systems for Adaptability by Means of Architecture Options. Systems Engineering, 11, 22.

FISHER, W. P. & WRIGHT, B. D. 1994. Introduction to probabilistic conjoint measurement theory and applications. International Journal of Educational Research, 21, 559-568.

FRANÇOIS, C. 1999. Systemics and Cybernetics in a Historical Perspective. Systems Research and Behavioral Science, 16, 17.

FRIEDMAN, G. & DSAGE, A. P. 2004. Case Studies of Systems Engineering and Management in Systems Acquisition. Systems Engineering, 7, 84-97.

GREEN, P. & GABOR, G. 2012. MisLeading Indicators: How to Reliably Measure Your Business, Praeger Pub Text.

GREENBERG, R. & COOK, S. C. 2006. A Generic BBN Safety Model. Product Safety Engineering Society Symposium. IEEE.

GUSTAFSSON, A., HERRMANN, A. & HUBER, F. 2007. Conjoint Measurement. Methods and Applications., Berlin Heidelburg New York, Springer.

HAINES, D. F. 2006. Regulatory failures and regulatory solutions: a characteristic analysis of meta-regulation. Proceedings, Annual Meeting, Law and Society Association. Baltimore, MD: Unpublished.

HAMANN, R. J., ZANDBERGEN, B. T. C. & ZIJDEMANS, P. J. 2009. Does Systems Engineering in Space Projects Pay? INCOSE International Symposium 2009. Singapore.

HARRIS, D. Year. Quantifying accident precipitating events. In: 6th International Australian Aviation Psychology Symposium, Sydney, Australia, 2003.

HELMSMAN INSTITUTE 2010. A Comparison of Project Complexity between Defence and other Sectors. The Helmsman Institute.

HITCHINS, D. K. 1992. Putting Systems to Work, Chichester, UK, Wiley.

HITCHINS, D. K. 1998. Systems engineering in search of the elusive optimum. Engineering Management Journal, 12.

HITCHINS, D. K. 2005. Systems Methodology. Conference on Systems Engineering Research. Hoboken, NJ,.

HOBDAY, M. 1998. Product complexity, innovation and industrial organisation. Research Policy, 26, 689-710.

HOLDEN, R. A. 2001. Qualifying Risk: Mission Ready Certification of Shipboard Systems. DTIC Document.

HONOUR, E. & VALERDI, R. 2006. Advancing an Ontology for Systems Engineering to Allow Consistent Measurement. Conference on Systems Engineering Research. Los Angeles, Ca.

HUBBARD, D. W. 2009. The failure of risk management: Why it's broken and how to fix it, Wiley.

JACKSON, D., THOMAS, M. & MILLETT, L. I. 2007. Software for Dependable Systems: Sufficient Evidence?, National Academies Press.

JIMÉNEZ, A., RíOS-INSUA, S. & MATEOS, A. 2003. A decision support system for multiattribute utility evaluation based on imprecise assignments. Decision Support Systems, 36, 65-79.

JØRGENSEN, M. 1999. Software quality measurement. Advances in engineering software, 30, 907-912.

KJAERULFF, U. B. & MADSEN, A. L. 2010. Bayesian Networks and Influence Diagrams: A Guide to Construction and Analysis, Springer.

KLUDZE, J. A. K. 2004. The Impact of Systems Engineering on Complex Systems. Conference on Systems Engineering Research. Los Angeles, Ca.: Stevens Institute of Technology.

LANGSETH, H. & PORTINALE, L. 2007. Bayesian networks in reliability. Reliability engineering & System safety, 92, 92-108.

LEE, J. C. & MCCORMICK, N. J. 2012. Risk and safety analysis of nuclear systems, Wiley.

LEVESON, N. 2003. A new accident model for engineering safer systems. Safety Science, 42, 34.

LINACRE, J. M., ENGELHARD, G., TATUM, D. S. & MYFORD, C. M. 1994. Measurement with judges: Many-faceted conjoint measurement. International Journal of Educational Research, 21, 569-577.

LIU, J., YUAN, C., GU, F. & BILLER, S. 2008. Functional Safety Certification: Practice and Issues. 4th IEEE Conference on Automation Science and Engineering. Key Bridge Marriott, Washington DC, USA.

LLOYD, S. 2002. Complex Systems: A Review. ESD Internal Symposium. Cambridge, MA: MIT Engineering Systems Division.

LUCE, R. D. 1980. Several possible measures of risk. Theory and Decision, 12, 217-228.

MACKENZIE, S. B., PODSAKOFF, P. M. & PODSAKOFF, N. P. 2011. Construct measurement and validation procedures in MIS and behavioral research: integrating new and existing techniques. MIS Quarterly, 35, 293-334.

MCDONOUGH, R. 2003. The Role of Government in Technical Regulation "Finding the Right Balance". APIA Technical Regulation Seminar. Brisbane.

MCINTOSH, M. K. & PRESCOTT, J. B. 1999. Report to the Minister for Defence on the Collins Class Submarine and Related Matters. Canberra, ACT: Commonwealth of Australia.

MURDOCH, J., CLARK, G., POWELL, A. & CASELEY, P. 2003. Measuring safety: applying PSM to the system safety domain. Proceedings of the 8th Australian workshop on Safety critical systems and software-Volume 33. Australian Computer Society, Inc.

NEIL, M., LITTLEWOOD, B. & FENTON, N. 1996. Applying Bayesian Belief Networks to System Dependability Assessment. Safety-Critical Systems: The Convergence of High Tech and Human Factors. Springer.

O'HAGAN, A., BUCK, C. E., DANESHKHAH, A., EISER, J. R., GARTHWAITE, P. H., JENKINSON, D. J., OAKLEY, J. E. & RAKOW, T. 2006. Uncertain judgements: Eliciting experts' probabilities, Wiley.

PERRY, C. 1994. A structured approach to presenting PhD theses: notes for candidates and their supervisors. ANZ Doctoral Consortium. University of Sydney.

PFANZAGL, J., BAUMANN, V. & HUBER, H. 1971. Theory of measurement, Physica-Verlag Würzburg.

REDINGER, C. F. & LEVINE, S. P. 1998. Development and Evaluation of the Michigan Occupational Health and Safety Management System Assessment Instrument: A Universal OHSMS Performance Measurement Tool. American Industrial Hygiene Association Journal, 59, 572-581.

ROBERTS, G., SMALLER, A., SIMPSON, R. & PLASKITT, M. 2013. The Assurance Afforded to a Risk-Based Naval Authority Certification Regime by Commercial and Naval Class. Available: http://media.bmt.org/bmt_media/resources/89/RINAPaperClassificationandSafetyCase17OctFinalDraft.pdf [Accessed 10 May 2013].

ROEDLER, G. J. & JONES, C. 2005. Technical Measurement, A Collaborative Project of PSM, INCOSE, and Industry. Practical Software and Systems Measurement, International Council on Systems Engineering Measurement Working Group, INCOSE-TP-2003–020–01, 9-10.

ROSS, A. M. & RHODES, D. H. 2008. Architecting Systems for Value Robustness: Research Motivations and Progress. SysCon 2008 - IEEE International Systems Conference. Montreal, Canada: IEEE.

ROSS, A. M., RHODES, D. H. & HASTINGS, D. E. 2008. Defining Changeability: Reconciling Flexibility, Adaptability, Scalability, Modifiability, and Robustness for Maintaining System Lifecycle Value. Systems Engineering, 11, 17.

ROWLAND, J. J. & ROWLAND, D. 1995. Professional competence in safety-related software engineering. Software Engineering Journal, 6.

RYAN, A. 2007. A Multidisciplinary Approach to Complex Systems Design. PhD, The University of Adelaide.

SHALUF, I. M., AHMADUN, F. L.-R. & SHARIFF, A. R. 2003. Technological disaster factors. Journal of Loss Prevention in the Process Industries, 16, 513-521.

SPROLES, N. 2000. Coming to grips with measures of effectiveness. Systems Engineering, 3, 50-58.

STEVENS, R. 2008. Profiling Complex Systems. SysCon 2008 - IEEE International Systems Conference. Montreal, Canada.

STEVENS, R. 2011. Engineering mega-systems: The challenge of systems engineering in the information age, CRC Press, Auerbach Pub.

TANG, D., HECHT, M., HECHT, H. & BRILL, R. 1996. Measurement-Based Dependability Evaluation for Safety-Grade Digital Systems. Proceedings of the 1996 American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Control and Human-Machine Interface Technologies.

TURBAN, E. & METERSKY, M. L. 1971. Utility theory applied to multivariable system effectiveness evaluation. Management Science, 17, B-817-B-828.

ULLRICH, J. R. & PAINTER, J. R. 1974. A conjoint-measurement analysis of human judgment. Organizational Behavior and Human Performance, 12, 50-61.

VALERDI, R. 2011. Convergence of expert opinion via the wideband delphi method: An application in cost estimation models. INCOSE International Symposium. Denver, CO.

VATN, J. 1998. A discussion of the acceptable risk problem. Reliability engineering & System safety, 61, 11-19.

VISSER, G. D., BROWN, K. B., WASANASIT & ROSNIZAM 2005. Objective Quality Evidence. Navy Engineering Bulletin March 2006 ed.: Royal Australian Navy.

WANG, C. & WULF, W. A. 1997. A Framework for Security Measurement. Proceedings of the Twentieth National Information Systems Security Conference. Baltimore, MD.

- THE END -

Appendix A Survey instrument: validity of the measurement construct

This appendix contains the information provided to respondents of the first phase of survey research. It contains the:

- Research Participant Information Sheet
- Survey Instrument (administered to respondents via web based survey tool).

UniSA | Defence and Systems Institute

Research Participant Information Sheet

Construct Validity of Technical Integrity Measurement

Researcher:	***Mr. Michael Edwards, M.Eng.Sci.***	***(M) 0400 493 024***
Supervisor:	*A/Prof. David Cropley, PhD*	*(B) 08 8302 3301*

Research Objective

The impact of formal technical regulation and certification frameworks on the technical integrity of complex engineered systems is poorly understood [1]. In researching this problem I am seeking a means by which to measure *technical integrity*.

Technical Integrity

In the technical regulations of the Australian Defence Force (ADF), technical integrity is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection. Technical integrity has various different meanings and has synonyms and related concepts occurring in other military and in civilian engineered system contexts. We see a recurrence of the concepts of safety, compliance, dependability, effectiveness, quality, maturity and reliability as technical objectives of regulation as well as business values like asset value and return on investment.

At a generalised level, *technical integrity* of a complex engineered system may be defined as compliance to technical standards and the minimisation of unintended consequences. An approach for formulating a composite measure of *technical integrity* has been developed from this definition [2] and this survey seeks to validate and refine this proposed measurement instrument via a consensus of expert judgements.

Your Role as a Judge

You have been requested to participate since you are *familiar* with the definition, design, integration, verification, operation and/or support of complex engineered systems. You do not have to consider yourself to be an *expert* to participate in this study as a judge.

Your Participation

Note: Participation in this survey research is voluntary.

As a participant, you are free to withdraw from this research at any stage now or in the future.

The researcher will take every care to remove responses from any identifying material as early as possible. Likewise individuals' responses will be kept confidential by the researcher and not be identified in the reporting of the research. However the researcher cannot guarantee the confidentiality or anonymity of material transferred by email or the internet.

This project has been approved by the University of South Australia's Human Research Ethics Committee. If you have any ethical concerns about the project or questions about your rights as a participant please contact the Executive Officer of this Committee, Tel: +61 8 8302 3118; Email: Vicki.allen@unisa.edu.au

What You Will Be Asked

You will be asked to provide basic demographic data (age, gender, education, professional qualifications and experience).

You will be asked to rate your agreement with a number of statements relating to the measurement of technical integrity. You will also be asked to provide other information that may assist the researcher to understand your reasons for these ratings or for you to provide additional ideas to the researcher.

What will be done with your answers

Survey data returns will be securely stored electronically on USB memory stick in office of supervisor, A/Prof D Cropley, SPRI Building, Second Floor, West Wing, Mawson Lakes Campus. Survey data will be kept for 5 years prior to erasure and will only be accessible by the researcher.

Data will be stored as non-identifiable records (i.e. no identification data about you will be stored and the demographic data supplied cannot be attributed to you personally). While information gained during the study may be published, it will not be personally identifiable and your personal information will remain confidential.

Data will be analysed to determine:

- how well judges agree with each other (inter-judge reliability rating),
- any interdependence between the judgement criteria,
- the level of support for the measurement construct, and
- the level of correlation between judge ratings and demographic characteristics of the judges.

Publications arising from this analysis will only report on aggregated results, not on any individual responses.

Outcomes Enabled by This Survey

It is intended that this survey will enable:

- validation of a measurement model for technical integrity, and
- establishment of a pragmatic tool for the measurement of technical integrity.

I will provide you with my results as they develop. These will be in the form of conference or journal papers and ultimately a thesis. I will notify you via e-mail when these reports are available and provide either links to the published site or electronic copies of the reports.

I hope you will find them of use in your professional endeavours.

References

1. Edwards, M., *The Impact of Technical Regulation on the Technical Integrity of Complex Engineered Systems*, in *INCOSE 2009*. 2009, International Council On Systems Engineering: Singapore.
2. Edwards, M., *What is Technical Integrity and How to Measure It?*, in *INCOSE International Symposium 2011*. 2011: Denver, CO, USA.

Construct Validity of Technical Integrity Measurement

1. Your Consent

To indicate your willingness to participate in this research, please read the following statements and check the appropriate answers.

This project has been approved by the University of South Australia's Human Research Ethics Committee. If you have any ethical concerns about the project or questions about your rights as a participant please contact the Executive Officer of this Committee, Tel: +61 8 8302 3118; Email: Vicki.allen@unisa.edu.au

If you select "NO" to any of these, and therefore do not give your consent, please exit the survey.

Please answer the following:

	Yes	No
I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to take part.	◯	◯
I understand the purpose of the research project and my involvement in it.	◯	◯
I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.	◯	◯
I understand that while information gained during the study may be published, I will not be identified and my personal results will remain confidential.	◯	◯

Construct Validity of Technical Integrity Measurement

2. Demographics

You are asked to provide basic demographic data (age, gender, education, professional experience).

Are you male or female?

- ◯ Male
- ◯ Female

Which category below includes your age?

- ◯ 17 or younger
- ◯ 18-20
- ◯ 21-25
- ◯ 26-30
- ◯ 31-35
- ◯ 36-40
- ◯ 41-45
- ◯ 46-50
- ◯ 51-55
- ◯ 56-60
- ◯ 61-65
- ◯ 66-70
- ◯ 71 or older

What is the highest level of technical education that you have completed?

- ◯ No tertiary technical education
- ◯ Certificate
- ◯ Bachelor's degree
- ◯ Master's degree
- ◯ Doctoral degree

Please indicate the main discipline of your technical education/qualification.

- ◯ Aeronautical engineering
- ◯ Electrical/electronic engineering
- ◯ Mechanical engineering
- ◯ Industrial engineering
- ◯ Chemical engineering
- ◯ Civil engineering
- ◯ Computer science
- ◯ Physics
- ◯ Mathematics
- ◯ Other (please specify)

How long have you been in professional practice? (i.e., full-time, relevant employment)

Years:

In which application domain has most of your professional experience been gained?

- ◯ Maritme defence systems
- ◯ Aerospace systems
- ◯ Land based defence systems
- ◯ Commercial ICT
- ◯ Other (please specify)

Construct Validity of Technical Integrity Measurement

In which sector are you currently or most recently employed in?

- ◯ Academia
- ◯ Government
- ◯ Commercial Industry
- ◯ Professional Service Contractor or Consultant

In which sector has most of your professional experience been gained?

- ◯ Academia
- ◯ Government
- ◯ Commercial Industry
- ◯ Professional Service Contractor or Consultant

Do you have or have had a engineering delegation under ADF or other technical regulatory arrangements?

- ☐ No/Don't Know
- ☐ Yes: NTRS - Level 2
- ☐ Yes: NTRS - Level 3
- ☐ Yes: NTRS - Level 4 or lower
- ☐ Yes: TAMM - SDE
- ☐ Yes: TAMM - DSDE
- ☐ Other (please specify)
- ☐ Yes: TAMM - DE1
- ☐ Yes: TAMM - DE2
- ☐ Yes: TRAMM-L - Level 6
- ☐ Yes: TRAMM-L - Level 5
- ☐ Yes: TRAMM-L - Level 4 or lower

3. Background and Context

The researcher has developed a measurement approach for the Technical Integrity of a Complex Engineered System.

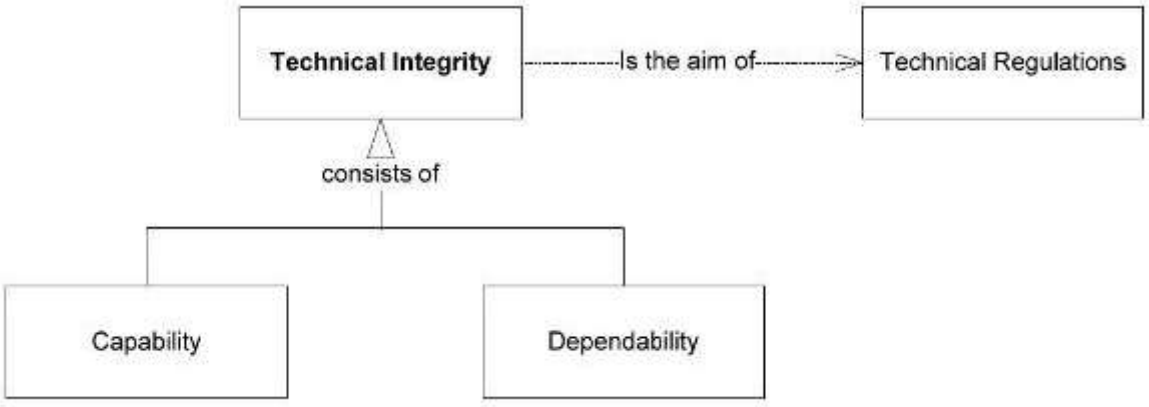
The proposed measurement technique hinges on the concept that the Technical Integrity of a Complex Engineered System can be considered in two orthogonal dimensions:

1. Capability - the achievement of the intended capabilities and attributes of a system. I.e., the system behaviours or characteristics that are:
a. explicitly required and
b. are the main focus of system design activities;

2. Dependability - the minimisation of unintended and/or undesired behaviours or characteristics of a system. I.e. , the system behaviours or characteristics that are:
a. not always explicitly identified for exclusion from the design,
b. the focus of activities like reliability analysis, system safety engineering and technical risk management that are undertaken to minimise the risk of undesired behaviours becoming extant in the completed system.

The remainder of this survey seeks your expert opinion in order to validate this construct or to identify a more valid construct for the measurement of technical integrity.

Model of Technical Integrity



Construct Validity of Technical Integrity Measurement

4. Your Viewpoint

The questions on this page seek to understand the point of view your responses represent. Your point of view is to be characterised by the role you most often undertake and the domain in which you work with complex engineered systems.

Please indicate the role that best matches the point of view you represent in responding to this survey

◯ Capability Manager (a manager or stakeholder of the entre capability for which a complex engineered system contributes. E.g. may be a senior defence force commander, the administrator of a large public utility, etc).

◯ Acquisition Manager (a person who is responsible or plays a role within the organisation that is responsible for the acquisition of a complex engineered system).

◯ System Operator (a person who primarily operates the complex engineered system).

◯ System Maintainer (a person who primarily maintains and repairs the complex engineered system).

◯ System Developer (a person who primarily designs, develops and integrates the complex engineered system).

Please indicate the domain of application of complex engineered systems that best fits the experience you represent in responding to this survey (indicate more than one if applicable)

☐ Military – Aviation

☐ Military - Naval

☐ Military - Land

☐ Military – Joint/C4ISR

☐ Civil – Aviation

☐ Civil – Maritime

☐ Civil –Medical

☐ Civil - Mining

☐ Civil – Chemical Process

☐ Civil – Industrial

☐ Civil - Telecommunications and IT

☐ Civil – Transportation

☐ Civil – Power Utility – Conventional

☐ Civil – Power Utility – Nuclear

☐ Civil – Water Utility

☐ Civil – Space

Other (please specify)

Construct Validity of Technical Integrity Measurement

5. Technical Integrity Scale

This set of questions seeks your judgments that will assist in determining the type of scale that is applicable for measuring Technical Integrity.

Please read the following statements carefully and indicate your level of agreement:

It is meaningful to say system A has higher Technical Integrity than system B.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

It is meaningful to ascribe a single number value to indicate the Technical Integrity of a system.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

It is meaningful to say that system A has, for example, twice the Technical Integrity of system B.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

Construct Validity of Technical Integrity Measurement

6. Capability

Recall the working definition of Capability - the achievement of the intended capabilities and attributes of a system.
I.e., the system behaviours or characteristics that are:
a. explicitly required and
b. are the main focus of system design activities

Assume that we can measure the Capability of a system on a scale of 0 to 1 where:
• 0 indicates no required capabilities or attributes of a system are achieved.
• 1 indicates all required capabilities and attributes of a system are achieved, and
• Intermediate values are determined as a proportion of required capabilities and attributes achieved.

Please read the following statements carefully and indicate your level of agreement:

Capability of a complex engineered system can be meaningfully measured on a scale of 0 to 1.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

Capability of a complex engineered system that exceeds the specified requirements is not of value.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

What, if anything, do you think may be a better measure of Capability for a complex engineered system?

◯ None better

◯ Something else better than achievement of required capabilities (please indicate)

Construct Validity of Technical Integrity Measurement

7. Dependability

Recall the working definition of Dependability - the minimisation of unintended and/or undesired behaviours or characteristics of a system. I.e. , the system behaviours or characteristics that are:
a. not always explicitly identified for exclusion from the design,
b. the focus of activities like reliability analysis, system safety engineering and technical risk management that are undertaken to minimise the risk of undesired behaviours becoming extant in the completed system.

Assume that we can measure the Dependability of a system on a scale of 0 to 1 where:
• 0 indicates a complete lack of system dependability (e.g., the system never works as intended, it frequently causes catastrophic accidents, etc).
• 1 indicates all risk of unintended and undesired behaviour of the system have been identified and either eliminated in the system or are acceptable to the owners, operators, and regulators (e.g., failure rates are tolerable, no failure ever cause catastrophic accidents, etc.), and
• Intermediate values represent the proportional level of risk between these extreme values.

Please read the following statements carefully and indicate your level of agreement:

Dependability of a complex engineered system can be meaningfully measured on a scale of 0 to 1.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

A complex engineered system with dependability that is better than the risk acceptance threshold of the system's stakeholders is of increased value.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

What, if anything, do you think may be a better measure of Dependability for a complex engineered system?

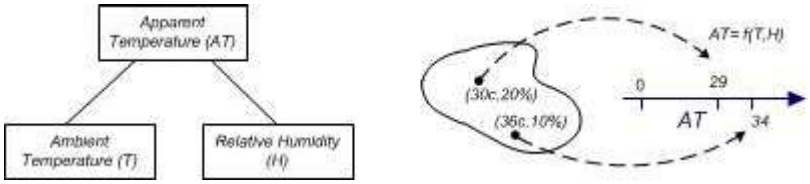
◯ None better

◯ Something else better than the risk of unintended, undesired behaviour (please indicate)

8. Conjoint Measurement of Technical Integrity

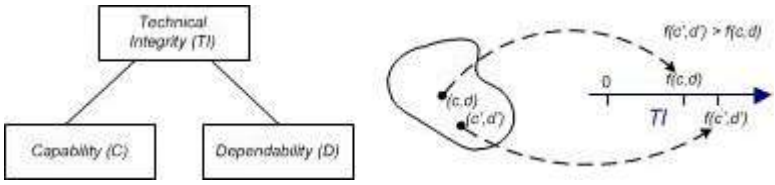
Conjoint measurement is a technique where 2 or more measurements are combined into a single measurement on an ordinary number scale that is meaningful to a range of interested parties. A common example is the combination of measurements of ambient temperature and relative humidity into "apparent temperature" – an indication of what the current climatic conditions feel like for a human. This is illustrated below showing how two combinations of temperature and humidity may be mapped to a single measure called apparent temperature.

Apparent Temperature: an example of conjoint measurement



The researcher intends to combine measurements of Capability and Dependability into a conjoint measurement for Technical Integrity in an analogous manner as illustrated below.

Concept of a Conjoint Measure for Technical Integrity



Please read the following statements carefully and indicate your level of agreement:

Constructing a conjoint measure for Technical Integrity from the component measures of Capability and Dependability is meaningful.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

A conjoint measure for Technical Integrity will allow meaningful comparison of the actual technical integrity of systems that are similar in type and application.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

A conjoint measure for Technical Integrity will allow meaningful comparison of the actual technical integrity of systems that are different in type and application.

Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Agree
◯	◯	◯	◯	◯

Construct Validity of Technical Integrity Measurement

9. Relative Importance of Capability and Dependability

This page seeks your judgment on the relative importance of capability and dependability across a range of domains you may have experience with.

Your reasoning for your judgment is also sought (note that this is at least as important to the research as the ratings).

In each domain in which you feel competent to judge, please indicate the relative importance of Capability or Dependability to systems within that domain.

	Capability Significantly More Important	Capability Much More Important	Capability More Important	Equally Important	Dependability More Important	Dependability Much More Important	Dependability Significantly More Important
Military – Aviation	◯	◯	◯	◯	◯	◯	◯
Military - Naval	◯	◯	◯	◯	◯	◯	◯
Military - Land	◯	◯	◯	◯	◯	◯	◯
Military – Joint, C4ISR, ..	◯	◯	◯	◯	◯	◯	◯
Civil – Aviation	◯	◯	◯	◯	◯	◯	◯
Civil – Maritime	◯	◯	◯	◯	◯	◯	◯
Civil –Medical	◯	◯	◯	◯	◯	◯	◯
Civil - Mining	◯	◯	◯	◯	◯	◯	◯
Civil – Chemical Process	◯	◯	◯	◯	◯	◯	◯
Civil – Industrial	◯	◯	◯	◯	◯	◯	◯
Civil - Telecommunications and IT	◯	◯	◯	◯	◯	◯	◯
Civil – Transportation	◯	◯	◯	◯	◯	◯	◯
Civil – Power Utility – Conventional	◯	◯	◯	◯	◯	◯	◯
Civil – Power Utility – Nuclear	◯	◯	◯	◯	◯	◯	◯
Civil – Water Utility	◯	◯	◯	◯	◯	◯	◯
Civil – Space	◯	◯	◯	◯	◯	◯	◯

What is there about your domain of practice or viewpoint that has influenced the way you have judged the relative importance of Capability and Dependability?

Construct Validity of Technical Integrity Measurement

What is there about the service(s) provided by the system(s) you were thinking of, that has an impact on the way you have rated the relative importance of Capability and Dependability?

Construct Validity of Technical Integrity Measurement

10. Any other comments?

Thank you for completing this survey.
This page is not mandatory, but if you would like to provide any of your observations having completed this judgment task, please do so here.

Any other comments or observations?

In particular, how has your understanding of "Technical Integrity" been changed by undertaking this survey?

Appendix B Survey instrument: measurement by consensual assessment technique

This appendix contains the information provided to respondents of the second phase of survey research. It contains the:

- Research Participant Information Sheet
- Survey Instrument (administered to respondents via web based survey tool).
- System descriptions for System A, System B and System C. Contrived examples of maritime radar systems.

UniSA | Defence and Systems Institute

Research Participant Information Sheet

The Measurement of Technical Integrity

Researcher:	***Mr. Michael Edwards, M.Eng.Sci.***	***(M) 0400 493 024***
Supervisor:	*A/Prof. David Cropley, PhD*	*(B) 08 8302 3301*

Research Objective

The impact of formal technical regulation and certification frameworks on the technical integrity of complex engineered systems is poorly understood [1]. In researching this problem I am seeking a means by which to measure *technical integrity*.

Technical Integrity

In the technical regulations of the Australian Defence Force (ADF), technical integrity is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection. Technical integrity has various different meanings and has synonyms and related concepts occurring in other military and in civilian engineered system contexts. We see a recurrence of the concepts of safety, compliance, dependability, effectiveness, quality, maturity and reliability as technical objectives of regulation as well as business values like asset value and return on investment.

At a generalised level, *technical integrity* of a complex engineered system may be defined as compliance to technical standards and the minimisation of unintended consequences. An approach for formulating a composite measure of *technical integrity* has been developed from this definition [2] and results from this survey are intended to validate and refine this proposed measurement instrument.

Consensual Assessment Technique (CAT)

The Consensual Assessment Technique (CAT) has been developed to measure the creativity embodied in a sample of products as judged by a range of judges. Amabile [3] established, over a large number of studies, that *creativity* can be reliably judged by appropriate judges (persons who are familiar with the domain, but not necessarily experts in the domain). Despite the fact that "creativity may be something difficult to describe, but it is still relatively easy for them [i.e., the judges] to identify [creativity] with a good degree of reliability".

While *technical integrity* of a complex engineered system and the *creativity* embodied in an artwork or other produced item are not direct analogues, there are enough similar characteristics of the problem space to investigate if the CAT technique can be successfully applied to establishing the *technical integrity* of a complex engineered system by a group of appropriate judges.

Similarities of the problem space include: both *technical integrity* and *creativity* are difficult to define precisely; they are not directly measurable; there is no complete nor agreed set of measurable components for either; both represent a quality of a product/system that people familiar with the domain are readily willing to subjectively judge in a relative sense (e.g. artwork A exhibits higher creativity than artwork B; system 1 has higher technical integrity than system 2).

Your Role as a Judge

You have been requested to participate since you are *familiar* with the definition, design, integration, verification, operation and/or support of complex engineered systems. You do not have to consider yourself to be an *expert* to participate in this study as a judge.

Your Participation

Note: Participation in this survey research is voluntary.

As a participant, you are free to withdraw from this research at any stage now or in the future.

This project has been approved by the University of South Australia's Human Research Ethics Committee. If you have any ethical concerns about the project or questions about your rights as a participant please contact the Executive Officer of this Committee, Tel: +61 8 8302 3118; Email: Vicki.allen@unisa.edu.au

What You Will Be Asked

You will be asked to provide basic demographic data (age, gender, education, professional experience).

You will be asked to review a single A3 page of technical data for each of 3 complex engineered systems.

Then you will be asked to rank these systems relative to one another across a range of judgement criteria.

What will be done with your answers

Survey data returns will be securely stored electronically on USB memory stick in office of supervisor, A/Prof D Cropley, SPRI Building, Second Floor, West Wing, Mawson Lakes Campus. Survey data will be kept for 5 years prior to erasure and will only be accessible by the researcher.

Data will be stored as non-identifiable records (i.e. no identification data about you will be stored and the demographic data supplied cannot be attributed to you personally). While information gained during the study may be published, it will not be personally identifiable and your personal information will remain confidential.

Data will be analysed to determine:

- how well judges agree with each other (inter-judge reliability rating),
- any interdependence between the judgement criteria,
- the level of correlation between judge ratings and objective characteristics of the systems, and
- the level of correlation between judge ratings and demographic characteristics of the judges.

Publications arising from this analysis will only report on aggregated results, not on any individual responses.

Outcomes Enabled by This Survey

It is intended that this survey will enable:

- validation of a measurement model for technical integrity,
- establishment of a pragmatic tool for the measurement of technical integrity, and
- identification of objectively measurable characteristics of a system that characterise or predict technical integrity.

I will provide you with my results as they develop. These will be in the form of conference or journal papers and ultimately a thesis. I will notify you via e-mail when these reports are available and provide either links to the published site or electronic copies of the reports.

I hope you will find them of use in your professional endeavours.

References

1. Edwards, M., *The Impact of Technical Regulation on the Technical Integrity of Complex Engineered Systems*, in *INCOSE 2009*. 2009, International Council On Systems Engineering: Singapore.
2. Edwards, M., *What is Technical Integrity and How to Measure It?*, in *INCOSE International Symposium 2011*. 2011: Denver, CO, USA.
3. Amabile, T.M., *Consensual Assessment Technique*, in *Creativity in Context*. 1996, Westview Press: Boulder, CO, USA. p. 41-79.

The Technical Integrity of Complex Engineered Systems

1. Your Consent

To indicate your willingness to participate in this research, please read the following statements and check the appropriate answers.

This project has been approved by the University of South Australia's Human Research Ethics Committee. If you have any ethical concerns about the project or questions about your rights as a participant please contact the Executive Officer of this Committee, Tel: +61 8 8302 3118; Email: Vicki.allen@unisa.edu.au

If you any "NO" to any of these, and therefore do not give your consent, please exit the survey.

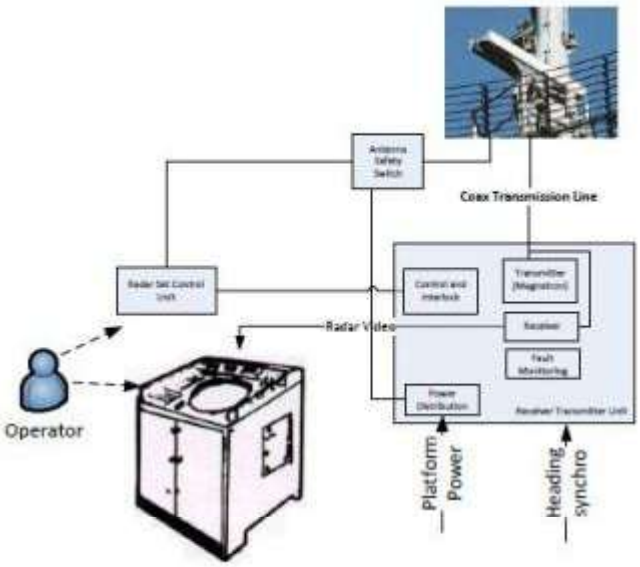
Please answer the following:

	Yes	No
I have read the Participant Information Sheet and the nature and purpose of the research project has been explained to me. I understand and agree to take part.	◯	◯
I understand the purpose of the research project and my involvement in it.	◯	◯
I understand that I may withdraw from the research project at any stage and that this will not affect my status now or in the future.	◯	◯
I understand that while information gained during the study may be published, I will not be identified and my personal results will remain confidential.	◯	◯

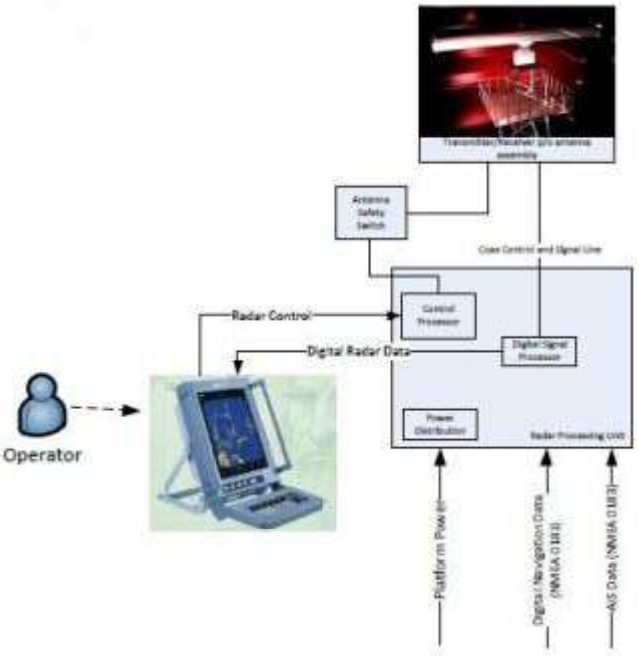
2. The Systems to be Evaluated

You have been supplied separately with a description of a number of systems which you will be asked to supply judgement on. At this stage you should have reviewed and be familiar with the information for each system. You should have the description pages for the systems at hand for the remainder of this survey.

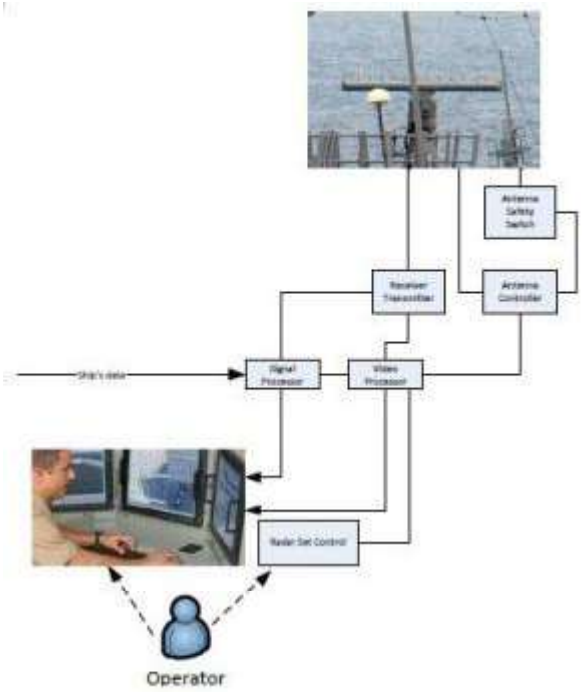
System A



System B



System C



Have you reviewed the system descriptions and have them on hand for the remainder of the survey?

- ◯ Yes
- ◯ No

3. Survey to Terminate

You have indicated you are not ready to complete the survey. You will need to exit this survey now by closing your browser window - please reattempt the survey when you have the information on the systems at hand.

4. Survey to Terminate

You have indicated you are not ready to complete the survey. You will need to exit this survey now by closing your browser window - please reattempt the survey when you have the information on the systems at hand.

5. Demographics

You are asked to provide basic demographic data (age, gender, education, professional qualifications and experience).

Are you male or female?

- ◯ Male
- ◯ Female

Which category below includes your age?

- ◯ 17 or younger
- ◯ 18-20
- ◯ 21-25
- ◯ 26-30
- ◯ 31-35
- ◯ 36-40
- ◯ 41-45
- ◯ 46-50
- ◯ 51-55
- ◯ 56-60
- ◯ 61-65
- ◯ 66-70
- ◯ 71 or older

What is the highest level of technical education that you have completed?

- ◯ Certificate
- ◯ Bachelor's degree
- ◯ Master's degree
- ◯ Doctoral degree

Please indicate the main discipline of your technical education/qualification.

- ◯ Aeronautical engineering
- ◯ Electrical/electronic engineering
- ◯ Mechanical engineering
- ◯ Industrial engineering
- ◯ Chemical engineering
- ◯ Civil engineering
- ◯ Computer science
- ◯ Physics
- ◯ Mathematics
- ◯ Other (please specify)

How long have you been in professional practice?
(i.e., full-time, relevant employment)

Years:

In which application domain has most of your professional experience been gained?

- ◯ Maritme defence systems
- ◯ Aerospace systems
- ◯ Land based defence systems
- ◯ Commercial ICT
- ◯ Other (please specify)

In which sector are you currently or most recently employed in?

- ◯ Academia
- ◯ Professional Service Contractor or Consultant
- ◯ Government
- ◯ Commercial Industry

In which sector has most of your professional experience been gained?

- ◯ Academia
- ◯ Professional Service Contractor or Consultant
- ◯ Government
- ◯ Commercial Industry

Do you have or have had a engineering delegation under ADF technical regulatory arrangements?

- ☐ No
- ☐ Yes: NTRS - Level 2
- ☐ Yes: NTRS - Level 3
- ☐ Yes: NTRS - Level 4 or lower
- ☐ Yes: TAMM - SDE
- ☐ Yes: TAMM - DSDE
- ☐ Yes: TAMM - DE1
- ☐ Yes: TAMM - DE2
- ☐ Yes: TRAMM-L - Level 6
- ☐ Yes: TRAMM-L - Level 5
- ☐ Yes: TRAMM-L - Level 4 or lower
- ☐ Other (please specify)

Your professional experience with complex engineered systems has included practice with (select all that apply):

- ☐ capability definition
- ☐ system specification
- ☐ system design
- ☐ hardware design and build
- ☐ software design and build
- ☐ system integration
- ☐ system testing
- ☐ system operation
- ☐ system maintenance
- ☐ acquisition management
- ☐ project management
- ☐ system sustainment
- ☐ technical publications
- ☐ operation and maintenance training

Other (please specify)

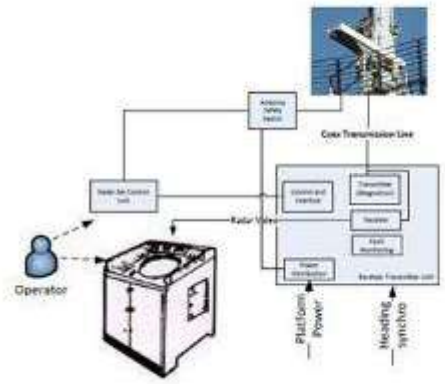
6. Your judgements

You are asked to rate each of the systems absolutely on the scale given for each of the judgement attributes.

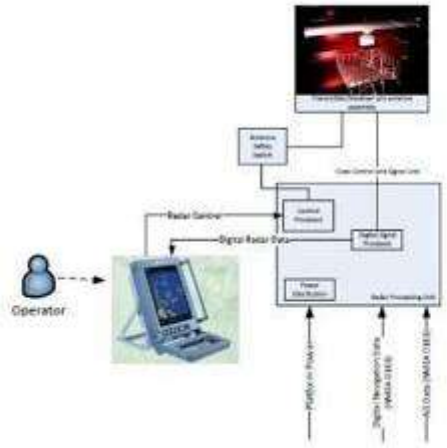
You can consider your input to be a discrete point on a continuous scale ranging from the extreme values displayed for each judgement attribute.

Since it is unlikely in reality that any two systems could display the same rating for the one attribute, you will be forced to rate each system differently from one another. Remember, you may go back and revisit the ratings for any system and/or judgement dimension prior to completing this page.

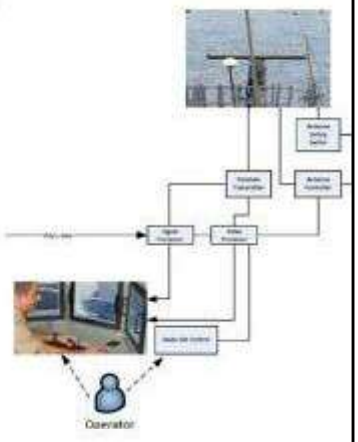
System A



System B



System C



Using your own subjective definition of "Technical Integrity", rate the degree to which each system exhibits Technical Integrity relative to the other example systems.

	No Technical Integrity															High Technical Integrity
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "Safety", rate the degree to which each system exhibits Safety relative to the other example systems.

	Unsafe															Safe
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "impact to the natural environment", rate the degree to which each system impacts the natural environment relative to the other example systems.

	Impact to Natural Environment															No impact to Natural Environment
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "impact to other systems", rate the degree to which each system impacts other systems relative to the other example systems.

	Impacts other systems															No impact to other systems
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "Fitness for Service", rate the degree to which each system exhibits Fitness for Service relative to the other example systems.

	Not Fit for Service															Fit for Service
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "Capability", rate the degree to which each system is capable relative to the other example systems.

	No Capability															High Capability
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "Effectiveness" rate the degree to which each system is effective relative to the other example systems.

	Ineffective															Highly Effective
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "Efficiency" rate the degree to which each system is efficient relative to the other example systems.

	Ineffecient															Efficient
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "Construction Quality" rate the degree to which each system has Construction Quality relative to the other example systems.

	Low Quality															High Quality
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "Operational Flexibility" rate the degree to which each system is operationally flexible relative to the other example systems.

	Inflexible															Flexible
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Using your own subjective definition of "Durability", rate the degree to which each system is durable relative to the other example systems.

	Not Durable															Durable
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

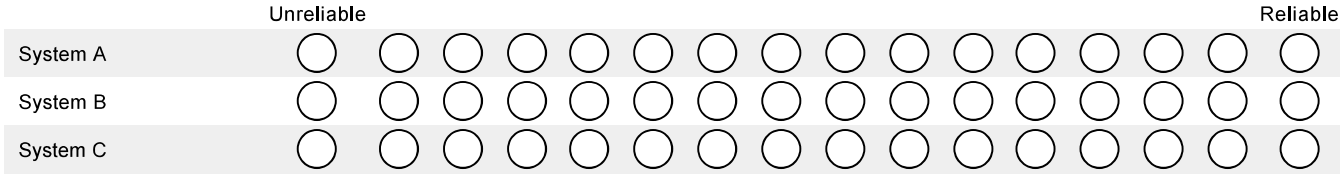
Using your own subjective definition of "Survivability", rate the degree to which each system is suvivable relative to the other example systems.

	Not Survivable															Suvivable
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

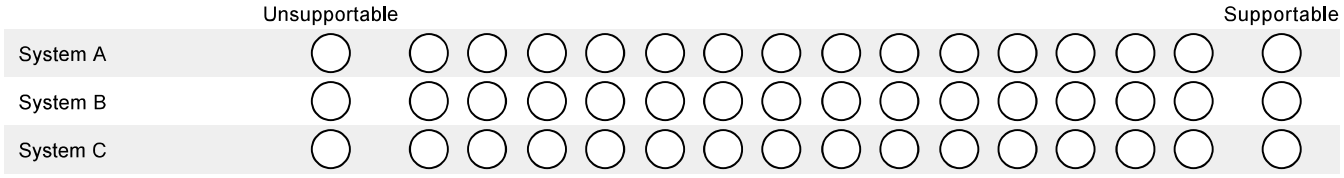
Using your own subjective definition of "Dependability", rate the degree to which each system is dependable relative to the other example systems.

	Not Dependable															Dependable
System A	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System B	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
System C	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

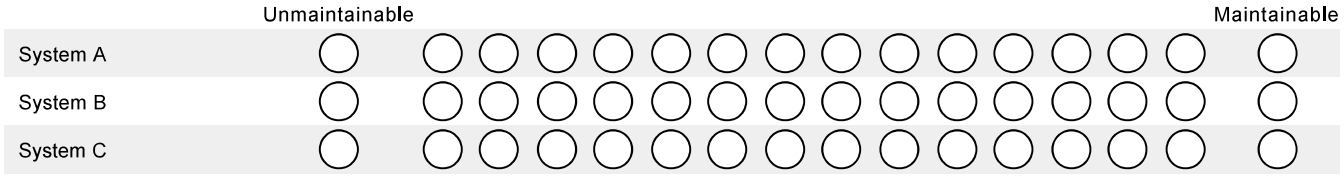
Using your own subjective definition of "Reliability", rate the degree to which each system is reliable relative to the other example systems.



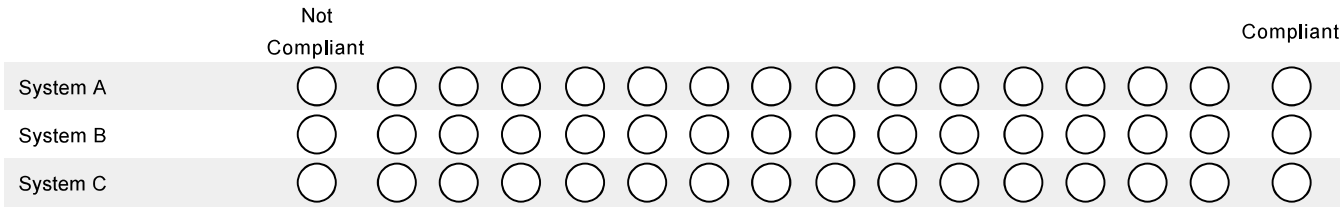
Using your own subjective definition of "Supportability", rate the degree to which each system is supportable relative to the other example systems.



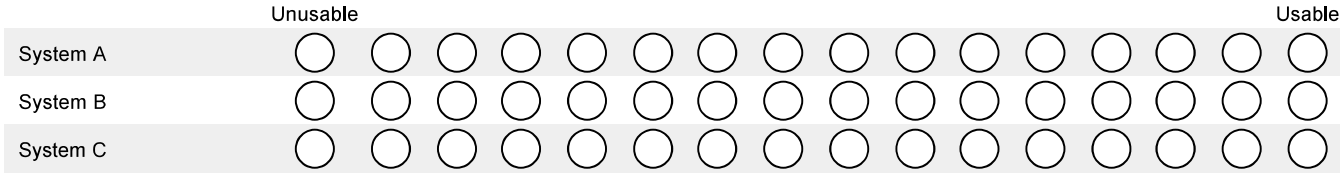
Using your own subjective definition of "Maintainability", rate the degree to which each system is maintainable relative to the other example systems.



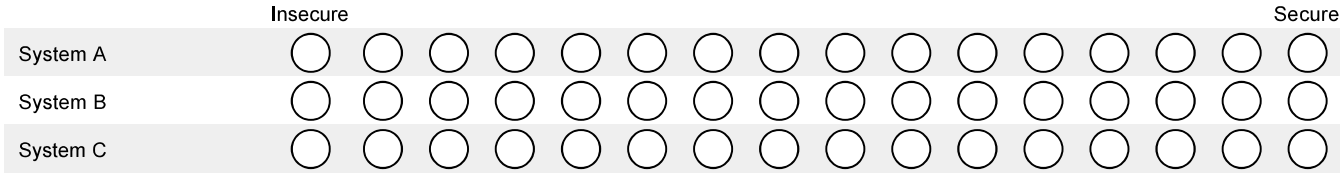
Using your own subjective definition of "Technical Compliance", rate the degree to which each system is technically compliant relative to the other example systems.



Using your own subjective definition of "Usability", rate the degree to which each system is usable relative to the other example systems.



Using your own subjective definition of "Security", rate the degree to which each system is secure relative to the other example systems.



Using your own subjective definition of "Technical Complexity", rate the degree to which each system is technically complex relative to the other example systems.

	Low Technical Complexity															High Technical Complexity
System A	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
System B	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
System C	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

Using your own subjective definition of "High Technology" rate the degree to which each system is considered to be High Technology relative to the other example systems.

	Low Technology															High Technology
System A	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
System B	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯
System C	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯

7. Your views and comments?

Thank you for completing the survey questions. The data you have provided will be used in quantitative analyses in order to establish a valid measurement tool for Technical Integrity.

In addition, your views, comments and insights as a result of undertaking this survey are anticipated to provide a valuable source of qualitative data to further inform this research.

This page is not mandatory, but your considered answers in response to the following questions would be much appreciated.

How has your opinion on the nature of "Technical Integrity" changed as a result of completing this survey?

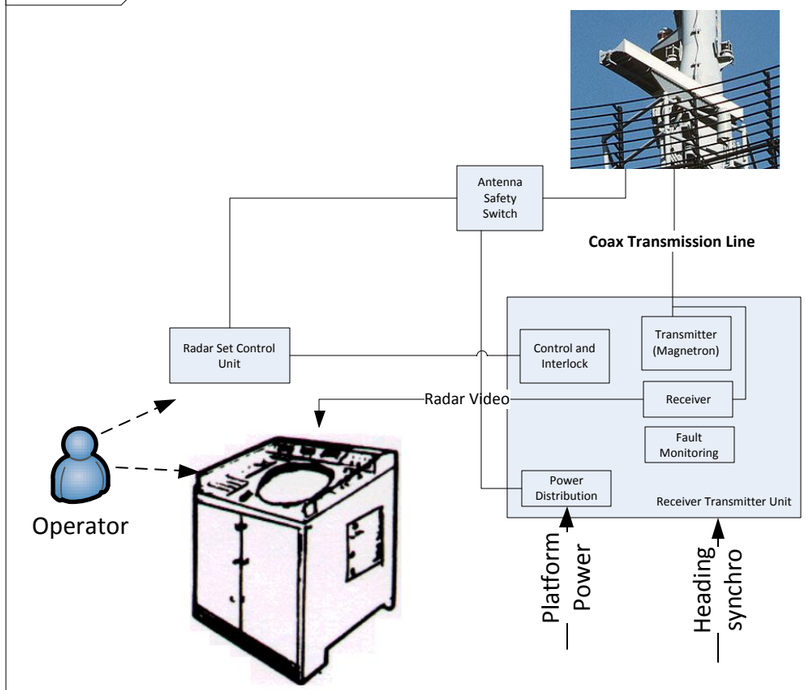
Can you comment on the extent to which you relied on objective data provided for the systems to make your judgements.

Can you reflect and comment on how you made judgements when you felt there was little or no objective data provided?

Any other comments you would like to make?

Radar A

Block Diagram



Capabilities & Performance

Primary Mission: detect ships, coastlines and hazards to navigation
Secondary Mission: low altitude air search
Instrumented Range: 190 km long pulse; 60 km short pulse
Range Accuracy: 22m, 198 m
Bearing Accuracy: 1.5º
Data rate: 16 rpm – 3.75sec
Auto Track Capacity: Nil; aids for manual tracking.

Interface to External Environment

RF Frequency: I-Band (X-band) tunable over 9.05 – 10.0 GHz
RF Bandwidth: 1.2 MHz long pulse; 10 MHz short puls
RF Power: 130-160kW peak
RF Waveform: 1µs pulse, 0.12 µs pulse
PRF: 750pps; 2250 pps

Military Threat Resistance

Unintentional jamming resistance: tunable frequency; polarisation
Intentional jamming resistance: polarisation
Counter Detection: N/A
Selective Emitter Identification: N/A

Shock and blast resistance: MIL-S-901

Technology

Antenna: two waveguide slotted arrays mounted back-to-back, one linearly horizontal polarised, one circularly polarised.
Transmission Line: coaxial
Transmitter: Magnetron
Receiver: Low noise RF front-end; sensitivity time control
Signal Processor: fast time constant filtering
Control: sector radiate control
Display: monochrome CRT; analog vide sweep

Supportability Features

Redundancy features: 5 display outputs
Maintenance Concept: on board repair by technicians
BIT Features: some, details not known
MTBF: 1200 hours
MTTR: 20 minutes

Provenance and Usage

Contracts: Originally contracted for in 1971.
Year first deployed: 1972
Number of units in service: >140
Platform classes: Frigates, destroyers, cruisers, minesweepers
Manufacturer Quality System: ISO 9001

Hazardous Materials:

Cadmium plated connectors used on all cable terminations.
No other hazmats disclosed.

Safety Features:

Antenna Safety Switch – when activated prevents rotation and radiation from the antenna by isolating power distribution to antenna assembly. Provides indication to operator.
Cabinet Interlock for electrical safety.

Security Features:

Physical security: only as provided by installation
Computer security: no general purpose computer

Specification/Standard Compliance

Specified Requirements	System A
Radar Performance compliant with IMO Resolution A.477 (XII) when installed on platform	✓
Shall operate in I-Band (3cm)	✓
Range performance to IMO Resolution A.477 (XII)	✓
Detect 60 m high coastline at 26 nautical miles	✓
Detect large surface ship at 15 nautical miles	✓
Detect small surface ship at 10 nautical miles	✓
Detect navigational buoy at 4 nautical miles	✓
Range scales to 50 nautical miles	✓
Adjust pulse width with range scale in use (auto or manual)	✓
Display range rings	✗
Range discrimination of two small target seperated by 25 yards	✓
Radar shall process GPS inputs of position, couse and speed over ground	✗
Radar shall incorporate an ARPA in accordance with IMO resolution A.823	✗
Ability to store and recall up to 20 electronic maps	✗
Provide blanking signal to platform electronic support equipment.	✓
Antenna rotation rate selectable in the range 24 to 40 rpm	✗
The transceiver shall be an integral part of the antenna system	✗
Power input shall be to MIL-STD-1399 Section 300A.	✓
Key lockable Safe to Transmit and Safe to rotate switches to be provided.	!
EMI/EMC to MIL-STD-461D.	!
Alerts shall be audible/visible to the operator if there is a loss of signal at any interface.	!
Extensive BIT shall allow maintainer to isolate fault to a LRU.	!
The radar shall allow the operator to choose between SI and imperial units for display.	✗
Compass input required for azimuth stabilisation.	✓
EM Log input for sea stabilisation.	✓

✗ Not compliant to requirement
! Partially compliant to requirement
✓ Compliant to requirement

Build Standard and Environmental Resistance

Temperature:
(Receiver/Transmitter Unit and Radar Set Control)
Non-Operating: -62°C to +75°C
Operating: OIC to +50°C
Antenna Pedestal: -28°C to +65°C

Humidity:
Up to 95% at +60°C (MIL-E-16400)

Enclosure:
Receiver/Transmitter Unit and Radar Set Control 45°, drip-proof.
Safety Switch, splash-proof.
Antenna Pedestal, submergible to 3 feet.

Vibration, and Inclination:
MIL-E-16400

Salt Spray:
MIL-E-16400

Heat Test:
MIL-E-16400

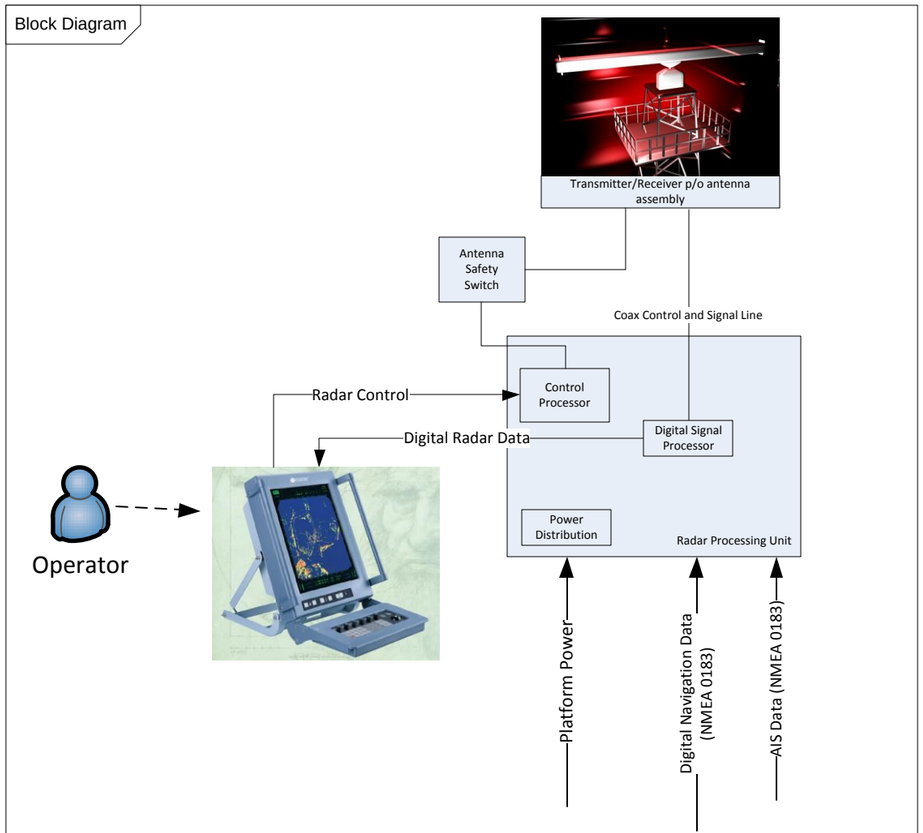
Platform Service Requirements

Power Type: 115 Vac/60Hz/1Φ
Power Draw: 1100W
Cooling Load to Air/Water: 100W
Data inputs: synchro bearing

Weight:
Display: 140kg
Receiver/Transmitter: 260 kg
Antenna Group: 131 kg

Radar B

Block Diagram



Capabilities & Performance

Primary Mission: detect ships, coastlines and hazards to navigation
Instrumented Range: 200 km long pulse
Range Accuracy: 1% of target range
Bearing Accuracy: 1.2º
Data rate: 60 rpm – 1 sec
Auto Track Capacity: 20 auto tracks

Interface to External Environment

RF Frequency: I-Band (X-band) tunable over 9.05 – 10.0 GHz
RF Bandwidth: 10 MHz
RF Power: 100kW peak
RF Waveform: 0.1µs pulse
PRF: variable iaw selected range scale

Military Threat Resistance

Unintentional jamming resistance: matched filtering providies immunity form most interfering signals
Intentional jamming resistance: N/A
Counter Detection: N/A
Selective Emitter Identification: N/A

Shock and blast resistance: Nil

Technology

Antenna: circularly polarised waveguide slotted array
Transmitter: Solid state tranmitter integrated into antenna pedestal
Receiver: Low noise, solid state RF front-end integrated into antenna pedestal
Signal Processor: digital signal processor implementing coherent matched filter
Control: sector radiate control; variable PRF set auto with selected range scale
Display: high resolution LCD colour; electronic chart underlay; integration with AIS tracks

Supportability Features

Redundancy features: digital displayand track outputs can be provided to other display systems
Maintenance Concept: repair by return to OEM
BIT Features: full go/no go test results available to operator
MTBF: 4000 hours
MTTR: 30 minutes

Provenance and Usage

Contracts: Originally contracted for in 2001.
Year first deployed: 2003
Number of units in service: >30
Platform classes: Naval patrol boats, customs vessels, motor yachts
Manufacturer Quality System: ISO 9001

Hazardous Materials:

No hazmats disclosed.

Safety Features:

Antenna Safety Switch – when activated prevents rotation and radiation from the antenna by isolating power distribution to antenna assembly. Provides indication to operator.

Security Features:

Physical security: only as provided by installation
Computer security: password security enforced on display and processing unit

Specification/Standard Compliance

Specified Requirements	System B
Radar Performance compliant with IMO Resolution A.477 (XII) when installed on platform	✓
Shall operate in I-Band (3cm)	✓
Range performance to IMO Resolution A.477 (XII)	✓
Detect 60 m high coastline at 26 nautical miles	✓
Detect large surface ship at 15 nautical miles	✓
Detect small surface ship at 10 nautical miles	✓
Detect navigational buoy at 4 nautical miles	✓
Range scales to 50 nautical miles	✓
Adjust pulse width with range scale in use (auto or manual)	✓
Display range rings	✓
Range discrimination of two small target seperated by 25 yards	✓
Radar shall process GPS inputs of position, couse and speed over ground	✓
Radar shall incorporate an ARPA in accordance with IMO resolution A.823	✓
Ability to store and recall up to 20 electronic maps	✓
Provide blanking signal to platform electronic support equipment.	✓
Antenna rotation rate selectable in the range 24 to 40 rpm	✗
The transceiver shall be an integral part of the antenna system	✓
Power input shall be to MIL-STD-1399 Section 300A.	!
Key lockable Safe to Transmit and Safe to rotate switches to be provided.	✗
EMI/EMC to MIL-STD-461D.	✗
Alerts shall be audible/visible to the operator if there is a loss of signal at any interface.	✓
Extensive BIT shall allow maintainer to isolate fault to a LRU.	!
The radar shall allow the operator to choose between SI and imperial units for display.	✗
Compass input required for azimuth stabilisation.	✓
EM Log input for sea stabilisation.	✓

Not compliant to requirement ✗
Partially compliant to requirement !
Compliant to requirement ✓

Build Standard and Environmental Resistance

Meets the requirements of:
IEC 60945 Maritime Navigation and Radiocommunication Equipment and Systems – General Requirements – Methods of Testing and Required Test Results

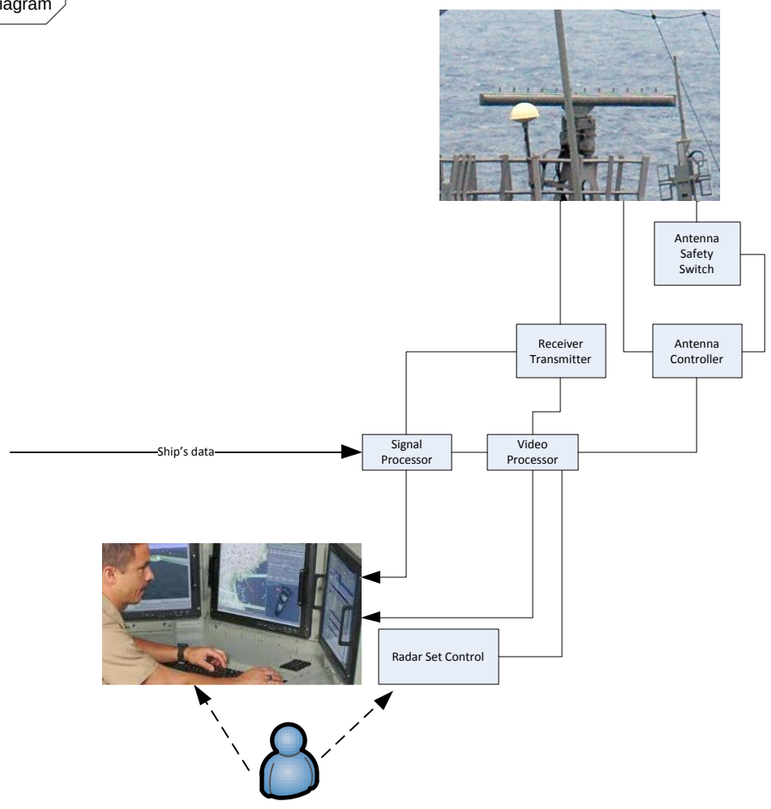
Platform Service Requirements

Power Type: 115 Vac/60Hz/1Φ
Power Draw: 1000W
Cooling Load to Air/Water:
Data inputs: digital – NMEA 0183 format

Weight:
Display: 20 kg
Processor Unit: 40 kg
Antenna Group: 90 kg

Radar C

Block Diagram



Capabilities & Performance

Primary Mission: surface and air surveillance
Secondary Mission: navigation
Instrumented Range: 56 nm
Range Accuracy: 30m, 75 m, 300m
Bearing Accuracy: 1.5º
Data rate: 2 or 4 sec
Auto Track Capacity: 128, auto track initiation

Interface to External Environment

RF Frequency: C-Band 5.5 GHz
RF Power: 280kW peak
RF Waveform: 0.1µs pulse, 0.25 µs pulse, 1 µs pulse
PRF: 750, 1200, 2400

Military Threat Resistance

Unintentional jamming resistance: coherent EMI suppression
Intentional jamming resistance: jam strobe detection, sector CFAR
Counter Detection: none
Selective Emitter Identification: none

Shock and blast resistance: MIL-S-901D Grade A

Technology

Antenna: slotted array with built in IFF antenna
Transmission Line: coaxial
Transmitter: Coaxial Magnetron
Receiver: Low noise RF front-end; sensitivity time control
Signal Processor: sector by sector CFAR, coherent on receive Digital Moving Target Indicatot
Control: Auto Frequency Control, sector radiate control
Display: colour LCD display console; raw video sweeps with auto-detect track symbology

Supportability Features

Redundancy features: multiple display and track data outputs
Maintenance Concept: on board repair by technicians
BIT Features: extensive to replaceable part level
MTBF: 3000 hours
MTTR: < 30 minutes based on field data

Provenance and Usage

Contracts: For current variant let in 2000.
Year first deployed: 2003
Number of units in service: >50
Platform classes: Frigates, destroyers, cruisers,
Manufacturer Quality System: ISO 9001

Hazardous Materials:

Cadmium plated connectors used on all cable terminations.
Zinc chromate primers used on antenna group.
No other hazmats disclosed.

Safety Features:

Antenna Safety Switch – when activated prevents rotation and radiation from the antenna by isolating power distribution to antenna assembly. Provides indication to operator.
Cabinet Interlock for electrical safety.

Security Features:

Physical security: only as provided by installation
Computer security: maintainer must use password to access computer elements of system.

Specification/Standard Compliance

Specified Requirements	System C
Radar Performance compliant with IMO Resolution A.477 (XII) when installed on platform	✓
Shall operate in I-Band (3cm)	✗
Range performance to IMO Resolution A.477 (XII)	✓
Detect 60 m high coastline at 26 nautical miles	✓
Detect large surface ship at 15 nautical miles	✓
Detect small surface ship at 10 nautical miles	✓
Detect navigational buoy at 4 nautical miles	!
Range scales to 50 nautical miles	✓
Adjust pulse width with range scale in use (auto or manual)	✓
Display range rings	✓
Range discrimination of two small target seperated by 25 yards	!
Radar shall process GPS inputs of position, couse and speed over ground	✓
Radar shall incorporate an ARPA in accordance with IMO resolution A.823	!
Ability to store and recall up to 20 electronic maps	✓
Provide blanking signal to platform electronic support equipment.	✓
Antenna rotation rate selectable in the range 24 to 40 rpm	!
The transceiver shall be an integral part of the antenna system	✗
Power input shall be to MIL-STD-1399 Section 300A.	✓
Key lockable Safe to Transmit and Safe to rotate switches to be provided.	✗
EMI/EMC to MIL-STD-461D.	✓
Alerts shall be audible/visible to the operator if there is a loss of signal at any interface.	✓
Extensive BIT shall allow maintainer to isolate fault to a LRU.	✓
The radar shall allow the operator to choose between SI and imperial units for display.	✗
Compass input required for azimuth stabilisation.	✓
EM Log input for sea stabilisation.	✓

✗ Not compliant to requirement
! Partially compliant to requirement
✓ Compliant to requirement

Build Standard and Environmental Resistance

Operating Temperature: -10°C to +50°C

Humidity:
Up to 95% at +45°C (MIL-STD-810)

Enclosure:
Inboard equipment: drip-proof to 15°
Antenna Pedestal: qualified for driving rain

Vibration:
MIL-STD-167-1

Platform Service Requirements

Power Type: 115 Vac/60Hz/1Φ
Power Draw: 2000W
Cooling Load to Air/Water: 1800W
Data inputs: synchro and digital navigation data form ship (heading, speed, position)
Weight:
Display: 200 kg
Receiver/Transmitter: 20 kg
Antenna Group: 180 kg

Appendix C Validity of technical integrity measurement construct: full text extracts referenced in qualitative analysis

This appendix contains the coded extracts[1] described in the body of this thesis. The extracts are presented under the corresponding headings as Qualitative Analysis section of Chapter 5 of the main thesis. Parts of these extracts that have been directly quoted in the main body of the thesis are highlighted.

Better capability measurement

No alternate capability measurement proposed

The following are the extracts that have been coded as "No Alt Capability Measure Proposed".

> Working definition is NOT that required for a 'capability'. ie The power to achieve a desired operational effect in a nominated environment, within a specified time, and to sustain that effect for a designated period. Commonwealth of Australia (2010) Capability Preparedness Level. Preparedness is the ability to respond to directed contingencies within a specified time frame for a given duration, it is a combination of readiness and sustainability of a capability. This may well be able to be 'measured' on such a scale, but unlike Tech Integrity it may be meaningful to say that Capability A is twice as good relative to Capability B. Care has to be taken on scaling 'specified reqts' - this can't be easily normalised as it is multidimensional and I believe that poor Conops to FPS is one of the root causes of current disfunctionality between assessing system A, against 'SoS B' and 'Force Level capabilities which have SoS 'working' together .
>
> A potential issue with measuring capability acheivement by a number is that capability elements will have priorities (currenty expressed as Essential, Desirable, etc) and that these priorities are subjective and can change over time, hence potentially invaliditing the assessment

[1] Typographical errors in extracts from survey responses are not corrected.

The Department of Defence defines ‘capability’ as the power to achieve a desired operational effect in a nominated environment, within a specified time, and to sustain that effect for a designated period (DCDM 2006, p. 5). Ascribing a single number to the combination of all of these is subjective, at best, and will differ on a project to project basis. e.g. Project A may place more weight on achieving the effect within the specified time, while project B considers being able to sustain that effect as more important.

I have concerned about the quality of the documented required capabilities, that might mean that only a measure of achievement of the required capabilities may an important part of the measurement

"Capability" is a multi-dimensional concept. Systems are designed to meet specific performance specifications, to meet specific reliability requirements, to meet specific production cost targets, etc. It may be possible to roll up an assessment of various capabilities into a single value; however, care must be taken in doing so since not all capabilities are of equal importance. The weighting of the many elements of a capability metric will vary between systems. The validity of comparisons of such metrics would need to be studied and validated across a number of systems.

Simplification of and more rigoruous requirements analysis Iterative measurement and involvement of end user

Concern is that complex systems, generated as systems of systems, surpass this assessment. Beneficial emergent behaviour can be a key goal. The measurement to a OSA architecture may apply better to the architecture, than the evolving system of systems based upon it.

Achievement of required capabilities is important, but the single number scale you have proposed is too lossy in information to be useful to a capability evaluator. Context has to be maintained, and as each solution will achieve each requirement in a potentially different way, then the single number measure looses insight into this. Also, it doesn't take into account dependability of each capability, or the relationships between them. It assumes each capability requirement is mutually exclusive, which they rarely are both as a capability and as an implementation of that capability.

Correct specification of necessary requirements and achievenemt of the same.

Context should always be required - even if a set of contexts is needed

My reservations regarding the proposed scale largely stems from a concern that the combination of constituent capability assessments (e.g. a 0.9 for deployment of troops, but a 0.6 for carriage of equipment) is not meaningful to represent as a single number.

Complex Engineering System - 7 Key Touch Points (7KTP) 1. Human (ie ease to use HMI) 2. Data (ie standard interfaces) 3. Application (ie HW/SW dev/support) 4. Adaptability (ie size, integration, growth) 5. Impact (ie effect in theater/operations) 6. Benefit (ie successes over life in service) 7. Innovation (ie distruptive [new] vs sustainable [reuse])

The definition is good but also need to gauge the ability of the system to evolve over time to address obsolescence with new technology but no change ib

functionality; also the ability of the system to evolve to increased functionality over time to meet changing needs of the User.

The system provides the capabilities as defined in the elements in a capability model or meets x/y elements. The first two questions use Capability (singular) yet this third question introduces Capabilities (plural).

A form of measurement that confirms achievement of required capability along with an assessment of technical risk

stretch targets, capability to meet perceived future requirements with minimal design change

Suggest use scale % scale. Its commonly used, e.g. acquired 95% capability and the other 5% was not practical or affordable. I recall ADF was considering figures some 10 years ago around 80% to make it even more affordable. Although this model tended to use the acquisition spec requirements with not weighting on the operational capabilities. Note that some defence acquisitions were already refined to reflect the min capability which means if you do not achieve 100%, the acquistion could be seriously deficient

capability as a function of cost would be a better measure with ability to range between marginally acceptable and a stretch goal of capablilty and then measure the value across that range to determine best for project

(a) and (b) above reflect the explicitly required, and the main focus of design - these are the designers expected/intended capabilities, however a system can include additional capabilities that may eventuate from the original intended design. For example, the (radar) system was designed to detect targets up to xxx nautical miles, but we found it could actually effectively detect much further, up to yyy nm. In the case of software, it may be able to handle higher throughput/more target tracks/more accurate measurements etc. but it can be constrained (often termed "dumbed down") to meet the requirements. Here the full capability is not implemented or exercised.

World's Best Practice

Ability to meet the purpose or purposes for which it is intended. It is generally intended that requirements should specify characteristics that allow the system to meet its intended purpose(s), but these are generally an approximation to the intended purpose. The expected may also change over time. It is also quite likely that once the system has been instantiated it may offer additional capabilities. (eg we purchased a second car just to transport people to work. but given the actual car, it now has the capacity to transport pets and shopping. we may then use the car for this purpose which is additional to the original intended purpose)

A value added assessment of the capability must be completed to see if achieving a greater than 1 scale is beneficial. Original requirements may need to be modified as capability is better understood and customer defines need more clearly. Costs vs effort must be considered.

The process of validating a complex system through in-use in the actual context (intended environments, users and operational scenarios) can demonstrate capability rather than measure capability. In this instance, the complex system is not measured against textual requirements but demonstrated with variable

scenarios over time. Feedback from all users to the developer directly is then used to develop what the capabilities and limitations of the system actually are. This is not a measure, but a statement of fact. Features missing can be added or superfluous features removed depending on cost. This method ensures that the complex system meets it's original intent. Unfortunately this process is time intensive so only applied in Commercial Industry where the customer requirements are poorly defined or unknown.

Need to be able to break the complex engineered systems into components, as the parts that will make up the system will have different reliability and capability ie need to understand where the spare capacity is in the system and what compoenents are under stress.

Set of components

The following are the extracts that have been coded as "Set of Components".

Potential better would be a simple capability "profile" against a small set of environments/modes.

Decisions made regarding capability are based on a combination of technical competence, such as operational advantage, commercial good sense and engineering precedent, all matched against capability gap analysis, concept assessment, product demonstration, and in-service evaluation. Distillation of these into a single figure of merit is meaningful for abstraction purposes at the highest levels, but only in the same sense that the ONR Technology Readiness Levels are helpful. A better measure of capability would comprise sub-ordinate grades against the metrics above

a range of measures (shown on, for example a radar chart) would facilitate display of multiple dimensions of capability

A series of measured points using input from a team of SME's from various disiciplines with detailed knowledge of the System

The capability of the system needs to be measured against different mission sets and there should be a measure for each mission set that the system is intended to perform in.

A better approach could be to use this 0-1 scale, against a small profile of the systems key capabilities, e.g. for a particular platform required to conduct anti-submarine warfare and air attack, then both of these key capabilities should be rated between 0-1 and expressed accordingly.

It can a 2D plot/bar chart of sub-system and level of effectiveness.

A measuerement system should go down to subsystem level,

Capability measure > 1

The following are the extracts that have been coded as "Capability Measure > 1".

If greater than the capabilitity can be/is achieved than greater than 1 is good.

Including in the scale (which may potentially be increased beyond a max of 1?) not only the required capabilities (which are usually tailored to a contract following an initial cost/performance tradeoff), but all of the other desirable capabilities that would exist in a "money-no-object" environment (which are often part of the set of requirements prior to option "down selection"

plus the additional capabilities that are inherent in the system even though not stated as a requirement

Suggest that alpha character could be appended to numeric to represent that capability exceeds baselined specification capability. (e.g. 0.8E could reflect that 80% of the particular capability is achieved against baseline specification however part or all of this provided capability exceeds specified performance requirements.)

Note that capability that exceeds requirments may add positive or negative value. Negative if, for example cost increases or dependibility decreases 2) From above, value could be greater than 1 if realy good or less than one if a dog

Achievement of required capabilities plus other capabilities that are inherent to the design/technology or as a result of unintended but beneficial behaviour/properties

capability better than 'Requirement' could be assigned greater than 1.0 provided that extra capability was agreed by operational staff as providing useful 'value'.

If greater than the capabilitity can be/is achieved than greater than 1 is good.

there is no room in such a model to meaningfully report a measure greater than 1, as if it is a valid element of that capability, it should be an appropriately weighted requirement (be it desirable vs essential etc.). If it cannot be contained within the definition of the system requirements, any value cannot be meaningfully attributed within the context of the capability as a whole

Need to also include additional benefits realised that were not originally intended or specified.

that is achieved beyond the required capability

it can be meaningful to state if it exceed the specified requirements - it can be useful to know if something is over-engineered and/or may have another application.

other features of a system that provide advantage or benefit should not be ignored just because they were not described in the original scope. the evaluation cannot exceed 1 because the secondary features/benefits are not described in scope

overall measure needs to take into account the emergent system level behaviour provided by the underlying capabilities, and the perceived usefulness of that behaviour in terms of the system operators/owners.

Qualitative assessment

The following are the extracts that have been coded as “Qualitative Assessment”.

A 'fit for purpose' assessment of the complex system by the operators in addition to 'the achievement of the intended capabilities and attributes of a system' is a better

overall measure 'Fit for purpose' assessment is based on requirements interpretation. It will be a multiple pass incremental improvement of complex system capability as requirements interpretation converges

A series of measured points using input from a team of SME's from various disiciplines with detailed knowledge of the System

Altenrate is a qualitative/relative assessment only, not a quantifiable one

Fitness for purpose is also relevent when it comes to determining the capability (i.e we can explicitly define and satisfy the requirements, but still end up with a system that has severly limited capability)

Qualitative measures.... for example, may be ratings given by a potential user group.

Measure applied by fulfilment of client need

overall measure needs to take into account the emergent system level behaviour provided by the underlying capabilities, and the perceived usefulness of that behaviour in terms of the system operators/owners.

the measure should be the System User's view on whether the system provides the expected experience in achieving the desired capability outcomes. (end user experience)

Operational capability measure

The following are the extracts that have been coded as “Operational Capability Measure”.

Does the system perform in the intended manner in its intended environment? Yes, the system passes Acceptance. System is capable. No, the system fails acceptance. Whilst even more binary, removes the intermediates that should not exist if designed and developed to meet the functional specification dictated.

I think a measure of capability should also explicitly refer to the "Need" or "Operational" aspects.

a measure of the overall effectiveness of the user's greater entity to achieve its goals would be useful

The capability of the system needs to be measured against different mission sets and there should be a measure for each mission set that the system is intended to perform in.

a weighted set of technical performance attributes that reflects a quantitative achievement of operational outcome

Measures Of Effectiveness (MOE) against the various stakeholders criteria of benefits of the system

Achievement of required capabilities plus other capabilities that are inherent to the design/technology or as a result of unintended but beneficial behaviour/properties

the "Capability" is better measured by the degree to which the operational system can be used by the operators to achieve the goals or objectives it was intended to support. This might be better represented by a ConOps or by Operational Trials of some sort.

Compliance against Measures of Performance and Sustainment (MOP and MOS).

Achieving the desired operational effect.

Any measure of Capability should be measured against the COIs and MOE/MOS.

Ability of the operator to utilise a complex engineered system to it's full capability.

Measures of Effectiveness or Measures of Performance

Measures of effectiveness or relative comparison to a reference capability

Weighted components

The following are the extracts that have been coded as “Weighted Component”.

Measured against the baseline requirements. These requirements can weighted hence capability achieved can be quantitatively assessed.

I would suggest that each required capability be qualified by say the identification of whether it is: Essential, Desired or Optional; and applicable weightings applied before the result is scaled to a value 0.. 1.

Some sort of weighting of the capabilities to reflect user preferences would help to discriminate.

a weighted set of technical performance attributes that reflects a quantitative achievement of operational outcome

explore a weighting factor in the components (set of requirements) of the desired capability

Intermediate values must be weighted as to the importance to the mission not just a requirement count.

perhaps the same model, but developed for each subsystem and then look at it as a whole, which means an architectural breakdown of teh overall system and some meanigful formula based on teh tree structure may need ot be developed

Measured against the baseline requirements. These requirements can weighted hence capability achieved can be quantitatively assessed.

you must have a clear, measurable baseline (with appropriate weightings etc. as all requirements are not equal and interdependencies may have significant impacts on overall capability beyond a binary measure of achievement per requirement.

The number of essential and optional attributes provided, each attribute being weighted

whilst a single capability could be rated on a scale of 0 to 1, the overall capability of the system (sum of capabilities) would not perhaps convey the best measurement as not all capabilities would be considered equal.

Importance (weighting) of each required capability should be factored into measurement of System Capability rather than simple proportion measure

Some capabilities are presumably critical while others may be 'nice to have'. It's unclear if there is an intent assign a higher weight to the more critical capabilities

explore a weighting factor in the components (set of requirements) of the desired capability

Other

The following are the extracts that have been coded as "Other".

An indication of level of achievement should be considered. In many cases requirements are set up to have a desired and an essential threshold. With a binary system a complex engineered system that fully complies with essential, but just misses the mark for desired, will get a score of 1, while it would be closer to 1.99 in reality

Number of fitness for service technical integrity risks with a Hazard Risk Index < 10

Suggest that alpha character could be appended to numeric to represent that capability exceeds baselined specification capability. (e.g. 0.8E could reflect that 80% of the particular capability is achieved against baseline specification however part or all of this provided capability exceeds specified performance requirements.)

Note that capability that exceeds requirments may add positive or negative value. Negative if, for example cost increases or dependibility decreases 2) From above, value could be greater than 1 if realy good or less than one if a dog

measurement of requirements effectiveness (whatever effectiveness means) could be a useful contributor to Capability measurement

A range of at least four measures would seem appropriate: fails to meet requried level of capability, achieves requried level of capability with work arounds, achieves required level of capability and exceeds required level of capability.

the 0 to 1 scale will become a decimal scale so why not have a scale that is 0 to 10 to start with. Take out as much human intepretation as possible

A combination of achieved capability, the potential for extended capability, the potential for colloration in achieving new capability and the increase in efficiency of achieving existing capability

differentiate between safety issues, mission critical issues etc

Better dependability measurement

No alt dependability measure proposed

The following are the extracts that have been coded as “No Alt Dependability Measure Proposed”.

> Interesting concept. It seems that this measure is a mixture of of measurable/calculatable things like availability and reliability, and risk probability values. Does it work to put them together or should they be two different measures ?

> It would be difficult to get the customer to accept something that is not at a value of 1 (ie. the risks have been identified, mitigated and accepted by the owners). So while the number may be useful during development to measure progress, a product with a technical integrity of >1 would not be sold off. It may be useful in-service however with the identification of engineering problems, however unless I am mis-understanding the point this is covered by the consideration of probability and consequence???

> Dependability is always dependent on context, and can never be simplified to a single ordinal measure. The is what Safety Integrity Levels and Design Assurance Levels try to do, but these are only targets and they never reflect achievement of dependability. There is a plethora of literature out there that criticises these levels, and so I recommend you study take these into account. Also, you're measure don't try to model uncertainty in our assessment of any of these parameters. There is never unlimited evidence, because there is never unlimited time or money, so there will always be uncertainty in our knowledge of capability and dependability.

> This is very much of a "It Depends" question wherein the cost of undependabilty could range from minor increased maintenance costs to putting lives at risk. e.g. if a plane's entertainment system occasionally didn't work it would be a minor annoyance versus a flight control system failure..

> Similar comment to previous in that A potential issue with measuring dependablity acheivement by a number is alos based on priorities and that these priorities are subjective and can change over time, hence potentially invaliditing the assessment

> The assignment of a numeric value can be compromised by the aggregation of the statistics that constitute the overall rating - for instance it may be that it has all risks of erratic behaviour have been mitigated to the acceptance of the customer (a score of 1) but reliability is poor - 60% availability, without proper weightings this could be construed as a dependable system as the overall score implies 0.8.

> Strongly recommend you read Nassim Talebs book on 'Anti-fragility' as I am not sure where exactly 'dependability' is situating this. Jackson is doing some good work on robustness etc too. Hoody, who is 'Director of DGTA Reform' is looking at Bow Tie analyses of Tech Integrity and the regulators views! Have you been talking to him / them yet - need to as they need some help assimilating that to regulatory obligations. Your work needs to address new OPS 2-2 views on Tech Integrity and 'safety'. Aviation Material includes technical equipment such things as: aircraft (manned and Remote Piloted/Uninhabited Air Vehicle, whether leased or owned by the Commonwealth), munitions, air defence equipment, air traffic control equipment, wide area surveillance systems, ground support and test

equipment, command and control systems, vehicles, small arms, training aids and the software associated with any of those systems. This includes Aviation Support Systems consisting of systems or services that are Defence-owned or operated on behalf of Defence; have a functional or physical interface with aircraft; and have the potential to affect the airworthiness of those aircraft. These can be ground-based, ship-based, aircraft-based or space-based and are to be approved by the ADF to operate when it is designed, constructed, maintained and operated to approved standards and limitations for its intended purpose, by competent and authorised individuals, who are acting as members of an approved organisation and whose work is both certified as correct and accepted on behalf of Defence. Aviation Support Systems are categorised by AAP 7001.048 Sect 2 Chap 6 Annex A as: • Category A. The output or function of the ASS directly affects the immediate performance of the aircraft (eg a flight control/data link system interface with an aircraft). The failure of a Cat A system could reasonably be expected to result in death or serious injury to personnel, or significant damage to property, with limited aircrew capacity to effect recovery action. This is the category that information used for targeting would conceivably need to be for 'Danger Close' JTAC Joint Fires applications. • Category B. The output or function of the ASS is relied upon by aircrew for the safe operation of the aircraft. The failure of a Cat B system could possibly result in injury to personnel or damage to property if aircrew are unaware of the problem and fail to execute normal recovery procedures. This would be the category most likely for targeting information assurance. • Category C. The output or function of the ASS is used by aircrew for advisory or non-safety purposes only. The failure of a Cat C system would not be expected to result in injury to personnel or damage to property. In SRG, the AP-3C Orion and E-7A Wedgetail and the ADFs ATC systems are categorised and design accepted at this level. I understand DGTA is working these 3 categories into their future Tech Reg framework (which is good) as to what it means (finally). Not sure how to answer the first question - anti-fragile and robust are easy, dependable should be but if you're try to just average things I vehemently disagree as the MOE/MOS/MOPs and emergent properties can all be driven to zero if one is missed when it counts! Most network systems and the people making the decisions are not guassian too! Your second question here is also probably far too convoluted for this country boy to wrap my head around fully I am sorry. The key problem is your finding earlier in APCOSE SETE 2012 that the multi-faceted 'system stakeholders' all have different risk acceptance paradigms/models, which of itself makes the problem non-linear in trying to work out who is 'right'. NOTE: complicated systems are NOT complex systems be it 'engineered' or not. DMO's definitions of complex systems is trivial, trite and undefensible!

As per my answer to the same question for capability

See comments on capability

I think using 1 as "all risks... have been identified...and....are acceptable" is setting the bar too low. In the end we often find ways to accept / tolerate risks that are not the peak of possible performance

I am not sure that it is meaningful to measure dependability on a numerical scale when the benchmark is something as soft and variable and risk acceptance threshold of system stakeholders.

Again not necessarily a better measure, but how well the the 'dependability' tools are applied (i.e reliability, risk managment etc) needs to be considered. Some such as reliability are easier to qunatify, and other less so (i.e. elimination of

undesiredbehaviours). Maybe the application of the 'depnedability' tools can be measured in asimilar manner to CMMI, but they still do not guarantee elimination of undesirble behaviour.

measurable but difficult to turn into a measure (we'd need to talk about it)

Agian the scale will become a decimal scale so define the acceptable figures, 1 to 10 etc thus minimising people spending time and effort on whether the system is a .54 or .55 which will happen in IMO. define the acceptable scale up front with meanings attached and state that the system must be placed in a box.

By including reliability and safety in one measure you are mixing apples and oranges. Whilst related, these 2 aspects are often in conflict; by raising the reliability (availability) of a system it may be at the expense of safety but by having one measure there may be no overall change in value. In effect, you are then hiding the true level of unsafeness which is your legal obligation to ensure.

Suggest use & scale. 100% = 1

The first is that safety should be a separate section and not included with dependability as whether it is safe or not has nothing to do with whether it is reliable or not.

dependable does not equal working as intended (that is a capability) but is how reliable the system is

Don't necessarily disagree but think the measure needs to be extended to take into account, supportability, maintainability, not just reliability analysis. Also consideration to the dependability and the operating environment needs to be addressed.

Once again this is a good definition, however also feel there needs to be an account of the systems ability to scale to address new situations. For example, susceptability to new cyber threats not know at the time or appreciated.

(D) Set of components

The following are the extracts that have been coded as “Set of Components”.

reliability must link to the components of the system, and have a scale to indicate how good the components of the system

This is inherently variable should a matrix should be used to characterise the dependability against the mission sets

a multi-dimensional concept

The aggregate risk is the value of the probabilistic expectation of an adverse outcome, due to a wide range of undesired system attributes, which may be expressed as a function of multiple variables rather than a single value. This could be also communicated as the likelihood of mission failure due to those undesired system attributes.

Dependability measure > 1

The following are the extracts that have been coded as “Dependability Measure > 1”.

> if better than risk acceptance threshold that is generally good unless it cost a lot.
>
> You could have greater than one-the thing I worked on was so dependable, it reduced demand on other systems
>
> Perhaps for a system with dependability better than the risk acceptance threshold, a figure >1 could be assigned using the same proportionality as for 0 - 1.
>
> Value for Money (VFM) must come into the equation. If I can get acceptable dependability for what I'm prepared to pay and additional dependability comes at a much higher price, then it has less value to me. If I'm faced with two systems that are otherwise equal and for the same price one meets my dependability requirement and the other exceeds it I would place a higher value on the latter (ie the one with the higher than required dependability).

(D) Qualitative assessment

The following are the extracts that have been coded as “(D) Qualitative Assessment”.

> Willingness of the operator to utilise a complex engineered system to it's full capability.

Risk assessment

The following are the extracts that have been coded as “Risk Assessment”.

> additional information would also be needed to understand the specific residual risks
>
> probability of accident leading to unintended harm to the environment, people or equipment
>
> The degree to which the system is available to perform the role or implement the capability it was intended to. Availability is a the key word here. If it has unintended behaviours that can be worked around or which don't impact the "mission" this may be acceptable. This is not dissimilar to what is proposed here. Frequent catastrophic failures would likely imply low availability. Some unintended behaviours are more serious than others. The measure of this seriousness should take into account the impact on ability of the users to perform their mission
>
> Perhaps the percentage of hazards which are not mitigated to ALARP within the hazard log.
>
> problem with dependability is that there could be unkown unknowns
>
> Dependability is multidimensional and using only one number makes trade off between the dimensions. For example system A may have more probability of failure than system B but may be safer than system B and both may have the same dependability index.

The process of defining that a system is safe should match the ALARP process once a level of acceptible risk is defined using the Hazard Risk Identifiers (HRIs).

Dependability also has to measure the uncertainty of a system - this should be an estimated "probabalistic" parameter.

(D) Weighted components

The following are the extracts that have been coded as "(D) Weighted Components".

The scale for intermediate values may not be linear and needs some thought of how it may be represented as accurately as possible

I think cost along with requirments weighting may want to be a part of your multivariate optimization problem

weighting of the many elements

the measurement of dependability of some systems is a complex issue (eg software) and if left with a measurement that is qulaitative, then perhaps weighting factors could be used in determining the level of dependability

What is dependability - reliability, adaptability, resilience, safety? Each term havs a meaning and requires a weighting for the project to be developed

Relative weightings covering priority and interdependency

(D) Other

The following are the extracts that have been coded as "(D) Other".

binary

Corect specification of necessary requirements and achievement of the same.

plus the opportunity cost of providing alternative capability to meet the same mission requirements

an alpha could be appended to reflect that dependability has been previously demonstrated on other programs (e.g. 0.9D - represents that some risk exists against the system, but that the system is operational - deployed - in other platforms/installations.

I would suggest that Dependability must be qualified over the entire life of the system and not just how the system is operating now - given that failure, obsolescence and required capability are all a function of time.

Failsafe Operation, Multiple Redundancy

Actual reliability metrics rather than a risk assessment

A time factor has to be involved, ie what is the probability of achieving a mission(or other succes measure) without critical failures?

Reflections on the measurement of technical integrity

Valid research problem

The following are the extracts that have been coded as “Valid Research Problem”.

More discussions of this topic can only help our industry.

Good luck with trying to quantify this complex issue. I'd be very interested in the outcomes of the research.

My undrstanding of 'technical integrity' has been clarified by this survey and I support the reasearch work being undertaken as a valid research subject.

I think that the research aim (if met) will have considerable benefit in quantifying an outcome that often can (at the moment) only be truely appreciated by highly experienced practitioners (who are most likely not the key decision makers!).

The challenge of 'measuring' both Capability and Dependability is significant and yet it is desirable that we can estimate/assign a comparative 'value' to Technical Integrity so that we can identify both the required Integrity and when a particular mission system falls above or below that level.

I think your research has a good goal

Interesting topic, will be interested in knowing how this idea evolves in time.

Anyway, on the whole I think it is a really interesting topic, which is probably difficult to distill into a 10 page survey, so great job. I'd be really interested to see the outcomes.

a measure of technical integrity against the ability to perform the real operational mission is probably far more interesting and valuable.

An intriguing framework. I look forward to seeing the function mapping the two variables to a 'goodness' score.

Very interesting question and one which the naval maritime field continues to struggle with.

The approach mentioned in the reference paper "What is Technical Integrity and How to Measure it" is very well presented and the concepts are definitely worthwhile exploring further.

A subject/approach that will help future large complex projects be more successful both technicially and programmaticly

Good luck with your research - a very important field of enquiry for today's complex systems of systems.

I lik ethe structured approach to this research.

I have mostly deemed Technical Integrity as fairly subjective, but this survey allowed me to think of it as a measurable assessment. Comparing similar yet simple systems is much easier than diverse complex systems. However, I believe

such an assessment is possible if perimeters are also complex and objective measurement is mutually agreed.

Some good questions requiring thought.

Tractability and usefulness of measurement

The following are the extracts that have been coded as "Tractability and Usefulness of Measurement".

Measuring the value of requirements satisfied and risk assessed may provide a reference, and possibily allow two projects undertaken within an oranisation with similar methods, application and personnel, I think there is far too much variability in how the requirements were developed and the 'dependability' elements assessed to make a clear comparison.

There is a real risk in simplifying such a difficult subject as Technical Integry to a single number where such a number is used by those who really don't understand the impact of experience with complex systems by organisations and individuals. This may drive the specification, development, integration or acceptance of such complex systems in unintended ways - comparing apples and oranges will often result.

By distilling Technical Integrity down to capability and dependability and associating these attributes with a numerical ranking is an interesting concept - but the criteria for applying the numerical values would need to be closely followed.

I think the approach you are proposing is too lossy to be useful. Actually it has a very high chance of being misleading.

I find pinning down dependability to one number difficult and therefore pinning both dependability and capability to one number even more challenging.

Technical Integrity has been collapsed into the here-and-now. This artificial limitation is OK for purposes of simplifying the problem. However, the implication that Capability and Dependability are equally quantifiable is perhaps not true.

In many instances measures are used as a short cut to understanding. In large, complex systems this is a dangerous game to play.

I feel that the example about combining two technical measures, like temperature and humidity, doesn't scale well up to the complexity of something like AWD or Wedgetail - i.e. the complexity of the system doesn't scale linearly with the number of inputs. Compressing complex information (in this case down to a single number) is a lossy process, but rarely do such approaches ever highlight this to the consumer, which leaves them with a false impression that this single number is sufficient basis for decision making.

Numbers alone are not always useful when comparing apples with non-apples. But it is better than no numbers at all.

This survey has made me consider methods for the measurement of TI, I've come to the conclusion that it is not an easy task.

Its a simple enough concept, like anti gravity, but rather hard to implement

Single measures of capability or dependability are useful in comparing design options so that the best value option may be selected. However, I don't see much benefit in comparing a technical integrity measure for two different systems serving two different purposes as this does not facilitate any meaningful decision making that I can envisage.

Reducing technical integrity to a numerical score can never substitute for a deep understanding of the inherent dynamics and compex structures of complex systems. All that the parameters can do is provide a quick look insight that should asist in prioritising work to address necessary changes requred to capability and depndability to meet a given mission

I doubt that capability and dependability can be regarded as 'orthogonal' aspects, as in my experience they may related to one another.

While I have never really considered assessing Technical Integrity numerically it would appear to be most difficult.

I can relate to a system being characterised as point on a two dimensional chart capability versus dependability. Not sure if a combined TI index will work.

My own view is that when you combine measures you lose context and other than giving a heads up to a problem are otherwise not useful. Technical Integrity is a difficult area with a lot of subjective elements.

The means of measuring is questionable.

It's ok to have different measures contributing to Technical Integrity, they don't need to be conjoined or balanced out in any way, possibly motivated because one contributor is very low and the other high. To use your own example, weather reports still report temperature and humidity seperate, rather than a single "comfort" number.

Never thought of "measuring" TI before and am sceptical that a meaningful measure can be derived.

I consider evaluation of TI levels as being somewhat subjective. The performance of a system can be dependent on variable factors such as environment (e.g. weather, time of day etc.) and the systems operators (level of training, experience, teamwork, fatigue and other human factors, communication etc.). The same system operated and maintained by different teams/crews can function differently, potentially impacting on the system's evaluated TI level.

No capability vs dependability prioritisation possible

The following are the extracts that have been coded as “No Cap vs Dep Prioritisation Possible”.

In any real world case, capablity/dependablity are conflicting in a cost/benefit/consequence, whethere benefit/consequence can be related to lives or dollars. it is difficult to see that a universal statement of which is more important, as the assessment can only be done in the context of the application.

I can't answer the above question because it assumes that capability and dependability are the same types of thing. While capability might be a continuous scale, dependability has minimums that have to be achieved in order to establish acceptable, tolerable or retainable risks. Your previous question doesn't recognise this, so it is not possible for me to answer it meaningfully. I suspect anyone that does answer it, doesn't appreciate dependability enough to provide an answer that is useful to you

Operational effectiveness needs both

They are not in competition, one wants 'robust, dependable capabilities - it is the specific MOS and cross-discipline/capabilities 'disconnects' that are likely to be complex and not be being actually addressed in the Tech Integrity assessments by 'domain experts' that will kill people.

In order to be effective from a capability perspective, the system must demonstrate a level of dependability. In some ways of thinking, dependability could be considered a subset of capability or share the decision on the level of capability a system can hold.

Defence must be capable and dependable. These qualities are equally as important.

I feel that capability is meaningless if it is not dependable. Highly capable systems must be available when required - this is why aviation and naval military applications have multiple levels of redundancy in systems.

Capability and dependability are related, as loss of dependability such as due to circuit failure means in many cases also loss of capability.

Without dependability the complex system will be unable to deliver the intended capability therefore equals no capability. Without correct capability equals no capability. Both aspects must be correctly specified and achieved in the completed system. Note this is regardless of complexity

It is all very well if a missile, for example, works correctly on a test range under controlled conditions, but it MUST work in the operational environment, possibly under unusual or unexpected conditions, first time, every time. There may be no second chance, so that is why both capability and dependability are important

Capability vs dependability always have a healthy tension that needs to be managed in the Defence world. The warfighter demands capability in order to be effective and competitive but always at the expense of being safe and reliable. Therefore equal.

In my opinion this question of relative importance is invalid since it the scope is far too general. What matters more for one mission may be quite different for a different mission. What does matter is that we have a good parametric measure of capability and aanother discrete measure of dependability - both are essential nd each must be compared individually with mission requirements. Depending on a composite (conjoint) parameter is dangerous as it removes the human perception of the meaning of the parameter to some abstract numerical figure. It was this kind of excessive abstraction that caused the sub-prime financial crisi in the USA. I have been influenced in my opinion by experience in the real world of large distributed heterogeneous complex systems where the whole is not purely a sum of the parts

Unhappy to put Capability and Dependability on a linear scale.

It is all well and good having a very capable asset, but if the asset is not reliable or safe then you do not know how to deploy that asset to the best of its abilities and may cause harm in a situation where you require a dependable asset. On the other hand a less capable asset which is highly dependable will at least give you know suprises and strategically you can decide to deploy that asset accordingly and make other measures to make up for that capability with other assets.

In all systems I have delivered including Naval military and civil Transport both Capability and Dependability are equally important.

This trade off isn't able to be consistently resolved within a broad domain. It depends on the platform, capability and the operational situation.

I object most strongly to nbeing forced to make an entry to the table in question 9, wich I regard as a flawed concept for reasons given. The entry in table 9 is NOT my opinion but was necessary to allow me to complete the survey In my opinion this question of relative importance is invalid since it the scope is far too general. What matters more for one mission may be quite different for a different mission. What does matter is that we have a good parametric measure of capability and aanother discrete measure of dependability - both are essential nd each must be compared individually with mission requirements.

There is generally a need to manage variety in complex system development and therefore the trade-off of capability vs dependability is of equal importance.

I actually think the above is flawed and same with the cojoint measurement unless the 2 are combined in some manner. Weighting needs to be considered between capability and dependability for different systems, for same systems in different operational environments, etc.

for the case of an aircraft collision avoidance system, I am not able to see how capability/dependability can be viewed at different levels of importance.

I doubt that capability and dependability can be regarded as 'orthogonal' aspects, as in my experience they may related to one another.

Valid technical integrity definition and structure

The following are the extracts that have been coded as “Valid TI Definition and Structure”.

Using the terms capability and dependability brings a more balanced focus of the two that are buried in the definition of Technical Integrity,

Model based approaches that are quantitative and simple will help inform higher management of the state of a system which otherwise may be had to express to higher management in a way they can gain understanding and meaning.

I think this is a very good approach to the issue.

To be sincere, I had never used the term "Technical Integrity". This is why I read the two white papers in your link to understand the meaning. Of course in my

professional life as an aeronautics complex system developers I have been dealing with all these concepts but not structured in such a way. Good job!

I still maintain that 'safety' is a specific measure of 'integrity' which does not lend itself to be 'rolled up' into other integrity measures

Add Safety in as a separate box and you may find that your proposal works better.

By distilling Technical Integrity down to capability and dependability and associating these attributes with a numerical ranking is an interesting concept - but the criteria for applying the numerical values would need to be closely followed.

It's a challenging topic; however the material presented seems quite relevant to the industry I am in and elevated my level of understanding of Technical Integrity.

The approach mentioned in the reference paper "What is Technical Integrity and How to Measure it" is very well presented and the concepts are definitely worthwhile exploring further.

I feel relatively comfortable viewing technical integrity as a culmination of a range of subordinate, "engineered" aspects of a complex system from the perspective of capability and dependability.

Appendix D Conjoint measurement aggregation: full text extracts referenced in qualitative analysis

This appendix contains the coded extracts[2] described in the body of this thesis. The extracts are presented under the corresponding headings as Qualitative Analysis section of Chapter 6 of the main thesis. Parts of these extracts that have been directly quoted in the main body of the thesis are highlighted.

Relative importance of capability and dependability

Domain of practice or viewpoint

Domain of practice influences

Defence systems practice influences

The following are the extracts that have been coded as "Defence Systems Practice Influences".

> : In military applications there is the risk that the system will be lost due to both the actions of the opposition and the dependability of the system. The ability to defeat the opposition is driven primarily by the capability of the system. Also, highly dependable systems come at a greater cost of acquisition with often a commensurate lesser capability due to the time taken to develop such systems. Often a more capable system can be deployed at a lower cost of ownership using COTS products and a simply throw away the COTS products when they fail.
>
> Safety & dangerous environments influenced military selections.
>
> Military systems tend to be accepted more on the capability presented, even if there is some level of undependability remaining in the system.
>
> Military systems typically have explicitly stated requirements for both capabilities and dependability

[2] Typographical errors in extracts from survey responses are not corrected.

Having served as a user of the complex system (and relied on those systems for remaining alive!) and experienced the impact of poor dependability (often in terms of poor reliability). One of my favourite sayings in this area is "Radar X is the best surface radar in the world - when it works". While you need a capability edge to increase success in modern warfare, that capability edge is useless if it does not work when required.

In military use, capability carries far greater reward in terms of military outcome and capability managers are constantly weighing the risk vs reward consideration when employing forces

The performance of the military systems is in my consideration when rating the relative importance of capability and dependability as this is a deciding factor in gaining an edge in battle

For systems that can kill people, it is more important that they do so only in a well known and controlled manner

In defence you need systems which are equally dependable as well as the required capability. The consequence of a shortfall in either can be catastrophic

Ability to fight and survive in a battle.

Experience in use of COP form C2/C3/C4I systems by ADF teams. There are significant implicit assuptions in utilisation of capabilities by teams in the field; joint, naval, land.

I feel that in a military design that dependability is slightly more important in performance viewpoint as the capablilty is lost without it.

Defence combat and communications systems and industrial monitoring and control systems have relatively high dependability requirements as compared with capability.

In the military environment, there may be a perception that Capability is more important but the requirements of technical integrity related to safety and reliability are now judged of equal importance ie that Dependability is of equal imporrtance in the gretaer Government outcomes perspective.

A capability level is not relevant unless there is some degree of dependability on that assessement. In wartime situations, however, a known capability that will operate as documented when required is more critical to the warfighter.

Capability vs dependability always have a healthy tension that needs to be managed in the Defence world. The warfighter demands capability in order to be effective and competitive but always at the expense of being safe and reliable. Therefore equal.

From a military perspective, capability and dependability go hand-in-hand. In theatre, you ant the best capability you can get; however, it cannot fail (must be dependable).

From a military systems perspective I think that they are inherently risky and therefore providing better capability should be more of a focus to maintain a technical edge

Work in defence where reliability and safety are very important but capability must be limited by space, weight and operator's ability to deal with information while in combat.

There's no point operating military sytems if they don't deliver the intended capability

Capability determines survivability in a threat environment. Whilst dependability is also very important, I lean towards capability as being of higher importance.

Military capability requirements paramount and dependability would be secondary in design selection, so long as selected design meets the acceptable risk threshold of the customer

For military applications, the capability is important to the warfighter but also, the system works as intended when required is mandatory. In a warfighter situation, it could be argued dependability is more important as a known capability that is known to behave as expected is more important than a great capability that may or may not work when required.

it is more important to be able to meet the Capability need (i.e. defend Australia) than say to deploy a non-Capable system that is 100% Dependable.

Defence systems need to be able to achieve both their mission aim as well as being dependable, as the implications usually involve injury or loss of life, or political implications.

I regard dependable access to information and communication as services. They can provide a sustainable advantage when they are highly reliable; enabling coordination eg tactical manoeurve and the ability to Observe, Orient, Decide and Act faster than your enemy.

it is imperetive that the system overall have enough dependbility to insure the safety of those personell. I view that dependabilty as more important

I have been involved in many different types of concept/design to manufacture projects for both civilian and military projects. This has given me the ability to compare many different methods of validating the capabilities of different complex systems.

In military environment the focus is on the mission and achieving the mission objectives

Military systems are required to operate when under threat or the potential of attack. Hence a shortfall in capability can be compensated by change of operational procedures to suit the capability, however a system that fails to operate when in a theatre of conflict may result in damage / injury to equipment or personnel or in the worst case, total loss.

Weapons systems have a limited life expectancy in a contingency. The higher capability they have the more likely it is that the fight can be won. Equally when I call on the system it must be sufficiently dependable to bring to action but assuming that this is achievable I want the highest capability I can employ so as to win

My view point is influenced by experience which suggests that the military rarely acquires exactly the capability they desire, but can develop workarounds to adapt to this. However, it is very difficult to adapt to poor dependability.

Aerospace defence practice influences

The following are the extracts that have been coded as "Aerospace Defence Practice Influences".

the dependability of an airframe is as if not much more important that the capability it projects

My military (combat) aircraft experience dealt with vehicle and major systems that needed to be highly capable and dependable to provide a reasonable probability of a 'win'.

I think context is a huge factor, and different contexts will make for different cases of whether capability or dependability is more important.

An undependable aircraft may be unairworthy and thus rates higher than being mission capability. Safety wins

For military aviation, there are some circumstances where the achievement of the mission (the capability) far outweighs the necessary dependability of the asset. This is a rare occasion though, and I think that on balance, particularly in peacetime the dependability argument will always win.

A large portion of my career has been in Over-the-horizon-Radar. An excellent capability that presents little safety/environmental risk. Systems like this are weited more towards the capability end for delivering TI. Initial career experience was with P-3C Orion aircraft, and although little exposure to aeronautical engineering, clearly getting this right is of high safety importance.

Through my experience in Naval aviation and the necessary interactions with very senior operators nad engineers, I have changed my thinking of the capability/dependability arguments.

In the context of aviation and naval miltary applications, the consequences of system failure can be catastrophic - eg unable to fly or keep afloat and/or unable to defend.

My answer in this one may be construed as sitting on the fence, however flying a military aircraft is inherently dangerous, and is only done because the capability outcomes are very important. However, if you are going to send a plane into harm's way to, say, gather intel, you need to know that those systems are going to work. If they aren't dependable, the capability doesn't exist anyway.

RAAF is more inclined to reduce capability to ensure dependability (ideally they want both) as flight safety and the underpinning technical and operational airworthiness regulations drive this

Civil and miltary aviation from my viewpoint due to the differing mission needs, i.e. civil the principal mission is one of dependability, whereas miltary is about capability of the aircraft in delivering it's mission.

Within the military domain, aviation and submarines would place a high importance on dependability as they are vunerable when failures do arise, whereas surface ships less so and land less again.

Within military aviation, the performance (capabilities) of a platform is the primary reason for employment in various operations. Dependability is important, but accomplishment of the mission is paramount. Within limited high-risk scenarios such as war or emergency, the capabilities of a system will be of paramount concern and over-ride concerns with reliability.

The criticality of dependability on equipment / technology /systems in an area of operations for Land systems and in an aerospace environment is essential to mission success and completion. The capability defeiciencies, if any, can be overcome with tactics and process in many environements.

For time-critical, safety critical functions such as flight controls, a lower level of performance (limitations on flight envelope etc) would be traded for a higher dependability due to the critical impact of failure and the limitations on backups/redundnacy in a timely manner.

Flight test and maritime aviation background also sub- sea operations

Maritime defence practice influences

The following are the extracts that have been coded as "Maritime Defence Practice Influences".

Naval applications, I feel that the capability is preferential to the dependability for there is time to address the dependability with stores and technical support onboard to manage the dependence issues.

Most operations involving naval platforms are conducted during peace time operations where focus is on safe operations - these activities are conducted over extended periods where access to support is limited and dependable operation is valued.

A ship at sea implicitly operates in a hazardous environment which can be unforgiving regardless of distance from shore, it only requires that the environment or situation it finds itself in is such that a catastrophic event occurs. A ship can sink, have catastrophic fire or capsize events either at sea or in harbour, although likelyhood will vary considerably

Serving in the RAN as a Weapons System maintainer, Combat System Maintenance Manager and Head of the Weapons Electrical department (responsible for the upkeep and optimal performance of all weapons and electronics) on a combatant warship.

In the Military Naval domain, the failure of a Capability can have severe consequences to the mission - e.g. not being able to defend against an attack. Accordingly, the viewpoint expresses is based on a 'risk to mission' approach as being of primary importance.

A ship of any kind is useless in its intended operational arena if it is gold plated in it's dependability but has no capability. It is similarly useless if it has gold plated capability and very poor dependability, i.e. breaks down all the time requiring to

often/always be in port. The relative importance of the two can be a little subjective, ultimately it's a matter of judgement by the owner/operator of the ship.

The hazards inherent in the operation of ships and the many detractors from reliability are part of the measure of the actual achievement of military outcome in the realistic total operational environment, not just the Capability measured against the requirements

I feel that capability is meaningless if it is not dependable. Highly capable systems must be available when required - this is why aviation and naval military applications have multiple levels of redundancy in systems.

In all systems I have delivered including Naval military and civil Transport both Capability and Dependability are equally important.

In the context of aviation and naval miltary applications, the consequences of system failure can be catastrophic - eg unable to fly or keep afloat and/or unable to defend.

Within the military domain, aviation and submarines would place a high importance on dependability as they are vunerable when failures do arise, whereas surface ships less so and land less again.

Flight test and maritime aviation background also sub- sea operations

Capability for naval projects has been more important than dependability as there has been availability of spares and trained personnel at sea - it is part of the naval business.

Land defence practice influences

The following are the extracts that have been coded as "Land Defence Practice Influences".

I feel that capability is meaningless if it is not dependable. Highly capable systems must be available when required - this is why aviation and naval military applications have multiple levels of redundancy in systems.

A large portion of my career has been in Over-the-horizon-Radar. An excellent capability that presents little safety/environmental risk. Systems like this are weited more towards the capability end for delivering TI. Initial career experience was with P-3C Orion aircraft, and although little exposure to aeronautical engineering, clearly getting this right is of high safety importance.

in the land environment, there is often greater flexibility/redundancy of systems than enables a greater focus on capability vs dependability.

Joint/C4ISR are generally built with redundancy inherent within the system, hence capability is important.

The criticality of dependability on equipment / technology /systems in an area of operations for Land systems and in an aerospace environment is essential to mission success and completion. The capability defeiciencies, if any, can be overcome with tactics and process in many environements.

Within the military domain, aviation and submarines would place a high importance on dependability as they are vunerable when failures do arise, whereas surface ships less so and land less again.

Civil aerospace practice influences

The following are the extracts that have been coded as "Civil Aerospace Practice Influences".

my work on space vehicles placed a very high importance on dependability & just meeting capability requirements was sufficient.

I was influenced by the consequence of failure. For example the safety of a civilian airliner (a dependability factor) would take a much higher prioirty over its ability to meet its seating capacity, range or fuel consumption (capability factors).

In the aviation domain, the consequences are potentially far more significant if undesired consequences remain in the system and therefore reliability and redundancy are more important.

Reductions in dependability have far-reaching and publicly unacceptable results to the public users of civil aviation complex systems.

Dependability is the driving influence for civil aviation applications given the use of it by the public.

Safety first, in particular in the aviation environment, but capability always. As such, I think they are equally as important to achieve.

Civil and miltary aviation from my viewpoint due to the differing mission needs, i.e. civil the principal mission is one of dependability, whereas miltary is about capability of the aircraft in delivering it's mission.

for the case of an aircraft collision avoidance system, I am not able to see how capability/dependability can be viewed at different levels of importance.

The Aviation environment is less forgiving with higher risk associated with failure.

Civil aviation has airworthiness as the paramount aim - even at the exclusion of payload/range (ie: redundancy is provided at the expense of capability).

Aviation is unique in its extreme balance of dependability (safety) and its capability (advanced technologies). Aviation has minimal maneuverability in its safety factors

Within civil aviation, airworthiness is the dominant concern of all aircraft design, operation and maintenance. An civil aircraft may have to abort its "mission" ie: passenger carriage for the sake of safety.

Other civil practice influences

The following are the extracts that have been coded as "Other Civil Practice Influences".

For civil application, the general public is less tolerant of failures and expects the capability to be present and usable all the time.

In IT systems, the balance is of more concern. Generally, commercial comms systems are somewhat more error-tolerant.

Mining is dominated by safety concerns and predictability of system performance.

Revenue and public perception of performance influenced Civil Teleco selection

failure is not an acceptable outcome from such systems

Experience working with customers and what they consider important to their business.

Defence combat and communications systems and industrial monitoring and control systems have relatively high dependability requirements as compared with capability.

From and IT and Telecomms perspective, technology is fast moving and capability is required to stay ahead of competitors

I have rated nuclear regarding high dependability due to my understanding of the safety aspects of that industry

From other civil applications, safety of the public users of systems is paramount.

In all systems I have delivered including Naval military and civil Transport both Capability and Dependability are equally important.

In space you only get one chance (this then also applies to medical, from the perspective of the individual, and for nuclear from the perception of the community). They have to work, there is no trade off of capability against dependability that would satisfy those involved.

In a commercial environment the enphasis is on capability with more forgiveness to something that is less dependable.

Safety has been the most critical aspect for transportation experience (rail, aircraft etc), hence dependability has been the primary driver.

Role in system viewpoint influences

System developer viewpoint influences

The following are the extracts that have been coded as "System Developer Viewpoint Influences".

From a V&V perspective the customer / stakeholders are predominantly interested in the fulfilment of the specified requirements - which translates as the capability. Dependability, being the areas defined by ILS disciplines often take second place.

I have been involved in many different types of concept/design to manufacture projects for both civilian and military projects. This has given me the ability to

compare many different methods of validating the capabilities of different complex systems.

Capability manager viewpoint influences

The following are the extracts that have been coded as “Capability Manager Viewpoint Influences”.

> In military use, capability carries far greater reward in terms of military outcome and capability managers are constantly weighing the risk vs reward consideration when employing forces

> From a military perspective, capability and dependability go hand-in-hand. In theatre, you ant the best capability you can get; however, it cannot fail (must be dependable).

Acquisition manager viewpoint influences

The following is the extract that has been coded as “Acquisition Manager Viewpoint Influences”.

> Consequences of dependability short falls in naval defence environment are high as customer (DMO) is higly sensitised to the risks professionally they perceive to their jobs.

System operator viewpoint influences

The following are the extracts that have been coded as “System Operator Viewpoint Influences”.

> I have a mix more biased to performace as I was a serving military officer who often had to work with systems that perfromed in a substandard way and perceived teh risk in combat that would result in.

> Many of the systems I worked with although dependable had performance that was significanlty inferior to those that I perceived the potential enemies had. I felt that should a war occur, the ships I served on would be very vunerable as was the USS Stark.

> Having served as a user of the complex system (and relied on those systems for remaining alive!) and experienced the impact of poor dependability (often in terms of poor reliability). One of my favourite sayings in this area is "Radar X is the best surface radar in the world - when it works". While you need a capability edge to increase success in modern warfare, that capability edge is useless if it does not work when required.

> Effort from the capability acquisition organisations was highly focussed on dependability at the expense of rigourously examining performance and putting in place high performance systems.

> Serving in the RAN as a Weapons System maintainer, Combat System Maintenance Manager and Head of the Weapons Electrical department

(responsible for the upkeep and optimal performance of all weapons and electronics) on a combatant warship.

Experience as an operator - vital that the system does what it is supposed to do, but also works when needed, doesn't malfunction, etc etc

Viewpoint is of the end user, not just the ability to design a product. What does the end user need and how often does he need it to work as advertised.

Being the end user

Services provided by the system

Dependable

The following are the extracts that have been coded as “Dependable”.

services that the customer has high dependency on being successful.

Having served as a user of the complex system (and relied on those systems for remaining alive!) and experienced the impact of poor dependability (often in terms of poor reliability). One of my favourite sayings in this area is "Radar X is the best surface radar in the world - when it works". While you need a capability edge to increase success in modern warfare, that capability edge is useless if it does not work when required.

the dependability of an airframe is as if not much more important that the capability it projects

No customer likes surprises. A dependable system is vital for confidence in the operation of a complex system.

Failures in services that are considered to be Safety Critical or Mission Critical would prevent them from deploying the system at all. The dependability of these services is usually far more important than how good those services are.

For military aviation, there are some circumstances where the achievement of the mission (the capability) far outweighs the necessary dependability of the asset. This is a rare occasion though, and I think that on balance, particularly in peacetime the dependability argument will always win.

In space you only get one chance (this then also applies to medical, from the perspective of the individual, and for nuclear from the perception of the community). They have to work, there is no trade off of capability against dependability that would satisfy those involved.

Dependability, (particularly based on safety and reliability) is always a critical factor in the success of a system.

risk to the system for catastrophic consequences

For systems that can kill people, it is more important that they do so only in a well known and controlled manner

Services that have a risk of catastrophic consequences with failure or lead to death would place more importance on dependability. Systems in these areas tend to be not at the leading edge of technology but more about what works and is reliable.

Dependibility is imporatant where safety/cost/reliability is paramount, say transportation

I feel that in a military design that dependability is slightly more important in performance viewpoint as the capablilty is lost without it.

The importance of having full capability is countered by its availability and assurance that it is able to be used safely - for that reason I have leant slightly to dependability.

Generally, I would place Dependability of greater importance than Capability, particularly where there is an overhead (e.g., because the system is remotely deployed) in repair of the system.

In the aviation domain, the consequences are potentially far more significant if undesired consequences remain in the system and therefore reliability and redundancy are more important.

Defence combat and communications systems and industrial monitoring and control systems have relatively high dependability requirements as compared with capability.

Most of the items listed have "safety critical" elements to them meaning that the dependability is more important.

a less capable system that they can really rely on would be far more desirable than a more capable system that often let them down. Predictability is more important

Experience as an operator - vital that the system does what it is supposed to do, but also works when needed, doesn't malfunction, etc etc

failure is not an acceptable outcome from such systems

it is imperetive that the system overall have enough dependbility to insure the safety of those personell. I view that dependabilty as more important

highy dependable system is not necessarily diminished if all capabilities are not delivered (especially if they are based on poorly formed requirements).

The criticality of dependability on equipment / technology /systems in an area of operations for Land systems and in an aerospace environment is essential to mission success and completion. The capability defeiciencies, if any, can be overcome with tactics and process in many environements.

I regard dependable access to information and communication as services. They can provide a sustainable advantage when they are highly reliable; enabling coordination eg tactical manoeurve and the ability to Observe, Orient, Decide and Act faster than your enemy.

Reliability and availability for use frist time every time when in an operational or similar environement.

For critical systems in a public environment (eg transport), dependability tends to be high

Civil aviation has airworthiness as the paramount aim - even at the exclusion of payload/range (ie: redundancy is provided at the expense of capability).

Within the military domain, aviation and submarines would place a high importance on dependability as they are vunerable when failures do arise, whereas surface ships less so and land less again.

RAAF is more inclined to reduce capability to ensure dependability (ideally they want both) as flight safety and the underpinning technical and operational airworthiness regulations drive this

Military systems are required to operate when under threat or the potential of attack. Hence a shortfall in capability can be compensated by change of operational procedures to suit the capability, however a system that fails to operate when in a theatre of conflict may result in damage / injury to equipment or personnel or in the worst case, total loss.

I think the rarity of actual operational need makes dependability important - it has to be available when the moment arives. Being able to operate at the limit of technological capability is a very infrequent requirement.

My view point is influenced by experience which suggests that the military rarely acquires exactly the capability they desire, but can develop workarounds to adapt to this. However, it is very difficult to adapt to poor dependability.

Civil and miltary aviation from my viewpoint due to the differing mission needs, i.e. civil the principal mission is one of dependability, whereas miltary is about capability of the aircraft in delivering it's mission.

Importance of or loss of capability

The following are the extracts that have been coded as “Importance Of or Loss of Capability”.

Many of the systems I worked with although dependable had performance that was significanlty inferior to those that I perceived the potential enemies had. I felt that should a war occur, the ships I served on would be very vunerable as was the USS Stark.

the context of miltary missions being capability based rather than dependability based

it is more important to be able to meet the Capability need (i.e. defend Australia) than say to deploy a non-Capable system that is 100% Dependable.

In a commercial environment the enphasis is on capability with more forgiveness to something that is less dependable.

The performance of the military systems is in my consideration when rating the relative importance of capability and dependability as this is a deciding factor in gaining an edge in battle

There's no point operating military sytems if they don't deliver the intended capability

ICT solutions now days need to be a careful balance of capability and reliability,

I have a mix more biased to performace as I was a serving military officer who often had to work with systems that perfromed in a substandard way and perceived teh risk in combat that would result in.

The subject services or systems are generally Capability centric which is why I've judged Capability as more important

The level of criticality of the system and the consequence of failure

Perceived relative impact (weight) of the two characterisitics on the value / success of the organisations that operate these systems.

some domains have variable levels depending on the circumstances - e.g. military sytems in war vs peactime.

Effort from the capability acquisition organisations was highly focussed on dependability at the expense of rigourously examining performance and putting in place high performance systems.

Ability to fight and survive in a battle.

Capability always tends to drive the market. Dependability tends to be assumed.

In defence you need systems which are equally dependable as well as the required capability. The consequence of a shortfall in either can be catastrophic.

Capability + if safety and performace requirements are critical and you must cope with them whatever it takes

Within military aviation, the performance (capabilities) of a platform is the primary reason for employment in various operations. Dependability is important, but accomplishment of the mission is paramount. Within limited high-risk scenarios such as war or emergency, the capabilities of a system will be of paramount concern and over-ride concerns with reliability.

he two main factors are the potential to impact on public safety and the potential impact of a system failure

Naval applications, I feel that the capability is preferential to the dependability for there is time to address the dependability with stores and technical support onboard to manage the dependence issues.

Military Comms - the ability to communicate is more important vs the extra features Poor performance & errors in a Civilian telco system can have large affects on revenue (e.g. billing systems)

In the Military Naval domain, the failure of a Capability can have severe consequences to the mission - e.g. not being able to defend against an attack. Accordingly, the viewpoint expresses is based on a 'risk to mission' approach as being of primary importance.

In the context of aviation and naval miltary applications, the consequences of system failure can be catastrophic - eg unable to fly or keep afloat and/or unable to defend.

From a military systems perspective I think that they are inherently risky and therefore providing better capability should be more of a focus to maintain a technical edge

Capability is important for gaining battlefield advantage, say sensors

Capability determines survivability in a threat environment. Whilst dependability is also very important, I lean towards capability as being of higher importance.

Military capability requirements paramount and dependability would be secondary in design selection, so long as selected design meets the acceptable risk threshold of the customer

From and IT and Telecomms perspective, technology is fast moving and capability is required to stay ahead of competitors

Weapons systems have a limited life expectancy in a contingency. The higher capability they have the more likely it is that the fight can be won. Equally when I call on the system it must be sufficiently dependable to bring to action but assuming that this is achievable I want the highest capability I can employ so as to win

The Fitness for Purpose of a highly capable system may not be diminished because it is not dependable.

For critical systems in a controlled/contained environment capability tends to be higher.

Capability for naval projects has been more important than dependability as there has been availability of spares and trained personnel at sea - it is part of the naval business.

in the land environment, there is often greater flexibility/redundancy of systems than enables a greater focus on capability vs dependability.

Joint/C4ISR are generally built with redundancy inherent within the system, hence capability is important.

Civil and miltary aviation from my viewpoint due to the differing mission needs, i.e. civil the principal mission is one of dependability, whereas miltary is about capability of the aircraft in delivering it's mission.

Safety

The following are the extracts that have been coded as “Safety”.

Each complex system I have been involved in has emphasized the importance of Capability, Dependability and SAFETY

A catastrophic failure aloft is potentially far more dire than at elevation 0.00 on the sea,

Aviation is unique in its extreme balance of dependability (safety) and its capability (advanced technologies). Aviation has minimal maneuverability in its safety factors

Fail- safe, back-up systems and "man-in-the-loop" operation with ability to "invert" complex systems to basics for continued safe operation in the event of failure of automation is becoming essential for safety of operation across military weapons platforms, public transport systems and nuclear/ space systems alike

An aviation background means that the safety of the systems is of more importance (in a regulatory environment) than the capability.

Most of the items listed provide services that people rely on to be safe, hence the expectation if that dependability is very important. Eg people would not fly in a new feature ridden plane that was not reliable.

Dependability is more critical where human factors are at stake, for instance in the provison of essential services or, in the case of aviation, where a critical failure will probably have catastrophic results

I believe that maintaining the capability (reliability) and ensuring that there no undesirable consequences are more important at in endeavours where there is a significantly higher risk to life (i.e aviation and medical).

If it flys D>C, otherwise C>D

Safety critical systems need to be dependable and definable

the two main factors are the potential to impact on public safety and the potential impact of a system failure

For time-critical, safety critical functions such as flight controls, a lower level of performance (limitations on flight envelope etc) would be traded for a higher dependability due to the critical impact of failure and the limitations on backups/redundnacy in a timely manner.

Within civil aviation, airworthiness is the dominant concern of all aircraft design, operation and maintenance. An civil aircraft may have to abort its "mission" ie: passenger carriage for the sake of safety.

Generally, those that have safety implications should be far more dependable

Three forces are at play: dread (nuclear), controlability (Space) and general impact (rest)

In defence you need systems which are equally dependable as well as the required capability. The consequence of a shortfall in either can be catastrophic.

airworthiness wins over mission

A ship at sea implicitly operates in a hazardous environment which can be unforgiving regardless of distance from shore, it only requires that the environment or situation it finds itself in is such that a catastrophic event occurs. A ship can sink, have catastrophic fire or capsize events either at sea or in harbour, although likelyhood will vary considerably

Consequence of the service not being provided on personal safety

The importance of having full capability is countered by its availability and assurance that it is able to be used safely - for that reason I have leant slightly to dependability.

it is imperetive that the system overall have enough dependbility to insure the safety of those personell. I view that dependabilty as more important

Civil systems need to be more dependable in order to protect people and the environment more so than achieving their capability aims

Reductions in dependability have far-reaching and publicly unacceptable results to the public users of civil aviation complex systems.

falling out of the sky vs sinking

Unanticipated behavior can result in the loss of life.

Dependability is more important where the consequences of failure are life threatening or would lead to a catastrophe, ie safety critical systems

Services that are integral to personnel safety is a strong factor

the consequences of being undependable in commercial aviation are catestrophic for example

When you are airborne there is less time to react and safety can take on more importance

There is a specific focus of safety and the need to demonstrate the assurance of safety so far as is reasonably practicable - the 'benefits' of capability and dependability form the context for such justifications (I therefore have a great issue with rolling safety into dependability)

Dependability is the driving influence for civil aviation applications given the use of it by the public.

Safety first, in particular in the aviation environment, but capability always. As such, I think they are equally as important to achieve.

In the aviation domain, the consequences are potentially far more significant if undesired consequences remain in the system and therefore reliability and redundancy are more important.

Most of the items listed have "safety critical" elements to them meaning that the dependability is more important.

Most operations involving naval platforms are conducted during peace time operations where focus is on safe operations - these activities are conducted over extended periods where access to support is limited and dependable operation is valued.

The consesquences of unexpected or undesired behaviour especially with regard to major personal/population safety risks, should be considered first in all domains. In no domain did I feel that capability is more important.

An undependable aircraft may be unairworthy and thus rates higher than being mission capability. Safety wins

I have rated nuclear regarding high dependability due to my understanding of the safety aspects of that industry

From other civil applications, safety of the public users of systems is paramount.

I was influenced by the consequence of failure. For example the safety of a civilian airliner (a dependability factor) would take a much higher prioirty over its ability to meet its seating capacity, range or fuel consumption (capability factors).

I still maintain that 'safety' is a specific measure of 'integrity' which does not lend itself to be 'rolled up' into other integrity measures

Add Safety in as a separate box and you may find that your proposal works better.

Safety has been the most critical aspect for transportation experience (rail, aircraft etc), hence dependability has been the primary driver.

Support concept

The following are the extracts that have been coded as “Support Concept”.

Military systems can afford to be at the cutting edge of Capability as they are often bundled with highly capable maintenance staff.

Operating remote from support services

ILS

Electronic systems for example works these days for several years without the need for repair. The again this raises another problem, there is no-one is around that remembers how to repair the bits when they start failing.

The ability of the system to be maintained during a mission. Or conversely, if there is no maintenance possible, the consequences of poor dependability.

Naval applications, I feel that the capability is preferential to the dependability for there is time to address the dependability with stores and technical support onboard to manage the dependence issues.

Generally, I would place Dependability of greater importance than Capability, particularly where there is an overhead (e.g., because the system is remotely deployed) in repair of the system.

Capability for naval projects has been more important than dependability as there has been availability of spares and trained personnel at sea - it is part of the naval business.

Appendix E Judgment of technical integrity and associated indicators: full text extracts referenced in qualitative analysis

This appendix contains the coded extracts[3] described in the body of this thesis. The extracts are presented under the corresponding headings as Qualitative Analysis section of Chapter 7 of the main thesis. Parts of these extracts that have been directly quoted in the main body of the thesis are highlighted.

Nature of technical integrity measurement

Subjective

I think that is has reinforced the thought that TI is very subjective and is largely based on the experiences of the ""judge"" more than simple compliance - as nothing complex is normally fully compliant and how the ""shades of grey"" are interpreted is quite subjective.

In my opinion it remains a quantitative measure of a system's suitability against criteria for Fitness for Service; Personnel Safety, impact to the Mission; and impact on the environment. Whilst it is a quantitative measure, there remains a level of subjectivity based on the experiences of the assessor as to where it sits on the quantitative scale.

Survey caused a thought process change to force a value for subjective quality assessments for systems with different capabilities and specifications.

as with any complex system there are trade-offs to be made, and accordingly the survey has highlighted that may also be the case when it comes to the three arms of technical integrity - safety, fitness for services, and environmental compliance. I think it has illustrated that because of these many potential trade-offs, an overall 'score' for technical integrity may be somewhat subjective.

the concept of quantifying technical integrity is an interesting one, however I believe it will always embody an element of subjectivity.

there is a subjective element (or a difficult to define complex of objective elements) that can only be measured through a thorough understanding of the

[3] Typographical errors in extracts from survey responses are not corrected.

system and the manner in which it was produced. Technical integrity is built into the system rather than something to be measured on completion of the system.

This really needs testing on many other examples of truly comparable systems- one of these (System A) was clearly a lot older, more robust (designed for Warships) heavier and prone to failure more often; whilst System B was current COTS technology adapted for Patrol Boats- but not designed for air surveillance (less capable- therefore not comparable from the outset?). Systems A & C had the same Primary mission (surface & air surveillance) and C was also heavy with high power demand. Nevertheless I felt intuitively biased toward the more modern COTS system B on many scores and thus, although it had no air surveillance capability I feel it had the highest technical integrity.

Ambiguous

I'm still confused to the point where I still could not give a definition that I would agree and accept on close scrutiny.

I still feel it is an ambiguous term.

Just made it fuzzier.

Quality factor correspondence

It was not possible to conduct true technical integrity analysis in the available time so I changed technical integrity to mean the same as capablity, Fitness for Service and Effectiveness. They all scored the same.

it was interesting to think about the terms used in the survey and what the differences between the terminology and how words with slightly different meanings may be assessed similarly.

Equated it more to requirements compliance than I previously thought

The challenge remains with defining what are the components of 'technical integrity' and how to objectively measure each. The separate criteria used in this latest 'survey' provided a good spread of useful criteria/indicators - well done.

Some attrributes are not directly proportional to TI. For example high technical complexity does not necessarily contribute to high technical integrity

Aggregation

It still remains a multi orthogonal collection of parameters that I am unable to reconcile down to an enumeration (e.g., high .. low).

The one point that comes from this is that many characteristics of the product contribute to the assessment of technical integrity,. They do however have different weightings of imporatnce that would somehow have to be taken into account.

A complex balancing act that really needs an algorithmic method of scoring technical attributes of competing systems

I am sure that you could develop an algorithmic (weighted) scoring system to integrate an overall score from the sum of sub-system scores.

Use of objective data

Strongly reliant

I relied very heavily on objective data where appropriate data was provided. It makes it very easy to rank capabilities. In some cases I would have liked to conduct more analysis to enable easier comparison of data.

I had the spec sheets in front of me the whole time eg I tried to look at efficiency in terms of power use, weight, on board spares

About half I looked for specific technical data, others ended up weighting two or three togther.

In order to provide a more comparative answers to the question provided, the systems specs had to be used as source data reference.

Tried to predominantly use objective data as opposed to overall interpretation and assumptions of the radars to make judgments.

If I was completing this survey as a proper engineering task I would spend the time to ensure I was using objective evidence and a complete understanding of each technical area to make my comparisons.

I relied heavily on the data provided, in particular the system's compliance combined with system features for the topic questions.

Approximately 70% of the time I relied on objective data provided for the systems to make my judgements.

I gave greater weighting to the systems which could provide OQE to standards. The higher the standard the better the score. This could be overkill if a high level of environmental compliance is not required.

I used all the available data to attempt my judgements

I read first the systems description and then use the computer to compare the three before answering the questions

I tried to make scores based on the information provided, but for some there was insufficient data, or unclear definitions of the term to be rated

I used the objective data as much as possible

Heavily. Objective data to compare specifications is needed to select measurable comparison. While empirical value was not important, its relative value to the other system is important.

About 80%.

Extensively

For each criteria I compared the characteristics of the 3 systems that supported the criteria under assessment. If this approach was not taken then it would, in my opinion, be very difficult to firstly justify the assessment or come back at a later

stage and repeat the assessment consistently. This is possibly a statement against the conservative nature of an engineer.

Approximately 70% of the time

most of my assessment were based strongly on evaluation of the data provided. I resorted to my prior experience and knowledge only when the data provided was insufficient

Most questions were supported by specific data groups (which provided the objective data). The decision to use specific data groups to respond to a question was born out of experience (and I've probably forgotten that 'I know that I know' - I just know).

Where the data was provided I tried to judge based on that data - even though my own ""instincts"" may have been different.

Totally as there was ample information on the systems.

Ojective data used as first port of call to establish criteria ranking. Where available, objective data used as primary indicator.

As much as possible albeit hampered by a lack of deep domain knowle3dge

Reliant but modulated

Context

I aimed to be able to point to at least some 'objective' data to be able to justify each decision. In addition, where data was conflicting or not conclusive, I used my general experience and knowledge in the context of likely operational uses to form an opinion.

I largely relied on objective data to make my assessments, although I also tried to go beyond purely the objective data to try and put assessment criteria into context/perspective.

Subjective data

Given that I am not a radar expert, I could only use the objective data in a relative sense, so my subjective definitions reflected a lack of absolutes.

The objective data is used to set the initial ratings, which are then moderated by considering the subjective data.

Experience

Usually two of three very close based on objective evidence and judgement had to rely on external knowlege outside the data presented.

I aimed to be able to point to at least some 'objective' data to be able to justify each decision. In addition, where data was conflicting or not conclusive, I used my general experience and knowledge in the context of likely operational uses to form an opinion.

I used the data available to the extent I thought it relevant. Where no data was available I relied on experience with similar systems.

most of my assessment were based strongly on evaluation of the data provided. I resorted to my prior experience and knowledge only when the data provided was insufficient

Most questions were supported by specific data groups (which provided the objective data). The decision to use specific data groups to respond to a question was born out of experience (and I've probably forgotten that 'I know that I know' - I just know).

in some cases I did use some objectivity - my experience is that MTBFs for commercial equipment (used in a Naval environment) are often far more optimistic (and usually from OEM calcs) than MIL.

Perhaps 75% - a lot of judgment calls are required.

As much as possible albeit hampered by a lack of deep domain knowle3dge

Other

The data sheets provided much, but not all, of the required information to help inform my judgements on each criteria

Some of the assessments were easier than others where specific objective data was provided that addressed the question. In just about all cases objective data was used to some extent - one or two might have been the exception.

I relied on objective data wherever possible, but had to rely on my own limited understanding for some areas.

about 50%

Difficult to rank the data provided as the portion missing was the value judgement weightings. Found it difficult to trade off criteria without the relative weights normally available when assessing relative values.

I found it difficult to justify using the same objective data to justify the rankings for different terms. Sometimes I spent time searching for another piece of objective data that I could use to differentiate the terms

Scored each comparative capability but often influenced by MTBF and MTTR

Environemental impact is difficult without a full lifecycle analysis of the development of the kit so I put them all close together but simply relied on the Hazmat declaration to separate them (but in reality I suspect construction techniques would make A easier to dismoantle and re-use, and given its weight would have a more substantial scrap value.

Combines data

About half I looked for specific technical data, others ended up weighting two or three togther.

Relied on a combination of the Specified Requirements compliance and the system characteristics.

In all cases I refered to the data provided for each system and factored this into my response.

Assessments were based on performance criteria, descriptive data and compliance with Mil-Stds

Environemental impact is difficult without a full lifecycle analysis of the development of the kit so I put them all close together but simply relied on the Hazmat declaration to separate them (but in reality I suspect construction techniques would make A easier to dismoantle and re-use, and given its weight would have a more substantial scrap value.

No objective data approach

Rate all equivalently

My judgements were based on aligning what I considered the relevant characteristics of the systems were that related to the judgement category. Very often where there was little to go on the rankings were clustered together, because we were not allowe to not differentiate.

In some cases I just ranked them all mid-range, and as close as possible to the same. In other cases, I used my best guess about what characteristics products of that type have.

The pictures actually provided me with a ""feeling"" about the system and that may have influenced the judging. Generally I tried to find at least one objective data item that allowed a justification of the judgment. If one system did not provide the objective data, then the distance from the others was not large as I assumed that they would be similar

Subjective evaluation

Context and use

Judgement was made on anticipated use of system based on requirements specified and experience of navigation/surface search radars; Access to standards to allow comparison of environmental factors; Experience with similar solutions that offer return to OEM repairs. This was the most difficult part.

Where there is little or no objective data, I tended to: 1. consider the system in the context of it's likely operational use scenarios; 2. aim to identify at least some objective data to justify the decision; and 3. use my experience and knowledge based as a 'sanity check'.

Where there was little or no objective data provided, judgements were made based on my understanding of the operational environment and the way the system is used operationally, taking into consideration the different attributes that may confer tactical and operational flexibility.

Experience

Where objective data was not available, assessment was based on prior experience, implict ""standards"" (e.g. age for technology, mil vs civilian for robustness) were used. Such assessments could not be justified; the rankings are opinions, and unlikely to be coherent. Given that the survey prohibited equal rankings, ""I don't know"" was excluded from the response possibilities. I had to make a decision, no matter how indefensible it was.

I used my limited knowledge of the 'real world' radars similar to your examples to assist my judgements.

With difficulty! Essentially through experience and analogy.

In areas where I felt I had experience and knowledge I had no problems making subjective judgements.

About 20% towards the end, I started to use gut feel rather than looking for specific data items

Judgement was made on anticipated use of system based on requirements specified and experience of navigation/surface search radars; Access to standards to allow comparison of environmental factors; Experience with similar solutions that offer return to OEM repairs. This was the most difficult part.

Relied on experience of similar systems capabilities and limitation

Where there is little or no objective data, I tended to: 1. consider the system in the context of it's likely operational use scenarios; 2. aim to identify at least some objective data to justify the decision; and 3. use my experience and knowledge based as a 'sanity check'.

I invariably fell back to experiences throughout my career where I had worked with similar systems and the results at the time.

Some judgements were made based on prior experience with the various technologies utlized in the three systems.

Based assessments on equivalent experience compared to implied performance / capabilities

Guesswork based on experience.

Where objective data was not available, ranking is made by experience and trust in previously utilsed systems encountered in my career, and experiences, results, and trust of these items and their performance. Again, this is entirely subjective, and not coherent.

Gut feel and experience

Radar A was older and has being used in larger ships and on what data was available constructed to Mil specs for the intended primary purposes. The lower technology level (but more mature technology) along with less 'bells and whistles' suggested a system that could and would be well understood. So where defintions were unclear (I'm not sure how to define durability and dependability for such systems) I went with the A is better than C is better than B approach.

Assumptions

Judgements were based on assumptions given the data provided in such scenarios.

Somewhat frustratingly, which I suppose can be reflective of reality. Best guess based on assumption or qualitative extrapolation of what I would expect from the systems.

Practical cosideration of likely performance

In some cases I just ranked them all mid-range, and as close as possible to the same. In other cases, I used my best guess about what characteristics products of that type have.

Based assessments on equivalent experience compared to implied performance / capabilities

Judgements were augmented by other data provided. For example reliability is not only influenced by MTBF but by repair concept, build standard, provenance etc. In addition factor was the previous experience with each of the difference genres of radar (""old"" MIL, commercial and ""current"" MIL) which provided assumptions (correct or otherwise) on likely performance/reliability.

Technology

Where objective data was not available, assessment was based on prior experience, implict ""standards"" (e.g. age for technology, mil vs civilian for robustness) were used. Such assessments could not be justified; the rankings are opinions, and unlikely to be coherent. Given that the survey prohibited equal rankings, ""I don't know"" was excluded from the response possibilities. I had to make a decision, no matter how indefensible it was.

Tried to interpolate-eg guessed, sorry judged, that a colour display may be more usable than monochrome

Civil versus military applications and Year of EiS (Entry into Service)

Radar A is older which I inherently associate with low technology, simple to understand, likely to break but simple to fix. Radar B seemed to be the highest technology. The need to go back to the OEM for repair cost it a lot of points multiple times. Radar C's fixed frequency and fixed PRFs implied a lack of flexibility

Impressions influenced decisions eg. how recent the technology and implementation suggested higher quality (given modern quality systems).

Using the relative ages of the technologies employed.

Tend to use the age of the product (year of first deployed) to judge age of technology. Also relied on nuber of units in service (even though no evidence on how well those units were performing)

Some judgements were made based on prior experience with the various technologies utlized in the three systems.

Judgements were augmented by other data provided. For example reliability is not only influenced by MTBF but by repair concept, build standard, provenance etc. In addition factor was the previous experience with each of the difference genres of radar (""old"" MIL, commercial and ""current"" MIL) which provided assumptions (correct or otherwise) on likely performance/reliability.

On comparative range and accuracy performance and fitness for purpose balanced against age of technology

Radar A was older and has being used in larger ships and on what data was available constructed to Mil specs for the intended primary purposes. The lower technology level (but more mature technology) along with less 'bells and whistles' suggested a system that could and would be well understood. So where defintions were unclear (I'm not sure how to define durability and dependability for such systems) I went with the A is better than C is better than B approach.

Risk

One of the aspects considered was envisaging the question application in the real world, supported by base knowledge of system operation with some assuptions including risk based conclusions.

This presents a higher risk as you then need to treat it as if it cannot meet a standard until it is proven otherwise.

Use related data

Collective assessment of related data provided was the only means to make an assessment. In a few cases there was very little to go on and relative ratings could have been different.

Somewhat frustratingly, which I suppose can be reflective of reality. Best guess based on assumption or qualitative extrapolation of what I would expect from the systems.

I tended to favour System B on the basis that it most closely met the requirements

Civil versus military applications and Year of EiS (Entry into Service)

Where there is little or no objective data, I tended to: 1. consider the system in the context of it's likely operational use scenarios; 2. aim to identify at least some objective data to justify the decision; and 3. use my experience and knowledge based as a 'sanity check'.

Tried to combine a number of objective data to that was similar to the judgement topic

Radar A is older which I inherently associate with low technology, simple to understand, likely to break but simple to fix. Radar B seemed to be the highest technology. The need to go back to the OEM for repair cost it a lot of points multiple times. Radar C's fixed frequency and fixed PRFs implied a lack of flexibility

My judgements were based on aligning what I considered the relevant characteristics of the systems were that related to the judgement category. Very often where there was little to go on the rankings were clustered together, because we were not allowe to not differentiate.

I gave higher weightings to specific data groups depending on the question (e.g., reliability I used the MTBF information almost to the exclusion of other data available). This was because I wanted to finish the questionnaire in a reasonable time. Other aspects required that more data groups be compared and these took longer to respond to the base question.

The pictures actually provided me with a ""feeling"" about the system and that may have influenced the judging. Generally I tried to find at least one objective data item that allowed a justification of the judgment. If one system did not provide the objective data, then the distance from the others was not large as I assumed that they would be similar

Judgements were augmented by other data provided. For example reliability is not only influenced by MTBF but by repair concept, build standard, provenance etc. In addition factor was the previous experience with each of the difference genres of radar (""old"" MIL, commercial and ""current"" MIL) which provided assumptions (correct or otherwise) on likely performance/reliability.

On comparative range and accuracy performance and fitness for purpose balanced against age of technology

Would not rate

Where objective data was not available, assessment was based on prior experience, implict ""standards"" (e.g. age for technology, mil vs civilian for robustness) were used. Such assessments could not be justified; the rankings are opinions, and unlikely to be coherent. Given that the survey prohibited equal rankings, ""I don't know"" was excluded from the response possibilities. I had to make a decision, no matter how indefensible it was.

I felt a bit uncomfortable about areas where there was little of no objective data, particularly areas that are not my expertise, like operation and flexibility of the system. If I was doing this for real I would consult subject matter experts for more appropriate judgements.

Making judgements agains subjective criteria was difficult. There was little objective data provided and this raised a lot of quiestions for me about context and interpretation of the assessment criteria.

Views on obtaining better ratings

Definitions

I did struggle in thinking about the differences between: Effectiveness vs Capability,Dependability vs Reliability vs Durability,and Technical Complexity vs High Technology

some terms / criteria are used in every day speech but will probably have different meanings to each respondent. For example, ""Durable"" was understood but was found to be difficult to assess.

Some of the terms need a definition or a context (durabilty for instance could be durability of technology/concept which would put A as high; in fact A could be durable but now obsolete, or it could refer to the physical durability of the system which comes down to build quality). With so many similar, undefined characteristics there is a tendency to revert to the behaviour described in my earlier comments above.

Group judgments

It would beneift from a group approach so that domain knowledge shrotfalls could be compensatede for. At the end of group discussion each member could then marked independently to allow for divergences of view to be catered for.

System data

On some of the criteria questioned/used, I was not able to make a definitive judgement as more information (eg on maintainability and supportability) would be required 'in the real world'.

The intended operational context of the radar might allow for better assessment?

Very difficult to make subjective decisions without knowing the purpose of the system. Some idea can be assumed from the requirements but still better to know its intended purpose and environment.

I think the missing piece of the puzzle is the price (i.e. market value, in dollars). Knowing the relative price of the three radars would have helped answer a number of the questions, especially some of the ones on capability and effectiveness. Extending that line of logic, if I wanted to check the level of techncial integrity of something I would look to the market price. How do I compare the value of an iPhone to a pencil? A house to an education? A horse to a bomb? Any one thing to something else? The answer will change depending on the customer, but that only reflects reality, I think. The challenge in your case is to work our the market value of something that has only one 'customer'.

Rating scales

It was difficult to determine how to score the assessments within the range. I placed most towards the upper end. Relative scores are perhaps more important than where they are placed on the scale.

Should ask about availability.

I felt the scale used was too broad for the nature of the questions

In several cases, because of a lack of data, I would have liked to rate 2 systems as equal. 16 Possible 'grades' seems quite a large span.

Your survey form did not accept ""a tie"" between two or more systems, thereby forcing differences between the evaluations of the three systems. I assume this was deliberate, but in one case, my evaluation for a question would have been A:B:C = 5:9:9 , but I was forced to submit an evaluation of A:B:C = 5:9:10 or 5:10:9

Use of a 16 value selection was a poor decision, should have used Likert's recommendation for five or seven values. The data was a bit intimidating, my responses were gut feelings after reading the specifications a few times - it felt like an IQ test.

I think the breadth of the scales used was probably too broad given the lack of information provided. While not an issue on each individual assessment, it may skew an overall assessment. Sometimes (usually) too much choice is counterproductive?

Appendix F The impact of technical regulation on the technical integrity of complex engineered systems

Peer reviewed paper in INCOSE 2009, International Council On Systems Engineering: Singapore.

This paper presented the broad background to the research problem and explicated particular research questions. It presented initial results of a literature review aimed at identifying relevant approaches to the measurement of technical integrity.

The Impact of Technical Regulation on the Technical Integrity of Complex Engineered Systems

Michael Edwards
Defence and Systems Institute, University of South Australia / Raytheon Australia
Michael.Edwards@postgrads.unisa.edu.au / medwards@raytheon.com.au

Copyright © 2009 by Michael Edwards. Published and used by INCOSE with permission.

Abstract. The explicit aim of the application of technical regulation is to ensure that the requisite Technical Integrity of the developed system is achieved. Technical Integrity encompasses fitness for service, system safety and minimised environmental impact. The beginnings of a program of research to test the hypothesis that "The application of technical regulation yields improved technical integrity of the complex engineered system" are established. It is shown that only the intended emergent properties of a complex engineered system and its unintended hazards are directly addressed by technical regulation. A candidate set of measures to assess technical integrity is proposed. The methods that may enhance technical integrity of complex systems are reviewed. The degrees to which technical regulatory frameworks actually enforce methods known to have a contribution to technical integrity are also reviewed resulting in a list of corollary questions to be considered by future research.

Introduction

"Technical Integrity" is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection, i.e. its compliance with technical standards and the minimisation of unintended consequences. The explicit aim of the application of technical regulation is to ensure that the requisite technical integrity of the system being developed is achieved. The implicit assumption is that in the absence of any technical regulation the requisite technical integrity may not be achieved.

Current and future technological systems are steadily increasing in complexity, both in terms of technology employed and the interactions they have with users and other stakeholders. Defence systems in particular are increasingly required to not only introduce higher organic capabilities within the system itself, but also to integrate into systems of systems to enable capabilities with higher performance across a wider spectrum of domains and environments. At the same time, the need for these systems to achieve ever increasing levels of technical integrity is paramount. Systems are expensive to develop and sustain, therefore there are less of them and each one that is realised must achieve higher and higher levels of availability and dependability. The stakeholders of the system, not least the general public (and consequently the political leaders of the nation) are less and less willing to accept mishaps – be they accidents to personnel, loss or damage of equipment or impacts to the natural environment.

Technical regulation seeks to provide more assurance to the acquirers and ultimately the users of engineered systems than just a reliance on system developers to unilaterally apply the "best" engineering and management processes. "Best" for a developer does not necessarily lead to a set of practices that will yield the required technical integrity. Engineering practices, like Systems Engineering for instance, do not alone assure technical integrity goals are a primary focus or that the processes will be executed by competent professionals with explicit responsibility for their design decisions and the system outcomes. Technical regulation

therefore generally attempts to assure that development is conducted to approved standards, by competent individuals in an organisation that is authorised to do so. There is also an explicit process to achieve certification; at a minimum by the developer formally attesting to the completeness and quality of its own work, but usually also encompassing some independent validation of the developer's claims of compliance with technical standards and processes.

It would be expected then, in some average sense at least, that the technical integrity of systems developed by an organisation outside an explicit technical regulatory system would be less than that developed inside a technical regulatory system. Indeed, this is accepted as a truism in the context of complex engineered systems for the Australian Defence Force (ADF) where acquisition and development under a Technical Regulatory Framework is now mandated for all procurements of defence materiel (ADF 2002).

The author is undertaking a program of research to test the hypothesis that "The application of technical regulation yields improved technical integrity of the complex engineered system". In particular the research intends to primarily consider the acquisition and development of complex engineered systems in the ADF context.

This paper defines a framework in which this research is intended to be conducted. It outlines the scope of technical regulation to be examined; it defines technical integrity and to establish limits to the research, defines the range of complex engineered systems of interest and current considerations in the development of these in the ADF context. Some initial research questions are posed and explored including the relationship of the emergent properties of complex engineered systems with technical regulation, it considers what set of measures may be appropriate for determining the technical integrity of a complex engineered system.. It also examines what engineering methods and related factors have been shown to or at least accepted as contributing to enhanced technical integrity of complex engineered systems. Finally it begins to establish the extent to which technical regulation enforces these methods.

Background

What is Technical Regulation?

In the ADF context, Technical Regulation is the means employed to assure technical integrity of materiel. The principles of technical integrity embodied in DI(G) LOG 08-15 (ADF 2002) are illustrated in Figure 1.

As a point of comparison, the development of civil aircraft and their systems are also governed by technical regulation. Airworthiness certification, as it is called by most national regulatory bodies is "the systematic process, during the design of an aircraft or airborne system of demonstrating conformance to a set of specific and predetermined airworthiness regulations (eg FAR 25) for a specific type and category of aircraft." (Kritzinger 2006).

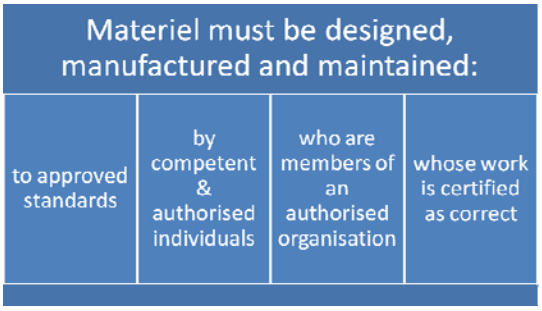


Figure 1. Principles of ADF Technical Regulation

What is technical integrity?

Technical Integrity is "An item's fitness for service, safety and compliance with regulations for environmental protection" (ADF 2002). The elements of technical integrity are shown in Figure 2 and are defined as:

- Fitness for service. "The materiel's ability to satisfy operational requirements. Hence it is a subset of technical integrity" (ADF 2002).
- Safety: "Freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment". (DoD 2000, MIL-STD-882D)
- Environment Protection: Poses no hazard to the environment.



Figure 2. The elements of technical integrity.

Influence of Technical Regulation on Technical Integrity

Technical Regulation is established as a means to assure technical integrity of developed systems and materiel. Figure 3 provides an influence diagram that summarises the key concepts. A technical integrity gap exists where the technical integrity being achieved by systems is less than that desired. It may be that extant systems have some performance or effectiveness (fitness for service) shortfalls, or have incurred accidents or environmental impacts that are beyond what can be tolerated by the users and stakeholders of the system. This gap influences the adoption of technical regulation as a means of achieving the requisite technical integrity. Technical regulation in an ADF context, requires that approved specifications and standards be in place for development, that development be conducted by competent individuals, working in authorized organizations and that all work is certified to be correct

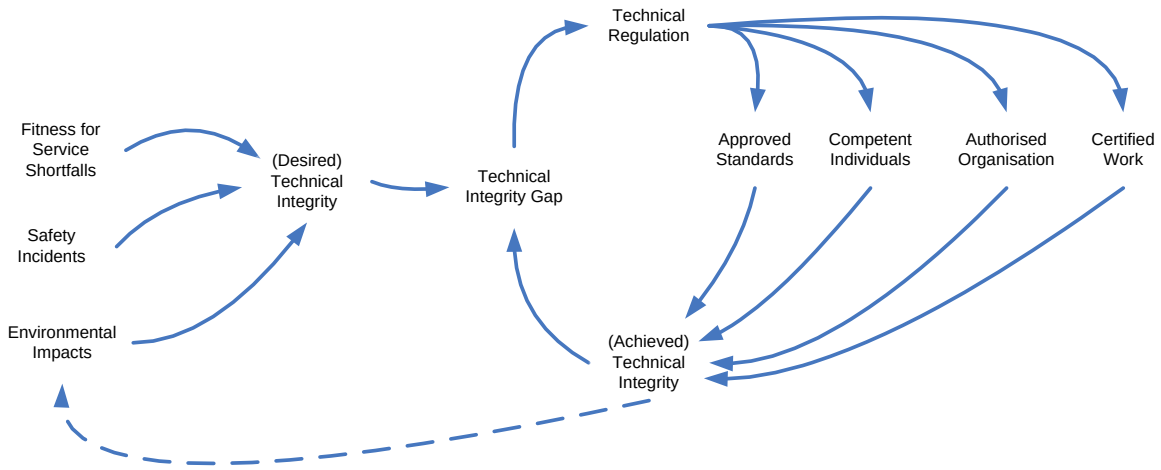


Figure 3. The influence of technical regulation on technical integrity.

Technical regulation is actualised by certifications which attest to the implementation of the regulations:

- The specifications developed to define the acquisition project must be certified as complying with the required standards (the list of required standards is usual maintained by a regulatory body). This certification is normally provided by the regulator (e.g. by endorsement of the specification and the plan for certification of the system).
- Secondly, the supplier must be certified as being competent to undertake the design, development and production of the system being acquired.
- Thirdly, the supplier certifies that delivered system has achieved the entire technical integrity requirement embodied in the specification.
- For significant mission and/or safety critical systems where the risks to technical integrity are reasonably high, a certification is generally required, usually by a party independent of the supplier, that the compliance claimed by the supplier has been verified (ie to provide assurance that the supplier's claim of compliance is valid).

Complex Engineered System Defined

The work of (Magee and de Weck 2002) attempts to classify a complex system as "a system with numerous components and interconnections, interactions or interdependencies that are difficult to describe, understand predict, manage, design and/or change." They also define an engineering system as "a system designed by humans having some purpose". Engineering systems are further classified as those with designed function that can transform, transport, store, exchange or control operands including matter, energy, information and value.

(DeRosa et al. 2008) assert that the essence of a complexity in engineering of complex systems is interdependence. Interdependence is where a reduction of components into a system's constituent parts is not alone sufficient to understand its behaviour. Behaviour is an emergent property which changes as elements of the system are added, removed or rearranged. A system can be designed for intended emergent properties, but it is very likely that some unintended properties will also be created.

Of particular interest, is the definition of complex engineered systems that are part technology and part human – termed composite-information/decision/action (Composite-

IDA) systems by (Hitchins 2003): “such systems operate at high stress levels, making great demands on operators and technologies not only for performance, but also for integrity and reliability in decision making”. In an ADF context, most Command and Control and weapons systems, which are of primary interest to the current research, fit well into this definition.

For the purpose of this paper then, a complex engineered system is defined to be a system that has been intentionally engineered from numerous components (technological (hardware and software) as well as human), each of which has several interconnections and is interdependent, yielding a system with both intended and unintended emergent behaviour. The primary interest therefore will be to determine measures for these emergent properties which can be considered to contribute to the technical integrity of the complex engineered system.

Complex Engineered Systems - Considerations in the ADF Context

The types of complex engineered systems of interest to this research include examples in the ADF like warship Combat Systems, airborne command and control, sensor and weapon systems, and ground based command and control systems for single service and joint operations. These may be compared and contrasted to similar systems in other services and complex systems like aviation systems, space systems and nuclear power systems in the civil sector.

As systems become bigger and more complex it is less feasible to design them from scratch. Rather, the current approach in the ADF context is to build and integrate systems from COTS/MOTS components in order to meet the objectives for the system. The system solution is therefore constrained by the MOTS components – not all initial objectives for the system may be met. But this is contrasted to a developmental solution where the attainment of system objectives may be constrained by the development cost and risk. Integration of a system from MOTS/COTS then, should be done with customer willing to be flexible in the requirements of the overall system - the vendors of the MOTS/COTS components are really driving a large fraction of the function, performance and technical integrity characteristics of the system.

Difficulties with designing and integrating the emergent behaviour and integrity of this type of system need to be addressed. These include the OTS design being relatively unknown and therefore difficult to establish its behaviour, faults and failure modes; addressing incompatibilities between interfacing products; products evolving over the systems lifecycle which may be beneficial or may become prematurely unsupportable.

How well does the technical regulatory system deal with these current trends in the development of more complex systems and systems of systems which increasingly are constrained by a philosophy of procuring MOTS components to reduce development risk?

Initial Questions and Concepts

What is the Relationship of Complex Engineered System Emergent Properties and Technical Integrity?

(Honour 2007) nominated a categorisation of emergent properties in two dimensions. Firstly “Designed” (ie intended emergent properties of the system) and “Surprise” (the potentially unintended emergent properties of the system). The second dimension encompasses “Useful” (desirable emergent properties), “Neutral” (neither desired nor undesired properties of the system) and “destructive” (undesirable properties of the system).emergent properties.

This framework has been extended in Table 1 to examine the relationship of each of the six categories of emergent properties with the extent to which technical regulation address the property.

Table 1. Categorisation of system emergent properties and the degree to which they are addressed by technical regulation.

EMERGENT PROPERTIES	**Useful**	**Neutral**	**Destructive**
Designed (ie intended)	Capability	Facts of Design	Accepted Trade-Off
Addressed by Technical Regulation	*Compliance – assuring that all desired capabilities are achieved.*	*Design is documented, understood and managed.*	*Risks to technical integrity are explicitly identified, minimised and residual risk accepted.*
Surprise (ie unintended)	Exploitable Feature	Facts of Existence	Fearful features (unanticipated hazards)
Addressed by Technical Regulation	*Not addressed*	*Not addressed.*	*Intent is that these are minimised or at least moved into an Accepted Trade-Off.*

Technical regulation clearly address the "Designed" range of emergent properties that a complex engineered system can exhibit. Capability is directly assured by efforts to assure compliance with requirements. Facts of Design: a focus on controlled engineering efforts, in particular configuration management and the adoption of systems engineering paradigms ensure the design is documented, understood and managed. Accepted Trade-Off: a technical risk management approach and in particular a focus on system safety engineering ensure that potential compromises to technical integrity are identified and made explicitly known to users and stakeholders of the system.

The "Surprise" ranges of emergent properties of the complex engineered system are not addressed so comprehensively by technical regulation. Unanticipated hazards and other unwanted features are intended to be identified by system safety engineering efforts; these efforts, to the extent they are successful, really move identified unintended hazards of the system into the "Accepted Trade-off" category. The categories of Exploitable features and Facts of Existence are not really addressed by technical regulation. For instance, technical regulation does not explicitly seek to enforce methods or other means to identify, let alone promote the generation of exploitable features which may create a system of much higher value to stakeholders of the system.

In summary, technical regulation would seem to be focussed on ensuring the intended features are achieved, the design is documented and controlled, risks are accepted and that to the maximum extent possible, detrimental unintended behaviours are mitigated. There is no focus on measuring or maximising the beneficial, yet unintended, exploitable features of a

complex engineered system. It would seem that there is great value in systems that have more beneficial features than were originally specified and these would be the features that would lend a complex engineered system to be evolved over its lifecycle. The lifecycle value of a system with more of these beneficial exploitable features should clearly be of more value to stakeholders than one with less. Such a system could be said to have a higher technical integrity than one that does not, all other factors being equal. Should technical regulation seek to operate in this space? Is it possible to mandate engineering methods that would seek this as a goal?

What can we measure to determine the Technical Integrity of Complex Engineered System?

It has long been accepted that there is no single overall measure for the quality or integrity of a complex engineered system (see for example (Boehm 1978)). "System Effectiveness" for a complex system, to the extent that it can be thought of as a partial proxy for Technical Integrity, most often appears to be a combination of capability, reliability and availability in a value hierarchy (Parnell, Driscoll, and Henderson 2008). A system effectiveness measure is usually constructed for supporting selection decisions between alternate systems. There is little in the way of comparative measures available to compare the utility of disparate systems.

In order to test the hypothesis of this research, it is necessary to settle on a set of measures that directly indicate the technical integrity of the emergent properties of the system. The measures need to be general enough so that they are applicable to across the class of systems of interest and be able to be supported by quantitative or qualitative data that can be practically obtained for those systems.

(Hitchins 2003) nominates a set of common candidates for measuring the value of a system. His candidates include: life cycle cost; efficiency; effectiveness; performance; availability; survivability; process (e.g. simplicity, resource consumption, ..); product ((e.g. utility, quality, fitness for purpose, …); entropy (the degree of disorder in the system).

(Barbacci et al. 1995) provide a useful set of software quality attributes encompassing performance, dependability, security, and safety. The subsequent methodology for determining the required quality attributes for a development, Quality Attribute Workshops (Barbacci et al. 2002), allows for the attributes to be developed for each system.

Table 2 presents an initial list of candidate measures to indicate the level of technical integrity achieved by a system. Potential sources of data to evaluate these measures are also included. It is intended to mature these measures over the course of the research.

Table 2. Measures for Assessing the Technical Integrity of a Complex Engineered System.

Technical Integrity components	Candidate measures	Sources of Data

Technical Integrity components	Candidate measures	Sources of Data
Fitness for Service The materiel's ability to satisfy operational requirements	Meeting operational requirements: • Compliance • Latent defects • Achieved performance • Other characteristics of the system: o Availability o Dependability o Robustness	Project verification records Project latent defect claims Results of operational evaluations In service RMA data
	Other attributes of a system that could be direct or indirect measures of technical integrity: • System operator satisfaction levels • Other stakeholder satisfaction levels • Public perceptions • Rate of problem reports discovered after acceptance	Operator surveys Stakeholder surveys Government committee reports. Press articles. In service trouble report records; warranty and latent defect records.
Safety Freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property	Freedom from hazards: • Accident rates • Accident severity • Incident (near miss) rates	Accident and incident databases. Accident and incident investigation reports.
Environment Poses no hazard to the environment	Compliance with environmental standard Freedom from environmental impact • Environment incident rates • Environment incident severity	Environment incident databases. Environment incident investigation reports.

A gap in the available literature is the lack of a theory for developing a composite measure of technical integrity of a complex engineered system. Such a theory would, as a minimum, provide a ranking scale of technical integrity for a particular class of complex systems. Ideally the theory could provide a ratio scale for a composite technical integrity measure and this measure could be used to compare and relate the technical integrity over a broad range of complex engineered systems. Value modelling (Parnell, Driscoll, and Henderson 2008) and conjoint measurement (Anon. 2009) are possible bases for the development of such a theory. Conjoint analysis techniques, in particular techniques for establishing preference functions and for conducting surveys of stakeholders (Gustafsson, Herrmann, and Huber 2007), also require examination.

What do research findings and heuristics point to as being the best methods for achieving Technical Integrity of a Complex Engineered System?

(ISO/IEC 26702 IEEE Std 1220-2005ISO/IEC 26702 IEEE Std 1220-2005 2007) is the current standard concerning the development of man-made systems that include one or more of hardware, software, human processes, procedures, facilities and entities in the environment. It forms the current accepted practice for the engineering of all such systems including the complex engineered systems of interest to this research. While the defined systems engineering processes (requirements analysis and validation, functional analysis and verification, synthesis, design verification, systems analysis and control) are not in dispute as being required for a successful system implementation project, there is no indication in the standard as to their relative priority or the contribution each process may have to the technical integrity of a complex system.

(Elm 2008) has conducted research that indicates which systems engineering processes have the strongest positive relationship to program performance. Technical Integrity is at least partly a subset of Program Performance (includes customer satisfaction, budget performance and schedule performance) as defined in Elm's research. The systems engineering processes that have a higher positive relationship with project performance than the SE capability as a whole are "Architecture", Trade Studies", "IPT Capability" and "Requirements Development and Management". (Boehm, Valerdi, and Honour 2008) also report evidence that the right level of systems engineering applied to a project, increase the likelihood of the project running to cost and schedule budgets, but generally treats technical performance or integrity outcomes as invariant (ie implicitly assumes they are achieved at completion of the project). However, this is rarely the case in practice. Many projects are deemed to be "complete" not having achieved all technical requirements and having to have accepted unresolved risks to technical integrity.

(McDermid 2001) suggests that rigour in engineering process is perhaps not the dominant contribution to the production of complex systems with the requisite technical integrity, but that relevant experience and competence of the designer is: "For airborne software developed to DO-178B, there seems little evidence that the more rigorous processes applied in software development for software of higher required Software Integrity Levels (SILs) actually yield a lower hazardous failure rate. The limited data does suggest that the most significant correlated factor in producing software with lower hazardous failure rates is domain experience. This is most linked to the consequent reduction in initial requirements errors that domain experience would bring."

(McDermid 2001) also casts doubt on the ability of some of the engineering standards to directly contribute to the assurance of technical integrity: "Do the software safety standards actually address safety issues? They seem to be mostly focused on quality and repeatability. You would think they would aim directly at potentially hazardous failure modes of the software. An evidence-based approach is proposed – eg provide analysis to show that data structures can never be corrupted; that a scheduled function can always run on time." It would seem that this criticism could be readily extended to systems engineering standards and even to the technical regulations and their supporting manuals – quality and the repeatability of process seem to be a primary focus rather than the imposition of particular management and engineering measures to attain technical integrity.

The prevalent view in the Australian defence industry at present remains that technical integrity can be assured through the use of systems engineering methodologies but "The

Systems Engineering discipline remains intrinsically tied to the individual skill, competence and notably the availability of senior systems engineers on projects" (Irving 2008). This accepted thinking places a focus on experience and competence both at an organisational and an individual level.

(Sheard and Mostashari 2008) posit a set of 26 principles for complex systems engineering. They include the categories of Systems architecting-type principles, Systems analysis principles, Problem-space-relevant-principles, Configuration management principles, Coordination principles, and Management-related principles. While these principles are generally known or thought to be valid heuristics for the development of complex systems and would be expected to have a positive effect on the technical integrity of a complex system, there is currently little research evidence to confirm such a conclusion.

How well do the technical regulation frameworks in place in the ADF enforce these methods?

The main elements of the ADF's technical regulation is to enforces the development of systems to approved standards, the work is to be undertaken by competent and authorised individuals, who are acting as members of an authorised organisation and whose work is certified as correct (ADF 2002). As an example, Figure 3 shows the main elements of the Navy Technical Regulatory Framework (NTRF). This distils the main elements that must be implemented in the acquisition and development of maritime materiel for the ADF.

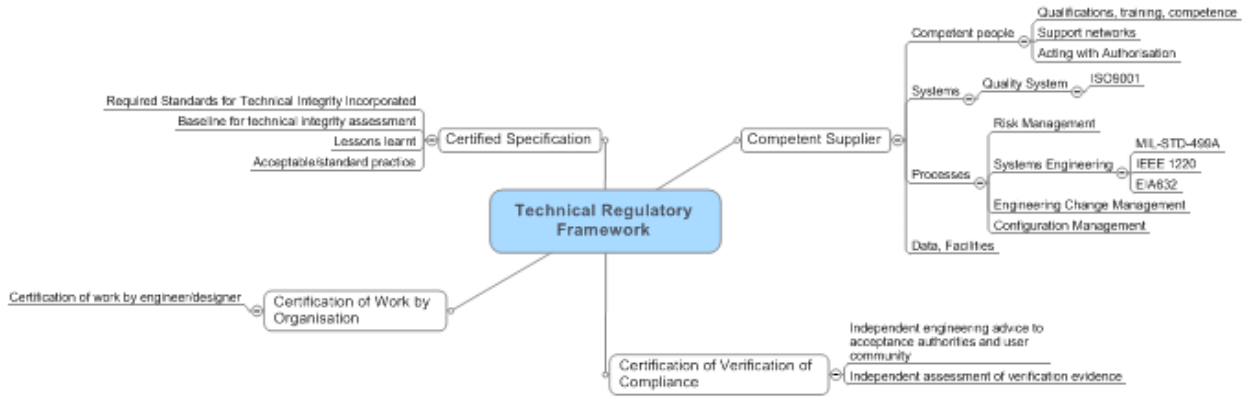


Figure 3. The main elements of the ADF's Navy Technical Regulatory Framework.

Table 3 takes the elements of the Technical Regulatory Framework highlighted in Figure 3 and provides an indication as to the extent that element is actually known (either through reported research evidence or compiled professional knowledge (e.g. standards)) to contribute to enhanced technical integrity outcomes for complex engineered systems.

Table 3. Elements of Technical Regulatory Framework and the Known Support for their contribution to Technical Integrity.

Technical Regulatory Framework	Support for Contribution to Technical Integrity
Competent Supplier	

Technical Regulatory Framework	Support for Contribution to Technical Integrity
• Competent people	Competency certifications assure minimum competency in place for significant design decisions. Certainly there are numerous case studies where the lack of competency has been directly linked to lack of required competency (eg HMAS Westralia). Also anecdotal evidence that the competence and domain experience of key designers is a determinant of project success. Not known if there is evidence that technical integrity is positively related to competent people.
• Systems (e.g. Quality System)	Standards like ISO 9001 capture industry wide knowledge. Certainly focuses on repeatability of processes to produce consistent products. Doe not focus specifically on technical integrity. Not known if there is direct evidence that technical integrity is positively related to quality systems.
• Processes	
Risk Management	The explicit consideration of risks to technical integrity embodied in the ADF Technical regulations does it would seem have to impose a mind set on the designers involved to minimise these risks. Not known if there is direct evidence that technical integrity is positively related to technical integrity risk management systems.
Systems Engineering	Standards like ISO/IEC 26702 capture industry wide knowledge for current projects. Research is beginning to provide evidence that the application of systems engineering process do contribute positively to project success and are worth the cost. Other research is indicating that the current processes requiring extension to cater for increasingly complex systems. The evidence for the relationship between these systems engineering processes and the technical integrity of complex systems is not yet established.
Configuration Management	Heuristics for complex system point out that CM must prepare for and accommodate design changes and design alternatives. The evidence for the relationship between the configuration management and the technical integrity of complex systems is not yet established.
• Data, Facilities	Research indicates that good decisions in complex situations are made on data (rather than instinct). It is therefore reasonable to extrapolate that data – correct and readily accessible to the designer – will allow good design decisions to be made. Not known if there is evidence that technical integrity is positively related to data, facilities and other supporting resource.
Certification of Verification of Compliance	
• Independent engineering advice to acceptance authorities and user community	Not known if there is support for this contributing to technical integrity.

Technical Regulatory Framework	Support for Contribution to Technical Integrity
• Independent assessment of verification evidence	Independent assessment is standard practice for high risk engineering work. Research indicates peer and expert review are effective means of eliminating defects in design.
Certification of Work by Organisation	
• Certification of work by engineer/designer	Not known if there is support for this contributing to technical integrity.
Certified Specification	
• Required Standards for Technical Integrity Incorporated • Baseline for technical integrity assessment • Lessons learnt • Acceptable/standard practice	Prescriptive regulatory regimes impose the list of standards that must be met to assure technical integrity. There are known issues (eg standards not able to keep up with pace of technology advances) and known benefits (available standards can embody previous lessons learnt) with the prescriptive approach. Not known if there is support to determine if there is any contribution to the technical integrity of complex engineered systems, which are typically unprecedented by requiring the use of extant standards.

Overall, it appears that there is a high level of experiential knowledge that the components of the technical regulatory framework support the attainment of technical integrity. They certainly seem necessary, but supporting evidence is generally scant in the literature. There is also no consideration is this analysis as to if they framework is a sufficient basis for technical integrity.

As mentioned previously in this paper, the technical regulatory framework does not address at all the attainment of emergent behaviour that may not have been intended but which is of benefit and value to the system. This is at best a nascent area of research.

Other considerations and questions that arise from this analysis that will need to be addressed by the intended research include:

Is certifying the specification any different from what standard SE standards require in terms of defining stakeholder requirements and formalising them before committing to development?

Is a supplier, as assessed as being competent under a technical regulatory framework, demonstrably better than one who isn't? Don't all suppliers of complex systems operate under an ISO 9000 or equivalent quality system accreditation? Don't all suppliers of complex systems seek to hire, train and retain competent staff to maintain their effectiveness and therefore there commercial advantage? Don't they all assure their staffs are qualified in order to mitigate liabilities for errors in design and to minimise things like insurance costs?

Don't all suppliers of complex systems strive to improve their processes and methods to achieve competitive advantage (eg most high technologies are strongly committed to seek best practice under schemes like CMMI for competitive advantage more so than because it was an imperative established by technical regulation)?

Don't all suppliers of complex systems assure they have access to and control all requisite information? Indeed don't they assiduously guard this data to protect their trade secrets and other elements of competitive advantage?

What does the explicit certification of compliance by the supplier achieve? Is not the supplier otherwise committed to achieve compliance with the specifications embodied in a contract anyway? They are also compelled to meet legislative requirements for product safety and compliance with environmental protection standards?

Is the certification of verification of compliance really any different than the technical contract oversight that would usually be applied by an educated customer function? Is it really any different to the independent verification and validation processes that would normally be applied to systems that were clearly safety critical?

Can prescriptive standards, embodied in certified specifications, really be expected to be highly applicable to unprecedented complex engineered systems? Does this mean we may really be left with mandated standard engineering processes to try to assure the technical integrity of these systems?

Conclusion

A framework to research the hypothesis "The application of technical regulation yields improved technical integrity of the complex engineered system" has been established. Initial research questions have been raised and initial concepts developed. It has been shown that only the intended emergent properties of a complex engineered system and it unintended hazards are directly addressed by technical regulation. A candidate set of measures and sources of data to assess technical integrity have been identified and it is anticipated that these will need to mature as the research progresses. The need for a composite measure of technical integrity has been identified and future research is intended to create a theory that could provide a ratio scale for a composite technical integrity measure and this measure could be used to compare and relate the technical integrity over a broad range of complex engineered systems.

The methods that may enhance technical integrity that are available to the designers of complex systems have been reviewed. They are largely heuristic at this time. Where primary methods like systems engineering have been quantitatively assessed in the literature, their impact on the overall project's performance is assessed rather than the level of technical integrity achieved on the resultant system. The degrees to which technical regulatory frameworks actually enforce methods known to have a positive contribution to technical integrity have also been reviewed. This review has resulted in a list of corollary questions to be considered by future research.

References

ADF. 2002. Defence Instruction (General) LOG 08-15. . In *Regulation of Technical Integrity of Australian Defence Force Materiel*. Canberra, ACT, Australia: Department of Defence.

Anon. 2009. *Theory of conjoint measurement*. Wikipedia 2009 [cited 9 March 2009]. Available from http://en.wikipedia.org/wiki/Theory_of_conjoint_measurement.

Barbacci, Mario, Mark H. Klein, Thomas A. Longstaff, and Charles B. Weinstock. 1995. Quality Attributes. Pittsburgh, Pennsylvania: Software Engineering Institute, Carnegie Mellon University.

Barbacci, Mario R., Robert Ellison, Anthony J. Lattanze, Judith A. Stafford, Charles B. Weinstock, and William G. Wood. 2002. Quality Attribute Workshops. In *Architecture Tradeoff Analysis Initiative*. Pittsburgh, PA: Carnegie Mellon Software Engineering Institute.

Boehm, Barry. 1978. *Characteristics of Software Quality, TRW Series of Software Technology*. New York: American Elsevier.

Boehm, Barry, Ricardo Valerdi, and Eric Honour. 2008. The ROI of Systems Engineering: Some Quantitative Results for Software-Intensive Systems. *Systems Engineering* 11 (3):14.

DeRosa, Joseph K., Anne-Marie Grisogono, Alex J. Ryan, and Douglas 0. Norman. 2008. A Research Agenda for the Engineering of Complex Systems. Paper read at IEEE International Systems Conference, April 7-10, 2008, at Montreal, Canada,.

DoD. 2000. MIL-STD-882D. In *Standard Practice for System Safety*: Department of Defense, USA.

Elm, J.P. 2008. A Study of Systems Engineering Effectiveness - Initial Results. Paper read at Systems Conference, 2008 2nd Annual IEEE, 7-10 April 2008, at Montreal, Canada.

Gustafsson, Anders, Andreas Herrmann, and Frank Huber. 2007. *Conjoint Measurement. Methods and Applications.* 4th ed. Berlin Heidelburg New York: Springer.

Hitchins, Derick K. 2003. *Advanced Systems Thinking, Engineering and Management*. Norwood, MA, USA: Artech House.

Honour, Eric C. 2007. Connecting an Architecture with its Emergent Properties. Paper read at Conference on Systems Engineering Research, 2007, March 14-16, at Hoboken, NJ , USA.

Irving, Ian. 2008. Emergent Challenges Facing Systems Engineering in the Defence Industry in Australia. *Systems Engineering Society of Australia Newsletter*.

ISO/IEC 26702 IEEE Std 1220-2005. 2007. In *Systems Engineering - Application and Management of the Systems Engineering Process*: ISO/IEC and IEEE.

Kritzinger, Duane. 2006. *Aircraft system safety. Military and civil aeronautical applications.* Cambridge, England: CRC Press, Woodhead Publishing Limited.

Magee, C. L., and O.L. de Weck. 2002. An Attempt at Complex System Classification. In *ESD Working Paper Series*: Massachusetts Institute of Technology, Engineering Systems Division.

McDermid, John A. 2001. Software Safety: Where's the Evidence? In *6th Australian Workshop on Industrial Experience with Safety Critical Systems and Software*. Brisbane, Australia: Australian Computer Society, Inc.

Parnell, Gregory S., Patrick J Driscoll, and Dale L Henderson. 2008. *Decision Making in Systems Engineering and Management*. Edited by A. P. Sage, *Wiley Series in Systems Engineering and Management*. Hoboken, New Jersey: John Wiley &Sons Inc.

Sheard, Sarah A., and Ali Mostashari. 2008. Principles of Complex Systems for Systems Engineering. *Systems Engineering*:17.

Biography

Michael Edwards holds a Bachelor of Engineering (Honours) and a Master of Engineering Science (Communications) from the University of NSW and a Master of Business Administration (Technology Management) from Deakin University. He is currently a part-time research student at the Defence and Systems Institute of the University of South Australia. He is a Fellow of the Institution of Engineers, Australia, a Chartered Professional Engineer, a member of the Systems Engineering Society of Australia, a Member of the Australian Institute of Program Management and a Master Project Director.

He is currently employed by Raytheon Australia where he is the Design Acceptance Representative for the RAN's Air Warfare Destroyer Combat System. His professional and academic interests build on an extensive experience base in the practice of engineering management and systems engineering largely in the domain of complex naval systems including applications in surface combatant and submarine combat systems, mine warfare systems, hydrographic systems and naval aviation systems.

Appendix G An ontology of technical regulation

Peer reviewed paper in Systems Engineering Test and Evaluation 2010: Adelaide, SA, Australia.

This paper presented the results of a broad review of technical regulations across of number of application domains of systems. It highlighted common and disparate elements of technical regulation. The ontology provides a structure for and definition of terms for subsequent research in this thesis and in the research area more generally. It provides a structure of the technical goals of technical regulation and the characteristics of complex engineered systems that are related to those goals. This supported the development of technical integrity concepts and measurement approaches.



An Ontology of Technical Regulation

Systems Engineering Test and Evaluation

Conference 2010

3-6 May, Adelaide

Abstract. Technical Regulation seeks the assurance that defined technical goals for a class of systems are attained during development and maintained throughout the system lifecycle. Establishing the extent to which this assurance is achieved is a focus of ongoing research by the author. Technical regulation of complex engineering systems is present in domains such as the aviation industry, nuclear power industry, biomedical devices, and defence; i.e., domains where the technical integrity (fitness for service, safety) of the system is paramount. Across these regulatory domains, the concepts and terminology of technical regulation vary. Even within the relatively narrow domain spread of the Australian Defence Force there are variations in technical regulation concepts and terms across the domains of land, air, sea and explosive ordnance. A difficulty therefore in quantifying the extent to which technical regulation achieves its goal across a range of domains is in first establishing a shared conceptualisation and terminology for the field, i.e., an ontology of technical regulation. This paper presents the analysis of a range of technical regulatory documents and develops an ontology for the study of technical regulation. The ontology is used to structure a knowledge base of technical regulations that will enable analysis and identification of points of difference and similarity between them.

INTRODUCTION

History of Regulation. Regulation is most generally defined as a sustained and focused control enacted by a public agency over activities which are valued by the community at large. Regulation usually acts to restrict behaviour associated with undesirable activities or outcomes. Regulation is not new. One of the earliest examples in contemporary history is the introduction in the USA in 1887 of an "Inter State Commerce Commission" to limit discriminatory pricing by the railroads. In the UK, USA and Australia, the 1950s and 1960s saw the introduction of regulation and regulatory bodies to deal with issues such as monopolies, gaming, industrial relations, civil aviation and workplace health and safety. The privatisation of government assets in the 80s and 90s led to more regulation as government sought to preserve some control over and to preserve the public's fair access to the services of what had until then been public utilities.

Note that there have also been balancing activities by governments seeking to de-regulate – or at least to reduce the "red tape" and inefficiency often associated with regulation. Since 1981 in the USA all regulations have been subject to cost-benefit analysis and since 1996 in the UK and European Community regulation has been subject to Cost of Compliance Assessment supported by a risk assessment (Baldwin and Cave, 1999).

Enforcing Regulation. Regulation may be enforced by formal or informal approach or a combination of the two. A formal approach is "insistent" and focuses on deterrence to assure its aims are achieved. It is often associated with compliance measures enforceable under law. An informal approach is "persuasive" in nature and focuses on seeking compliance though means of education, advice, persuasion and negotiation. When in time a particular regulation aims to intervene is another consideration. Intervention may be preventative, act-based or harm-based (Baldwin and Cave, 1999).

Technical regulation in the Australian Defence Force (ADF) context has aspects of formal and informal approaches, but appears to be more

biased toward the persuasive approach. Intervention is mostly aimed at being in the preventative stage.

Regulatory Standards. The standards enforced by regulation are related to the timing of the intended intervention:

A regulatory approach using specifications or design standards is related to preventative intervention aims. They are relatively easy to enforce and compliance costs are readily estimated. But they are highly intrusive in that the regulator is involved in defining process and design and this may be a disincentive for the supplier to innovate. Normally, this approach should be limited to cases where the regulator has available expertise and information that is not available to private suppliers.

A regulatory approach using Performance or Output Standards is related to act-based intervention aims. In this approach the standards demand a level of delivery or focus on the level of acceptable risks. The standards are not technologically restrictive.

A regulatory approach using Target Standards is related to harm-based intervention aims. Goals are stated directly, e.g., the absolute avoidance of certain harmful consequences. Suppliers are completely free as to how to meet the goals, but it may be socially inefficient if the suppliers do not have the capacity and expertise to understand how to achieve these goals (Baldwin and Cave, 1999).

Why Technically Regulate? In complex systems, failure is almost considered inevitable; systems operate on the edge of instability, and often require operators to keep the system from the brink of catastrophic mishaps.

As discussed by (Cook, 2000) "interesting complex systems" are almost always inherently and unavoidably hazardous. The frequency of hazard exposure may be minimised but the presence of hazards remains. The high consequence of failure of complex systems leads to multiple layers of defence being incorporated into the system. These include technological defences (back-up/redundant systems, 'safety' controls, etc) as well as organisational, institutional and regulatory defences (e.g., policies and procedures, certification, work rules, training).

A way of thinking of Technical Regulation is as a layer of Swiss cheese in Reason's model (Reason, 2000). Figure 1 shows an adaptation of Reason's "Swiss Cheese Model" that considers Technical Regulation as one of the layers of defence in the construction of a complex system. A failure in construction or the application of technical regulation may leave a 'hole' in that layer of defence but it would ideally be unlikely to 'line-up' with failures in the other defensive layers of the system as is shown here.

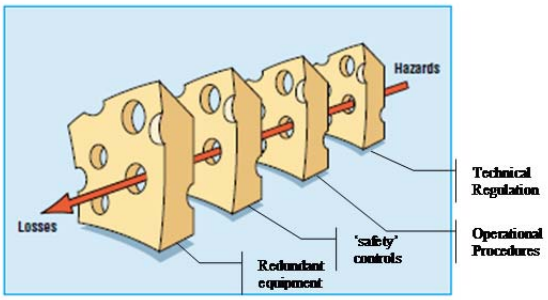


Figure 1. Reason's "Swiss Cheese Model" adapted to show Technical Regulation as one of the layers of defence in the development of a complex system.

Figure 2 illustrates a high level view of the intended influences of a technical regulation. The aim is to assure technical goals are achieved and, in this example, the technical regulation does this by requiring the application of materiel standards and enforcing the use of particular technical processes.

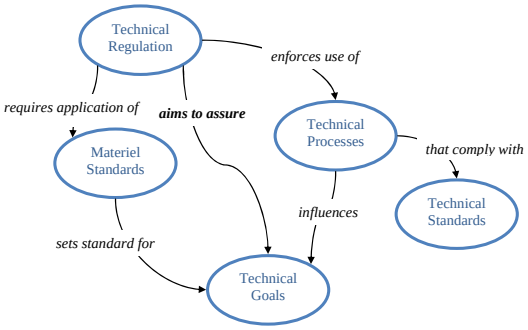


Figure 2. Intended influences of Technical Regulation.

Technical Regulation in the ADF Context. "Technical Integrity" is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection, i.e., its compliance with technical standards and the minimisation of unintended consequences. The explicit aim of

the application of technical regulation is to ensure that the requisite technical integrity of the system being developed is achieved and that technical risks are explicitly identified and reduced as far as practicable to acceptable levels. The implicit assumption is that in the absence of any technical regulation the requisite technical integrity may not be achieved.

Defence systems are steadily increasing in complexity and are increasingly required to not only introduce higher organic capabilities within the system itself, but also to integrate into systems-of-systems to enable capabilities with higher performance across a wider spectrum of domains and environments. At the same time, the need for these systems to achieve ever-increasing levels of technical integrity is paramount. Systems are expensive to develop and sustain, therefore there are less of them and each one that is realised must achieve higher and higher levels of availability and dependability. The stakeholders of the system, not least the general public (and consequently the political leaders of the nation) are less and less willing to accept mishaps – be they accidents to personnel, loss or damage of equipment or impacts to the natural environment.

Technical regulation seeks to provide more assurance to the acquirers and ultimately the users of engineered systems than just a reliance on system developers to unilaterally apply the "best" engineering and management processes. "Best" for a developer does not necessarily lead to a set of practices that will yield the required technical integrity. Engineering practices, like Systems Engineering for instance, do not alone assure that technical integrity goals are a primary focus during development, or that the processes will be executed by competent professionals with explicit responsibility for their design decisions and the system outcomes. Technical regulation therefore generally attempts to assure that development is conducted in compliance with approved standards, by competent individuals in an organisation that is authorised to do so. There is also an explicit process to achieve "certification"; at a minimum by the developer formally attesting to the completeness and quality of its own work, but usually also encompassing some independent validation of the developer's claims of compliance with technical standards and processes.

It would be expected then, in some average sense at least, that the technical integrity of systems developed by an organisation outside an explicit technical regulatory system would be less than that developed inside a technical regulatory system. Indeed, this is accepted as an axiom in the context of complex engineered systems for the ADF where acquisition and development under a Technical Regulatory Framework is now mandated for all procurements of defence materiel (ADF, 2004).

The Impact of Technical Regulation. The author is pursuing a research agenda to address the following problem statement. To what extent do formal regulation and certification frameworks, applied to complex system acquisition and development programs in the Australian defence context, yield improved technical integrity of the delivered system? (Edwards, 2009a)

WHAT IS AN ONTOLOGY?

According to Princeton University's Wordnet (Miller, 2009), an ontology is "([*in the discipline of]* computer science) a rigorous and exhaustive organization of some knowledge domain that is usually hierarchical and contains all the relevant entities and their relations".

"An ontology provides a shared vocabulary, which can be used to model a domain, that is, the type of objects and/or concepts that exist, and their properties and relations" (Arvidsson and Flycht-Eriksson, 2002).

NEED FOR AN ONTOLOGY AND INTENDED USE.

Define a Basis for a Field of Study. A difficulty in quantifying the impact of technical regulation on the development of complex engineered systems is the lack of a clearly defined basis to the field of study.

Technical regulation clearly exists as evidenced by the numerous laws, policies, standards and regulatory bodies. Therefore, the first step in quantification is to define the field in terms of the nature and relations of the activities that are commonly considered to be "technical regulation."

Structure a Knowledge Base. The ontology for technical regulation that is presented in this paper will be used to capture and characterise the breadth and depth of regulatory requirements imposed by any particular instance of a technical

regulatory framework. The ontology is intended to be used in future research to provide a structure in which data can be collected about technical regulation as it has been applied and support correlation of this data with technical integrity outcomes for complex engineering system developments.

Questions to be answered using the Ontology. The ontology needs to be designed with the end purpose in mind. The resultant ontology should therefore support the answering of questions like:

What engineering processes are mandated by a particular technical regulation to ensure a technical goal is achieved?

What is the range of engineering processes that are mandated by a number of technical regulations to ensure a technical goal is achieved?

Do the engineering processes mandated by a set of technical regulations vary across the types of systems being regulated or according to the system life-cycle phase?

Do the engineering processes mandated by a set of technical regulations achieve their goal?

ONTOLOGY DEVELOPMENT

Methodology. The methodology used broadly followed the development process of (Noy and McGuinness, 2001) and has been adapted as:

1. Determine the domain and scope of the ontology.

2. Enumerate the main terms of the ontology by a review of a representative sample of technical regulations.

3. Normalise the terms used in the ontology to terms already defined in international standards (in particular ISO/IEC Guide 2 and ISO/IEC 17000 as presented in (Hunter, 2009)). Use and adapt those definitions for the ontology, and document new class definitions where required.

4. Define the classes and class hierarchies using a combination approach – initially a top down structuring based on the generally understood concepts of the technical regulation domain; then a bottom up structuring based on the identification of instances of the classes within a set of actual technical regulations.

5. Define the properties of the classes including the relationships between them.

6. Validate the ontology by using it to structure a knowledge base populated from a sample of technical regulations.

In practice, recursion through steps 2 through 6 has been necessary to construct a valid ontology.

Domain and Scope. The domain considered for the ontology, selected in support of research objectives, is the technical regulation of complex engineered systems where the purpose of the regulation is to ensure achievement of the stated technical objectives.

The breadth dimension primarily considers technical regulation in the ADF for the purpose of ensuring that the developed ontology provides a complete coverage of this area. In addition, technical regulations from other militaries and from civil regulatory domains related to complex engineered systems have been considered to increase the breadth of considerations and the resultant scope of the ontology.

Technical regulation of complex engineering systems is present in domains such as the aviation industry, nuclear power industry, biomedical devices, and defence; i.e., domains where the technical integrity (fitness for service, personnel safety and minimisation of impact to the environment) of the system is paramount. Across these regulatory domains, the concepts and terminology of technical regulation vary. Even across the constrained range of the ADF system applications and technologies there are variations in technical regulation concepts and terms across the domains of land, air, sea and explosive ordnance.

The author's continuing research is focused on the impact of technical regulation on the technical integrity of complex engineered systems primarily in the context of the ADF. Therefore, the full scope of ADF technical regulatory frameworks is considered. In addition, to assure a fuller development of the ontology a range of other frameworks – from overseas militaries, from domestic civil regulators and overseas civil regulators of complex engineered systems – have been considered.

Implementation. The resultant ontology has been used to structure a simple, but extensible knowledge base as illustrated in Figure 3. The ontology has been documented using the

CORE® systems engineering tool (Vitech_Corporation, 2009) and this will be utilised in future work to build the knowledge base. It will contain the actual technical goals, provisions, regulated system types and life cycle phases from instance of Technical Regulations captured in a database that is structured in accordance with the ontology. It will also contain case study results for instances of application of a technical regulation that capture technical goals achieved, provisions implemented, and other research data.

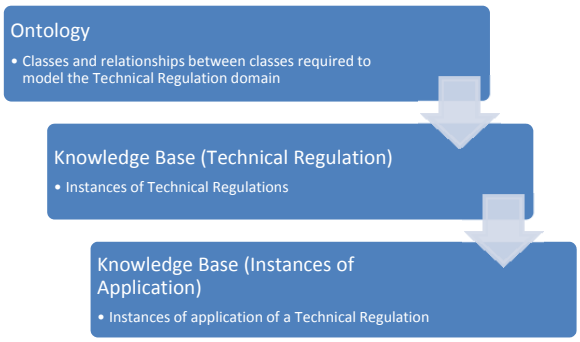


Figure 3. Implementation of the Ontology and Knowledge Base.

RESULTANT ONTOLOGY OF TECHNICAL REGULATION

Technical Regulation. The first level of the ontology is shown in Figure 4. It has been designed to focus on the aspects shown: the technical goals that the application of regulation is seeking; the provisions (or requirements) that the regulation imposes; the types or classes of systems and the life cycle phase of those systems for which the regulation is intended to apply.

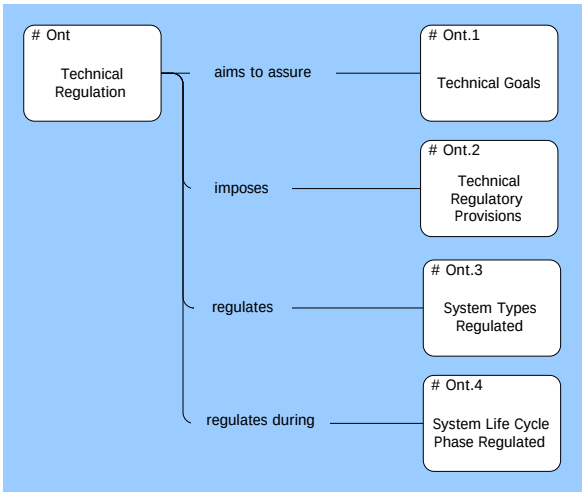


Figure 4. First Level of Ontology.

[1] Technical Goals. At the second level of the ontology, the types of technical goals that may be the assurance aim of a Technical Regulation are illustrated in Figure 5. The first 7 of these are defined consistent with those identified as "Aims of standardization" in ISO/IEC Guide 2 (Hunter, 2009). The eighth, Supportability, has been added since regulatory goals associated with reliability, maintainability and other aspects of system support that are evident in some regulations examined are not explicitly covered in the definitions of ISO/IEC Guide 2.

[2] Technical Regulatory Provisions. The types of provisions that a Technical Regulation may impose are illustrated in Figure 6. This is the most complicated part of the ontology due to the intended use. It is intended that in developing the knowledge base consistent with this part of the ontology, points of difference between different instances of technical regulations will be able to be quickly differentiated and examined. The establishment of relationships between these provisions and the attainment of technical goals is a key use for the ontology.

[2.3] Provisions for Process. Provisions for Process is a key area and has been broken down into Systems Engineering Processes (taken to mean the generic engineering management and technical processes that may be undertaken to develop a complex system, e.g., the processes defined in ANSI/EIA-632) and Technical Discipline processes (taken to mean the specific processes undertaken within an engineering discipline). Examples of provisions for process include: how to design a structure for an airframe or for a ship; how to design electronic circuit assemblies; or how to design and implement computer software.

[2.6] Provisions for Conformity Assessment. Another key area is Provisions for Conformity Assessment. The types of provisions that may be imposed for conformity assessment are illustrated in Figure 7. Most Technical Regulations examined in developing the ontology, devote considerable provisions to the way objects such as products, data, processes, individual competence and organisational capability are verified and how and by whom compliance with other provisions of the regulation is attested to.

[2.6.3] Attestation. Attestation is defined as the issue of a statement, based on a decision following review, that fulfilment of specified

requirements has been demonstrated. It is a particular area where regulations vary in the use of terminology and where this ontology seeks to establish a clear classification. ISO/IEC 17000 (Hunter, 2009) has been used as a basis for the ontology in this area. The definitions for each class of attestation are:

[2.6.3.1] Declaration. A first party attestation. That is, a conformity assessment activity that is performed by the person or organization that provides the object (e.g., a supplier's Certificate of Conformance).

[2.6.3.2] Certification. A third-party attestation related to products, processes, systems or persons. That is, a conformity assessment activity that is performed by a person or body that is independent of the person or organization that provides the object and of the user interests in that object.

Note that certification of a management system is also sometimes called "registration".

[2.6.3.3] Accreditation. A third-party attestation related to a conformity assessment body. That is, an assessment body conveying formal demonstration of its competence to carry out specific conformity assessment tasks.

[2.6.3.4] Second Party Attestation. There is no alternate defined term for this class of attestation in ISO/IEC17000. It is a conformity assessment activity that is performed by a person or organization that has a user interest in the object.

It can be thought of as the acceptance of a product or service by the purchasers or users of the system. In technical terms only, the term "Design Acceptance" is an example of a Second Party Attestation that used in some ADF technical regulations.

[3] System Types Regulated. Figure 8 illustrates the hierarchy for the types of systems to which the technical regulation may apply. In this ontology, consideration is limited to "complex engineered systems". The focus of ongoing research is in the military system domain, therefore there is an emphasis in the hierarchy to those types of systems and to systems in the civil domain where there is a similar technology employed and/or the critical technical goals for the regulatory regime are similar (e.g., safety and security).

[4] System Life Cycle Phase Regulated. The hierarchical breakdown of system life cycle phases is shown in Figure 9. These life cycle phases have been taken from (Honour and Valerdi, 2006) with the intention being to leverage from and compare to published findings of their research on COSYSMO and the Systems Engineering – Return on Investment (SE-ROI) projects.

Other Considerations. Other aspects of technical regulation have been examined, but are not currently included in the ontology at this time since they do not directly support the types of questions that are to be researched.

As an example, the types of authorities under a Technical Regulatory regime that may be responsible for creating and maintaining the regulations ("Regulatory Authority") and enforcing the regulations ("Enforcement Authority").

Definitions for each of the elements in the ontology are documented in the CORE model but are not completely reproduced herein for the sake of brevity.

VALIDATION OF THE ONTOLOGY

Instances of Technical Regulations. The ontology has been validated for its intended use as a basis for a knowledge base of the domain by analysing a set of technical regulations of interest to the research agenda. The regulations considered in developing and validating the ontology are:

- The Australian Defence Force's overall policy and regulations for assuring the technical integrity of materiel (ADF, 2004).
- The Australian Defence Force's regulations for assuring the technical integrity of land materiel (Army, 2003).
- The Australian Defence Force's overall policy and regulations for assuring the technical integrity of maritime materiel (RAN, 2003).
- The Australian Defence Force's overall policy and regulations for assuring the technical airworthiness of state aircraft and aviation systems (RAAF, 2005).
- The Australian Defence Force's overall policy and regulations for assuring the technical integrity of explosive ordnance (Australian_Government, 2007).
- The United States Navy's platform level

and strike force (formally battle force) level Naval Warfare Systems Certification Policy (Department_of_the_Navy, 2005).

- The U.S. Department of Energy's policies and guidelines for acquisition of capital assets which include the regulations for acquisition of highly safety and security critical nuclear power assets (US_Department_of_Energy, 2008, US_Department_of_Energy, 2006).

The goals, provisions, and system types and life-cycle phases regulated have been identified in each regulation and captured in the knowledge base. In practice, this validation activity has led to recursive changes to the ontology, with a validated ontology being the outcome presented previously.

Technical Goals and Quality Attributes. Some of the technical regulations considered above express their technical goals in terms of quality attributes and it is neither always straightforward nor objective to allocate these goals to the technical goals of the ontology. To further validate the ontology and to refine the descriptions of each of the Technical Goal classes, two generic lists of quality attributes for systems were considered. The first (Wikipedia, 2010) is an exhaustive list of possible quality attributes for a system. The second (Firesmith et al., 2009) contains a more constrained hierarchy of quality attributes characterised into focus areas to support system architecture and design. Both these sets of attributes have been mapped against the technical goals to ensure that the technical goals are a sufficient set for the ontology and to assist in refining the definitions. Table 1 shows the result of this activity. Quality attributes sourced from (Wikipedia, 2010) are in lowercase while those sourced from (Firesmith et al., 2009) appear in Titlecase in the third column. This activity confirmed the need to add the Technical Goal "Supportability" to the Ontology as mentioned previously.

Technical Goal	Description	Associated Quality Attributes
Fitness for Purpose	Ability of a product, process, or service to serve a defined purpose under specific conditions.	Compliance Performance Correctness Functionality Operability Usability accuracy affordability Correctness dependability deployability effectiveness fidelity mobility precision relevance responsiveness stability timeliness usability
Compatibility	Suitability of products, processes or services for use together under specific conditions to fulfill relevant requirements without causing unacceptable interactions.	Interoperability compatibility installability interoperability
Interchange-ability	Ability of one product, process, or service to be used in place of another to fulfill the same requirements.	Configurability adaptability configurability customizability flexibility interchangeability modularity
Variety Control	Selection of the optimum number of sizes or types of products, processes, or services to meet prevailing needs.	Configurability Capacity Efficiency configurability efficiency scalability
Safety	Freedom from unacceptable risk of harm.	Safety safety
Product Protection	Protection of a product against climatic or other adverse conditions during its use, transport or storage.	Robustness Security Survivability deployability durability mobility portability survivability

Technical Goal	Description	Associated Quality Attributes
Supportability	Ability of a product, process, or service to be supported to serve its defined purpose under specific conditions. Support may include maintenance and training.	Availability Predictability Reliability Stability administrability availability learnability maintainability manageability recoverability reliability repeatability reproducibility reusability serviceability stability sustainability testability upgradability

Table 1: Technical Goals of the Ontology associated with Quality Attributes of a System.

DISCUSSION

Earlier in the paper, samples of the types of questions to be answered using the ontology were posited. In this section insights relative to these questions and other insights that have been gained from the validation work are discussed. These insights derive from a body of analysis work that is resident within the knowledge base.

Classes That Vary Across Regulations. Some instances of classes defined in the ontology vary across instances of technical regulations since they related to particular application domains or system and technology types. Differences in provisions for product are, as would be expected different across the domains, e.g., the structural engineering standards required for aero-structures and for maritime platforms are different. Similarly, provisions for technical discipline processes differ across system types and application domains.

Variations Due to Different Regulatory Approaches. Some instances of classes defined in the ontology are expected to vary due to different regulatory approaches: "insistent" versus "persuasive" enforcement approaches; or the timing of regulatory intervention ("preventative", "act-based", or "harm-based"). Most technical regulations considered to date use a mixed enforcement approach but generally are all "preventative" in terms of intervention timing. Thus we see both the presence of mandatory standards and requirements as well as guidelines and guidebooks (the former indicative of the insistent approach, the latter indicative of the persuasive approach) within the same regulatory framework,

Classes That Do Not Vary Across Regulations. Some classes in the ontology are not expected to have instances that vary greatly over the range of technical regulations associated with complex engineered systems. As an example, consider the class Provisions for Systems Engineering process. All technical regulations examined thus far rely on a Systems Engineering process at either an "insistent" level or at least at a "persuasive" level with Systems Engineering guidelines that are in accordance with international standards. Consider a second example, the class Provisions for Organisations, where accreditation of an organisation's quality system to a recognized standard is always required.

This discussion provide additional assurance that the ontology is valid for its intended purpose in defining a basis for a field of study, structuring of a knowledge base and the ability to provide insights into questions raised in the field.

CONCLUSION AND FUTURE RESEARCH

Conclusion. The domain of technical regulation has been discussed and defined. An ontology of technical regulation has been developed and validated for use as a basis for future research. A knowledge base of technical regulations has been populated.

Future Research. The population of the knowledge base with instances of technical regulations will enable the objective comparison of different technical regulations. This will provide objective data for answering questions like "what engineering processes are mandated for achieving particular technical goals, and is their variance across the types of systems regulated?".

It is intended that the population of the knowledge base will deepen to beyond the level of instances of technical regulations to instances of application of technical regulation to complex engineered systems. This will provide a basis for the collection of quantitative and qualitative data in support of the data collection and analysis

phase of planned research (Edwards, 2009b).

REFERENCES

ADF (2004) Defence Instruction (General) LOG 08-15 (AM 1). *Regulation of Technical Integrity of Australian Defence Force Materiel.* Canberra, ACT, Australia, Department of Defence.

ARMY (2003) Defence Instructions (Army) LOG 1/2003. *Regulation of the Technical Integrity of Land Materiel.* AMNDT 1 ed. Canberra, ACT, Department of Defence, Army Headquarters.

ARVIDSSON, F. & FLYCHT-ERIKSSON, A. (2002) Ontologies I. Homepage of PhD Course: Semantic Web, Linkoping University, Sweden.

AUSTRALIAN_GOVERNMENT (2007) DEOP 102 (AM1). *Technical Integrity of Explosive Ordnance.* Canberra, ACT, Australia, Department of Defence.

BALDWIN, R. & CAVE, M. (1999) *Understanding Regulation. Theory, Strategy and Practice,* Oxford, UK, Oxford University Press.

COOK, R. I. (2000) How Complex Systems Fail. Cognitive Technologies Laboratory, University of Chicago.

DEPARTMENT_OF_THE_NAVY (2005) NAVSEA Instruction 9410.2. *Naval Warfare Systems Certification Policy.* Washington Navy Yard, DC, Naval Sea Systems Command.

EDWARDS, M. (2009a) The Impact of Technical Regulation on the Technical Integrity of Complex Engineered Systems. *INCOSE 2009.* Singapore, International Council On Systems Engineering.

EDWARDS, M. (2009b) Research Proposal: The Impact on the Technical Integrity of Complex Engineered Systems by Technical Regulation. *Defence and Systems Institute.* Adelaide, SA, University of South Australia.

FIRESMITH, D. G., CAPELL, P., FALKENTHAL, D., HAMMONS, C. B., LATIMER, D. & MERENDINO, T. (2009) *The Method Framework for Engineering System Architectures,* Boca Raton, FL, CRC Press, Taylor and Francis Group.

HONOUR, E. C. & VALERDI, R. (2006) Advancing an Ontology for Systems Engineering to Allow Consistent Measurement. *Conference on Systems Engineering Research.* Los Angeles, Ca.

HUNTER, R. D. (2009) *Standards, Conformity Assessment, and Accreditation For Engineers,* Boca Raton, FL, USA, CRC Press, Taylor and Francis Group.

MILLER, G. A. (2009) WordNet - About Us. *http://wordnet.princeton.edu*. Princeton University.

NOY, N. F. & MCGUINNESS, D. L. (2001) Ontology Development 101: A Guide to Creating Your First Ontology. *Online.* Stanford, CA, Stanford University.

RAAF (2005) Australian Air Pubication 7001.053(AM1). *Technical Airworthiness Management Manual.* Director General Technical Airworthiness.

RAN (2003) Defence Instructions (Navy) LOG 47-3. *Regulation of Technical Integrity of Australian Defence Force Maritime Materiel.* AMDNT NO 1 ed. Canberra ACT, Department of Defence (Navy Headquarters).

REASON, J. (2000) Human Error: Models and Management. *British Medical Journal,* 320**,** 768-70.

US_DEPARTMENT_OF_ENERGY (2006) DOE O 413.3A Program and Project Management for the Acquisition of Capital Assets. Washington, D.C.

US_DEPARTMENT_OF_ENERGY (2008) DOE G 413.3-1 Managing Design and Construction Using Systems Engineering. Washington, D.C.

VITECH_CORPORATION (2009) CORE Software. *http://www.vitechcorp.com/*. Virginia, USA.

WIKIPEDIA (2010) List of system quality attributes. *http://en.wikipedia.org/wiki/List_of_system_quality_attributes*.

BIOGRAPHY

Michael Edwards holds a Bachelor of Engineering (Honours) and a Master of Engineering Science (Communications) from the University of NSW and a Master of Business Administration (Technology Management) from Deakin University. He is currently a part-time research student at the Defence and Systems Institute of the University of South Australia. He is a Fellow of the Institution of Engineers, Australia, a Chartered Professional Engineer, a member of the Systems Engineering Society of Australia and a Member of INCOSE.

He is employed by Raytheon Australia where he is the Design Acceptance Representative for the RAN's Air Warfare Destroyer Combat System. His professional and academic interests build on an extensive experience base in the practice of engineering management and systems engineering largely in the domain of complex naval systems including applications in surface combatant and submarine combat systems, mine warfare systems, hydrographic systems and naval aviation systems. He may be contacted at Michael.Edwards@Raytheon.com.au or Michael.Edwards@postgrads.unisa.edu.au.

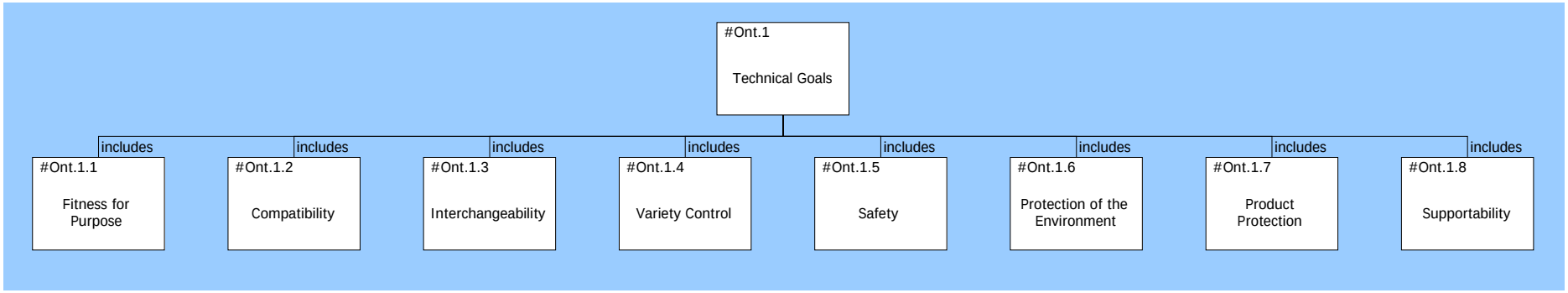


Figure 5. Second Level of Ontology – Technical Goals

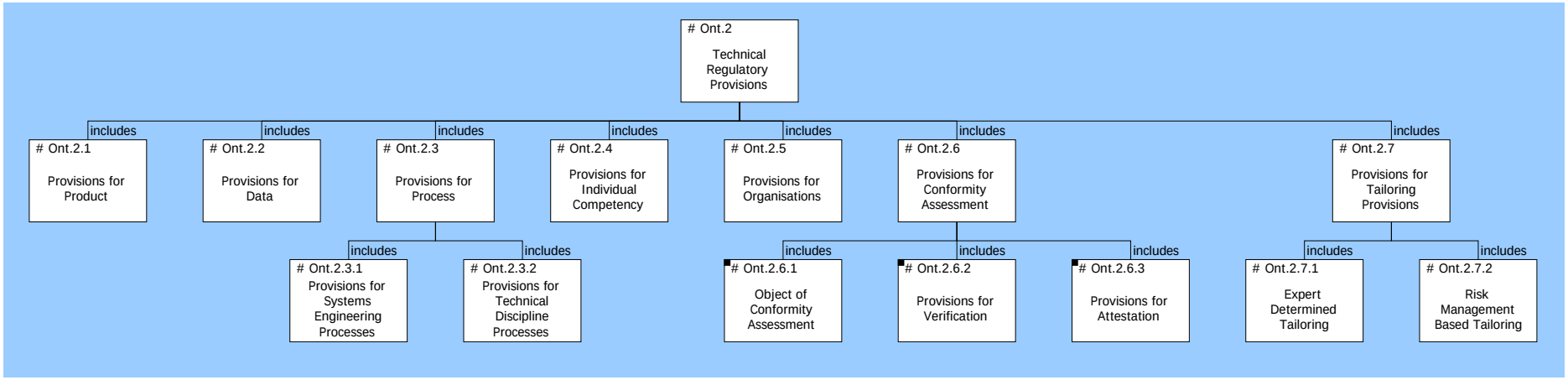


Figure 6. Second Level of Ontology – Technical Regulatory Provisions

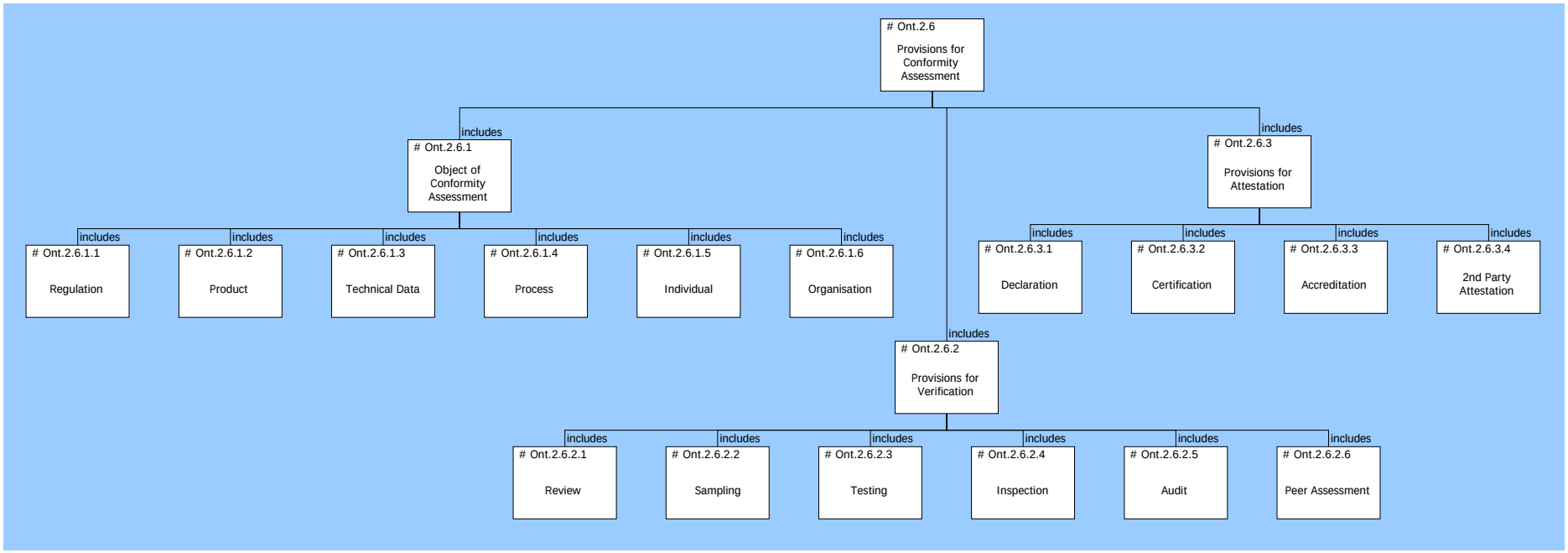


Figure 7. Third Level of Ontology – Provisions for Conformity Assessment

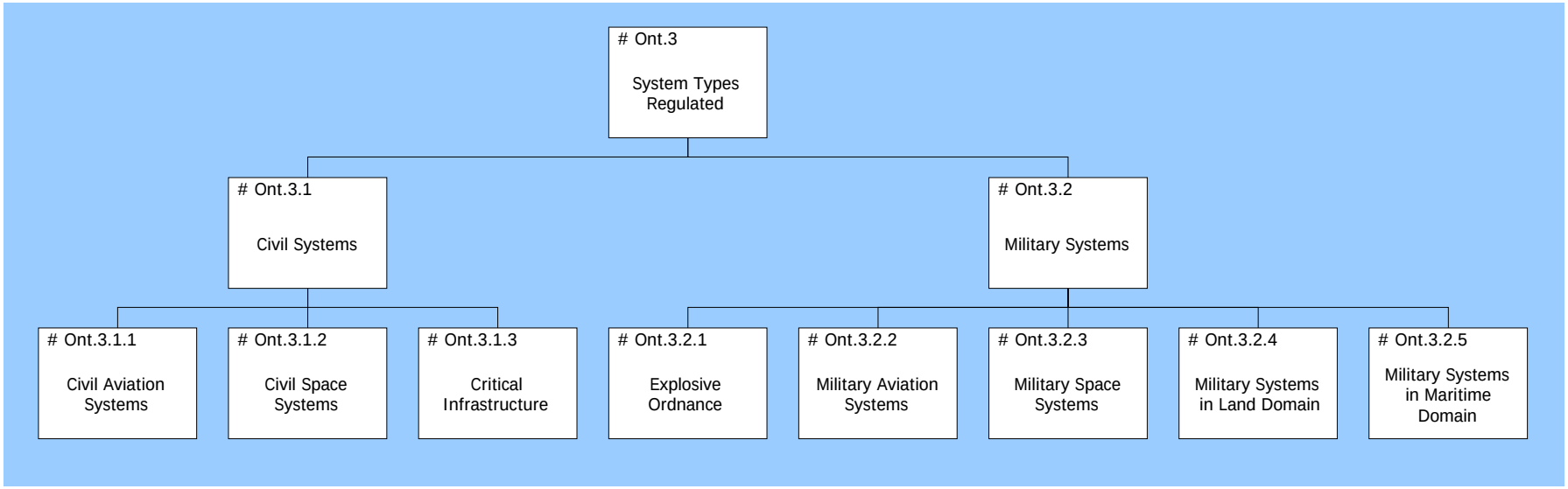


Figure 8. Second Level of Ontology – System Types Regulated.

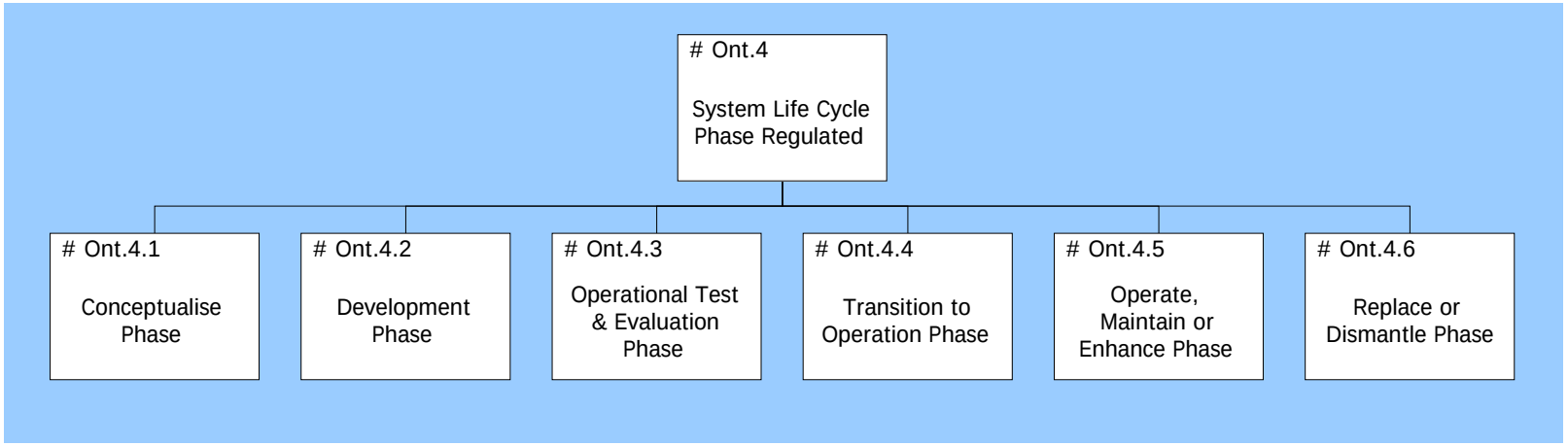
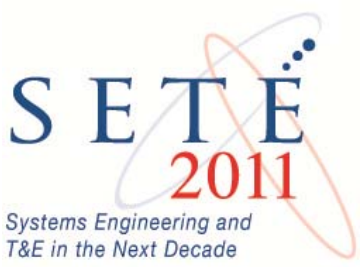


Figure 9. Second Level of Ontology – System Life Cycle Phase Regulated

Appendix H Is systems engineering necessary for the achievement of technical integrity?

Peer reviewed paper in Systems Engineering Test and Evaluation 2011, Systems Engineering Society of Australia: Canberra, ACT.

This paper explored in detail one aspect of the research problem: if the provisions of technical regulations actually have some evidentiary basis for improving technical integrity. The paper focused on Systems Engineering being an almost universal provision of technical regulations. It concludes that there is little objective evidence to support the effectiveness of Systems Engineering on the technical outcomes of systems projects. The outcome supports the research objective of being able to measure technical integrity in order to build support for Systems Engineering and other provisions of technical regulations.



Is Systems Engineering necessary for the achievement of Technical Integrity?

Michael Edwards

Raytheon Australia / UniSA DASI

edwmt002@mymail.unisa.edu.au

Copyright © 2011 by Michael Edwards. Published and used by SESA with permission.

Abstract. Recently published research indicates the effectiveness of Systems Engineering for improving the outcomes of complex projects. However, the research does not clearly demonstrate support for a positive relationship between the application of Systems Engineering effort and the technical integrity, quality or effectiveness of the complex engineered system. This paper presents the outcomes of a Relational Analysis of selected literature to establish what the systems engineering development methods and associated factors that are well validated as necessary for the development of complex engineered systems with high technical integrity. The identified systems engineering methods are contrasted with ADF technical regulations that seek to assure technical integrity. Insights are identified that may improve the specificity of technical regulations for assurance of technical integrity of complex systems. Gaps in current knowledge are highlighted for future inquiry.

INTRODUCTION

The author's current research is concerned with complex engineered systems and inquiry into the impact of technical regulation on the technical integrity of these systems particularly during the design and development phase of the system lifecycle. In the Australian Defence Force (ADF), *technical integrity* is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection, i.e. its compliance with technical standards and the minimisation of unintended consequences. Technical regulations within the ADF, to varying degrees, require the application of Systems Engineering (SE) methods as some of the measures to assure *technical integrity*.

This introduction summarises what is understood about the effectiveness of SE, defines and discusses *technical integrity* and related concepts, shows how the issue of linkage of SE methods to *technical integrity* and technical quality outcomes has not been well established, and identifies a categorisation of SE methods for use in this study. The paper then moves on to describe the relational analysis approach derived for this study, then presents and discusses the results from that analysis.

THE EFFECTIVENESS OF SYSTEMS ENGINEERING

The widespread adoption of SE by organisations undertaking the acquisition and development of complex systems provides significant support for the hypothesis that systems engineering is effective in improving project outcomes. There is certainly a well recognised and documented myriad of pitfalls that can befall a project that does not apply SE. The response to these pitfalls leads to heuristics which implicitly underlie the SE methods outlined in international standards and authoritative bodies of knowledge. However, quantifying how effective SE is remains a subject of ongoing research.

What Systems Engineers Believe

Perhaps not surprisingly, most systems engineers consider that SE has a positive impact on project and product outcomes. (Kludze 2004) reported results of his research that indicates that an overwhelming majority of systems engineers consider that SE provides a positive impact to overall project outcomes (95% of those surveyed at NASA and 99.6% of those INCOSE members surveyed). In terms of "Technical Performance" of the delivered system, again there is very strong belief from

the systems engineers surveyed that SE provides a positive impact (84% of those surveyed at NASA and 93% of those INCOSE members surveyed). A high percentage of survey respondents also indicate a belief that SE plays a strong role in reducing the risk of unintended consequences (91% from NASA and 94% from INCOSE).

The Systems Engineering – Return on Investment Project

(Honour 2010) has found during his research on Systems Engineering Return On Investment (SE-ROI) that while SE leads to better programs, SE does not necessarily lead to better systems ("an unexpected result")!

Honour's research finds that applied SE effort, correlates positively for project outcomes like "Schedule performance", "Cost performance" and to overall "Project Success" (the latter as indicated by stakeholder acceptance of the project outcomes), but is not correlated with what he terms "Technical Quality".

"Technical Quality" in Honour's research is related to the concept of Key Performance Parameters (KPPs) that are of primary importance to the system's stakeholders. For each KPP there is a Threshold level and an Objective level defined. The Threshold level corresponds to the specified minimum requirements that a system must achieve for Acceptance while the Objective level corresponds to performance levels of value to the stakeholders of the system.

It is apparent in most projects that the pursuit of KPPs beyond a threshold level is not normally an objective of SE based development. Consequently it is not surprising that SE Effort does not positively correlate with Technical Quality. Of the 8 constituent elements of SE used by Honour, none correlate positively with Technical Quality. In fact Requirements Engineering shows a slight negative correlation with Technical Quality. This could be interpreted as Requirements Engineering being a process with a focus on ensuring that specified requirements are met and that precludes the project team (or at least diverts the team) from looking at building better (i.e., higher value, higher performance) systems.

System Engineering in Space Projects

(Hamann, Zandbergen, and Zijdemans 2009) have examined the relationship between SE methods applied and the resultant technical quality, cost and schedule outcomes from a number of space projects. The results show a positive effect of SE on cost and schedule outcomes. However, there is little positive correlation between SE effort and the technical result of the project. Rather, most space projects are driven to deliver the required technical result. Application of SE seems to make sure the technical result is achieved at less cost and schedule that if it had not been applied.

Study on Systems Engineering Effectiveness

The Software Engineering Institute (SEI) and the National Defense Industry Association (NDIA) has undertaken surveys of projects for the US DOD and the subsequent analysis broadly supports the hypothesis that more projects are successful if they are executed by organisations utilising higher SE capability. The result is more pronounced when the challenge presented by the project is higher (Elm et al. 2007).

"Project Performance" in this study is decomposed into performance elements for cost, schedule and scope, each of which are derived from measures commonly available within a project (Cost Performance Index (CPI), Schedule Performance Index (SPI), KPPs achieved, requirements verified). However, the analysis intentionally did not evaluate the relationships between SE capabilities and scope or technical performance. So while the reported results in (Elm et al. 2007) support the notion that higher levels of SE capability lead to better overall project outcomes, it is not possible to infer anything about the technical performance of systems produced from organisations exhibiting higher levels of SE capability.

In order to dig a little deeper to see if there is a relationship between SE capability and the technical outcomes for a developed system, the author has corresponded with Joe Elm at the SEI who has been kind enough to provide some further analysis and results from the data gathered previously by the

SEI/NDIA study (Elm 2011). These results are included below noting they should be interpreted with an understanding of the methodology reported in the SEI/NDIA report (Elm et al. 2007).

"I re-examined the data from the earlier study, and performed a quick analysis of the relationship between the variables of TOTAL SE Capability (TTL-SEC) and satisfaction of requirements (Perf-Scope).

The distributions of TTL-SEC and Perf-Scope for the sample are shown below" in Figure 1. Note in the TTL-SEC distribution the 46 projects in the survey sample reported SE capability in the medium to high range. In the Perf-Scope Distribution a response of 4.00 means there is "strong agreement" that requirements have been satisfied as originally planned, 3 means "agreement", 2 means "disagreement" and 1 means "strong disagreement".

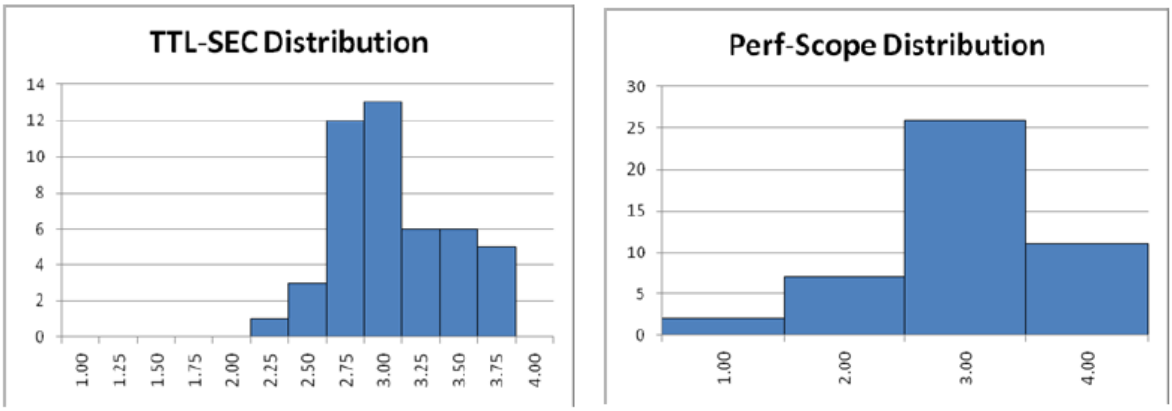


Figure 1. The Distribution of Total SE Capability and the Distribution of Performance-Scope (Elm 2011).

The relationship between the variables of TOTAL SE Capability (TTL-SEC) and satisfaction of requirements (Perf- Scope) is shown in

Figure 2.

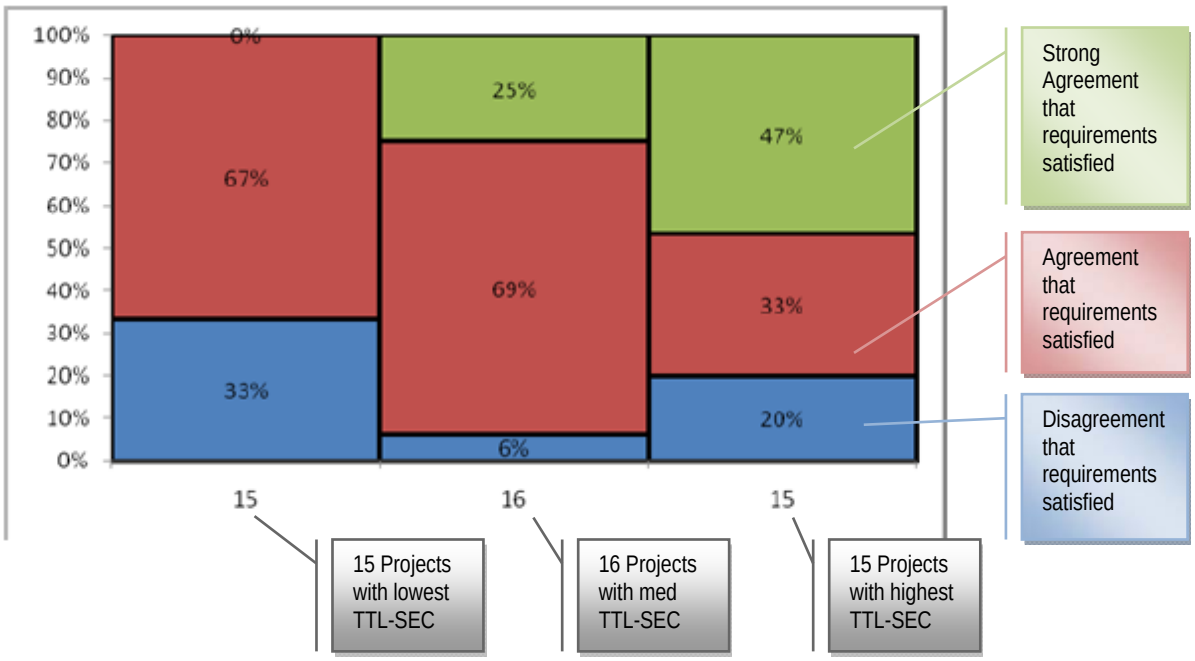


Figure 2. Total SE Capability related to Performance-Scope (Elm 2011).

"The left column represents those projects (quantity 15) with lower TTL-SEC capabilities (TTL-SEC <= 2.72) the right column represents those projects (quantity 15) with higher TTL-SEC capabilities (TTL-SEC <= 3.1), and the middle column represents the remaining projects (quantity 16).

Due to the coarse resolution of the Perf-Scope variable, establishment of breakpoints to evenly distribute projects between Low, Moderate, and High Perf-Scope was not possible. Thus, breakpoints were set as follows:

Low Perf-Scope < 3.0 (9 projects)

Moderate Perf-Scope = 3.0 (26 projects)

High Perf-Scope > 3.0 (11 projects)

Thus, within each column of the chart above, the lower block represents those projects with Perf-Scope of 1 or 2, the middle block of each column represents those projects with Perf-Scope of 3, and the top block represents those projects with Perf-Scope of 4.

There appears to be a reasonably strong relationship between TTL-SEC and Perf-Scope. Goodman and Kruskal's Gamma is calculated at 0.51".

This result shows support for the notion that organisations applying higher levels of SE capability to the development of a system are judged to more frequently and better achieve the requirements established at the commencement of the project.

DEFINING TECHNICAL INTEGRITY

The author defines *technical integrity* as consisting of two attributes *Capability* and *Dependability*, as shown graphically in Figure 3, where:

Capability is taken to mean that the system meets all specified requirements for performance, function and quality attributes in the defined environment and when used as intended. *Capability* is the set of emergent properties of a system that is designed (i.e. intended) and useful (i.e. required); and

Dependability is taken to mean that the risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. It includes concepts of robustness (system is tolerant of usage that may exceed specified conditions without catastrophic consequences), reliability (failures do not happen often), and system safety (risk of unidentified hazards and failure modes resulting in undesirable outcomes is minimised).

This definition is consistent with the objective of technical regulations in the ADF and of many other technical regulatory systems in place for the development of complex and high risk consequence systems (refer (Edwards 2011) for a fuller exposition of this definition). The definition can be readily seen to encompass the scope of *technical integrity* as used in the ADF. Fitness for Service is equivalent to *Capability* and Safety and Environment Protection are hyponyms of *Dependability*.

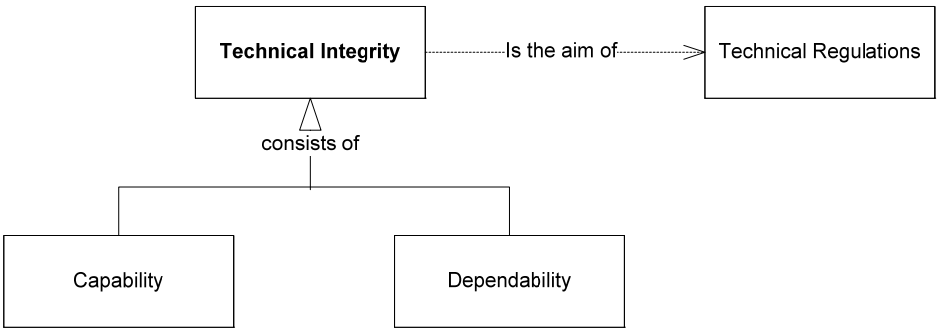


Figure 3. The Definition of Technical Integrity.

Technical integrity differs from the concepts of technical quality, technical scope and requirements achievement used in the SE effectiveness studies reviewed above although there is some overlap. To the extent that *technical integrity* is in part composed of *capability* which is defined in terms of the achievement of requirements in defined specifications and standards, there is definite overlap with the technical quality or technical outcome measures for which estimates in turn rely on proxy measure like KPP achievement and percentage of requirements achieved. But in the area of *dependability* or system safety more commonly, this does not seem to be taken into account in the measures of technical quality used to date. This is not unexpected since often the level of safety to be achieved by a system is not specified. Rather work statements require safety programs to be undertaken where identified risks are reduced to acceptable or As Low as Reasonably Practicable (ALARP) levels. This is really a subjective outcome and it is difficult to compare the system safety outcomes of one system against another, particularly if they are disparate systems destined for use in different environmental and operational contexts.

It is evident then that while the ongoing research work being undertaken to show the project benefits of SE, and that these efforts have some value to established support for the notion that the application of SE results in systems with improved *technical integrity*, there is not yet clear support for this in the published research.

CATEGORISING SYSTEM ENGINEERING METHODS

The aim of this study is to establish what SE methods are directly linked to *technical integrity* goals for systems. But how should SE methods be categorised?

SE is generally defined to include both technical processes and technical management processes. Indeed ISO/IEC 15288:2008 (ISO/IEC 2008) defines "Technical Processes" as well as the technical management processes of "Agreement", "Organizational Project-Enabling" and "Project" processes. At first blush it would seem likely that the technical processes would be most likely to positively influence the attainment of *technical integrity* goals and technical management processes would be more likely to directly influence the cost and schedule outcomes of a project rather than having a direct influence on *technical integrity*. However many technical regulations require both technical processes and technical management processes to be applied to system design and development. So the influence of the non-technical processes cannot be ignored in this study.

Is there an existing categorisation of SE methods that may be re-used for this study?

The ISO/IEC 15288:2008 (ISO/IEC 2008) defines processes that are invoked throughout the lifecycle stages of a system. At first it seemed like this would be the most obvious set of categories to use. However, some of the key studies of SE benefits available do not strictly adhere to the categories that could be directly inferred from the processes in ISO/IEC 15288:2008. The SEI/NDIA Survey of Systems Engineering Effectiveness (Elm et al. 2007) identified 12 areas of SE Capability. By establishing, through expert consensus, the most important artefacts from the (Capability Maturity Model - Integration (CMMI)) and categorising them into 12 identified areas of SE Capability broadly in accordance with CMMI Process Areas. (Honour and Valerdi 2006) presented an ontology that categorises SE Effort into the 8 categories. These categories were derived from a consideration of the information in notable SE standards: ANSI/EIA-632 (1999), IEEE-1220 (1998), ISO/IEC 15288 (2002), CMMI (2002) and MIL-STD-499C (2005). It is noted that some of these standards have been superseded as of the date of this paper. Looking at SE as it may be in the future for increasingly complex systems, (Sheard and Mostashari 2008) have identified potential principles for the development of complex systems that may need to be added to today's SE practices. The INCOSE Systems Engineering Handbook (SEH) (INCOSE 2010) directly follows the ISO/IEC 15288 but in addition covers a number of speciality engineering areas that must be integrated by the SE effort. A comparison of the categories used in the aforementioned sources is presented in Table 3 categorising methods/processes by primary classes of technical, technical management and speciality engineering.

After consideration of this comparison, noting the sources considered by (Honour and Valerdi 2006) and taking its subsequent usage on the SE-ROI project as validation, it was decided to adopt that

ontology for this study. This also allows some ability to directly compare the results of the SE-ROI project to this study. The SE categories used are "Mission/purpose definition", "Requirements engineering", "System architecting", "System implementation", "Technical analysis", "Verification and validation", "Technical management/leadership" and "Scope Management". These categories are described in Table 4.

STUDY METHODOLOGY

What does literature say about SE methods that are necessary and validated to achieve *technical integrity* in the design, development and implementation of a complex engineered system? In this section we identify an analysis approach for the study including a list of key publications defining SE methods and *technical integrity* goals, and a framework for objectively analysing them.

A RELATIONAL ANALYSIS APPROACH

This study follows a formal research methodology termed Relational Analysis, a type of Content Analysis (an excellent overview of Relational Analysis and Content Analysis more generally may be found at (*Writing Guide: Content Analysis*)). The approach derived for this study consists of;

- Identifying the question,
- Choosing the samples for analysis,
- Determining the relationships to be examined,
- Define the concepts and words to be searched for in the samples (coding),
- Defining the relationships between the concepts and the characteristics of the relationships to be evaluated,
- Examination of each of the samples to extract concepts and relationships, and
- Analysis and discussion of the results.

The goal of this study is really about finding the existence of concepts and the characterisation of relationships within the sample documents. The analysis does not seek to find out how often a concept or relationship is mentioned in a given sample, so statistical analysis commonly associated with content analyses will not be produced by this study.

The Analysis Question

The fundamental question to be addressed by the analysis is: *What does the literature indicate as the SE methods that support the attainment of technical integrity of a complex engineered system?*

Samples for Analysis

The categories of literature chosen for the analysis include:

• *Studies* – peer reviewed papers and authoritative reports. The basis for selection is that they provide some rigorous support, ideally quantitative, that SE method(s) lead to the attainment of *technical integrity* goals.

• *Best Practice*s – standards of national and international bodies; handbooks from professional societies, industry and government bodies. The basis for selection is that they provide some support, usually heuristics that are agreed to by a knowledgeable group of practitioners, that SE method(s) lead to the attainment of *technical integrity* goals.

• *Technical Regulations* – regulations, regulatory frameworks and regulation manuals that have as a primary objective the attainment of *technical integrity* goals.

Table 1 shows the samples chosen in each of these categories for this study.

Relationships to be Examined

The aims of the study is to establish what SE methods are related to the attainment of *technical*

integrity goals for systems, where are these SE methods documented and what is the basis for knowing that they do contribute to the attainment of *technical integrity* goals. Further, to establish if Technical Regulations are enforcing the "right" SE methods to attain the *technical integrity* goals. The types of relationships that will be examined are derived as shown in Figure 4.

Coding for Systems Engineering and Technical Integrity Concepts

The *Studies*, *Best Practices* and *Technical Regulations* documents are manually analysed to extract SE methods into the categories that are coded as per Table 4 and to extract SE outcomes that relate to *technical integrity* into the categories that are coded as per Table 5.

Table 1: Samples Selected for this Study.

Studies	Systems Engineering Return on Investment (SE-ROI)	(Honour 2010)
	Does SE In Space Projects Pay?	(Hamann, Zandbergen, and Zijdemans 2009)
	SEI/NDIA: A Survey of Systems Engineering Effectiveness	(Elm et al. 2007)
Best Practices	ISO/IEC-15288:2008 Systems engineering - System life cycle processes	(ISO/IEC 2008)
	Systems Engineering Handbook. A Guide for System Life Cycle Processes and Activities	(INCOSE 2010)
	NASA Systems Engineering Handbook	(NASA 2007)
Technical Regulations	Navy Technical Regulations Manual (NTRM)	(Navy 2003)
	Technical Airworthiness Management Manual (TAMM)	(RAAF 2005)
	Technical Regulation ADF Materiel Manual- Land (TRAMM-L)	(Army 2009)
	Technical Integrity of Explosive Ordnance (DEOP 102)	(Department of Defence 2007)

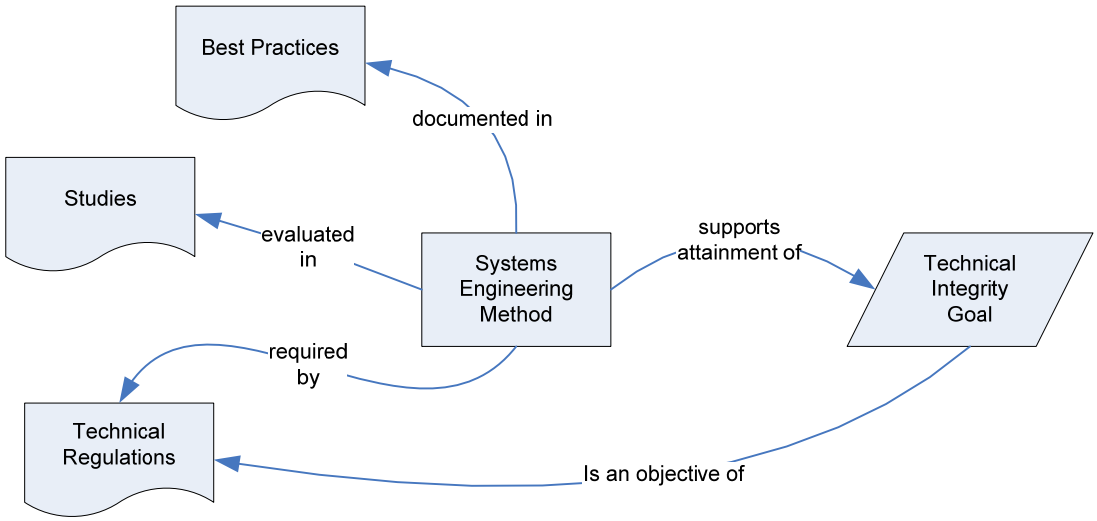


Figure 4. Relationships to be examined.

Relationship Coding, Characteristics and Evaluation

A manual form of proximity analysis is conducted iteratively with the previous analysis step. This really means that extracted SE methods and extracted *technical integrity* goals are examined to see if there is a relationship between the terms expressed in the text.

Relationships in Studies or Best Practices Documents

If a relationship is evident between a SE Method and a *technical integrity* goal in a *Studies* or *Best Practices* document then it is characterised as per Table 2

Table 2: SE Method – Technical Integrity Goal Relationship Characterisation.

Direction	D	**Does the SE Method support or detract from attainment of the *technical integrity* goal?**
	D= +1	Positive support - the SE Method supports the attainment of the *technical integrity* goal.
	D= 0	No support – there is no clear support that the SE Method supports the attainment of the *technical integrity* goal.
	D= -1	Negative support - the SE Method detracts from the attainment of the *technical integrity* goal.
Justification	J	**How well justified is the relationship that the SE Method support or detract from attainment of the *technical integrity* goal?**
	J=e	Evidentiary. There is research support, usually quantitative the SE Method contributes to the attainment of the *technical integrity* goal.
	J=h	Heuristic. There is experience-based support that the SE Method contributes to the attainment of the *technical integrity* goal.
Existence	E	**How does the relationship exist in the document analysed?**
	E=e	Explicit. There is a direct documenting or claim of the relationship between the SE Method and the *technical integrity* goal in the text.
	E=i	Implicit. There is an indirect claim, subjectively discernable by the researcher from the text, of the relationship between the SE Method and the *technical integrity* goal.

The relationship can then be evaluated for its "influence". Where there is a relationship between the SE method and a *technical integrity* goal the *influence*, *I* is defined as:

$$I = f(D, J, E)$$

The *influence*, *I*, if not null (0) can be considered to fall in one of the 8 cells of the chart depicted in Figure 5.

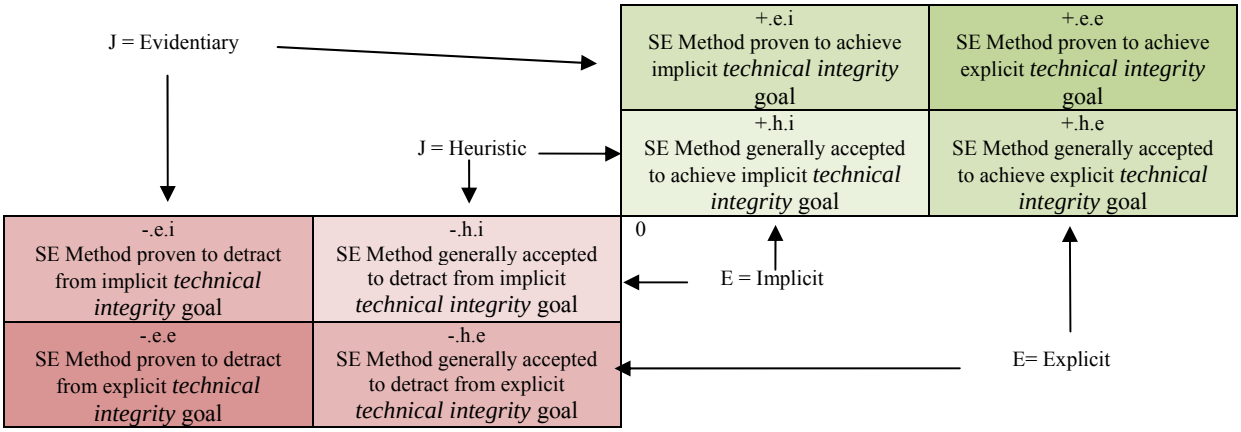


Figure 5. Influence Chart, showing the possible values of Influence, $I = f(D, J, E)$

As implied by the shading, the green squares on the upper right hand side of Figure 5 indicate a positive *influence*. The most positive *influence*, indicated by the darkest green shading, is where we have evidence that a SE method positively contributes to the attainment of an explicitly intended *technical integrity* goal. The red squares on the lower left hand side of Figure 1 indicate a negative *influence*. The most negative and least desirable *influence*, indicated by the darkest red shading, is where we have evidence that a SE method detracts from the attainment of an explicitly identified *technical integrity* goal.

Relationships in Technical Regulations

The relationships then examined in *Technical Regulations* identify what SE methods and processes are prescribed to have a stated or inferred relationship to the *technical integrity* goals nominated in the regulatory document. The extent or coverage of the SE methods that have been prescibed in each of the regulations is coded as per "*Yes*" indicates the method is fully prescribed in the regulations and associated guidance, "*Part*" means some aspects of the SE method are prescribed while "*No*" indicates there is no prescription of the SE method.

Approach to Analysis of Results

SE methods invoked by the *Technical Regulations* will be correlated with the corresponding methods that were found in the analysis of the *Studies* and *Best Practices* documents.

Where there is a SE method invoked by the *Technical Regulations* and there is strong support for the influence of that method on *technical integrity* goals, then we may conclude that the invocation by the regulation is justified.

Where SE methods are not invoked by the *Technical Regulations* but we have found there is strong support for the influence of that method on *technical integrity* goals, then we may conclude that a revision to the *Technical Regulation* could be justified.

Where there is a SE method invoked by the *Technical Regulations* and there is not strong support, or worse negative influence of that method on *technical integrity* goals, then we may conclude that the SE method is a candidate for removal from the regulation and/or a candidate for future research to justify inclusion.

Where a technical integrity goal has no SE methods that influence it, or if the influence is very low, we may conclude that SE (at least as defined by this study) does not contribute to the *technical integrity* goals. In this case other engineering methodologies may be needed to achieve the *technical integrity* goals.

Limitations of the Method

The degree of subject area knowledge and judgement that needs to be applied by the researcher may preclude this method from producing truly objective results if researchers from different backgrounds

used this technique. However, it is asserted that the insights gained from this method, while strictly qualitative and unable to be statistically validated, are of value commensurate with the analysis effort.

ANALYSIS, RESULTS AND DISCUSSION

RELATIONAL ANALYSIS OF SYSTEMS ENGINEERING STUDIES

In this section, the result of the relational analysis for each sample in the Studies is presented. A brief overview of the coverage and methodology employed in the sample document is presented. The outcome of the assessment of the sample's coverage of the SE methods is documented, the relationship between the documented SE methods and technical integrity goals and the assessment of the Influence of that relationship is presented.

SE-ROI

The SE-ROI study (Honour 2010) takes the amount and quality of SE effort applied to a project in each of the SE activity categories and assesses the correlation of that effort against project success factors (cost, schedule, technical quality and overall perception of success).

The SE method categories used in the SE-ROI results are the same as used in this paper (as indicated earlier in this paper) so the coverage is complete.

There is no explicit linkage to the *technical integrity* goals that are the focus of this study. Rather if we assume that the 'technical quality' (being an evaluation of the achievement of key stakeholder requirements) assessed by Honour's study, as a proxy for the *capability* attribute of *technical integrity* we can consider the results as being an evidentiary relationship between the SE method and *capability*. There is no such proxy for *dependability* outcomes in the study.

Assessing the *Influence* of SE then considers that the vast majority of the projects assessed in the SE-ROI study show that specified technical quality - termed the threshold level - was achieved and this achievement was not correlated with the overall amount of SE effort applied (see (Honour 2010) Figure 13). Further, there is no correlation between the SE effort associated with any of the SE methods and technical quality (see (Honour 2010) Figure 14). Small amounts of SE effort applied seem to yield systems that largely meet the specified requirements. The other result is that the attainment of objectives valuable to the stakeholder (e.g., a faster plane) also does not show any correlation to the level of SE effort applied. There appears to be only 2 projects that produce a system with technical quality <1.0 - the threshold acceptability level. There are not any projects examined that have not spent at least ~4% of overall project cost on SE effort. So, from this study we cannot directly ascertain what the technical quality outcome may be if no effort was expended on SE. The relationship between SE effort and technical quality is not shown to be negative, but nor is there a proven positive impact of SE effort on technical quality and thereby to *capability* and overall *technical integrity*.

While there is an evidentiary aspect to the relationship, there is little support to claim a positive influence. Therefore all the SE methods evaluated in the SE-ROI are assessed as *Influence* rating of *I = 0* (SE Method shows no relationship to attainment of the *technical integrity* goals).

SE in Space Projects (SE-Space)

The SE-Space study reported by (Hamann, Zandbergen, and Zijdemans 2009) looked at a relatively small sample set of 9 European research satellite systems. The independent variable in the study was the percentage of SE tasks undertaken from a standard SE checklist. "Technical quality" was assessed on a 5 point scale with 0=Sufficient quality, 2= Very Good Quality and -2=Serious Rework. Overall a result broadly consistent with the results of the SE-ROI project was reported: "positive effects mainly show up in the cost and schedule results of the project, the technical quality of the project result being generally of a rather satisfactory level".

In terms of coverage for the purpose of this study, all the SE methods with the exception of the categories of *Mission/purpose definition* and *Scope management* were assessed in the relational analysis as having been covered by the SE-Space study.

While there are no explicit linkages to the *technical integrity* goals that are the focus of this study, if we assume that the "technical quality" (being an evaluation of the quality of the produced system) as a proxy for the *capability* element of *technical integrity* we can consider the results as being a evidentiary relationship between the SE method and *capability*. There is no such proxy for *dependability* in the SE-Space study.

Assessing the *Influence* of SE then considers that over the small sample set of ~9 European research satellite systems. 8 of the systems resulted in technical quality of 1 or 2 over a range of 20-80% SE tasks undertaken. One resulted in -1 (slight rework) corresponding to about 45% of SE tasks undertaken (see (Hamann, Zandbergen, and Zijdemans 2009) Figure 6). "This result is not surprising: space (scientific) projects are performance driven, that is, the technical result shall always be good at the expense of additional cost and schedule slip". As for the SE-ROI study, the relationship between SE tasks undertaken and the resultant technical quality is not shown to be negative, but nor is there a proven positive impact of increasing the number of SE tasks on technical quality and thereby to *capability* and overall *technical integrity*.

While there is an evidentiary aspect to the relationship, there is little support to claim a positive influence. Therefore the Overall SE in SE-Space is assessed as *Influence* rating of *I = 0* (SE Method shows no relationship to attainment of the *technical integrity* goals). The sub-categories of SE have been rated indeterminate.

SEI/NDIA Survey of SE Effectiveness (SE-Effect)

As discussed in the introduction to this paper the SE-Effect study (Elm et al. 2007) evaluates the effectiveness of SE by looking at the level of SE capabilities applied to a project (where the capabilities are derived from CMMI process areas) and the resultant "Project Performance".

In terms of coverage for the purpose of this study, all the SE methods with the exception of the categories of *Mission/purpose definition* and *Scope management* were assessed in the relational analysis as having been covered by the SE-Effect study.

There are no explicit linkages to the technical integrity objectives that are the focus of this study, however the 'Perf-Scope' (being an evaluation of the degree to which the produced system has satisfied its original requirements) as a proxy for the *capability* element of *technical integrity* we can consider the results as being a evidentiary relationship between the SE method and *capability*. There is no such proxy for *dependability* in the SE-Effect study.

The findings against Perf-Scope presented in the introduction to this paper, indicate that more projects executed with a High Total Systems Engineering Capability score more highly on Perf-Scope than those executed with Low Total SE Capability. However there are no findings currently available to assess the SE method categories individually.

This result provides support for a positive evidentiary aspect to the relationship between overall SE capability applied and a *capability* outcome. Therefore data from the SE-Effect study is assessed as *Influence* rating of *I = +.e.i* for Overall SE (SE Method proven to achieve implicit *technical integrity* goals). The sub-categories of SE have been rated indeterminate at this time.

RELATIONAL ANALYSIS OF SYSTEMS ENGINEERING BEST PRACTICES

In this section, the results of the relational analysis for each sample of the SE Best Practices are presented. A brief overview of the purpose of the sample document is presented. The outcome of the assessment of the sample's coverage of the SE methods is documented, the relationship between the documented SE methods and technical integrity goals and the assessment of the Influence of that relationship is presented.

ISO/IEC 15288:2008

ISO/IEC 15288 (ISO/IEC 2008) "establishes a common framework for describing the life-cycle of systems created by humans," it "defines a set of processes and associated terminology" with "the ultimate goal of achieving customer satisfaction."

Analysis of this standard shows that there is equivalent coverage of each of the SE methods in Table 4.

There is no direct claim (or evidence provided or referenced) for the contribution of these processes to the attainment of *technical integrity*. Rather, implicit in this standard is an experientially derived paradigm, that if not applied is known to often lead to poor project and system outcomes.

The assurance of *technical integrity* that could be obtained by utilising the methods defined in this standard can only be realised if the "Stakeholder Requirements Definition Process" and the "Requirements Analysis Process" completely and correctly identify the requirements that must be achieved in system development to produce a system with technical integrity, i.e., the stakeholder requirements must define all criteria for *capability* and *dependability.*

Therefore the Overall SE category for ISO/IEC 15288 is assessed as Influence rating of I = +.h.i (SE Method generally accepted to achieve implicit *technical integrity* goals). There is little objective means to evaluate the sub-categories of SE methods so these have been evaluated as indeterminate.

INCOSE Systems Engineering Handbook (INCOSE-SEH)

The INCOSE-SEH (INCOSE 2010) "provides an authoritative reference to understand the discipline in terms of content and practice", "is consistent with ISO/IEC 15288:2008" and "further elaborates the processes and activities to execute the processes."

The coverage of the SE method categories defined in Table 4 is complete. Notably this handbook also includes a large section elaborating on Speciality Engineering Activities whose applicability is generally determined by the application domain. The only quantitative justification for the effectiveness of SE is based upon earlier reports from Honour's SE-ROI project that show the effectiveness of SE "in controlling cost and schedule overruns and reducing the uncertainty of project execution".

Similar to analysis of ISO/IEC 15288, the assurance of *technical integrity* by utilisation of the processes within this handbook may only be inferred if the technical integrity criteria are well defined in the requirements processes. Therefore, as for ISO/IEC 15288, the Overall SE for INCOSE-SEH is assessed as Influence rating of I = +.h.i (SE Method generally accepted to achieve implicit *technical integrity* goals). There is little objective means to evaluate the sub-categories of SE methods so these have been evaluated as indeterminate.

NASA Systems Engineering Handbook (NASA-SEH)

The NASA-SEH (NASA 2007) "provides information on systems engineering best practices and pitfalls to avoid." It is not intended to be directive and is built on the basis of a lessons learned philosophy (e.g., adopting recommendations from sources like the Columbia accident investigations). "The objective of systems engineering is to see that the system is designed, built, and operated so that it accomplishes its purpose safely in the most cost-effective way possible considering performance, cost, schedule, and risk."

The coverage of the SE method categories in Table 4 is complete. Notably this handbook also includes a considerable scope on speciality engineering areas like system safety, reliability, human factors to name a few. There is no quantitative justification for the effectiveness of the SE processes elaborated in the NASA-SEH. Rather the implicit justification for the application of SE methods is what has occurred previously if the methods are not followed.

A point of difference to the ISO/IEC 15288 standard and to the INCOSE-SEH is a much stronger emphasis to the goals of the SE processes. The goals explicitly stated cover all of the aspects of *technical integrity* defined earlier. Therefore the NASA-SEH is assessed as Influence rating of I = +.h.e (SE Method generally accepted to achieve explicit *technical integrity* goals). However, there is little objective means to evaluate the sub-categories of SE methods so these have been evaluated as indeterminate.

RELATIONAL ANALYSIS OF ADF TECHNICAL REGULATIONS

In this section we present findings of what SE methods and processes are prescribed by the ADF *Technical Regulations* analysed. Figure 6 shows the coverage of the SE methods that have been found in each of the regulations. 'Yes" indicates the method is fully prescribed in the regulations and associated guidance, "Part" means some aspects of the SE method are prescribed while "No" indicates there is no prescription of the SE method.

Consideration of Figure 6 and its underlying analysis leads to the following observations:

- There is effectively no specificity in any of the regulations as to what element of *technical integrity* they are prescribed to affect. It is therefore to be presumed that all prescribed SE methods are meant to influence all components of *technical integrity.*
- Overall SE is only called out in NTRM and TRAMM-L – both these are focussed somewhat more on design of materiel than is the DEOP 102 and differ from the TAMM which is focussed mainly on "airworthiness".
- Requirements engineering and Verification and Validation are required by all of the regulations. This is explainable if we can assume all the materiel requirements that will, if satisfied, lead to technical integrity have been completely and correctly defined.
- Technical management/leadership is required by all the regulations. Particular emphasis is found on configuration management and risk management and a fair amount on project planning and monitoring processes. It was not anticipated at the outset of this study that the management processes would be seen as key to developing systems with high technical integrity.
- Scope management does not appear at all in the technical regulations. This is understandable since there is presumably no desire amongst technical regulators to regulate commercial/agreement relationships or processes.
- Mission purpose/definition, Systems architecture, System implementation, and Technical Analysis are not all required or not strongly prescribed by any of the technical regulations. This is somewhat surprising since it would be expected that the quality of the architecture and of the implementation would be significant determinants of the technical integrity of the product.

	Required by			
Systems Engineering Processes	**ADF Technical Regulations**			
SE Category	**TAMM**	**NTRM**	**TRAMM-L**	**DEOP**
Overall Systems Engineering	No	Yes	Yes	No
Mission purpose/definition	Yes	No	No	Yes
Requirements engineering	Yes	Yes	Yes	Yes
System architecting	Part	No	Part	No
System implementation	No	No	Part	No
Technical Analysis	No	Part	No	No
Technical management/leadership	Yes	Yes	Yes	Part
Scope management	No	No	No	No
Verification and Validation	Yes	Yes	Yes	Yes

Figure 6. Coverage of SE Methods in ADF Technical Regulations

ADF TECHNICAL REGULATIONS COMPARED WITH VALIDATED SE METHODS

In this section, the previous results are compared. In particular we look to see to what extent SE and

its categorised methods can be shown to objectively influence attainment of *technical integrity*. We then compare this against what the technical regulations require.

Influence of SE on Technical Integrity

Figure 7 shows the *Studies* and *Best Practices* mapped on an Influence Chart according to the overall influence as assessed in the preceding sections. It primarily confirms that there is no direct, objective evidence that the application of SE explicitly and positively influences all aspects of *technical integrity*. The SE-Effect study provides some support for the influence on *capability* but the other studies do not provide any clear support. The *Best Practices* provide mostly implicit support for the notion that SE helps attain required *technical integrity* goals. The NASA-SEH provided somewhat more explicit detail about the objectives of the SE processes it defines which align directly to *capability* and *dependability* aspects of *technical integrity*.

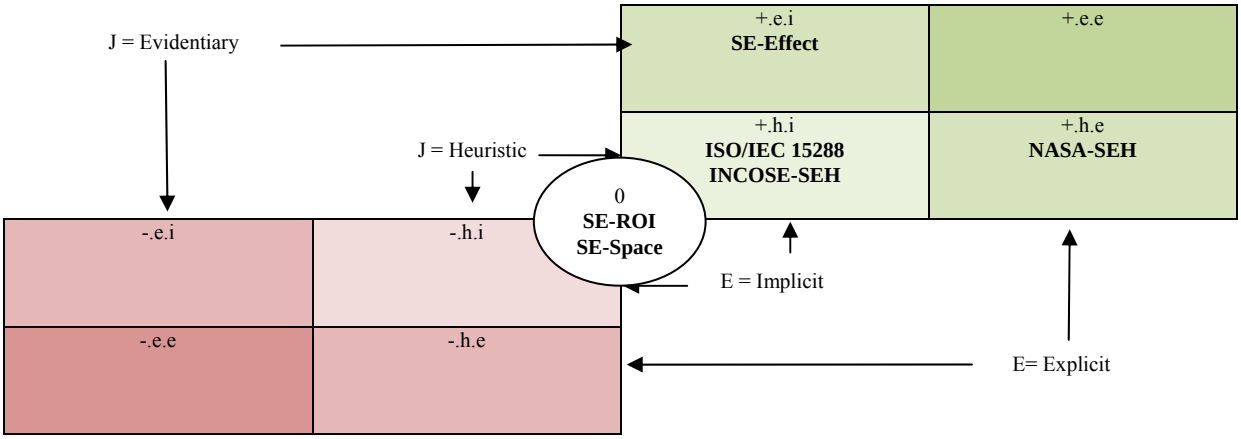


Figure 7. Influence Chart for Studies and Best Practices

SE Influences Compared to Technical Regulation Provisions

Figure 8 presents the assessed *Influence* of the *Studies* and *Best Practices* at a lower level of detail, i.e., against the SE method categories identified for this study. The influences are presented alongside the SE methods required by the ADF technical regulations.

	Required by				Influence, I					
Systems Engineering Processes	**ADF Technical Regulations**				**Studies**			**Best Practice**		
SE Category	**TAMM**	**NTRM**	**TRAMM-L**	**DEOP**	SE-ROI	SE-Space	SE-Effect	15288	INCOSE	NASA
Overall Systems Engineering	No	Yes	Yes	No	0	0	.e.i	h.i	h.i	h.e
Mission purpose/definition	Yes	No	No	Yes	0	x	x	?	?	?
Requirements engineering	Yes	Yes	Yes	Yes	0	?	?	?	?	?
System architecting	Part	No	Part	No	0	?	?	?	?	?
System implementation	No	No	Part	No	0	?	?	?	?	?
Technical Analysis	No	Part	No	No	0	?	?	?	?	?
Technical management/leadership	Yes	Yes	Yes	Part	0	?	?	?	?	?
Scope management	No	No	No	No	0	x	x	?	?	?
Verification and Validation	Yes	Yes	Yes	Yes	0	?	?	?	?	?
						x = Not addressed				
						? = No result available, indeterminate				

Figure 8. Influence Results mapped to SE Methods required by ADF Technical Regulations.

What is first apparent is the relatively sparse result. At the level of the SE method categories there are, in general, no objectively discernable relationships between the method and *technical integrity* goals. Where there are some results at this level, in the SE-ROI results, they do not reveal support for a positive relationship.

The positive, yet weaker than may have been expected, influence of *Overall Systems Engineering*

justifies the provisions for a SE approach in the NTRM and TRAMM-L.

The common focus of the ADF technical regulations for *Requirements Engineering*, *Technical management/leadership* and *Verification and Validation* has indeterminate support in the sampled *Studies* and *Best Practices*.

The main conclusions to be drawn from this analysis are:

- SE methods, in an overall sense, are not strongly, objectively validated to support technical integrity goals. There is some support in the literature mainly based on the heuristic that the application of SE avoids pitfalls in complex system projects.
- There is no negative influence uncovered that suggests SE or any of its practices may detract from technical integrity.
- There is no valid support that any particular category of SE method influences technical integrity more than any other. Therefore it is questionable why some of the SE activities are explicitly required by ADF technical regulations and others are not.

Questions that follow from these analyses then include:

- If there is not strong objective evidence for the impact of SE on technical integrity, how do we know the prescriptions within technical regulations for SE are really required?
- Given the strong support in the literature that (the right amount of) SE supports the achievement of complex system development projects at less cost and schedule than may otherwise be the case, wouldn't it be reasonable to assume that all results-focused development organisations would already be using SE methods? Of course we know they are, so the question is really one of, is it necessary for technical regulations to prescribe SE methods?
- If we take the position that the SE methods have been derived from experience to avoid pitfalls associated with the development of complex systems, then why don't the technical regulations make prescriptions for all the categories of SE methods? Why for instance would System architecting and System implementation methods not be expected to play a significant role in the development of a system with high technical integrity?
- Can an engineering process, consisting mainly of requirements engineering, verification, configuration management and risk management really be enough to assure the technical integrity of a complex system development?

OTHER FACTORS THAT MAY IMPACT TECHNICAL INTEGRITY

Of course, the ADF technical regulations do not prescribe SE (or parts thereof) alone. There is at least as strong a prescription in all ADF technical regulations for

- Materiel standards/specifications to contain the requirements which, if conformed to, will provide technical integrity;
- Competence of design personnel and the design organisation;
- An explicit certification process where the designer is clearly made responsible for the technical integrity of its design by means of formal attestations and that the certification is independently validated; and
- Speciality engineering methods, particularly system safety engineering methods and also domain specific engineering methods in some technical regulations (e.g. engineering for airframe structural integrity in the TAMM).

Certainly the competence and expertise of designers will impact the *technical integrity* of complex systems and lack of requisite competence has been demonstrated to be a contributor to catastrophic failures like Piper Alpha (Bale and Edwards 2000). Similarly it seems reasonable that the other methods prescribed should also contribute to *technical integrity*, but how strong exactly are there contributions?

Technical management/leadership methods as defined in the SE literature are highly overlapped with Project Management (PM) processes. Clearly a project must be funded and managed to completion and transitioned to service for it to have any value to the stakeholders and PM techniques are widely accepted and deployed to this end. There must be some interplay between PM processes that seek to find an optimum balance between cost, schedule and product quality and *technical integrity* goals.

CONCLUSION

This paper posed the question "Is systems engineering necessary for the achievement of *technical integrity*?" to which the answer is a somewhat meek *probably!* There is certainly no evidence to show that it hurts.

Systems Engineering, in constant use and refinement for well over 50 years, is clearly accepted as being the effective methodology for complex system development. Recently, research has been revealing to us that it can be objectively shown to positively impact on cost and schedule outcomes of a project. The impact on technical quality of the developed system is less clear. Despite this lack of clear and strong objective support of technical outcomes, technical regulations in the ADF have in whole or in part, prescribed systems engineering methods as part of a regulatory system to achieve technical integrity of complex defence systems.

So perhaps the question should have been restated as: "Do technical regulations need to prescribe systems engineering for the acquisition and development of complex systems?"

While there is a strong professional belief amongst practitioners that SE positively influences the technical quality of a complex system, the analysis presented in this paper shows there is little objective evidence to justify prescription of SE methods in technical regulations for the goal of *technical integrity*. Rather, when the cost and schedule benefits of SE are considered, all outcome focused, acquisition and design organisations have and will continue to use SE to deliver complex systems projects, it seems of little benefit in having technical regulation also require what is already common practice.

FUTURE RESEARCH DIRECTIONS

Future research will look at complementary and speciality engineering methods to see what ones are recognised and/or proven to contribute to *technical integrity* goals. Risk management techniques, particularly Systems Safety Engineering will be an obvious, initial focus since it is generally invoked by all the ADF technical regulations. The effectiveness of technical governance (the certification process) also requires research as does the impact of individual and organisational competence on the *technical integrity* of complex systems.

REFERENCES

Army. 2009. Technical Regulation ADF Materiel Manual- Land (TRAMM-L): Director of Technical Regulation - Army, Commonwealth of Australia.

Bale, E.A., and D.W. Edwards. 2000. Technical Integrity - An Engineer's View. *Trans IChemE* Vol 78, Part B (September 2000):355-361.

Capability Maturity Model - Integration (CMMI). Software Engineering Institute, Carnegie Mellon University.

Department of Defence, Australia. 2007. Defence Ordnance Publication 102 (AM1) Technical Integrity of Explosive Ordnance.

Edwards, Michael. 2011. What is Technical Integrity and How to Measure It? In *INCOSE International Symposium 2011*. Denver, CO, USA.

Elm, Joseph. 2011. email to author, January 6, 2011. Pittsburgh, PA: Carnegie Mellon University, Software Engineering Institute, CERT.

Elm, Joseph P., Dennis R. Goldenson, Kaled El Emamm, Nicole Donatelli, Angelica Neisa, and

NDIA SE Effectiveness Committee. 2007. A Survey of Systems Engineering Effectiveness - Initial Results: Software Engineering Institute, Carnegie Mellon University and the National Defense Industrial Association.

Hamann, R.J., B.T.C. Zandbergen, and P.J. Zijdemans. 2009. Does Systems Engineering in Space Projects Pay? In *INCOSE International Symposium 2009*. Singapore.

Honour, Eric. 2010. Systems Engineering Return on Investment. In *INCOSE International Symposium 2010*. Chicago, USA.

Honour, Eric, and Ricardo Valerdi. 2006. Advancing an Ontology for Systems Engineering to Allow Consistent Measurement. Paper read at Conference on Systems Engineering Research, at Los Angeles, Ca.

INCOSE. 2010. Systems Engineering Handbook. A Guide for System Life Cycle Processes and Activities.

ISO/IEC. 2008. ISO/IEC-15288:2008 Systems and software engineering - System life cycle processes.

Kludze, Jr. Ave K. . 2004. The Impact of Systems Engineering on Complex Systems. In *Conference on Systems Engineering Research*. Los Angeles, Ca.: Stevens Institute of Technology.

NASA. 2007. Systems Engineering Handbook.

Navy. 2003. ABR 6492 Navy Technical Regulations Manual (NTRM): Commonwealth of Australia.

RAAF. 2005. Australian Air Publication 7001.053(AM1) Technical Airworthiness Management Manual: Director General Technical Airworthiness, Commonwealth of Australia.

Sheard, Sarah A., and Ali Mostashari. 2008. Principles of Complex Systems for Systems Engineering. *Systems Engineering*:17.

Writing Guide: Content Analysis. 2011. Colorado State University [cited 12 Jan 2011]. Available from http://writing.colostate.edu/guides/research/content/index.cfm.

BIOGRAPHY

Michael Edwards holds a Bachelor of Engineering (Honours) and a Master of Engineering Science (Communications) from the University of NSW and a Master of Business Administration (Technology Management) from Deakin University. He is an Engineering Fellow with Raytheon Australia and is a part-time research student at the Defence and Systems Institute of the University of South Australia. He is a Fellow of the Institution of Engineers, Australia, and a Chartered Professional Engineer, a member of the Systems Engineering Society of Australia and a Member of INCOSE. His professional and academic interests build on an extensive experience base in the practice of engineering management and systems engineering largely in the domain of complex naval systems including applications in surface combatant and submarine combat systems, mine warfare systems, hydrographic systems and naval aviation systems.

APPENDIX - REFERENCED TABLES

Large tables referenced in the text are provided in this Appendix.

Table 3: Comparison of Categorisations of Systems Engineering Methods.

Systems Engineering	ISO/IEC 15288	INCOSE Handbook V3.2	SEI SE Capability Elements	Honour/Valerdi Ontology	Sheard/Mostashari Complex SE Principles
Technical Processes	Stakeholder Requirements Definition Requirements Analysis Architectural Design Implementation Integration Verification Transition Validation Operation Maintenance Disposal	Stakeholder Requirements Definition Requirements Analysis Architectural Design Implementation Integration Verification Transition Validation Operation Maintenance Disposal Cross-Cutting Technical Methods Modelling, Simulation, and Prototyping Functions-Based Systems Engineering Object□Oriented Systems Engineering Method	Requirements Development and Management Trade Studies Product Architecture Technical Solution Product Integration Verification Validation Configuration Management	Mission/purpose definition Requirements engineering System architecting System implementation Technical Analysis Verification and validation	Systems Architecting (System Designer) Systems Analysis (System Analyst) Problem-Space-Relevant (Requirements Owner) Configuration Management (Information Manager)
Technical Management Processes	Agreement - Acquisition - Supply Organisational Project-Enabling - Life Cycle Model Management - Infrastructure management - Project Portfolio Management - HR Management - Quality Management Project - Project Planning - Project assessment and Control - Decision Management	Agreement - Acquisition - Supply Organisational Project-Enabling - Life Cycle Model Management - Infrastructure management - Project Portfolio Management - HR Management - Quality Management Project - Project Planning - Project assessment and Control - Decision Management	Project Planning Project Monitoring and Control Risk Management IPT Related Capability	Technical management/ leadership Scope Management	Coordination (Coordinator) Management-Related (Technical Manager)

Systems Engineering	ISO/IEC 15288	INCOSE Handbook V3.2	SEI SE Capability Elements	Honour/Valerdi Ontology	Sheard/Mostashari Complex SE Principles
	- Risk Management - Configuration Management - Information Management - Measurement	- Risk Management - Configuration Management - Information Management - Measurement			
Speciality Engineering Activities	*Not covered*	Design for Acquisition Logistics – Integrated Logistics Support Cost-Effectiveness Analysis Electromagnetic Compatibility Analysis Environmental Impact Analysis Interoperability Analysis Life-Cycle Cost Analysis Manufacturing and Producibility Analysis Mass Properties Engineering Analysis Safety & Health Hazard Analysis Sustainment Engineering Analysis Training Needs Analysis Usability Analysis/Human Systems Integration Value Engineering	*Not covered*	*Not covered*	*Not covered*

Table 4: Systems Engineering Method Categories (after (Honour and Valerdi 2006)) and Coding.

SE Method Category	Description of Category	Coding for Category
Mission purpose/definition	The starting point for the creation of a new system, or the modification of an existing system, is to define the mission or purpose of the new/changed system. This mission is typically described in the language of the system users rather than in technical language	Mission definition Purpose Problem definition Mission statement User requirements Operating concept Operational concept
Requirements	A long-recognized core discipline of systems	Requirements engineering

SE Method Category	Description of Category	Coding for Category
engineering	engineering has been the creation and management of requirements, formal technical statements that define the capabilities, characteristics, or quality factors of a system.	Requirements management Requirements definition Requirements analysis Requirements validation Requirements traceability Requirements tracking Requirements change management
System architecting	The design aspect of systems engineering is to define the system in terms of its component elements and their relationships. This category of effort has come to be known as "architecting," following the practice of civil engineering in which the structure, aesthetics and relationship of a building are defined before doing the detailed engineering work to design the components.	System architecture System concept System design Solution definition Technical solution Solution representation Functional design Requirement derivation Requirement allocation Generation of alternatives Selection of alternatives Trade-off Synthesis
System implementation	The broader view that systems engineering is responsible for the technical development of the system, usually including delivery and installation of the first or prototype version. With this broader view, there are specific efforts related to the system implementation documented in the standards, such as component implementation, system integration, and transition to use.	System implementation System build System integration Component implementation Component build System installation Transition to use System commissioning Product integration
Technical Analysis	System-level technical analysis, particularly as related to assessment of system performance against the requirements. What characterizes each of these types of analysis is the multi-disciplinary scope of the analysis, focused on system emergent properties.	Performance analysis Systems analysis Functional analysis Decision analysis System effectiveness analysis Cost analysis Technical risk analysis Trade off Trade analysis
Technical management/leadership	All standards recognize the inherent need for technical management as a part of systems engineering, an effort required by the size of the engineering teams involved in many system design programs.	Systems Engineering management Technical management Technical leadership

SE Method Category	Description of Category	Coding for Category
	Activities focused on the internal team.	Technical planning Technical control Technical metrics Technical assessment Technical resource management Technical risk management Monitoring and control Baseline control Configuration management Configuration control Change control Interface management Data management Technical review Technical audit Knowledge management
Scope management	Technical definition and management of acquisition and supply issues. Two standards specifically refer to acquisition and supply, while two others explicitly cover subcontract or supplier management. This area of effort applies to the contractual relationships both upward and downward.	Acquisition management Supplier management Technical subcontract management Scope management Statement of work management Supplier agreement Agreement process
Verification and Validation	The standards also agree on the inclusion of verification and validation in the scope of systems engineering. Verification is described as the comparison of the system (or the developmental artefacts) with its requirements through the use of examinations, analysis, demonstrations, tests, or other objective evidence. Validation is described as the comparison of the completed system (or artefacts) with the intended mission or purpose of the system.	Verification Validation Test Trial Demonstration Evaluation Acceptance test

Table 5: List of Technical Integrity and Coding.

Technical Integrity Goal Category	Description of Category	Coding for Category
Capability	Taken to mean that the system meets all specified requirements for performance, function and quality attributes in the defined environment and when used as intended. Capability is the set of emergent properties of a system that is designed (i.e. intended) and useful (i.e. required). The coding presumes the quality attributes are to be objectively specified (e.g. Maintainability in terms of Mean Time to Repair, Maximum Time to Repair, etc).	Fitness for service Fitness for purpose Requirement compliance Standards compliance Performance Accuracy Precision Fidelity Correctness Timeliness Functionality Operability Usability Affordability Compatibility Interoperability Robustness Survivability Durability Supportability Availability Reliability Maintainability Stability Administrability Manageability Recoverability Testability Repeatability Reproducibility Sustainability Upgradeability.
Dependability	Taken to mean that the risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. It includes concepts of robustness (system is tolerant of usage that may exceed specified conditions without catastrophic consequences), reliability (failures do not happen often), and system safety (risk of unidentified hazards and failure modes resulting in undesirable outcomes is minimised).	Safety Personnel safety Product safety System safety Safety risk Mishap risk Accident risk Environmental damage Environmental harm Environmental accident Environmental risk Environment preservation

Appendix I A framework for the evaluation of verification evidence supporting the assessment of environmental qualification

Peer reviewed paper in SETE APCOSE 2012, Systems Engineering Society of Australia: Brisbane, QLD.

This paper examines a relatively narrow aspect of assuring the technical integrity of complex engineered systems. ADF technical regulation requires the establishment of objective evidence to qualify the technical integrity of systems and equipment. In practice, this evidence is often not available and engineering authorities responsible for certification are left with making subjective judgments. This paper contributes a framework for making those judgments consistently.

Moreover, this study provided the researcher with a constrained problem u base the development of survey research instruments, analysis and reporting methods and to enrol expert judges into the problem domain for later research phases. The methods developed by this research were refined and used subsequently to validate and develop the measurement instrument for technical integrity.



A Framework for the Evaluation of Verification Evidence Supporting the Assessment of Environmental Qualification

Michael Edwards

Raytheon Australia / DASI UniSA

michael.edwards@raytheon.com.au

Abstract. The verification evidence available to assess the compliance of Defence materiel against requirements that define the ADF's intended operational environment often does not provide complete objective assurance. Design authorities and certification authorities must then make subjective assessments as to the acceptability of this OQE and the residual risk associated with the materiel when used in the ADF's intended operational environment. This paper contributes a framework, validated by a consensus of expert judgments, which will assist these authorities in consistently making these risk-based engineering judgments.

BACKGROUND

Motivation

Systems acquired for the Australian Defence Force (ADF) may be MOTS/COTS or an integration of the same into a new system. Such programs seek value for money and on-time delivery as an imperative and a justification, in part at least, for a MOTS/COTS approach. Government policy directs that the ADF must consider COTS/MOTS and any decisions to modify these off-the-shelf systems must be with a view to the costs and benefits involved:

"Defence must also take into account what the market can deliver 'off-the-shelf'. The Government needs to make informed decisions about the appropriate mix of cost, risk and capability. Consequently, as highlighted in the White Paper, the Government has decided that Military-off-the-Shelf (MOTS) and Commercial-off-the–Shelf (COTS) solutions to Defence's capability requirements will be the benchmark. Decisions to modify MOTS or COTS solutions will need to be based on a conscious acknowledgement of the costs and benefits to the military outcomes, and the risks involved." (Faulkner 2009)

Environmental qualification of defence materiel is essential to assure the dependability of the materiel when employed in the intended operational environment of the ADF (including climatic, platform induced, threat induced, and electromagnetic environments). Many systems and equipment available to the ADF as off-the-shelf purchases are likely to have been qualified by the parent defence force or original designer during system and engineering development.

Designers integrating systems with MOTS/COTS components and certification authorities responsible for identifying and advising the operational users about the residual technical integrity risks associated with these systems, commonly face the following questions and issues:

- Is the parent defence force operational environment the same or similar enough to the ADF's intended operational environment?
- Equipment has often evolved through use with parent defence force and may not have environmental qualification data available for its currently provided configuration.
- Does a long field history with a parent defence force mean the equipment is fit for service

with the ADF?

- The original evidence and test reports from the environmental test program may not be readily available to evaluate the verification results against the ADF's operational environment.
- Many of the benefits of COTS/MOTS (i.e., cost and schedule) are eroded if the equipment has to be requalified.

Consequently, there is commonly a cost-technical integrity risk trade-off in accepting MOTS/COTS for use. The risk is often categorised as not knowing with high certainty if the equipment is capable of dependable operation in the ADF operational environment. This paper addresses a need for a consistent framework that designers and certification authorities can use to evaluate the available verification evidence for qualification of environmental requirements.

Environmental Requirements

In this paper we use the term "environmental requirements" to mean the specified requirements that define the intended operational environment that the ADF intends to use the system in. Environmental requirements include:

- Climatic conditions (e.g., temperature ranges, humidity ranges, wind, precipitation, lightning, etc.)
- Platform induced conditions (e.g., vibration, platform motion, etc.)
- Threat induced conditions (e.g., shock, blast, fire, etc.), and
- Electromagnetic conditions (interference and susceptibility limits).

Guidance is available for establishing a set of environmental requirements, e.g., MIL-HDBK-2036 (DoD 1999) is commonly used as a starting point for defining the environmental requirements for maritime, land, air and space based systems.

Environmental Qualification Programs

Standards like MIL-STD-810 (DoD 2000) define how to plan and conduct an environmental qualification program for new equipment.

Many systems developed for large militaries like the USA undergo a development lifecycle that may include a sequence of prototypes, Engineering Development Models, an environmental qualification program, Techeval, Opeval, with different levels of verification recursion after modifications. Once the capability is mature, a production contract is let followed by operational use which can lead to ongoing post-production modifications. Typically, many years later these systems may be made available as MOTS to other nations. But what is the environmental qualification status at this point? Often the environmental qualification status cannot be recovered or provided to give objective assurance to the acquisition authority for it to consider the qualification limits of the equipment versus their intended usage of the system.

In another common occurrence, equipment may be offered as MOTS, but the manufacturer takes the opportunity of a new order to modify the base MOTS configuration to rectify obsolescence or as part of a product line evolution. While these are desirable or necessary modifications, it does have the side effect of moving the equipment away from its environmental qualification basis.

But to reap the cost and schedule benefits of COTS/MOTS we do not want to have to repeat an environmental qualification program unless the level of risk (the unknown fitness for service) warrants the cost.

Evidence of Environmental Qualification for COTS/MOTS?

So, for a COTS/MOTS procurement, what is typically available as evidence to substantiate compliance with an environmental requirement?

Typically there are no explicit test results for the exact condition and exact equipment configuration

being acquired. From no evidence in some cases, to a range of "evidence" including attestations of acceptability for other defence forces, anecdotes of service history, test results from previous versions of the system, extension analyses or analyses from first principles of varying applicability.

The perennial issue is how do we know if this "evidence" is enough or adequate? How much assurance does it give us for certifying the system for use in the ADF's intended operational environment?

The objective of the remainder of this paper then is to develop a framework to help objectify the engineering risk-based decision making that goes on when considering the acceptability of offered "evidence", the residual technical integrity risk, and the determination of whether any specific qualification testing is required to characterise the COTS/MOTS for employment by the ADF.

RESEARCH APPROACH

It is hypothesised that there exists a framework that describes a spectrum of confidence in ascribing compliance to an environmental requirement. The level of confidence is related to the type and extent of the "evidence" provided in the claim of compliance. The aim of this framework is to better identify the level of risk in design certification decisions and to explicate the assumptions made by the design certifier and design acceptance authorities.

The general approach is to conduct survey research that will provide the data to form the framework (effectively survey research feeding a grounded theory study (Leedy and Ormond 2005)). In particular:

- Identification of the types of evidence most commonly offered for claims of compliance and construction of a survey instrument, and
- Elicitation of expert judgements from senior engineers practiced in the technical certification of materiel for the ADF to evaluate the types of evidence in terms of level of confidence with which they would ascribe a finding of compliance.

Expert Judgement

An expert is a person who is suitably qualified and whose background and practice in the subject area is recognised by his or her peers. Expert judgements are often sought when it is impractical to directly measure the quantities of interest. The approach in this paper is to use expert opinion to organise qualitative and quantitative data into a framework for making consistent compliance assessments. The methods for eliciting and analysing expert judgement are well developed (see, e.g., (Meyer and Booker 2001)) and will not be explained in detail in this paper.

Survey Instrument

The survey instrument developed consists of a web-based questionnaire seeking demographic information and a set of ratings and rankings against questions that relate an expert judgement of confidence in ascribing compliance versus the level or type of evidence offered for verification. The full survey instrument is provided at Annex A. The order of the cases within each question was randomised for each judge to minimise bias that could be induced by consideration of each alternative in a set order.

SURVEY RETURNS

Return Rate

Sixty-seven experienced engineers, professionally known to the researcher, were asked to participate in the web-based survey. At survey closure, 38 of 67 (57 %) had fully completed the survey.

Professional Experience

A histogram of the professional experience of the respondents is shown in Figure 1. The average experience of the respondents is 24 years and the median is 24.5 years. This represents a very experienced set of professional engineers.



Figure 1. Histogram of the Professional Experience of Survey Respondents

Employment

The categorised current employer of each the respondents is shown in Figure 2. This demonstrates a representative spread of engineers from government and commercially employed engineers.

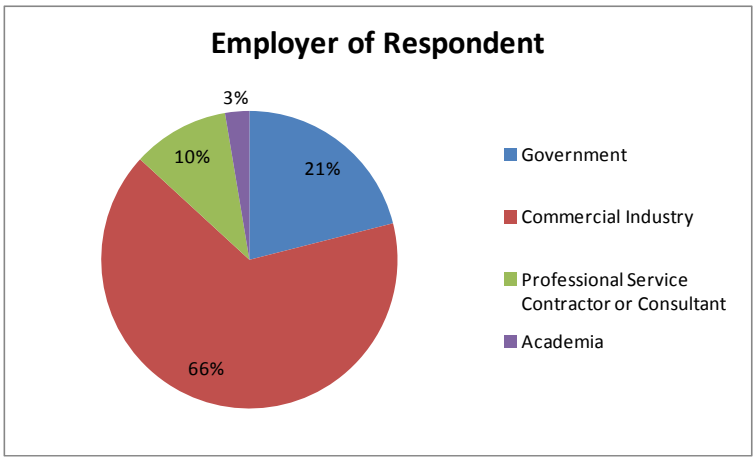


Figure 2. Percentage of Respondents by Current Employer Category

ADF Engineering Authority Delegations

Seventy-nine percent of the respondents indicated that they hold or have held formal delegations of Engineering Authority in organisations governed by ADF technical regulations. The delegations are shown in Figure 3. This again demonstrates the respondents represent a set of very experienced engineers and that they are, or have previously been, entrusted to make significant engineering decisions for the ADF. There is also representation across the three technical regulatory domains of the ADF (maritime materiel– regulated under the Naval Technical Regulatory System (NTRS); aerospace materiel – regulated by the Technical Airworthiness Regulator (TAR); and land materiel under the Technical Regulation of ADF Material – Land (TRAM-L)).

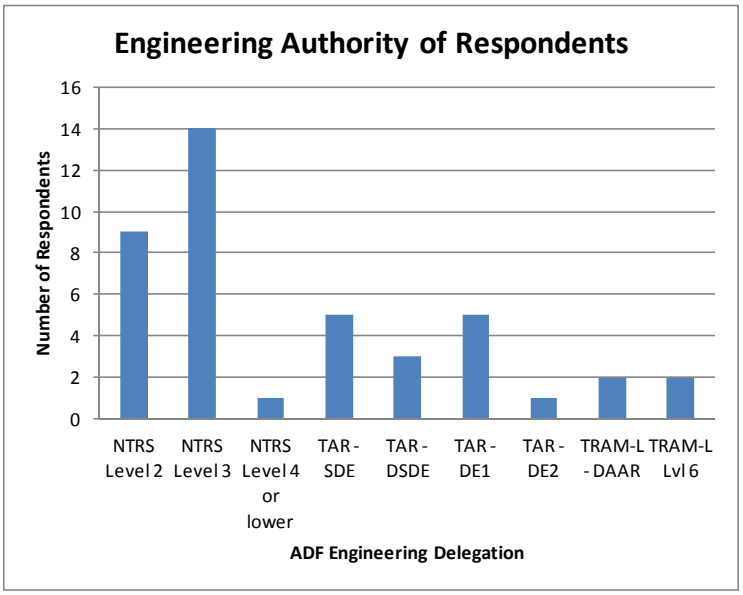


Figure 3. Histogram of ADF Engineering Authority Held by Respondents

DATA ANALYSIS

Note that the survey data collected is rating or rankings. These form an ordinal scale of measurement and therefore the analysis of them requires the application of nonparametric statistics. The estimates of central tendency we use are the median and mode and the estimate of variability is the interquartile range (the range in which the middle 50% of ratings fall). Averages and variances have no valid meaning on ordinal data.

Confidence in Service History

The judges were each asked "If you were given a verification analysis that cited service history of an equipment as the only evidence, how would you assess the compliance of the equipment?" (see the survey instrument at Annex A for the exact cases they were asked to consider – a short form of each appears in the figures and discussion here). shows a raw view of the judge's ratings. It gives a good feeling for the spread of the judge's responses and which types of service history evidence are considered more confidently for arguments of compliance than others.

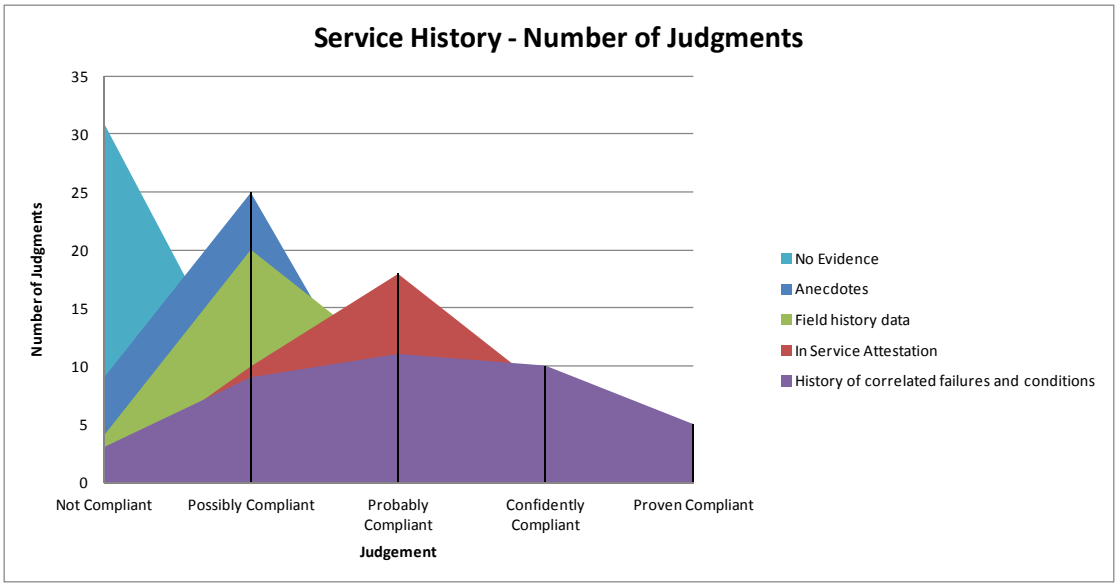
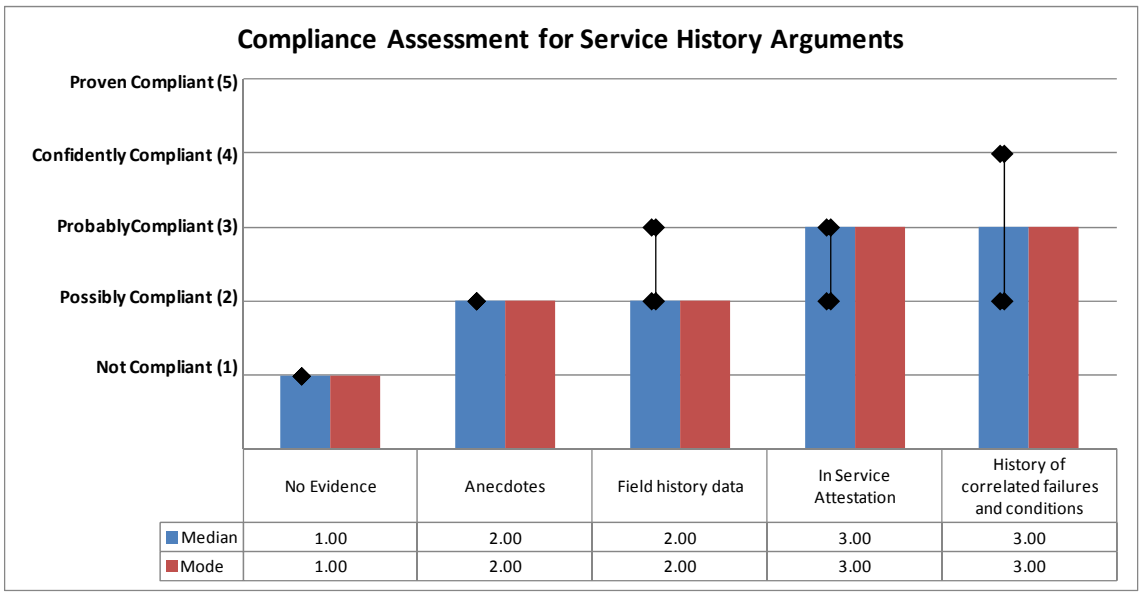


Figure 4. Number of Judgements versus Type of Service History Evidence.

Figure 5 presents an aggregated view of the judgements. This chart shows the mode and the median of the judges' responses. The error bars indicate the interquartile range around the median.



	No Evidence	Anecdotes	Field history data	In Service Attestation	History of correlated failures and conditions
Median	1.00	2.00	2.00	3.00	3.00
Mode	1.00	2.00	2.00	3.00	3.00

Figure 5. Aggregated Judgements on Compliance Assessment for Service History Evidence (Median with Interquartile Range and Mode)

The aggregated judgment suggests that even the best form of service history data – evidence that includes field history data where failures are correlated against experienced environmental conditions – is only considered in the "Possibly Compliant" to "Confidently Compliant" range by a consensus of expert judgment. This indicates that Service History Arguments as verification for compliance do not lead to high confidence in a compliant finding.

Confidence in Verification by Testing

The judges were each asked "If you were provided test reports that claimed to verify an environmental condition requirement how would you rate the compliance of the equipment if the report(s) showed a …" (see the survey instrument at Annex A for the exact cases they were asked to consider – a short form of each appears in the figures and discussion here). Figure 6 shows a raw view of the judge's ratings. It gives a good feeling for the spread of the judge's responses and which types of test outcomes and evidence are considered more confidently for arguments of compliance than others.

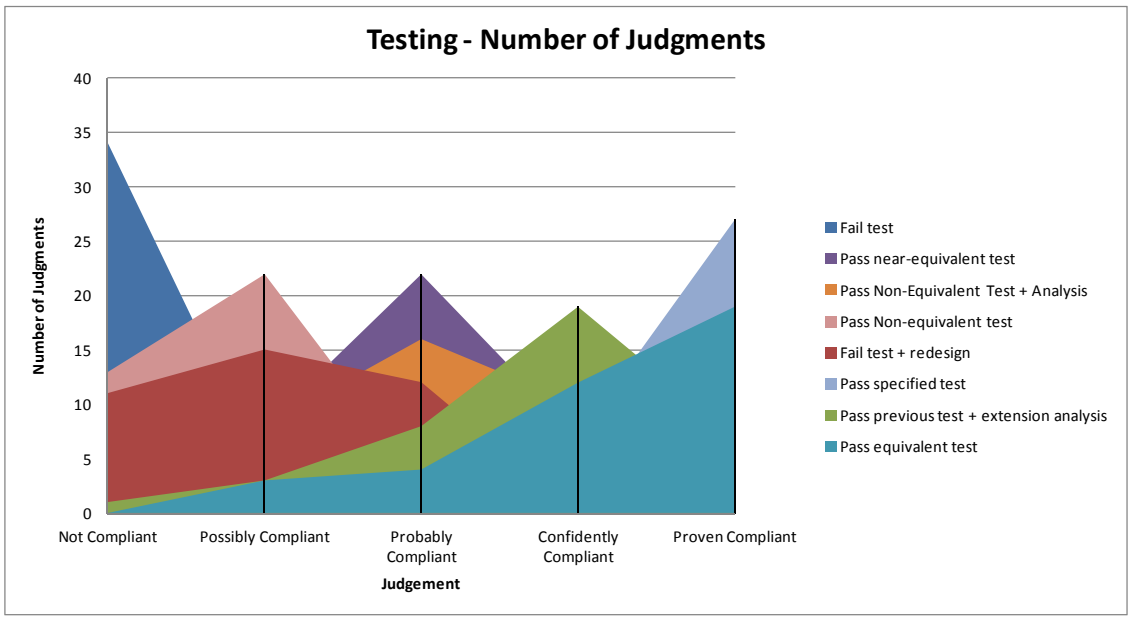


Figure 6. Number of Judgements versus Type of Testing Evidence.

Figure 7 presents an aggregated view of the judgements. This chart shows the mode and the median of the judges responses. The error bars indicate the interquartile range around the median.

The aggregated judgment data suggests that only passing a specified test or an equivalent test are considered as "Proven Compliant". These are the only types of verification evidence considered in this survey that attain this highest level of compliance confidence. As expected, a test failure culminates in a judgement of "Not Compliant". Redesign after test failures or forms of extension analysis provide for intermediate values of confidence in finding compliance.

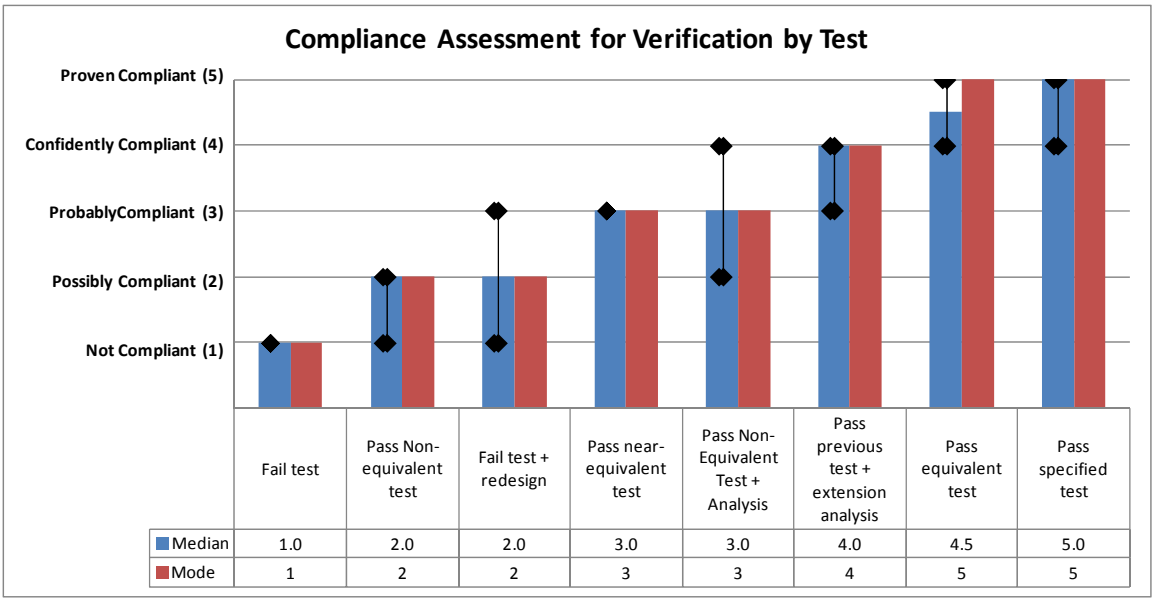


Figure 7. Aggregated Judgements on Compliance Assessment for Verification by Test (Median with Interquartile Range and Mode)

Confidence in Attestation

Attestation is defined as the "Issue of a statement, based on a decision following review, that fulfilment of specified requirements has been demonstrated" (Hunter 2009). In this paper we use the term Attestation to include forms of evidence including product data sheets, warranties and certifications.

The judges were each asked "If an environmental condition requirement is verified by audit of certificates or other forms of attestation, how would you rate the compliance of the equipment if the audit showed:" (see the survey instrument at Annex A for the exact cases they were asked to consider – a short form of each appears in the figures and discussion here). Figure 8 shows a raw view of the judge's ratings. It gives a good feeling for the spread of the judge's responses and which types of attestation are considered more confidently for arguments of compliance than others.

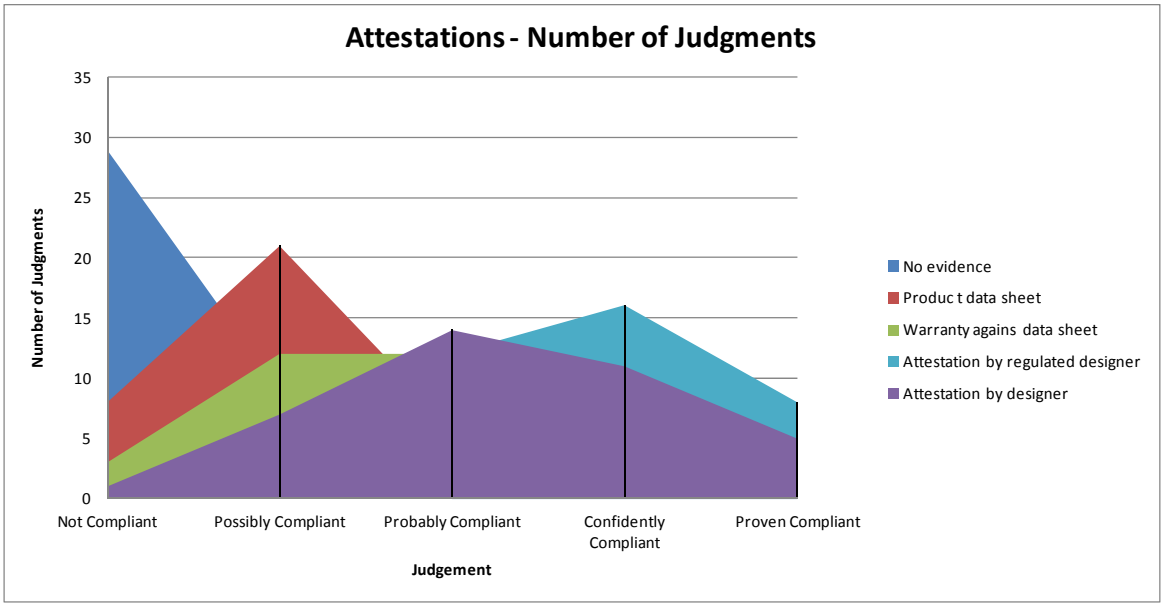


Figure 8. Number of Judgements versus Type of Attestation Evidence.

Figure 9 presents an aggregated view of the judgements. This chart shows the mode and the median of the judges' responses. The error bars indicate the interquartile range around the median.

The aggregated judgment suggests that the best form of attestation available is that "Documented attestation by a regulated designer (e.g. by an ADF accredited AEO)". This reflects the positive influence the respondents feel that technical regulation affords over a design certification not covered by technical regulation. The wide interquartile range on supplier warranties reflect that some respondents consider a warranty as quite confidence building while most consider it only "Possibly Compliant". There is almost complete agreement that no evidence yields a finding of Not Compliant.

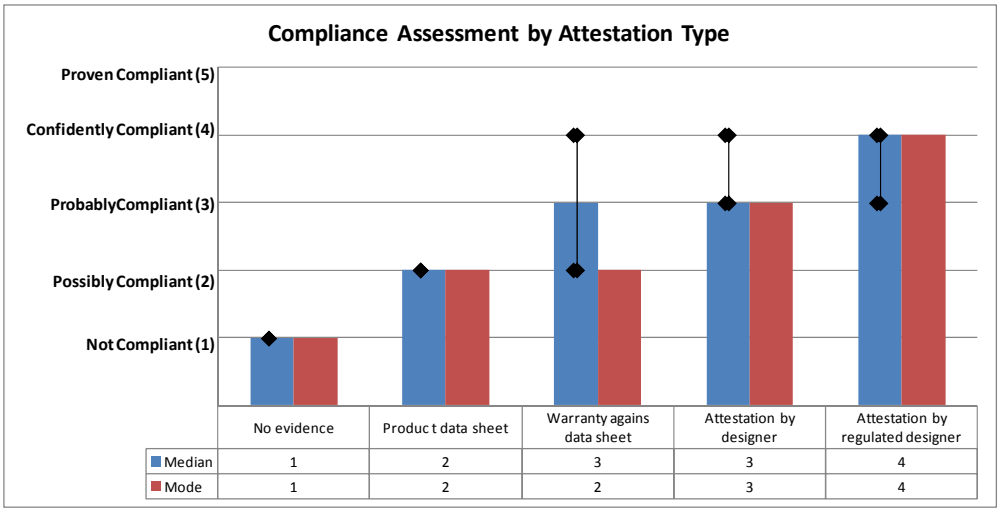


Figure 9. Aggregated Judgements on Compliance Assessment by Attestation (Median with Interquartile Range and Mode)

Confidence in Verification by Types of Analysis

The judges were each asked "How would you rank (highest to lowest) the quality of the following analysis approaches:" (see the survey instrument at Annex A for the exact cases they were asked to consider – a short form of each appears in the figures and discussion here). Figure 10 presents an aggregated view of the judgements.

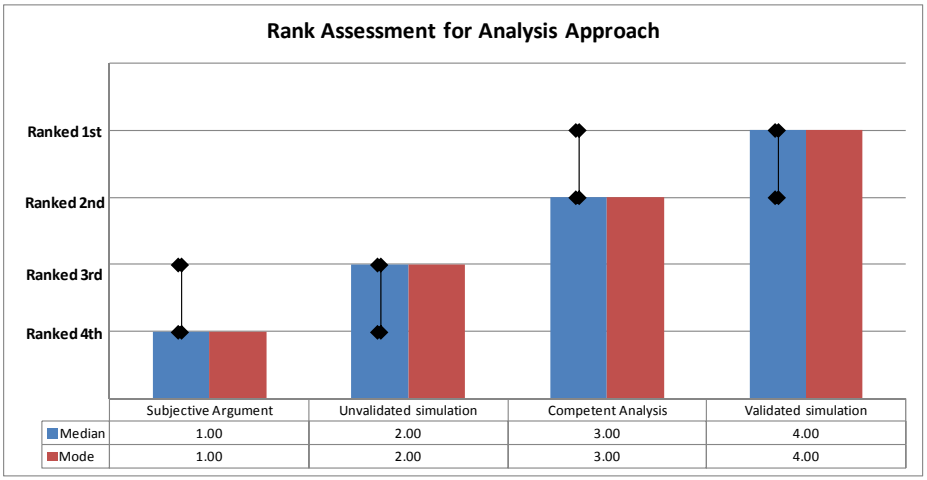
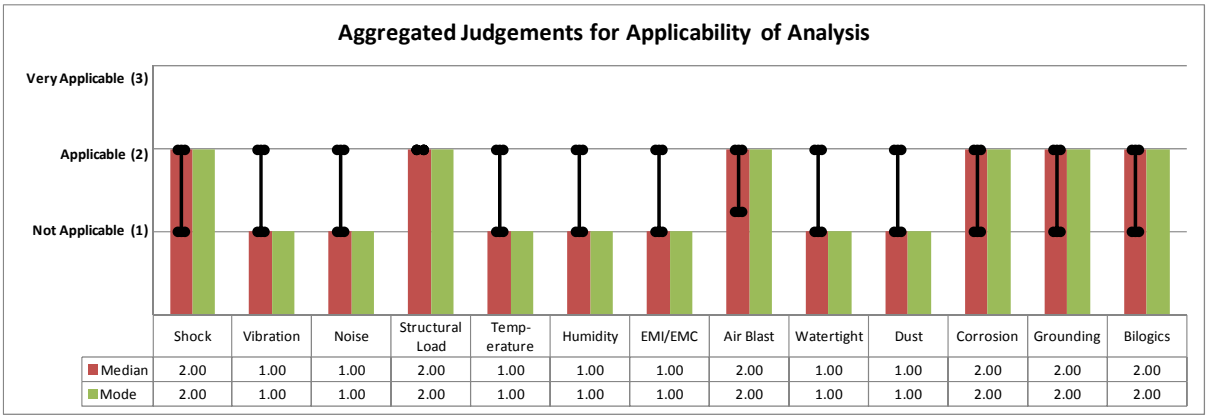


Figure 10. Aggregated Rank Judgements on Confidence in Types of Analysis Approach (Median with Interquartile Range and Mode)

"Validated computer simulation" and "Mathematical analysis undertaken by competent professional" are clearly assessed as the most highly considered analysis approaches. Not surprisingly "Subjective argument offered as analysis" was ranked the lowest as measured by the mean but "Bespoke, unvalidated computer simulation" was only considered a little more highly by the judges (higher median value but the same interquartile range).

Appropriateness of Analysis to Environmental Requirement Qualification

The experts were asked to rate the applicability of analysis as a verification technique to various types of environmental conditions (shock, vibration, etc, as shown in Figure 11). Applicability was requested on a 3 point rating scale of Not Applicable, Applicable and Very Applicable. Aggregate ratings are considered using, median (shown with interquartile range) and mode of the entire set of judgements.



	Shock	Vibration	Noise	Structural Load	Temp-erature	Humidity	EMI/EMC	Air Blast	Watertight	Dust	Corrosion	Grounding	Bilogics
Median	2.00	1.00	1.00	2.00	1.00	1.00	1.00	2.00	1.00	1.00	2.00	2.00	2.00
Mode	2.00	1.00	1.00	2.00	1.00	1.00	1.00	2.00	1.00	1.00	2.00	2.00	2.00

Figure 11. Aggregated Judgements for the Applicability of Analysis for Various Types of Environmental Conditions.

The results imply that the consensus view is that verification by analysis is not applicable for environmental conditions of vibration, noise, temperature, humidity, EMI/EMC, watertightness and dust. Environmental conditions where the consensus view is that verification by analysis is applicable includes shock, structural load, air blast, corrosion, electrical grounding and resistance to biological degradation. There is no consensus that verification by analysis is "very applicable" for any of these environmental conditions (i.e., no interquartile range, where 50% of the judgements lie within, include "very applicable").

There are some counter intuitive results revealed. For example, the consensus view that analysis is applicable for verification of shock but that it is not applicable for vibration is at odds with the increasingly common approach to create finite element analysis models of structures which can equally well demonstrate both shock and vibration performance of the design. This may indicate either: that there is an opportunity for more education about the role of analysis as a valid verification technique for some environmental requirements; or, indicates the experienced designers and certification authorities really need to be more insistent on the provision of evidence other than analysis in order to obtain high confidence in compliance finding.

Confidence in Independent Engineering Assessment

The experts where asked "If you were provided with an engineering judgement independent from the designer/manufacturer, affirming the claim of the designer/manufacturer, how would this affect your assessment of compliance?" and asked to respond against a three point rating scale of Lower my assessment, No effect and Raise my assessment. Figure 12 shows a clear majority of the respondents believe that independent engineering assessments provide additional confidence to judgements of

compliance.

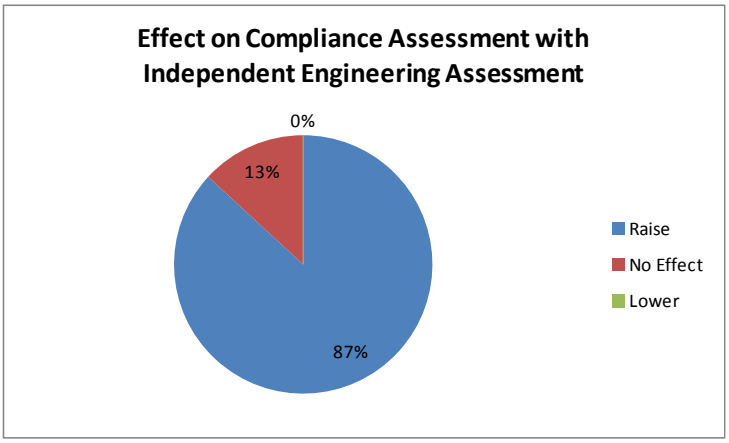


Figure 12. Effect on Compliance Assessment of Independent Engineering Assessment.

Inter Judge Agreement

A statistical measure of the degree to which the judges agree in Kendall's coefficient of concordance (W). This is appropriate to ordinal data: ranks or ratings transformed to ranks. It also allows the ready calculation of the average Spearman correlation coefficient ($\overline{r_s}$) and the null hypothesis (i.e. that there is no agreement between the judges) may be readily tested using the Chi-square distribution as defined in Section 10.4 of (Howell 2010).

Table 1 shows the calculated values for W and $\overline{r_s}$ for each judgment group. A value of W=0 indicates no agreement between judges and a value of W=1 indicates complete agreement. $\overline{r_s}$ may also range between -1 and 1, where 1 indicates complete positive correlation of all the judge ratings, 0 indicates no correlation in the ratings and -1 indicates a complete negative correlation of all the judges ratings. The final column indicates if the null hypothesis (i.e., that is there is no agreement between the judges) can be rejected at the 0.01% level. i.e. "Yes" means that the data supports us being 99.99% confident there is some level of agreement between the judges.

Table 1. Kendall's Coefficient of Concordance, the Average Spearman Correlation Coefficient and Null Hypothesis Testing Results for each Group of Judgments.

Judgement Group	W	$\overline{r_s}$	Null Hypothesis Rejected (0.01%)
Service History	0.65	0.64	Yes
Verification by Testing	0.69	0.68	Yes
Attestation	0.71	0.70	Yes
Types of Analysis	0.82	0.82	Yes
Applicability of Analysis	0.21	0.19	Yes
Independent Engineering Assessment	0.66	0.65	Yes

It is generally considered that $0.6 < W < 0.8$ indicates a strong agreement and $W > 0.8$ indicates a very strong agreement among the judges. We have very strong agreement for the ratings on Types of Analysis and strong agreement for all other judgement groups with the exception of Applicability of Analysis. While the null hypothesis has been rejected, the level of inter judge agreement at W = 0.21 is quite modest. This is not surprising given the somewhat counter-intuitive results we saw earlier for

Applicability of Analysis and the relatively wide spread of judgements shown in Figure 11.

Overall, again excepting Applicability of Analysis, the inter judge agreement results provide confidence that we can use the aggregated judgements to synthesise a valid framework.

THE RESULTANT FRAMEWORK

The aggregated expert judgements presented in the earlier sections of this paper, have been synthesised into a framework that presents the confidence in ascribing compliance of a system to an environmental requirement based on the type of evidence presented. The resultant framework is shown at Figure 13. Confidence in ascribing compliance is rated on an ordinal scale from *Not Compliant* to *Proven Compliant*. The ratings of *Possibly Compliant*, *Probably Compliant* and *Confidently Compliant* are the areas where engineering judgment is required. The validated framework provides a tool for consistently making those judgements, enabling a consistent identification of the level of risk in design certification decisions and explicates the assumptions made by the design certifier and design acceptance authorities.

Service History Evidence is shown as the weakest form of evidence. Anecdotes of succesful field history provide little confidence in compliance assessment. Even the provision of actual field data for the reliable operation of the system in defined environmental conditions only leads to a assessment confidence of *Probably Compliant*.

Most notable in the framework is the fact that only verification by test evidence is considered to lead us to the highest confidence in finding compliance (*Proven Compliant*), and then only when it is for the specified or quivalent test conducted on the same configuration of equipment. Testing of previous configurations with extension analyis extrapolating the results yields *Confidently Compliant*. Non-equivalent testing gives some lower level of assurance – *Probably Compliant*. Failed tests, if the fault is rectified, but the equipment is not re-test only provide *Possibly Compliant* assurance. Failed test without rectification of design or construction confirms a level of *Not Compliant*.

Little weight is given to the attestation offered by a product data sheet; only an assurance level of *Possibly Compliant*. Commercially enforceable provisions against a product data sheet does increase the confidence in making a compliant assessment. Formal attestation by a designer provides more confidence and infers a belief amongst the expert judges that a professional designer's integrity in making an attestation of compliance gives them more assurance than a commercial warranty. Further still, an attestation by a designer working under a technical regulatory environment (e.g., within an ADF Authorised Engineering Organisation (AEO)) provides assurance at the *Confidently Compliant* level.

Analysis Evidence provides reasonable confidence when it is shown to be calculations conducted by competent professionals or the result of a validated simulation. Subjective arguments passed in as analyses and unvalidated simulations or models lead to low confidence – *Possibly Compliant* – in finding compliance.

Significantly, the framework demonstrates that many of the commonly proffered items of evidence (unquantified service history data, subjective arguments, vendor's product data sheets, etc.) provide little confidence to support a finding of compliant.

LIMITATIONS AND ACCEPTABILITY

Of course, this framework is limited to giving the designer or certification authority a valid tool to support his judgment on finding a system to be compliant with an environmental requirement or not. It does not attempt to define if, for example *Probably Compliant*, is sufficient for acceptance of the system or not. To reach this acceptability decision, the decision maker needs to consider the risk of the system failing or being non-dependable in the intended operational environment (since there is a finite chance of the system being not compliant) versus the costs of gaining more confidence in compliance finding (by conducting further tests, seeking more data or conducting more analyses). This is a regular cost-benefit trade-off where the benefit is reduction in uncertainty of compliance.

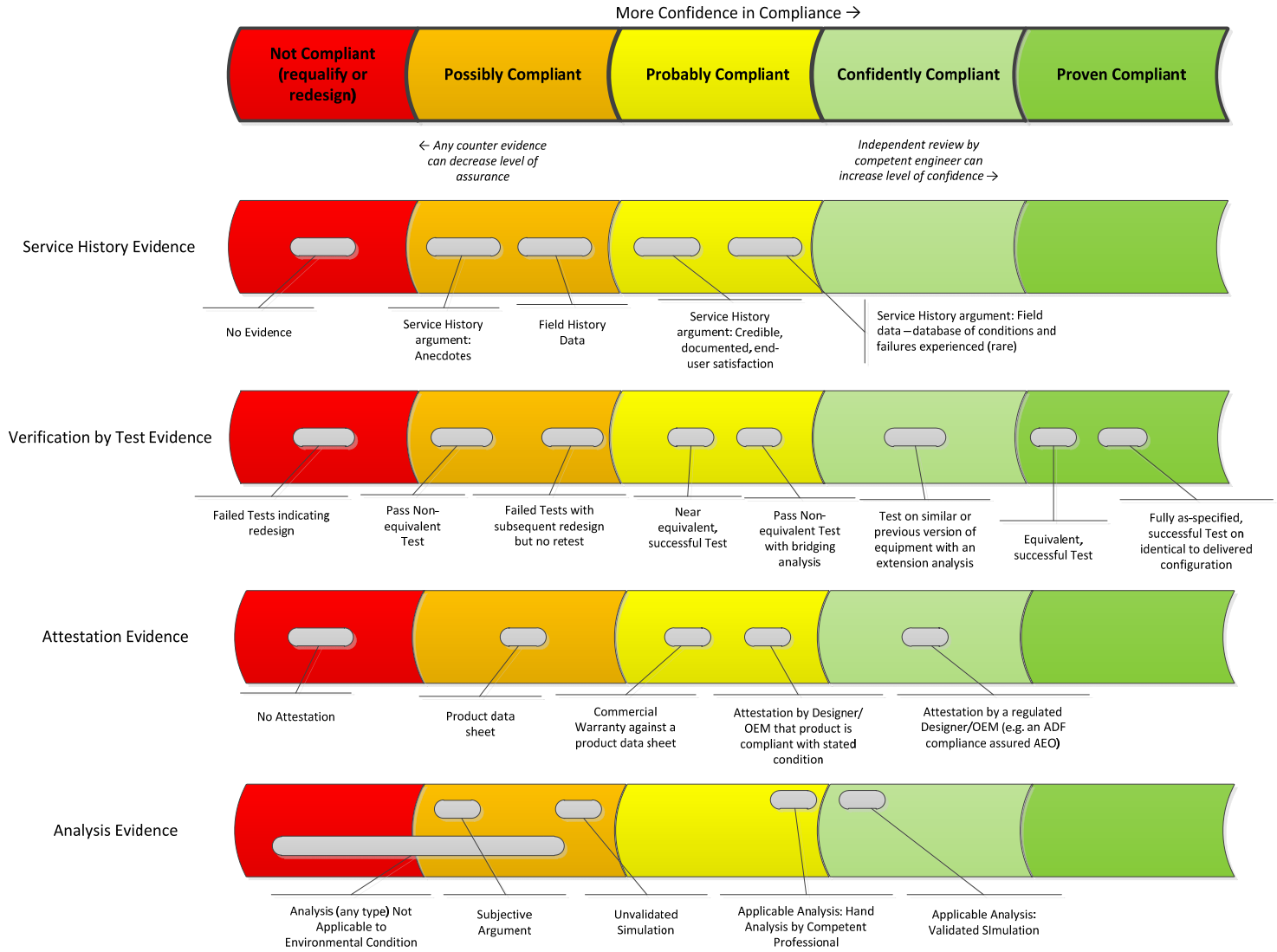


Figure 13. A Framework for the Evaluation of Verification Evidence Supporting the Assessment of Environmental Requirement Qualification

CONCLUSION

The more assured we can be of a system's compliance, the less risk there is that the system will fail to operate as required in the intended operational environment.

Conversely if we can more objectively identify the level of uncertainty we judge in that compliance assessment, the better we can assess the need for more verification and the more clearly we can communicate the residual risk to the end users of the system.

This paper contributes a framework, validated by a consensus of expert judgments, which will assist system designers and design certification authorities in consistently making risk-based engineering judgments of a system's compliance to specified environmental requirements.

REFERENCES

DoD (1999), 'MIL-HDBK-2036', *Preparation of Electronic Equipment Specifications* (Department of Defense, USA).

--- (2000), 'MIL-STD-810F', *Test Method Standard for Environmental Engineering Considerations and Laboratory Tests* (Department of Defense, USA).

Faulkner, John 'Minister of Defence's Forward to "Defence Capability Plan 2009"', <http://www.defence.gov.au/dmo/id/dcp/html/foreword/index.html>, accessed 19 Nov 2011.

Howell, David C. (2010), *Statistical methods for psychology* (7th edn.; Australia Belmont, CA: Thomson Wadsworth) xxi, 768 p.

Hunter, Robert D. (2009), *Standards, Conformity Assessment, and Accreditation For Engineers* (Boca Raton, FL, USA: CRC Press, Taylor and Francis Group).

Leedy, Paul D. and Ormond, Jeanne Ellis (2005), *Practical Research. Planning and Design* (8 edn.; Upper Saddle River, NJ: Pearson Education International).

Meyer, Mary A. and Booker, Jane M. (2001), *Eliciting and Analyzing Expert Judgment*, eds Richard F. Gunst, et al. (ASA-SIAM Series on Statistics and Applied Probability: ASA SIAM).

BIOGRAPHY

Michael Edwards holds a Bachelor of Engineering (Honours) and a Master of Engineering Science (Communications) from the University of NSW and a Master of Business Administration (Technology Management) from Deakin University. He is an Engineering Fellow with Raytheon Australia and is a part-time research student at the Defence and Systems Institute of the University of South Australia. He is a Fellow of the Institution of Engineers, Australia, and a Chartered Professional Engineer, a member of the Systems Engineering Society of Australia and a Member of INCOSE. His professional and academic interests build on an extensive experience base in the practice of engineering management and systems engineering largely in the domain of complex naval systems including applications in surface combatant and submarine combat systems, mine warfare systems, hydrographic systems and naval aviation systems.

Annex A- Survey Instrument

System Environmental Requirements Compliance Assurance

Question 1 -

Years of Professional Experience

Question 2 -

Current employer

Question 3

Do you have or have had an engineering delegation under ADF technical regulatory arrangements?

Question 4

If you have held a delegation of engineering authority, please indicate all the delegations you hold or have held.

Question 5 - Rating Scale

If you were given a verification analysis that cited service history of an equipment as the only OQE, how would you assess the compliance of the equipment?

- OQE contains service history anecdotes only
- OQE contains credible end-user attestation of acceptable service history
- OQE includes field history data
- OQE includes field history data where failures are correlated against experienced environmental conditions.
- No OQE

Question 6 - Rating Scale

If you were provided test reports that claimed to verify an environmental condition requirement how would you rate the compliance of the equipment if the report(s) showed a:

- Pass of a validated, equivalent test
- Fail of a test without subsequent rectification
- Pass of the specified test on previous equipment versions with credible extension analysis
- Fail of a test with consequent redesign, but no retest.
- Pass of a near-equivalent test
- Pass of a non-equivalent test with a bridging analysis

- Pass of the specified test on the as-delivered equipment configuration
- Pass of a non-equivalent test

Question 7 - Rating Scale

How appropriate do you believe verification by analysis can be for the following environmental conditions? (Not Appropriate, Appropriate, Very Appropriate - as good as test)

- Mechanical shock
- Vibration
- Noise
- Structural Loading
- Operational Temperature Range
- Operational Humidity Conditions
- EMI/EMC
- Air blast effects
- Water tightness
- Dust resistance
- Corrosion resistance
- Electrical Grounding
- Resistance to degradation by biologics

Question 8 - Ranking

How would you rank (highest to lowest) the quality of the following analysis approaches:

- Mathematical analysis undertaken by competent professional
- Subjective argument offered as analysis
- Validated computer simulation
- Bespoke, unvalidated computer simulation

Question 9 - Rating Scale

If an environmental condition requirement is verified by audit of certificates or other forms of attestation, how would you rate the compliance of the equipment if the audit showed:

- No attestation evidence

- Product data sheet only
- Commercially enforceable warranty against a product data sheet
- Documented attestation by the designer/manufacturer that the equipment is operable in stated environmental condition
- Documented attestation by regulated designer (e.g. by an ADF accredited AEO)

Question 10 - Choice

If you were provided with an engineering judgement independent from the designer/manufacturer, affirming the claim of the designer/manufacturer, how would this affect your assessment of compliance?

- Lower my assessment
- No effect
- Raise my assessment

Appendix J What is technical integrity and how to measure It?

Peer reviewed paper in INCOSE Symposium 2011, International Council On Systems Engineering: Denver, CO.

This paper presented the initial outcome of the concept development phase of the research. A conjoint measurement construct, formulated from the concepts of capability and dependability, and approaches to measure capability and dependability are identified and a list of indicators that may be used in a measurement instrument for technical integrity have been refined.

.

What is Technical Integrity and How to Measure It?

Michael Edwards
Defence and Systems Institute, University of South Australia
edwmt002@mymail.unisa.edu.au

Copyright © 2011 by Michael Edwards. Published and used by INCOSE with permission.

Abstract. The impact of formal technical regulation and certification frameworks on the technical integrity of complex systems is poorly understood. In researching this problem the author is seeking a means by which to measure *technical integrity*. *Technical integrity* is defined and discussed, in both the context of complex systems for the Australian Defence Force and in wider civilian and military contexts. *Technical integrity* is defined as being composed of two separable attributes that reflect the intended, desired emergent properties of a system and the minimisation of unintended, undesired emergent properties. An approach for formulating a composite measure of *technical integrity* is developed from this definition and is presented for critical review.

Introduction

The author's area of research is concerned with complex engineered systems and inquiry into the impact of technical regulation on the technical integrity of these systems particularly during the design and development phase of the system lifecycle. A framework for this research was reported in (Edwards 2009) and this paper presents part outcomes of that ongoing research.

In the Australian Defence Force (ADF), *technical integrity* is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection, i.e. its compliance with technical standards and the minimisation of unintended consequences. *Technical integrity* has various different meanings and has synonyms and related concepts occurring in other civilian and military system contexts. The first part of this paper briefly reports the results of a survey of these various meanings and related concepts in order to establish a working *technical integrity* definition that is useful across a range of complex engineered systems.

At this time, the intended positive impact of technical regulation on the technical integrity of a complex engineered system remains largely based on the application of heuristics of what should be done but without quantitative support. One of the key issues in establishing such support is the need to establish some sort of measure for *technical integrity*. With a measure established it would be possible to establish the benefit (i.e., the amount of additional *technical integrity*) obtained by the application of technical regulation, or the degree thereof. The current phase of the research is thus focused on finding support for the hypothesis: *Systems developed under Technical Regulation have higher Technical Integrity*.

The second part of this paper therefore tackles the problem of establishing a measure of technical integrity by first assessing existing methods for measuring similar properties of a complex engineered system and then by selecting a form of measure and formulating a new

approach for a composite measure of technical integrity. This approach is presented for critical review.

What Is Technical Integrity of a Complex Engineered System?

There are many definitions for *technical integrity*, many related concepts, usages of terminology and stakeholders for the desired and undesired emergent properties of complex engineered systems. Some of the related terms and concepts include: suitability for service; safety; certified products; system effectiveness; airworthiness (in the aerospace domain); systems assurance; fitness for purpose; technical assurance; product quality; compliance with class rules (in the maritime domain); compliance with build standards; mission assurance; and dependability among many others.

Some definitions include concepts of cost or economic value, while others are independent of cost or economic value. Typically all involve the concept of avoidance of unintended negative consequences (particularly safety). Typically most involve the concept of doing what was intended to be done (fitness for purpose/service, quality, mission assurance) and to have that done for most of the time (available, reliable).

Technical Integrity in the Australian Defence Force

ADF Materiel. In the Australian Defence Force, *technical integrity* is the term used to encompass the concept of a system's fitness for service, safety and compliance with regulations for environmental protection (ADF 2010). Materiel is defined as "All items including ships, tanks, self-propelled weapons, aircraft etc and related spares, repair parts and support necessary to equip, maintain and support military activities without distinction as to its application for administrative or combat purposes."

The three main components of *technical integrity* are further defined as:

- Fitness for service: The systems ability to satisfy operational requirements.
- Safety: Freedom from those conditions that can cause death, injury, or occupational illness.
- Environment Protection: Poses no hazard to the environment.

Concepts of Technical Integrity in a Wider Context

The term *technical integrity* is not commonly used across all industries or government enterprises where complex systems are engineered, maintained and put to use. This section initially looks at a few examples where the term technical integrity is used and compares these with the usage in the ADF. Then a number of related concepts and the terminology used is presented and contrasted.

Technical Integrity in the Commercial Industry. *Technical integrity* or just *integrity* is not widely used in commercial industry. One area where it is used is in the area of process plants. For example, Woodside uses the definition "Integrity – achieved when risk of failure that endangers personnel, environment or asset value is tolerable and is reduced to ALARP" (Hackett 2008) and PI Asset Integrity Ltd state "Technical Integrity in any process plant is essential to ensure that there will be no harm done to people or the environment. Assuring Technical Integrity of your facility by regular interval audits makes good business sense". (*Technical Integrity Audits* 2007). (Bale and Edwards 2000) definition is "Technical integrity is concerned with the development of the design such that it is carried out by well trained

personnel, who have been assessed to be competent, in accordance with recognized, sound practices and procedures and such that there is adequate provision by way of reviews and audits, to ensure the design intent is unimpaired in any way that could cause undue risk or harm to people or damage to the environment."

Interesting in these definitions is the direct parallel with the ADF's components of personnel safety and environmental compliance as well as the contrast between the ADF's Fitness for Service and the concepts of "asset value" and "good business sense". Clearly the latter may be explained by the mission outcomes as required by a commercial entity. That is, creation and maintenance of value to further the business.

Objectives in Regulations for Design and Operation of Systems. The design and operation of significant and complex systems is often regulated to prevent unwanted, unintended consequences by governments or by international organisations with multi-national government support. The objectives of these regulations may be thought of as technical integrity objectives. Examples include:

- Lloyds Rules Ship Classification: Compliance with a set of "rules" or other "technical standards" that have been shown to be appropriate for the objective of safe and compliant ships.
- The U.S. Department of Energy's policies and guidelines for acquisition of capital assets which include the regulations for acquisition of highly safety and security critical nuclear power assets (DoE 2006, 2008).
- The United States Navy's platform level and strike force (formally battle force) level Naval Warfare Systems Certification Policy (NAVSEA 2005) has as its stated technical objectives "Mature, Safe and Reliable Warfare Systems".

In the above few examples we see again the concepts of safety, compliance, maturity and reliability as technical objectives of regulation. A larger range of technical regulations have been examined in order to define a basis for study and to structure an ontology of technical regulation as reported in (Edwards 2010).

MIL-STD-882C Safety Compared to ADF Technical Integrity.

MIL-STD-882C (DoD 1993) defines Safety as "Freedom from those conditions that can cause death, injury, occupational illness or damage to or loss of property or damage to the environment" and identifies "operational effectiveness, time and cost" as constraints. There is close to an exact parallel when compared to the ADF's *technical integrity:* "damage to the environment" and "Freedom from those conditions that can cause death, injury, occupational illness" in MIL-STD-882C is the same respectively as Environment Protection and personnel safety as defined in the ADF's technical regulations (ADF 2010).

The one point of difference is between the "freedom from those conditions that can cause ... damage to or loss of property" of MIL-STD-882C and the "Fitness for Service" concept of the ADF's technical regulations. The latter is fully inclusive of the MIL-STD-882C concept, but as we have seen in previous sections of this paper also considers the achievement of regulatory requirements and capability requirements. Failure to meet the latter requirements in particular can be treated as a risk to mission achievement or as a mishap where the outcome is the inability to conduct certain missions or aspects of the intended mission.

A Definition of Technical Integrity

Consideration of the technical integrity and system safety concepts reviewed previously lead to a conclusion that there are two separable aspects to technical integrity of a complex engineered system. The bases for the separation are the emergent properties of the system – "useful" and "destructive" emergent properties (see (Edwards 2009), Table 1 that extends the work of (Honour 2007) to consider the influence of technical regulations on system emergent properties). The "useful" basis implies that the system meets all the requirements and standards that have been prescribed for it. We term this property of the system as the *Technical Integrity* attribute *Capability*. The "destructive" basis implies the broad notion of minimising the unspecified, but clearly undesired properties of the system. We term this property of the system as the *Technical Integrity* attribute *Dependability*. This definition is illustrated in Figure 1.

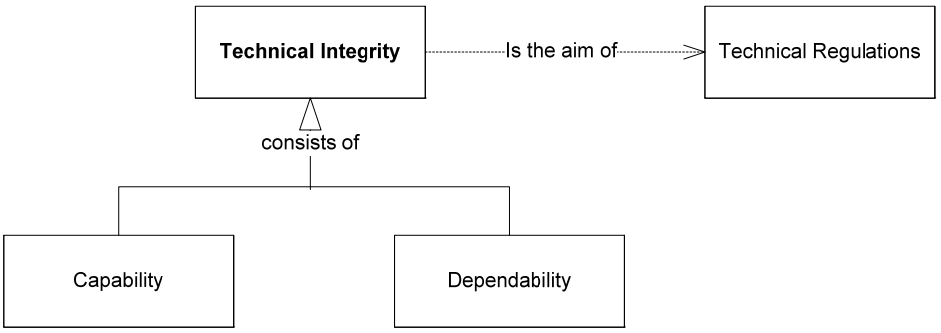


Figure 1. The Definition of Technical Integrity

More directly, *Technical Integrity* is defined as consisting of *Capability* and *Dependability*, where:

Capability is taken to mean that the system meets all specified requirements for performance, function and quality attributes in the defined environment and when used as intended. *Capability* is the designed set of emergent properties of a system that are intended and useful (i.e., required); and

Dependability is taken to mean that the risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. It includes concepts of robustness (system is tolerant of usage that may exceed specified conditions without catastrophic consequences), reliability (failures do not happen often), and system safety (risk of unidentified hazards and failure modes resulting in undesirable outcomes is minimised).

It should be noted that these terms may be used differently in different contexts. *Dependability* is often used in the literature but without a universally agreed definition. For instance, (Prasad 1998) reviews many definitions of *dependability* before arriving at a definition useful for his work, albeit with some differences to that defined herein.

The definition proposed in this paper can be seen to encompass the scope of *technical integrity* as used in the ADF. Fitness for Service is equivalent to *Capability* and Safety and Environment Protection are hyponyms of *Dependability.*

Can Technical Integrity Be Measured?

Motivation to Measure Technical Integrity

Having defined *technical integrity*, the remainder of this paper is inspired (or perhaps cursed?) by Lord Kelvin who is reported to have said: "I often say that when you can measure what you are speaking about, and express it in numbers you know something about it, but when you cannot measure it, when you cannot express it in numbers your knowledge is of a meagre and unsatisfactory kind: it may be the beginning of knowledge, but you have scarcely in your thoughts, advance to the stage of science whatever the matter may be".

The hypothesis being investigated by the author is that: *Systems developed under technical regulation have higher technical integrity*. To find support for this hypothesis, it could be imagined that a number of case studies could be evaluated for the *technical integrity* of the produced system and this could be plotted against the amount of technical regulation that was applied to the system development. Data supporting the hypothesis may look like that presented in **Error! Reference source not found.**.

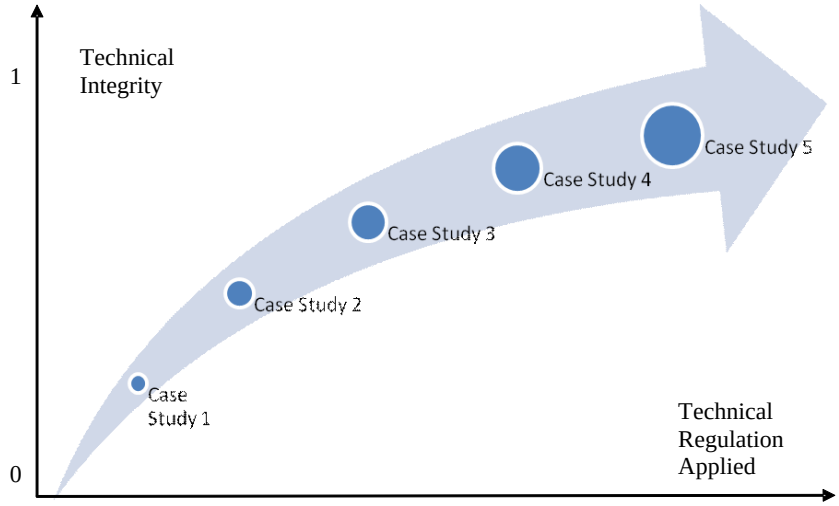


Figure 2. Concept of data supporting the hypothesis: *Systems developed under technical regulation have higher technical integrity*.

Fundamental therefore to any attempt of quantitative research into the impact of technical regulation is the ability to measure in some sense the resultant technical integrity of a complex engineered system.

Can we settle on a measure, a set of measures or some composite measure that directly indicate the *technical integrity* of the system? The measures need to be general enough so that they are applicable across the class of systems of interest and be able to be supported by quantitative or qualitative data that can be practically obtained for those systems. The measurement goals that have been set are: the measure(s) is objective; ideally the measure forms a ratio scale, but at least forms an ordinal scale; the measure is applicable across a wide range of complex engineered systems.

What Support for Such a Measure Already Exists?

Measures of System Properties. It has long been posited that there is no single overall measure for the quality or integrity of a complex engineered system (see for example (Boehm 1978)). "System Effectiveness" for a complex system, to the extent that it can be thought of as a partial proxy for *technical integrity*, most often appears as a combination of capability, reliability and availability in a value hierarchy (Parnell, Driscoll, and Henderson 2008). While, a system effectiveness measure is usually constructed for supporting selection decisions between alternate systems there is little in the way of comparative measures available to compare the utility of disparate systems.

(Hitchins 2003) nominates a set of common candidates for measuring the value of a system. His candidates include: life cycle cost; efficiency; effectiveness; performance; availability; survivability; process (e.g., simplicity, resource consumption, ...); product (e.g., utility, quality, fitness for purpose, …); entropy (the degree of disorder in the system). (Barbacci et al. 1995) also provides a useful set of software quality attributes encompassing performance, dependability, security, and safety. In "Measuring the Life-Cycle Value of Enduring Systems", (Browning and Honour 2008) take the view that a system's value can only be determined by the stakeholders who receive some utility from the system. These examples all present a view that there is a range of system qualities or attributes that should be able to be measured in some sense, but cover a broader concept than *technical integrity*.

The increasing focus in recent years on systems architecture as a technique in the development of complex engineered systems often cites as a self justification that the measurement of system attributes akin to *technical integrity* is, if not impossible, at least not tractable. For example, (Maier and Rechtin 2009) state that "architecture complements science where science is weakest – dealing with immeasurables of highly complex systems", and "most important factors are not measureable: worth, safety, affordability, political importance, environmental impact ... and have no realistic basis for numerical analysis." The problem of dealing with the attributes of "ultra-quality" systems is also salient, for instance how to measure 100% availability, zero defects, etc. It is posited in this case that measurement can only be indirect, that an analytic design approach alone does not achieve ultra-quality, and heuristic approaches must also be applied

"Qualities do not have to be expressed 'qualitatively' – they should be quantified". (Gilb 2005) dedicates a chapter of his book, about the methodology of "Planguage", to scales of measurement. He tackles concepts often claimed to be immeasurable like "user-friendly" "Environmentally-friendly" and "Customer satisfaction" by specifying scales of measure. He posits principles including "If you can't define a scale of measure, then the goal is out of control" and "If you can't quantify it, you can't control it". I would therefore speculate with some confidence that Gilb's views on technical integrity would be that "defining a scale of measure is a reasonable practical process".

Proxies for Technical Integrity. In considering the measurement of *technical integrity* it is worth considering if there are any other measures that are offered as a proxy for *technical integrity* or at least could be considered as a valid full or partial proxy. One particularly interesting set of data suggests consideration of accident rates of a complex system in service be a proxy for *technical integrity*? Figure 3 shows an example of the yearly number of accidents being used to demonstrate the positive impact of technical and operational regulation on the fatal accident rate of ADF aircraft.

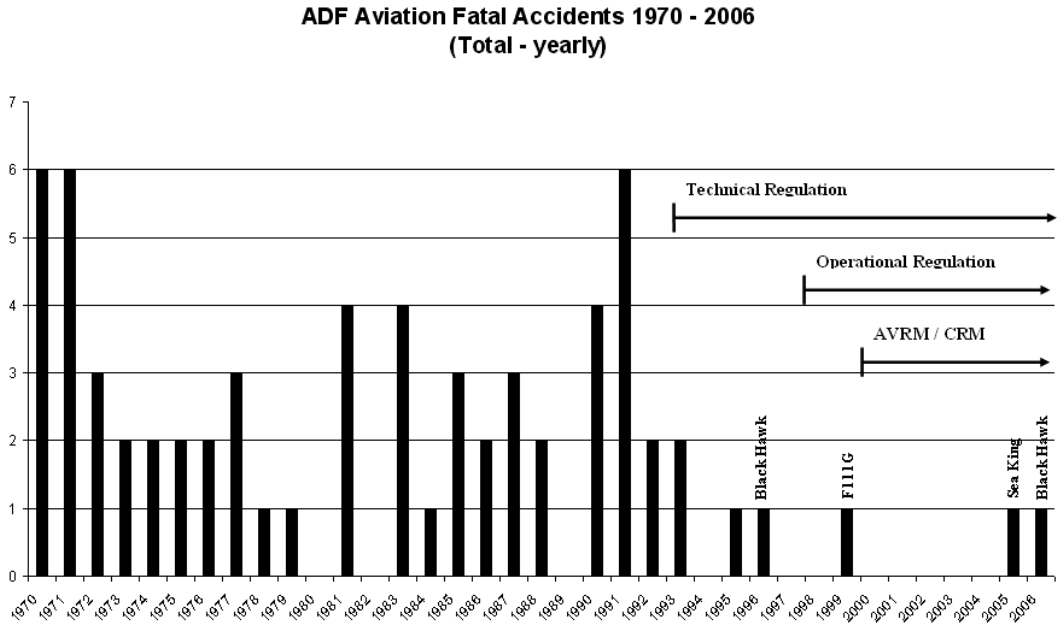


Figure 3. Total number of ADF fatal aviation accidents over the last 37 years (RAN 2007)

It is feasible that these numbers could be taken to form a measurement of the accident rate of the entire ADF aviation capability over a defined time frame and that this could have some validity as a proxy measure for *technical integrity.* However, it is worth noting that this type of measure may be easily confounded by issues like the numbers of aircraft in service in any particular year, the technology of the aircraft in service, the experience level of aircrews, the type of flying (operations versus training), sustainment spending and perhaps other variables. All these would need to be carefully controlled for in using this as a proxy measure.

Measurement of Technical Integrity is Feasible. There is established precedent for decomposition approaches to evaluate system properties and proxy measurements for properties of complex systems that are broadly similar to *technical integrity*. While it may not be straightforward, we have taken the view that measurement of *technical integrity* is feasible and desired and we therefore progress to seeking such a measure.

Selecting a Form of Measurement

This section looks at selecting a form of measurement for *technical integrity.* Measurement theory (see for example (Sydenham 1982)) has been considered in the development of the approach but the formal mathematical background and constructs are omitted for brevity in this account.

Forms of Measurement. The following forms of measurement are available for consideration:

- Direct measurement. The class of measurement including extensive measurement where something like a physical attribute of an object can be directly observed (e.g., a linear measurement).
- Indirect measurement, including:

- Derived. Establishing a relation between two or more things that can be directly measured (e.g., density which is established from the direct measurement of volume and mass).
- Associative. Measurement of at least one other property and the relationship of that to the desired property (e.g., measure resistivity to determine temperature).
- Multi-dimensional. A collection of component vectors without combination (e.g., a vector that defines the lengths of the 3 sides of a triangle).
- Conjoint. A combination of component properties which establishes an ordering of the desired property based on measurable component properties (e.g., the establishment of a "comfort" measurement by combining temperature, wind speed and relative humidity measurements).

Selection of a Measurement Form. We now consider the available forms of measure and the suitability of those for *technical integrity.* Clearly a direct extensive measurement of *technical integrity* is not possible so direct measurement is ruled out. Consideration of indirect measurement forms concludes that:

- Derived measurement is rejected since there is no known law for combining any measurable properties of a system to evaluate technical integrity.
- Associative measurement is rejected since there is no empirically known relationship between technical integrity and any other directly measureable property of a system.
- Multi-dimensional measurement is rejected since it cannot produce an ordinal, interval or ratio scale that is desired for the intended research.
- Conjoint measurement is accepted as the most appropriate measurement form for *technical integrity*. It requires some decomposition or at least identification of factors that contribute to *technical integrity* and the definition of mathematical relationships to combine these measurements into an empirical order.

Conjoint Measurement Approach

It is to be noted that we have not identified any empirically and objectively determined combination of system attributes that we can say represents *technical integrity.* But this should not be an obstruction to the formulation of a useful measure. We do however need a formalism based on measurement theory and logical arguments. The theory for conjoint measurement is not new, having been established by (Luce and Tukey 1964) and extended to include polynomial conjoint measurements by (Tversky 1967). There are no hard and fast rules for arriving at a conjoint measurement, but the usual method is to form a product structure (Roberts 1984) and assess that structure against well defined appropriateness conditions.

Product Structure. The product structure for this conjoint measurement is represented by the components of *technical integrity* previously defined, i.e., *technical integrity* consists of the system emergent properties of *capability* and *dependability*. If we represent the measures for *technical integrity* to be *TI*, *capability* to be *C* and *dependability* to be *D*, then we expect (assuming for now *capability* and *dependability* are measurable[1]) that:

$$TI = f(C, D)$$

[1] And if they are not measurable the product structure may be decomposed further.

We can think of this function as transforming any ordered pair of measurements *(c,d)* into a number on an ordinal, interval or ratio scale for TI.

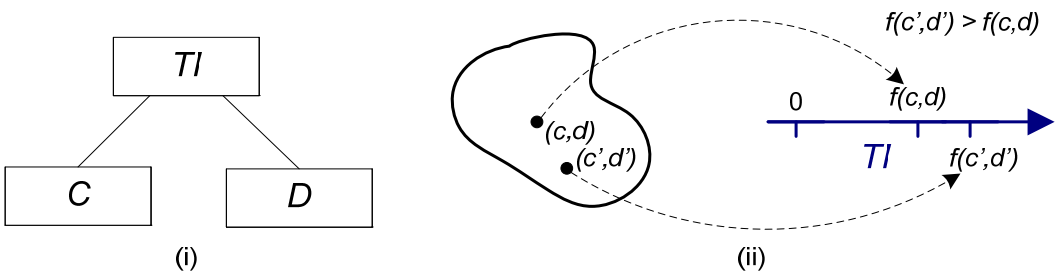


Figure 4. (i) Product Structure for TI.
(ii) Mapping of measured attributes to ordinal scale of TI.

Appropriateness Conditions. Decision theory provides advice on conditions that should be satisfied by the attributes selected for the product structure (Keeney and Raiffa 1976). They include *completeness*, *operationality*, *decomposability*, *non-redundancy* and *minimum size*. The *completeness* criterion requires consideration of the adequacy of the measurands as a set in including all the areas of concern for the intended conjoint measurement. This criterion can only be considered as a guide. The *operationality* criterion requires assurance that each measurand has a real contribution to the final measurement value. The *decomposability* criterion guides us to ensure that the measurands are independent (a necessary mathematical condition for additive conjoint). The *non-redundancy* criterion requires the elimination of double counting. The same parameter should not be measured more than once. The *minimum size* criterion guides us to limit the number of attributes chosen - more than 5 measurands is considered too many.

In order to assess the proposed product structure against these criteria we next examine what we expect the characteristics of the *TI* conjoint measurement to be, then compare this construct against the criteria.

Expected Characteristics of TI Measure. What might a measurement function for $TI = f(C,D)$ look like?

Consider a system that has met all its specified requirements. This would have a high or maximal value for *C*. However if this same system exhibited emergent behaviour that caused the death or injury of personnel, it would have a low or minimal value of *D*. What would we expect the measure of *TI* to be in this case? Clearly this would be a low value for *TI* – a system that meets its capability objectives yet often kills unintended personnel is next to useless! At the other end of the spectrum consider a system that never kills, injures nor poses any health hazard to personnel, equipment or the environment – this would have a high or maximal value of *D*. This system however never achieves its service requirements. It may be that it is just largely non-compliant with all capability requirements. This system would have minimal value on the *C* scale. Clearly this also would exhibit a low value for *TI* – a system that does not meet any performance objectives is worthless. Clearly, the desired outcome of a high *D* value and a high *C* value would exhibit a high *TI* value. These examples lead us to expect that on a graph of *D* versus *C* there would be a region of high *TI* values (desired) and a region near the axes of low *TI* values. There are then is range of intermediate values for *TI*. Figure 5 shows what such a graph could look like. The lines on the graph represent isolines of *TI*.

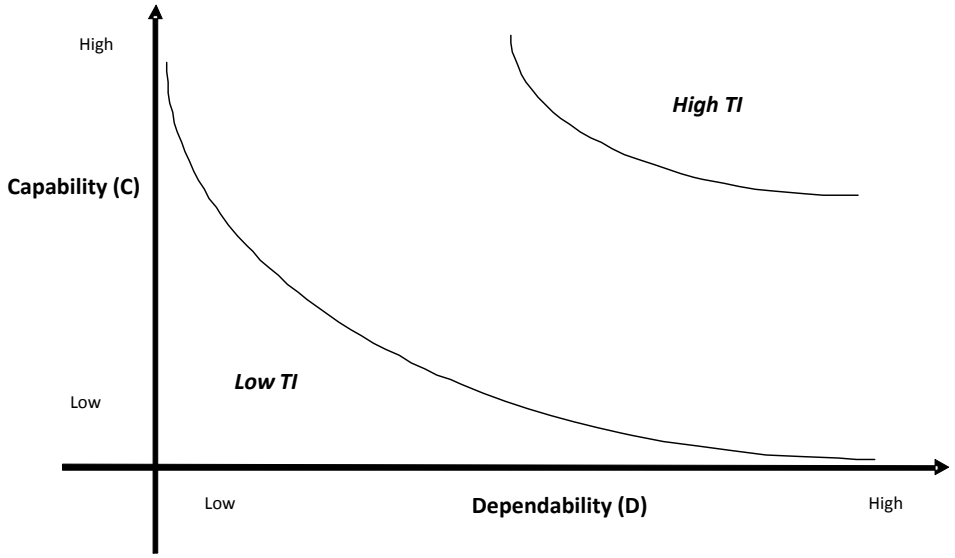


Figure 5. Expected form of *TI = f(FS,PS)*

We now assess expected measurement outcome against our completeness criteria *completeness*, *operationality*, *decomposability*, *non-redundancy* and *minimum size*. Based on work presented previously in this paper that defined *technical integrity*, we conclude that *C* and *D* form a complete set of measurands for TI. Each measurand is also clearly operational on the outcome of *TI*, e.g., a low value of either measurand is required to yield a low value of *TI.* Are the measurands independent? At this stage we assert that they are based on the intention that they separately assess intended versus unintended properties of the system. Double-counting is a potential difficulty with this approach. For instance, if a reliability requirement is well specified as Mean Time Between Failure in the system's specification then it could be double counting to consider reliability as a component of the dependability measure in this context. This will need further consideration in the definition of measurements for *C* and *D.* Minimum size has been achieved at this level of the product structure by having simply two contributing attributes. We conclude therefore that it is appropriate to form a conjoint measurement of *TI* from the attributes *C* and *D.*

Establishment of the Conjoint Measurement Model. Commonly, a conjoint measurement takes an additive form. An additive conjoint measurement for *TI* would be of the form:

$TI = \alpha.C + \beta.D + \gamma$ where α, β, γ are constants.

If we consider the boundary conditions of the measurement where we require that: if *C*=0 then *TI*=0; if *D*=0 then *TI*=0; and if *C*=maximum and *D*=maximum then *TI*=maximum, then the additive form is rejected since it will not yield these values irrespective of the values defined for α, β, γ.

The boundary conditions really suggest a multiplicative model to ensure *TI* = 0 if either *C* = 0 or *D*= 0. The mathematical foundations for multiplicative conjoint measures were first established by [Roskies 1967] as an extension of the additive conjoint model.

The following conjoint measurement model is proposed for *TI*:

$TI = C^{\alpha}.D^{\beta}$ where α, β, are constants > 0, values to be determined.

In order to demonstrate this model we constrain *C, D* to be real numbers in the range [0,1][2] which in turn constrains TI to the range [0,1]. If we consider when $\alpha, \beta = 1$ as a starting point we can visualise the resulting mapping of *C, D* to *TI* when it is equally influenced by its attributes. This is shown in Figure 6.

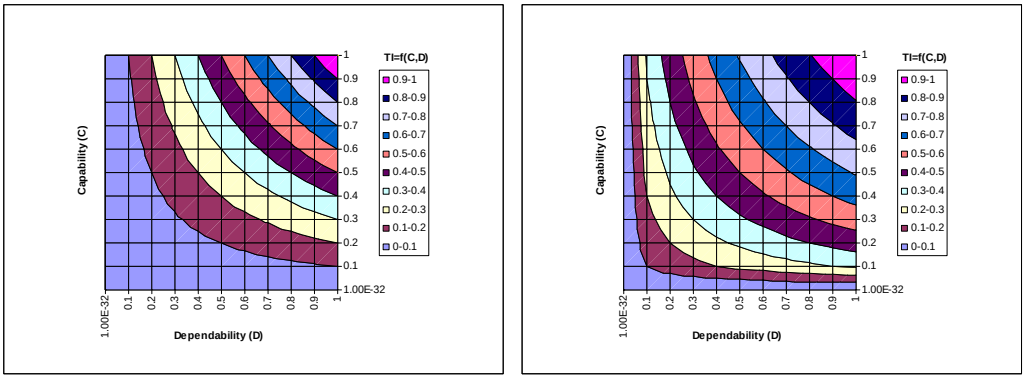


Figure 6. Capability and Dependability with Equal Influence
((i) *α=β=1, TI = C.D. (ii) α=β=0.5, TI = √(C.D)*)

To confirm the multiplicative conjoint makes sense at the boundaries we consider in Figure 7 the settings for α, β to model extreme conditions where TI is influenced by only one of its attributes.

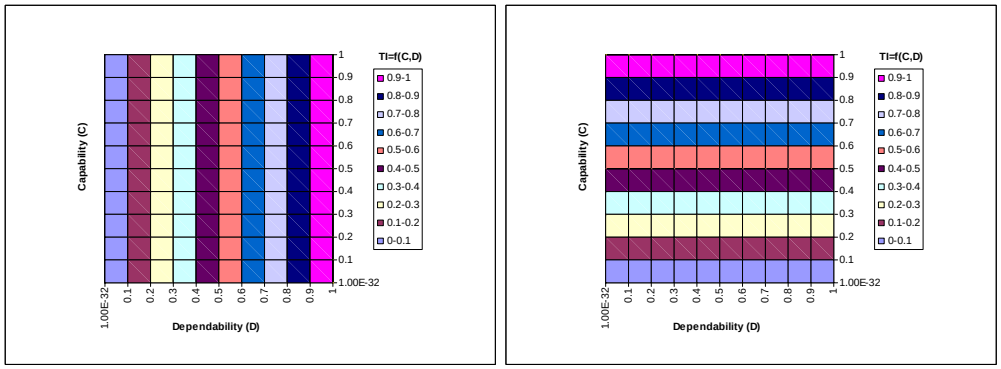


Figure 7. Extreme cases: (i) TI Not influenced by Capability (*α=0, TI = D*). (ii) TI Not influenced by Dependability (*β=0, TI = C*)

The measurement model caters for the more general case where there is a different relative utility of *C* and *D*. If this difference in utility is modeled by appropriate selection of values of α, β the shape of the TI function can be skewed for a dependability-biased measure of *technical integrity* or a capability-biased measure of *technical integrity*. This is illustrated for two selections in Figure 8.

[2] A not unreasonable constraint as we shall see when we look at establishing measures for C&D.

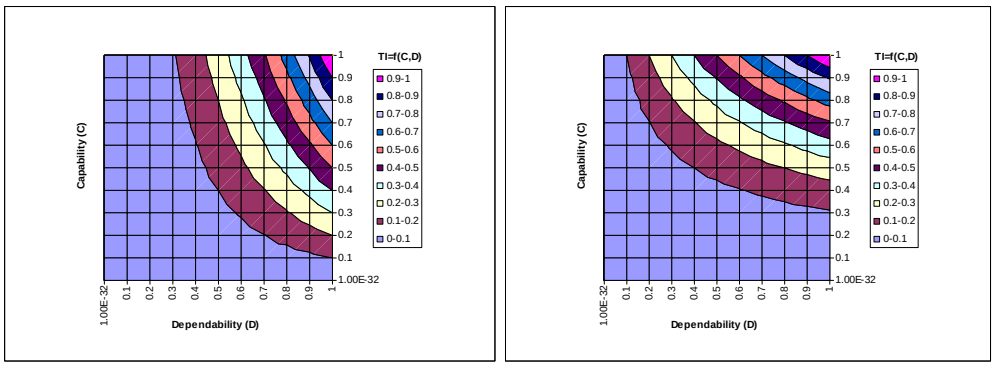


Figure 8. Differing Relative Utilities of Capability and Dependability:
(i) Dependability-biased ($\alpha=1, \beta=2, TI = C.D^2$).
(ii) Capability-biased ($\alpha=2, \beta=1, TI = C^2.D$).

Further research is intended to define the approach for establishing α and β empirically with stakeholders. It is noted these utilities may be different for different stakeholders of the system under consideration. An Inter-rater Reliability test will be used to evaluate the chosen values and to establish if consistent values across a range of disparate systems can be arrived at. This would support a finding that the conjoint measurement for *TI* is useful for comparisons across a range of disparate systems.

Notwithstanding the selection of this measurement model, we still have the problem of defining measures for *capability* and *dependability* that may be no more tractable than TI to measure. How are these components to be measured such that the TIs of different systems may be compared? We require either a previously determined measurement scale for each or to determine a simple ordinal scale for each measurand.

A Measurement Approach for Capability

The definition for *capability* arrived at earlier is "that the system meets all specified requirements". The measure should then be the fraction of the originally intended requirements actually achieved. This would equate to failure to meet any requirements being assigned $C = 0$ and all requirements being achieved as $C=1$ and intermediate values measured on the interval (0, 1) as a ratio scale.

While it is common for detailed metrics to be kept about requirements changes and achievements in most system development projects, these may not be readily disclosed to researchers. It may be that for practical purposes, subjective assessments of *capability* may also need to be mapped to values on the scale.

There is precedent for using an ordinal scale relating to requirement achievement. The Study of Systems Engineering Effectiveness (Elm et al. 2007) uses requirements verified as the measure for "Scope", one of the defined elements of "Project Performance" in that study. The study uses a 4 point ordinal scale to represent a subjective assessment of requirement verification. The SE-ROI project (Honour 2010) considers "Technical Quality" as being an ordinal scale where satisfaction of requirements is the "Threshold" level and delivered performance beyond that specified for minimum acceptability, but of value to the stakeholders, is termed the "Objective" level.

A Measurement Approach for Dependability

The definition of *dependability* derived earlier is equated to the absence of risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. How do we go about measuring the totality of known and unknown risks in a complex system?

Feasibility of Measurement. It seems clear that we should be able to measure *dependability,* at least on an ordinal scale. We should be able to say that System A is more dependable than System B. Measurement on a ratio scale should not be beyond reach. E.g., we can state how much safer (based on historical accident statistics) air travel from City A to City B is compared to a car journey between the same cities.

Required Form of Measurement. Consistent with the structure of the conjoint measurement established earlier, we require a scale $D \in [0..1]$ where 0 indicates the system is not at all dependable and there is certainty in undesired outcomes occurring and 1 indicates the system is 100% dependable and there is no risk of undesired outcomes. Ideally this will be a ratio scale. How to arrive at measurement values for real systems remains a task for future research.

Approaches under Investigation. The literature in the area of measurement of overall risk of unintended consequences of a system largely concentrates on the identification of risks and the qualitative (more often) and quantitative (less often) assessment of each of those risks.

One approach may be to consider dependability as two separately quantified components. One component would be the identified risks and the other the unidentified risks. The identified risks could be combined into a single vale where each consequence identified is associated with a cost (i.e., dollar) value and that value weighted by likelihood of its occurrence. All these weighted values may be summed together to form an overall risk. How to quantify the unidentified risks remains an issue. Some proxy measurement following approaches established for the estimation of faults remaining undetected in complex software intensive systems is possible but arguably not pragmatic to be able to collect across a range of disparate systems. Collection of a range of related measurements and combining them in an assessment tool based on a Bayesian belief network is also possible (Fenton et al. 1998).

Another approach is to look at the number of incidents that breach dependability in the operational life of a system. While this may lead to a more direct form of measurement, there are inherent difficulties that include the infrequent occurrence of high consequence breaches of dependability in critical systems. This makes inference of dependability difficult. A proxy measure where the dependability of the system may be inferred by the more frequent occurrences of low consequence breaches may be a path to further this approach (e.g. (Harris 2004) finds that in aviation the relationship between frequency of occurrence and consequence magnitude for safety incidents follow a power law indicative of self organising critical systems).

A subjective estimating approach for dependability may be to formulate a survey instrument where stakeholders and/or experts in the system may be questioned about dependability characteristics of the system and their responses combined to form a dependability measure. Precedent for this as a pragmatic approach includes the System Usability Scale (Brooke 1996).

Conclusion

A definition of *technical integrity* has been established that combines two attributes of a complex engineered system. The attributes reflect the intended, desired emergent properties of a system and the minimisation of unintended, undesired emergent properties of a system. These attributes have been termed *capability* and *dependability* respectively. It is desirable and feasible to measure *technical integrity* and an approach using a conjoint measurement from measurements of *capability* and *dependability* has been formulated.

Future research is required to further this measurement approach by:

establishing and validating an approach to measure *dependability*. The approach needs to make a pragmatic estimate of the uncertainty of undesired behaviour that is objective, but not too onerous or costly to preclude its use across a wide range of disparate systems.

assessing the relative utilities of *capability* and *dependability* in the multiplicative conjoint measure and seeking to determine if there exists a form of the measurement that may be used across disparate classes of complex engineered systems,

Critical review, ideas and guidance are welcomed by the author to shape this approach into a practical measurement across a broad range of systems to investigate the effectiveness of governance, management and technical processes applied to the development and operation of complex engineered systems.

References

ADF. 2010. Defence Instruction (General) LOG 4-5-012 Regulation of Technical Integrity of Australian Defence Force Materiel. Canberra, ACT, Australia: Department of Defence.

Bale, E.A., and D.W. Edwards. 2000. Technical Integrity - An Engineer's View. *Trans IChemE* Vol 78, Part B (September 2000):355-361.

Barbacci, Mario, Mark H. Klein, Thomas A. Longstaff, and Charles B. Weinstock. 1995. Quality Attributes. Pittsburgh, Pennsylvania: Software Engineering Institute, Carnegie Mellon University.

Boehm, Barry. 1978. Characteristics of Software Quality, TRW Series of Software Technology. New York: American Elsevier.

Brooke, John. 1996. SUS - A Quick and Dirty Usability Scale. In *Usability Evaluation in Industry*, edited by P. W. JORDAN, B. THOMAS, B. A. WEERDMEESTER and I. L. McCLELLAND. London.: Taylor & Francis.

Browning, Tyson R., and Eric C. Honour. 2008. Measuring the Life-Cycle Value of Enduring Systems. *Systems Engineering* Vol. 11, No. 3:16.

DoD. 1993. MIL-STD-882C System Safety Program Requirements: Department of Defense, USA.

DoE. 2006. DOE O 413.3A Program and Project Management for the Acquisition of Capital

Assets. Washington, D.C.: US Department of Energy.

———. 2008. DOE G 413.3-1 Managing Design and Construction Using Systems Engineering. Washington, D.C.: US Department of Energy.

Edwards, Michael. 2009. The Impact of Technical Regulation on the Technical Integrity of Complex Engineered Systems. In *INCOSE 2009*. Singapore: International Council On Systems Engineering.

———. 2010. An Ontology of Technical Regulation. In *Systems Engineering Test and Evaluation 2010*. Adelaide, SA, Australia.

Elm, Joseph P., Dennis R. Goldenson, Kaled El Emamm, Nicole Donatelli, Angelica Neisa, and NDIA SE Effectiveness Committee. 2007. A Survey of Systems Engineering Effectiveness - Initial Results: Software Engineering Institute, Carnegie Mellon University and the National Defense Industrial Association.

Fenton, N., B. Littlewood, M. Neil, L. Strigini, A. Sutcliffe, and D. Wright. 1998. Assessing dependability of safety critical systems using diverse evidence. *IEE Proceedinps - Software* Vol. 145, No. I. (February 1998):5.

Gilb, Tom. 2005. Scales of Measure. How to Quantify. In Competitive Engineering. A Handbook for Systems Engineering, Requirements Engineering and Software Engineering using Planguage: Butterworth-Heinemann.

Hackett, Mary. *Technical Integrity and Process Safety*. National Offshore Petroleum Safety Authority 2008 [cited 20 October 2010. Available from http://www.nopsa.gov.au/presentation/.

Harris, David. 2004. Is Aviation Safety a Self-Organising Critical System. Paper read at SETE 2004 Focussing on Project Success, 8-10 November 2004, at Adelaide, SA, Australia.

Hitchins, Derick K. 2003. *Advanced Systems Thinking, Engineering and Management*. Norwood, MA, USA: Artech House.

Honour, Eric C. 2007. Connecting an Architecture with its Emergent Properties. Paper read at Conference on Systems Engineering Research, 2007, March 14-16, at Hoboken, NJ , USA.

Keeney, R.L., and H. Raiffa. 1976. Decisions with Multiple Objectives: Preferences and Value Tradeoffs. New York: John Wiley & Sons.

Luce, R. Duncan, and John W. Tukey. 1964. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology* 1 (1):1-27.

Maier, Mark W., and Eberhardt Rechtin. 2009. *The Art of Systems Architecting*. Third Edition ed. Boca Raton, FL: CRC Press, Taylor & Francis Group.

NAVSEA. 2005. NAVSEA Instruction 9410.2 Naval Warfare Systems Certification Policy. Washington Navy Yard, DC: Department of the Navy, Naval Sea Systems Command.

Parnell, Gregory S., Patrick J Driscoll, and Dale L Henderson. 2008. *Decision Making in Systems Engineering and Management*. Edited by A. P. Sage, *Wiley Series in Systems*

Engineering and Management. Hoboken, New Jersey: John Wiley &Sons Inc.

Prasad, Divya Kumari. 1998. Dependable Systems Integration using Measurement Theory and Decision Analysis, Department of Computer Science, University of York.

RAN. 2007. Nias Island Sea King Accident Board of Inquiry Report: Royal Australian Navy.

Roberts, Fred S. 1984. Measurement theory with applications to decision making, utility, and the social sciences. In *Encyclopedia of Mathematics and its Applications*. Cambridge Cambridgeshire and New York, NY, USA: Cambridge University Press.

Sydenham, P.H. 1982. *Handbook of Measurement Science*. Chichester, UK: John Wiley and Sons.

Technical Integrity Audits. PI Asset Integrity Ltd 2007 [cited 20 October 2010. Available from http://www.piai-ltd.com/Technical_Audits.htm.

Tversky, Amos. 1967. A general theory of polynomial conjoint measurement. *Journal of Mathematical Psychology* 4 (1):1-20.

Biography

Michael Edwards holds a Bachelor of Engineering (Honours) and a Master of Engineering Science (Communications) from the University of NSW and a Master of Business Administration (Technology Management) from Deakin University. He is an Engineering Fellow with Raytheon Australia and is a part-time research student at the Defence and Systems Institute of the University of South Australia. He is a Fellow of the Institution of Engineers, Australia, a Chartered Professional Engineer, a member of the Systems Engineering Society of Australia and a Member of INCOSE. His professional and academic interests build on an extensive experience base in the practice of engineering management and systems engineering largely in the domain of complex naval systems including applications in surface combatant and submarine combat systems, mine warfare systems, hydrographic systems and naval aviation systems.

Appendix K The content validity of a technical integrity measurement construct

Peer reviewed paper in SETE 2013, Systems Engineering Society of Australia: Canberra, ACT.

This paper presented the initial results of the survey research aimed at establishing the content validity of the technical integrity measurement construct. It presents the findings of the relative importance of the concepts of capability and dependability and how these are considered to vary across the domains of system application.

.



The Content Validity of a Technical Integrity Measurement Construct

Michael Edwards
Raytheon Australia / UniSA DASI
michael.edwards@raytheon.com.au

Abstract. The impact of technical regulation on the technical integrity of complex engineered systems is poorly understood. In researching this problem, a means by which to measure technical integrity is necessary. A conjoint measurement construct for technical integrity was previously theorised and published for critical review. This paper presents the results of Survey Research that has gathered expert opinion to validate the measurement approach and to provide data to set the initial parameters of a measurement instrument.

INTRODUCTION

Background

The author's area of research is concerned with complex engineered systems and inquiry into the impact of technical regulation on the technical integrity of these systems particularly during the design and development phase of the system lifecycle. A framework for this research was reported in (Edwards 2009) and this paper presents a summary of results of recent progress of that research.

In the Australian Defence Force (ADF), technical integrity is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection, i.e. its compliance with technical standards and the minimisation of unintended consequences (Australian Defence Force 2010). Technical integrity has various different meanings and has synonyms and related concepts occurring in other civilian and military system contexts. (Edwards 2011) has surveyed those various meanings and related concepts in seeking to establish a universal technical integrity definition.

At this time, the intended positive impact of technical regulation on the technical integrity of a complex engineered system remains largely based on the application of heuristics of what should be done but without quantitative support. One of the key issues in establishing such support is the need to establish a measurement construct and a corresponding measurement instrument for technical integrity.

Technical integrity construct

Consideration of the varied definitions of technical integrity, parallels with system safety concepts and the relationship inferred between technical regulation and the emergent properties of a complex engineered system were analysed by (Edwards 2011) with the conclusion that there are two separable or orthogonal[1] aspects to technical integrity of a complex engineered system. The bases for the separation are the emergent properties of the system; the 'useful' and 'destructive' emergent properties as described by (Edwards 2009).

The 'useful' basis implies that the system meets all the requirements and standards that have been prescribed for it. This property of the system is termed as the concept *capability*. The 'destructive' basis implies the broad notion of minimising the unspecified, but clearly undesired properties of the system. This property of the system is termed as the concept *dependability*. The theory of technical integrity relating these concepts is illustrated in Figure 1.

[1] meaning the characteristic of being independent, not redundant, not overlapping.

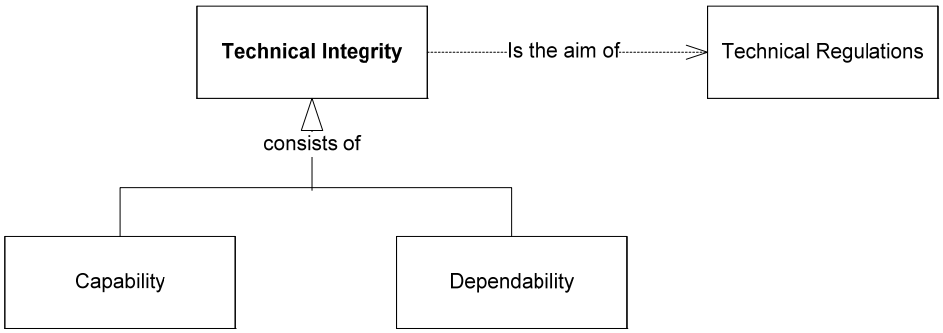


Figure 1. The theorised definition of technical integrity.

More completely then, *technical integrity* is defined as consisting of capability and dependability, where:

Capability is defined to mean that the system meets all specified requirements for performance, function and quality attributes in the defined environment and when used as intended. Capability is the designed set of emergent properties of a system that are intended and useful (i.e., required); and

Dependability is defined to mean that the risk associated with unintended, undesired emergent behaviour of the designed system is acceptably low and that undesirable limitations resulting from intentional trade-off during design are acceptable. It includes concepts of robustness (system is tolerant of usage that may exceed specified conditions without catastrophic consequences), reliability (failures do not happen often), and system safety (risk of unidentified hazards and failure modes resulting in undesirable outcomes is minimised).

Conjoint measurement approach

Technical integrity of a complex engineered system is not measurable directly and no means for measurement via other indirect means has previously been determined. However this should not be an obstruction to the formulation of a useful measure. A formalism based on measurement theory and logical arguments is required. The theory for conjoint measurement is not new, having been established by (Luce and Tukey 1964) and extended to include polynomial conjoint measurements by (Tversky 1967). There are no hard and fast rules for arriving at a conjoint measurement, but the usual method is to form a *product structure* (Roberts 1984) and assess that structure against well defined appropriateness conditions.

Product structure

The product structure for this conjoint measurement is represented by the components of technical integrity previously defined, i.e., technical integrity consists of the concepts of capability and dependability. Representing the measure for technical integrity to be *TI*, capability to be *C* and dependability to be *D*, the product structure is illustrated in Figure 2 (i) and defined mathematically as a function by:

$$TI = f(C, D)$$

This function transforms any ordered pair of measurements *(c,d)* into a number on an ordinal, interval or ratio scale for *TI* and is illustrated by Figure 2 (ii).

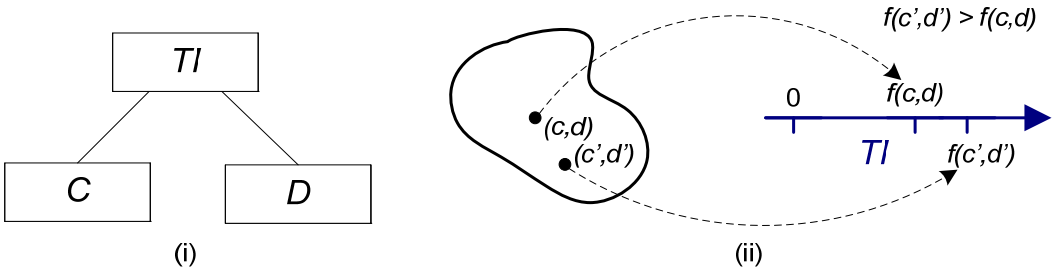


Figure 2. (i) Product Structure for TI.
(ii) Mapping of measured attributes to ordinal scale of TI.

The approach for formulating a conjoint measure of technical integrity, the anticipated characteristics of the measure and candidate practical approaches for a measurement instrument have previously been presented for critical review (Edwards 2011).

Research objectives

This primary research objective reported in this paper is to establish the content validity of the technical integrity measurement construct. Content validity refers to the extent to which a measure represents all facets of the phenomenon of interest and may be assessed via a consensus of expert judgements (Allen and Yen 1979, 95-96) (Tashakkori and Teddlie 1998, 83).

An additional research objective reported herein is to determine coefficients to use for the aggregation of capability and dependability into the conjoint measurement of technical integrity via quantitative analysis of judges' ratings.

METHOD

Strategy

To establish content validity, a survey research strategy was used to collect expert judgements in the form of ratings reflecting the levels of agreement with aspects of the proposed measurement construct. The rating data is analysed quantitatively in order to determine the level of support for the measurement construct. Effectively, the quantitative analysis procedure is establishing the content validity of the proposed measurement construct by testing the following hypotheses:

- Technical integrity can be measured on an ordinal scale.
- Technical integrity can be measured by a single numeric value.
- Technical integrity can be measured on a ratio scale.
- Capability of a complex engineered system can be measured on a scale of range 0 to 1.
- Capability of a system in excess of specified requirements is valued.
- The proposed capability measure is best.
- Dependability of a complex engineered system can be measured on a scale of range 0 to 1.
- Dependability of a complex engineered system better than the risk acceptability threshold is valued.
- The proposed dependability measure is best.
- Technical integrity can be measured by a conjoint measurement from the components capability and dependability.
- The conjoint measurement of technical integrity allows for comparison of similar systems.
- The conjoint measurement of technical integrity allows for comparison of dissimilar systems.

The survey research strategy also collected expert judgments in the form of relative importance ratings of capability and dependability across a range of 16 military and civil system domains. The question asked: 'In each domain in which you feel competent to judge, please indicate the relative importance of capability and dependability to systems within that domain.' The 16 domains for which judgments were requested were:

- Military – Aviation
- Military – Naval
- Military – Land
- Military – Joint, C4ISR
- Civil – Aviation
- Civil – Maritime
- Civil –Medical
- Civil - Mining
- Civil – Chemical Process
- Civil – Industrial
- Civil – Telecommunications and IT
- Civil – Transportation
- Civil – Power Utility – Conventional
- Civil – Power Utility – Nuclear
- Civil – Water Utility
- Civil – Space

The rating data is analysed quantitatively in order to arrive at aggregation weightings for the capability and dependability elements of the multiplicative conjoint equation. Effectively the quantitative analysis procedure is testing the hypotheses that *the ratings of the relative importance of capability versus dependability are the same across the different system domains*.

The survey research strategy also collected expert judgments in the form of written responses from open-ended questions that are the subject of qualitative and mixed-method analysis.

Survey sample

The identification of target respondents is best described as 'purposive sampling' (Tashakkori and Teddlie 1998, 76). Survey respondents were requested to participate based on their familiarity with the definition, design, integration, verification, operation and/or support of complex engineered systems. Intentionally in the survey design, they did not need to be considered to be an expert to participate in this study as a judge since demographic data collected allows examination of variations in judgements against criteria that may be proxies for expertise.

The survey responses were collected in July through September 2012. 207 requests were made, with 125 of those commencing the survey and 116 completing the survey, representing a 56% completion rate.

Demographics of respondents

This section is summary of the demographic data provided by the respondents.

There is a preponderance of male respondents in the group (94.6%). This is not particularly unrepresentative in the engineering of defence systems. 94.8% of the respondents have a university level education. This may be compared to the general Australian population where 24% people aged

between 15 and 64 years hold a bachelors level degree or higher (Australian Bureau of Statistics 2011) and demonstrates that the respondents as a group are significantly more educated than the general population as expected in the engineering and usage of complex systems. A majority of respondents (59.5%) identified their main technical education as being in electrical/electronic engineering.

The group exhibits significant professional experience (most with greater than 10 years as shown in Figure 3) and 52% with formal engineering delegations in ADF regulatory frameworks (refer Figure 5); an experienced and responsible group of professionals.

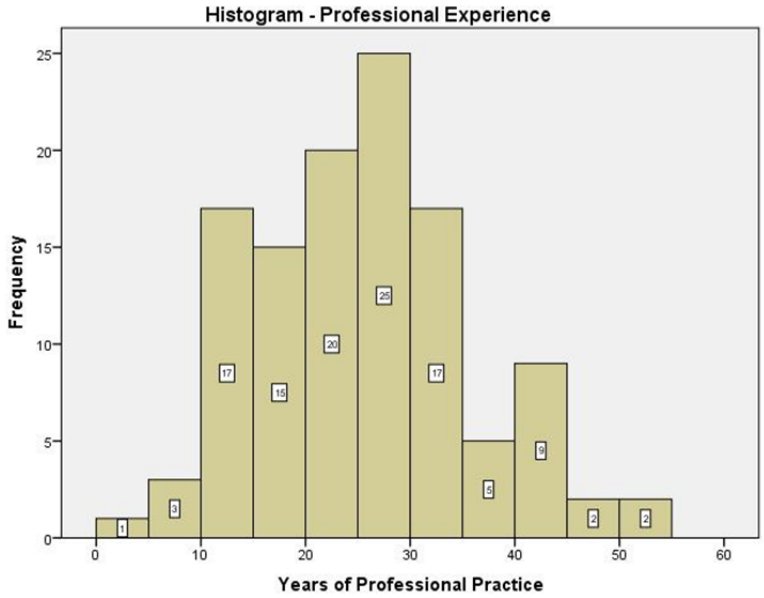


Figure 3. Histogram of the years of professional practice of the respondents.

Figure 4 shows the system application domain where the respondents have their main experience. A majority of respondents have their main experience in the application domains of Maritime Defence Systems and Aerospace Systems. It is possible that this reflects the researcher's own professional network from which most respondents were invited and also that the area of technical regulation and technical integrity is of high importance and is currently in active debate amid regulatory reform initiatives within the Australian Air Force and Navy particularly. The relatively small number of respondents, about 11%, in 'Other' or Commercial ICT may limit any generalisations that can be made across non-military application domains.

Sixty one (52.6%) of the respondents indicated that they hold or have held formal delegations of Engineering Authority in organisations governed by ADF technical regulations or other regulatory systems. The delegations held are shown in Figure 5. This demonstrates that many of the respondents represent a group of very experienced engineers and that they are, or have previously been, entrusted to make significant engineering decisions for the ADF. There is also representation across the three technical regulatory domains of the ADF (maritime materiel– regulated under the Naval Technical Regulatory System (NTRS); aerospace materiel – regulated by the Technical Airworthiness Regulator (TAR); and land materiel under the Technical Regulation of ADF Materiel – Land (TRAM-L)).

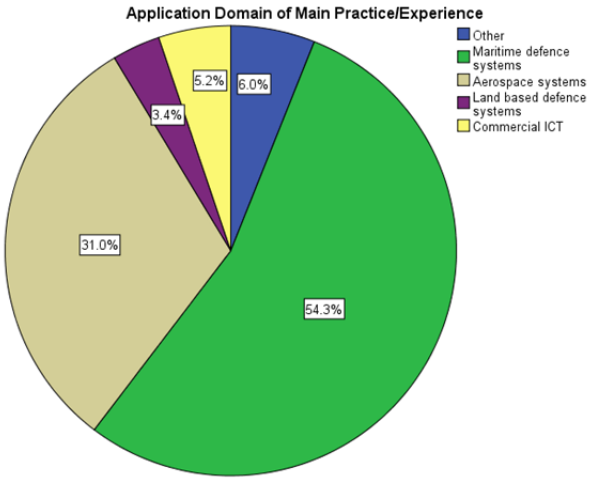


Figure 4. Pie chart of the system application domain where the respondent has most experience.

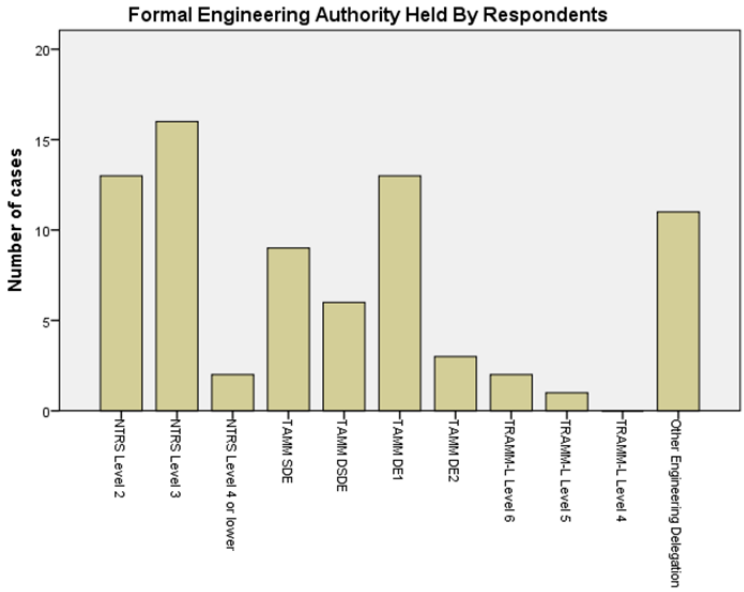


Figure 5. Bar chart of the engineering delegation held by respondents.

Analysis procedures

The analysis procedure for each set of ratings data was to:

- Tabulate the frequencies of the level of agreement ratings and produce a bar chart to assist in visualising the data.
- Re-code the data into categories to allow testing of the hypothesis.
- Significance of the hypothesis testing result is established using both a chi-square test and Lawshe's Content Validity Ratio (CVR) (Lawshe 1975) with significance values as amended by (Wilson, Pan, and Schumsky 2012).

The analysis procedure to determine coefficients to use for the aggregation of capability and dependability into the conjoint measurement of technical integrity followed the following steps:

- For each system domain, descriptive statistics (mode, mean[2], median range and interquartile

[2] Note, the rating scale is ordinal, and therefore the mean is not strictly valid, but it is useful for initially assessing trends in the results.

range) and bar charts of the frequency of each rating are generated to characterise the response data.

- The mean, median and interquartile range is exported to a spreadsheet for further inspection, ordering and visualisation in order to assist in interpretation of the results.
- The significance of the sample median values against set rating values is conducted using the non-parametric One-Sample Wilcoxon Signed Rank Test which provides a test to see if a sample median differs significantly from a hypothesised value. The only assumption for the data needed is that it is an ordinal variable which is true for this data set (UCLA: Academic Technology Services).
- Variations across groups, constructed from the demographic data, are explored using the Independent-Samples Kruskal-Wallis test. This non-parametric test allows for the comparison of 'scores on a continuous variable for three or more groups' (Pallant 2010, 232-5).

The qualitative analysis procedure essentially followed that of (Leedy and Ormond 2005, 140).

RESULTS

Quantitative analysis: content validity of measurement construct

The raw results and statistical analysis of the ratings data are summarised in this section. The first subsection provides a complete report against the first hypothesis with the following section only being a summary of results of the latter hypotheses (for brevity in this paper).

Technical integrity – measurement on an ordinal scale

The survey instrument posed the assertion 'It is meaningful to say system A has higher technical integrity than system B' and the respondent was requested to rate their agreement on a 5-point scale. Results obtained are shown in Figure 6.

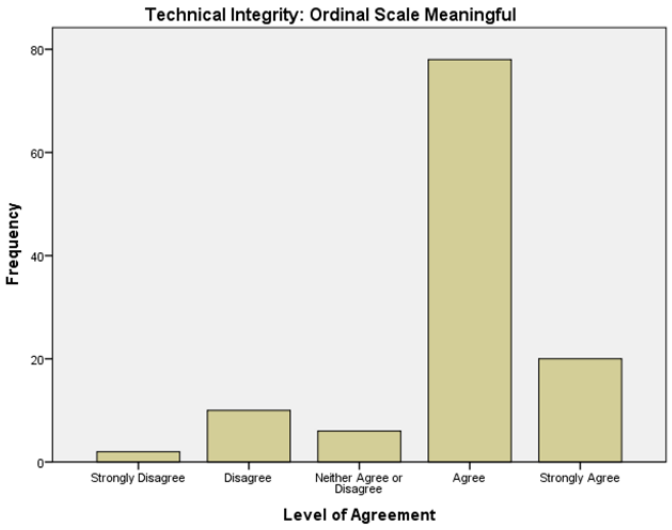


Figure 6. Bar chart of the level of agreement ratings of 'ordinal scale meaningful'.

To find if there is support for the hypothesis 'Technical integrity can be meaningfully measured on an ordinal scale', the 5 point scale was re-coded into two categories: Supported (Agree or Strongly Agree) and Not Supported (Neither Agree or Disagree, Disagree, Strongly Disagree) with analysis results being:

- A chi-square test indicates that the 84.5% of judgements in support of the ordinal scale for technical integrity is significant, χ^2 (1, n=116) = 55.2, p<0.001.
- Lawshe's CVR is calculated at 0.69 which indicates that at least half the respondents agree and is significant at the 0.005 level (one tailed test).

Summary of quantitative analysis results

Table 1 summarises the quantitative results in terms of Lawshe's CVR where this has been calculated. The results for the measurement construct appropriate for technical integrity indicate support for an ordinal scale but they are not in support of a ratio scale. While the data also indicates non-support of measuring technical integrity using a single numerical value, it is at odds with results obtained on conjoint measurement.

Measurement Concept	Specific hypothesis tested	Not Supported	Indeterminate	Supported
Technical Integrity Scale	Ordinal			0.69
	Numeric (single value)	-0.45		
	Ratio	-0.71		
Capability Scale	[0..1] capability measurement meaningful			0.35
	Capability beyond specification is valued			0.61
	Proposed capability measurement is Best	-0.28		
Dependability Scale	[0..1] dependability measurement meaningful		0.24	
	Dependability greater than risk acceptance threshold is valued			0.59
	Better dependability measurement possible		0.21	
Conjoint for Technical Integrity	Conjoint from capability and dependability is meaningful			0.36
	Measure useful for comparison of similar systems			0.38
	Measure useful for comparison of dissimilar systems	-0.44		
Values are Lawshe's CVR.				

Table 1. Summary of quantitative results.

Judgment variation across demographic categories

The ratings data was explored to see if any of the responses were significantly associated with demographic categories. Each of the rated questions was analysed against the categorical demographic information by forming a cross-tabulation and using the 'Eta' calculation within SPSS as the appropriate association statistic for ordinal versus nominal data. 10 demographic categories were analysed for associations across the 8 judgement questions from the survey The highest value from Eta obtained was 0.379 with most falling well below 0.3 indicating no significant association of ratings across these demographic categories. The other demographic category information is the age of respondents and the years of professional practice, both ordered values. A total of 16 associations were examined via Spearman's correlation coefficient, rho, as the appropriate association statistic for rank orders. None of the rho values were large with most below 0.1 and none statistically significant. It can therefore be concluded that none of the ratings results are significantly associated with any demographic category.

Derivation of conjoint aggregation weights

In (Edwards 2011) it was theorised that $TI = C^a \times D^b$ (i.e., a multiplicative conjoint). The objective of this aspect of the survey research was to derive data that would allow the establishment of the initial aggregation weights for the conjoint equation.

Figure 7 shows an example of the survey results for one of the 15 system domains for which the relative importance of capability and dependability were rated.

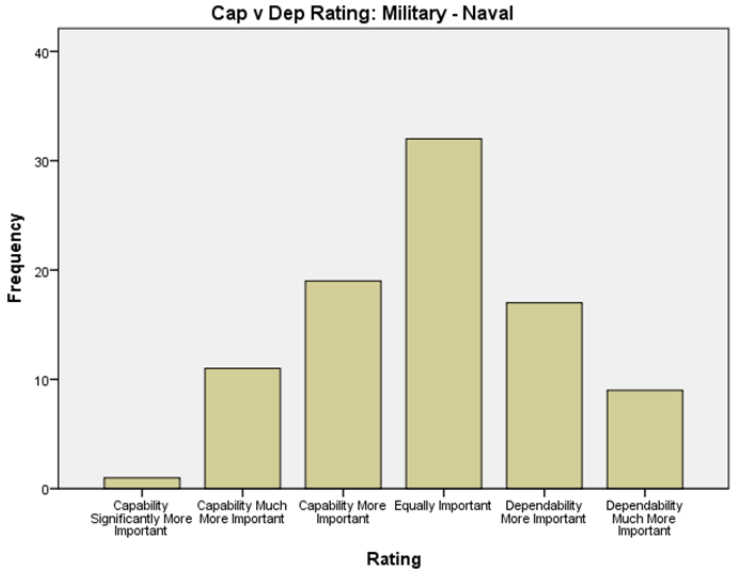


Figure 7. Frequency of ratings of relative importance of capability and dependability in the system domain of 'Military – Naval'.

Examination of the median values from the judge's ratings suggested that the judgments on the relative importance of capability and dependability may be clustered around 3 values. In order to see if this is statistically valid, a Wilcoxon Signed Rank Test was conducted to test if the sample median of each variable differs significantly from 4, 5, and 6 (Capability and Dependability Equally Important, Dependability More Important, and Dependability Much More Important). Note that the Wilcoxon Signed Rank Test does not assume the data is interval and normally distributed (which is the case with this data set). The test was conducted at a Confidence Interval of 95% at significance of 0.05.

Plotting the statistically significant median values in descending order against the system domain types shows that the system domain types clearly fall into three significant categories – A, B and C as shown in Figure 8.

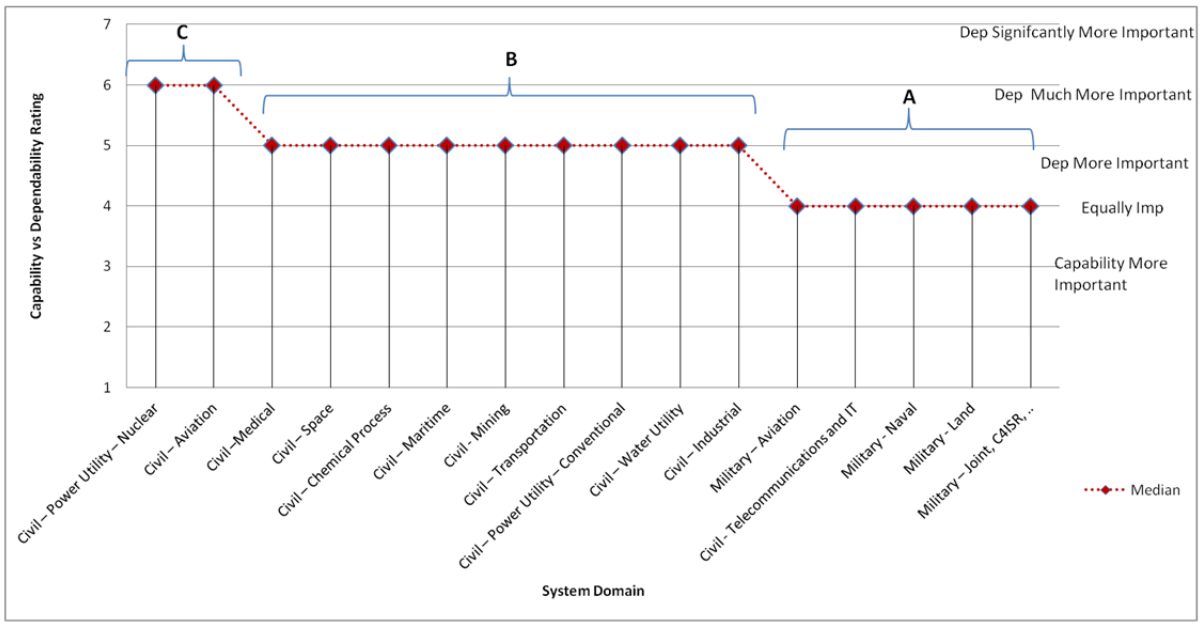


Figure 8. Significant median ratings of relative importance of capability and dependability versus system domains.

The following categories and characteristics are evident from this analysis:

- Category A: for all the Military system domains and the Civil ICT system domains, capability and dependability are rated of equivalent performance.
    - Although for Military – Aviation domain many ratings start emphasising the increased importance of dependability.
- Category B: in the Civil system domains of Water Utility, Power Utility, Transportation, Maritime, Mining, Chemical Process, Space and Medical dependability is considered more important than capability.
- Category C: in these domains we see a marked increase in the relative importance of dependability over capability through the domains of Civil Aviation and Nuclear Power. It is likely that this reflects a perception by the judges as to the magnitude of the catastrophe that may arise from these system domains.

The hypothesis that *the ratings of the relative importance of capability versus dependability are the same across the different system domains* is rejected by these results.

The results above indicate that for Category A system domains (Military system domains and the Civil ICT system domains) the coefficients are a=b. An arbitrary setting of a=b=1 could be assumed and this seems like a reasonable starting point. Further, the results suggest that if a=1 is arbitrarily set, then b>1 for Category B Civil system domains and b>>1 for Category C Civil system domains (civil aviation and nuclear power systems).

DISCUSSION

This discussion section combines consideration of quantitative and qualitative analysis results (the former provided earlier in summary form and the latter not included herein for the sake of brevity).

Technical integrity measurement proposal validity

The qualitative responses provide some overall insight into the validity of the proposal to measure technical integrity. A dichotomous response was observed where 15% of respondents indicated the research objective (i.e., pursuit of a measurement construct for technical integrity) was valid and it

would be of benefit to practice and another 9% commented that they agreed with the theorised definition and structure; and 18% made comments that indicated doubts as to the tractability of the measurement proposal or its usefulness.

It is noteworthy that many of the comments seem to consider an application of the measurement instrument beyond that of the primary application intended. Note that the primary intent is to validate a measurement instrument in order to evaluate the impact of technical regulations on the technical integrity of complex engineered systems. Many of the doubting comments reflect an application where the measurement may be used by decision makers to determine the acceptability of a system. The latter would be an additional benefit of the measurement construct if it were found to be valid and useful for that application. It is concluded from these considerations that the overall proposal to measure technical integrity is valid for the primary purpose, but extension to other applications would need to be the subject of further validation.

Technical integrity measurement scale

The form of scale that technical integrity is to be measured on is now considered. The objective for scale was that the measure forms a ratio scale, but at least forms an ordinal scale.

The quantitative analysis indicates that an ordinal scale for a technical integrity measurement is valid but that a ratio scale is not valid. This means that the majority of respondents agreed that complex engineered systems could be ranked according to a measure of their technical integrity (i.e., it is valid to say that system A has greater technical integrity than system B). They do not agree that the technical integrity of complex engineered systems can be measured on a ratio scale (i.e., it is not valid to say that system A has twice the technical integrity of system B).

Based on the quantitative analysis, no significant view as to the representation of the technical integrity measurement as a single numeric measurement emerged. An examination of an anomaly between this outcome and positive results from the validity of the conjoint measurement question and concluded that the former result should be discounted in favour of the significant majority that considered a conjoint approach to technical integrity, and therefore implicitly a numeric value, to be valid.

It is concluded that an ordinal measurement scale for technical integrity is valid and that technical integrity should not be defined on a ratio scale.

Conjoint measurement for technical integrity

The proposed conjoint measurement approach is now discussed. The objective was to validate that the combination of the defined capability and dependability measures form a meaningful and valid measure. The quantitative analysis indicates that a conjoint measurement constructed from components capability and dependability is considered valid and the use of this measure to compare similar types of systems is also considered valid. However, the majority of respondents considered that such a measure would not be valid in comparing the technical integrity of dissimilar systems.

It had been hoped that the derivation of different relative weightings for capability and dependability for different types of systems may allow for comparison across dissimilar system domains. Perhaps, in hindsight, a limitation of the survey instrument is that it did not directly pose a question to the respondents to see if they would consider the use of appropriate weightings to each category of system types would allow for comparison of dissimilar systems. In any event, accepting the result as obtained, it can only be considered that the measure is valid for the comparison of similar systems. This may not be a severe limitation for the purpose of this overall research. For example, it is to be considered valid to use a conjoint measure to compare the technical integrity of different surface combatant ships and in doing this assess the impact of the Naval Technical Regulatory System. Likewise we can consider it valid to use a conjoint measure to compare the technical integrity of different aircraft in the air force in order to assess the impact of regulations for Technical Airworthiness Management.

Measurement of capability

A statistically significant majority of respondents supported the concept that capability could be meaningfully measured on such a 0 to 1 construct (CVR = 0.35). Further, the significant majority

supported the concept that capability that exceed the specified requirements was of value (CVR = 0.61). This combined with qualitative responses indicated an extension of the capability scale beyond 1 is warranted but there is no basis suggested as to how to approach the evaluation of these capabilities beyond those defined by capability and/or technical requirements.

But, a significant majority of respondents indicated they did not think the capability 0 to 1 measurement construct was the best measure available (CVR = -0.44). The open-ended responses provided with these judgements showed a large number of respondents could not indicate what the better measurement approach may be.

Many suggestions for an alternate measurement indicated using evaluation against operational criteria (i.e., detail in an OCD like MOPs, MOEs or capability in the operational environment as evaluated by the end-users). The open responses indicate that some respondents do not see and/or have confidence in the linkage between end-user capability measures and the system's more technically based requirements and there is doubt about how well requirements for a system represent the capability actually required by the system's stakeholders.

Many of the open responses indicated the respondent was not satisfied with the definition of Capability given or had defaulted to their own pre-conceived idea of capability. This led to responses that did not propose an alternate measurement scale but rather an alternate definition for capability. This indicates there may be a need to re-label c*apability* to another term that does not carry such preconceptions. 'Fitness for Service', for instance, had been considered (particularly since in the ADF's Regulation of Technical Integrity (Australian Defence Force 2010) it is defined as 'fitness for service is the materiel's ability to satisfy operational requirements') but comes with other different preconceptions in different domains. Often, 'Fitness for Service', is not considered to be tightly linked to the system requirements which a developer is focused on realising in the complex engineered system.

Measurement of dependability

A majority of respondents supported the concept that dependability could be meaningfully measured on a 0 to 1 measurement construct (Chi-square test significant ($p<0.010$), but CVR = 0.24 not significant at 0.05 level).

A significant majority supported the concept that dependability that was better than the stakeholders' risk acceptability threshold was of value (CVR = 0.59). This combined with qualitative responses from four respondents indicated an extension of the dependability scale beyond 1 is warranted but with only one suggested measurement approach for this: 'using the same proportionality as for 0 – 1'.

60% of respondents indicated they did think the dependability 0 to 1 measurement construct was the best measure available (CVR = 0.21, not a statistically significant majority). The open responses provided with these judgements indicated an additional 18% (20 respondents) who did not really offer an alternative or better measure for dependability despite answering they thought there was one. Adding these responses back into support for the proposed scale gives a CVR = 0.57, a statistically significant result at the 0.05 level that adds confidence to the claim of validity for the proposed measure of dependability.

To summarise, it is concluded that for dependability: the proposed 0 to 1 measurement construct is valid (a majority of respondents concur that it will be meaningful) and is the best scale available (a majority of respondents concur that there is no better measurement construct available); and Extension of the construct beyond 1 to reflect additional value is considered desirable, but there is no consensus as to how that should be measured.

Conjoint measurement aggregation weights

The quantitative analysis results indicate that there are three sets of aggregation weights of relative importance of capability and dependability. Each set of weights covers three categories of systems:

- Category A - 'Military systems' and 'Civil ICT systems' where capability and dependability are of equal importance,

- Category B - 'Important Civil systems' where dependability is considered more important than capability, and
- Category C - 'Critical Civil systems' where dependability is considered much more important than capability.

It is noteworthy that for no class of system do the consensus judgements indicate that capability is more important than dependability. This is a key outcome considering the practice in systems development over the recent decades.

In the researcher's professional experience, predominately in the design and production lifecycle phases of military systems, there has clearly been a bias in process and behaviour toward achieving specified requirements (capability in our definition) over the minimisation of unintended consequences (dependability in our definition). Consider that the Military acquisition processes in Australia make payments to contractors (usually the designers and producers of these systems) when requirements have been verified for system acceptance and that there is often little or no payment associated with design evidence supporting a claim of the dependability of the system (Defence Materiel Organisation 2007, 46-79). However it is evident that the introduction of technical regulation over the last decade has had some effect on increasing the importance of dependability in design by increasing emphasis on techniques like system safety engineering and risk management more generally and ensuring overall Technical Integrity outcomes are explicitly tied to professional responsibilities of designers through formal certification of design and product. However, there is still room for improvement since it is believed 'that the organisations as a whole do not place technical integrity at the same level of importance as short-term operational imperatives' (Rizzo 2011, 67).

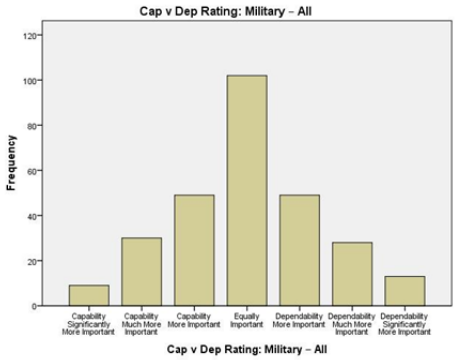
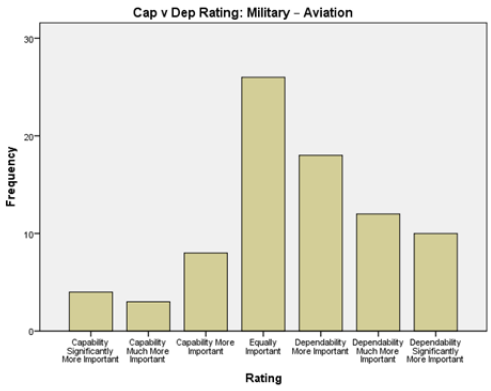
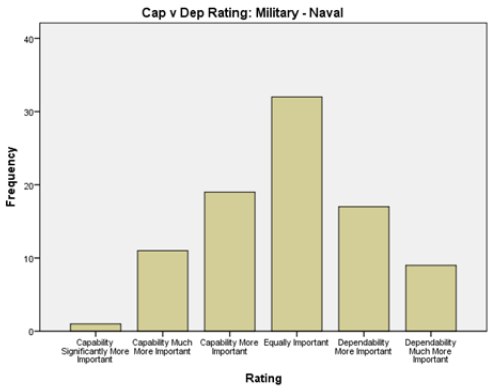
The result that has been termed 'Critical Civil systems' likely reflects a perception by the judges of the magnitude of the catastrophe that may arise from these system domains and the public perception of risk for these type of systems (nuclear power systems being the prime example).

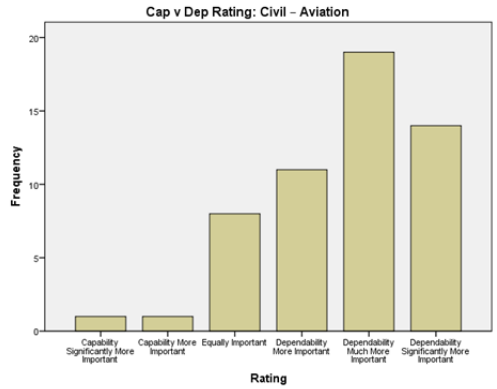
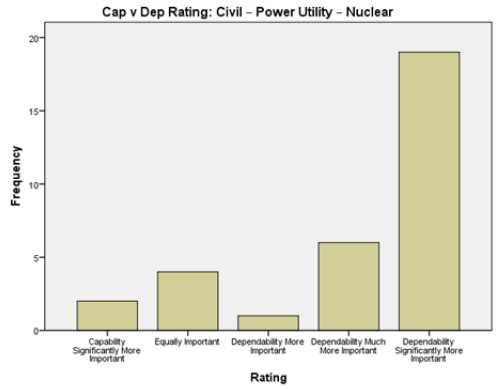
Table 2 compares the ratings data obtained by the survey and compares it to the number of comments received that have been coded to corresponding system domains. The table indicates that the spread of qualitative comments made in justifying relative importance judgements is consistent with the distribution of the ratings data.

The qualitative analysis of comments supports expected reasoning for the relative importance of capability and dependability. Dependability becomes a primary consideration of a system when the consequence of failure of the system is high, if there are no second chances, or if stakeholder confidence in the system is paramount. For military systems the importance of meeting capability requirements for defeating a threat in a military environment is a key consideration but dependability is equally important, i.e., having the capability available to use when required. Some viewed that dependability may be "assumed" or dependability shortfalls readily accepted since workarounds such as increased maintenance staff or on-hand spares can be applied to achieve an acceptable 'operational dependability' but there is no such ready fix considered to be available for capability shortfalls.

A theme that is evident in several comments that may impact on the validity of the conjoint measurement construct is that "Safety" should be considered as an explicit or separate element of technical integrity. This view, if accepted, would have the construct of technical integrity changed to have a third element in the conjoint measure. Related comments indicate a view that Safety figures much more highly than other risks of unintended consequence in the consideration of dependability and thus safety consequences drives the rating of dependability. Acceptance of this view may see that the estimation of dependability is adequately approximated by an estimation of safety related risks alone. Judgement data that may support one or other of these approaches is being obtained in the subsequent phase of the research agenda.

Table 2. Capability versus dependability importance judgements: comparison of quantitative and qualitative data.

Quantitative Ratings Data	Qualitative Comments			Comments
	C>D	C=D	C<D	
				Military - All is combined from ratings of Military - Aviation, Military - Naval, Military - Land, Military – Joint, C4ISR.
	9	5	7	Number of comments made on defence systems generally.
	14	**8**	**18**	Number of comments made generally plus specific comments from Military - Aviation, Naval and Land. Even spread of comments made is consistent with distribution quantitative responses, i.e., symmetric about capability and dependability being equally important.
	1	2	9	The larger number of comments made on importance of dependability is consistent with the skew in the distribution of quantitative responses, toward dependability.
	2	1	2	Even spread of comments made is consistent with distribution quantitative responses, i.e., symmetric about capability and dependability being equally important.

Quantitative Ratings Data	Qualitative Comments			Comments
	C>D	C=D	C<D	
	0	2	8	The larger number of comments made on importance of dependability is consistent with the distribution of quantitative responses centred around dependability being much more important.
	0	0	1	The quantitative ratings of Civil - Power Utility - Nuclear exhibit the most extreme response in favour of the priority of dependability. Only one comment has been made that correlates with this result: “rated nuclear regarding high dependability due to my understanding of the safety aspects of that industry”.

CONCLUSION

Findings

Most generally, the conjoint measurement construct for technical integrity is considered valid for the ranking of similar systems for the purpose of assessing the impact of technical regulation. More specifically it has been found that:

- Measurement of technical integrity on an ordinal scale is valid.
- Constructing a conjoint measurement from measures of capability and dependability is valid.
- Measurement of capability on a 0 to 1 range is valid.
- Capability in excess of that specified is valued, but no practical basis for extending the measurement of capability above 1 has been identified.
- Measurement of dependability on a 0 to 1 range is valid.
- Dependability better than the acceptability threshold of the system’s stakeholders is valued, but there is no consensus opinion offered as to what to base measurement of dependability above 1 upon.

The relative importance of capability and dependability for the purpose of establishing aggregation weights in a multiplicative conjoint equation, has been demonstrated to fall into three categories:

- ‘Military systems’ and ‘Civil ICT systems’ where capability and dependability are of equal importance,

- 'Important Civil systems' where dependability is considered more important than capability, and
- 'Critical Civil systems' where dependability is considered much more important than capability.

Significantly, it is also noted that in no system domain considered does the consensus view indicate that capability outweighs dependability.

Limitations

It is not considered valid to measure technical integrity on a ratio scale. This may limit future research goals that will use this construct (e.g., it may never be able to produce results that indicate x% more spent on technical regulation has increased technical integrity of systems by y%).

Other applications of the conjoint measurement construct for technical integrity may not be valid (e.g., for the purpose of judging the acceptability of a system). This will not impact the intended research usage, but may detract from the potential value of the approach in other applications.

The obvious starting point for the measurement of capability is to measure the fraction of the originally intended requirements actually achieved in the system. Doubts and scepticism as to the validity of the system's written requirements as a basis for measurement have been expressed in the survey results. These views may limit the acceptance of the capability measurement by some stakeholders.

Numeric weightings for the multiplicative conjoint measure have been established only for the category of 'Military systems' and 'Civil ICT systems' where both the weighting parameters to unity since the consensus opinion is that capability and dependability are of equal importance. Numeric weightings for the other categories of systems cannot be estimated from the available data.

REFERENCES

Allen, M.J., and W.M. Yen. 1979. Introduction to measurement theory: Brooks/Cole Monterey, CA.

Australian Bureau of Statistics. 2011. 6227.0 - Education and Work, Australia, May 2011 Australian Bureau of Statistics.

Australian Defence Force. 2010. Defence Instruction (General) LOG 4-5-012 Regulation of Technical Integrity of Australian Defence Force Materiel. Canberra, ACT, Australia: Department of Defence.

Defence Materiel Organisation. 2007. DMO Acquisition and Sustainment Manual. Canberra. ACT: Department of Defence, Australian Government.

Edwards, Michael. 2009. The Impact of Technical Regulation on the Technical Integrity of Complex Engineered Systems. In INCOSE 2009. Singapore: International Council On Systems Engineering.

Edwards, Michael. 2011. What is Technical Integrity and How to Measure It? In INCOSE International Symposium 2011. Denver, CO, USA.

Lawshe, C. H. 1975. A Quantitative Approach to Content Validity. Personnel Psychology 28 (4):563-575.

Leedy, Paul D., and Jeanne Ellis Ormond. 2005. Practical Research. Planning and Design. 8 ed. Upper Saddle River, NJ: Pearson Education International.

Luce, R. Duncan, and John W. Tukey. 1964. Simultaneous conjoint measurement: A new type of fundamental measurement. Journal of Mathematical Psychology 1 (1):1-27.

Pallant, J. 2010. SPSS survival manual: A step by step guide to data analysis using SPSS: Open University Press.

Rizzo, Paul J. 2011. Plan to reform Support Ship Repair and Management. Canberra, ACT: Defence, Commonwealth of Australia.

Roberts, Fred S. 1984. Measurement theory with applications to decision making, utility, and the

social sciences. In Encyclopedia of Mathematics and its Applications. Cambridge Cambridgeshire and New York, NY, USA: Cambridge University Press.

Tashakkori, A., and C. Teddlie. 1998. Mixed methodology: Combining qualitative and quantitative approaches. Vol. 46: Sage Publications, Incorporated.

Tversky, Amos. 1967. A general theory of polynomial conjoint measurement. Journal of Mathematical Psychology 4 (1):1-20.

UCLA: Academic Technology Services, Statistical Consulting Group. 2012. Statistical analyses using SPSS [cited 5 November 2012 2012]. Available from http://www.ats.ucla.edu/stat/spss/whatstat/whatstat.htm.

Wilson, F. Robert, Wei Pan, and Donald A. Schumsky. 2012. Recalculation of the Critical Values for Lawshe's Content Validity Ratio. Measurement and Evaluation in Counseling and Development 45 (3):197-210.

BIOGRAPHY

Michael Edwards holds a Bachelor of Engineering (Honours) and a Master of Engineering Science (Communications) from the University of NSW and a Master of Business Administration (Technology Management) from Deakin University. He is an Engineering Fellow with Raytheon Australia and is a part-time research student at the Defence and Systems Institute of the University of South Australia. He is a Fellow of the Institution of Engineers, Australia, and a Chartered Professional Engineer, a member of the Systems Engineering Society of Australia / INCOSE. His professional and academic interests build on an extensive experience base in the practice of engineering management and systems engineering largely in the domain of complex naval systems including applications in surface combatant and submarine combat systems, mine warfare systems, hydrographic systems and naval aviation systems.

Appendix L The consensual assessment technique for the measurement of technical integrity

Peer reviewed paper in SETE 2013, Systems Engineering Society of Australia: Canberra, ACT.

This paper presented the initial results of the survey research aimed at assessing the reliability of using CAT for the indirect assessment of the technical integrity of a complex engineered system.



The Consensual Assessment Technique for the Measurement of Technical Integrity

Michael Edwards
Raytheon Australia / UniSA DASI
michael.edwards@raytheon.com.au

Abstract. The Consensual Assessment Technique (CAT) is a well-developed and validated technique for creativity assessment. It establishes that creativity can be reliably judged by appropriate judges (persons who are familiar with the domain, but not necessarily experts in the domain). While technical integrity of a complex engineered system and the creativity embodied in an artwork are not direct analogs, there are enough similar characteristics of the problem space to investigate if the CAT can be successfully applied to establishing the technical integrity of a complex engineered system by a group of appropriate judges. This paper reports the results of such an investigation by looking at the results of judgements made on the technical integrity of three contrived radar systems.

INTRODUCTION

Background

The author's area of research is concerned with complex engineered systems and inquiry into the impact of technical regulation on the technical integrity of these systems particularly during the design and development phase of the system lifecycle. A framework for this research was reported in (Edwards 2009) and this paper presents part outcomes of that ongoing research.

At this time, the intended positive impact of technical regulation on the technical integrity of a complex engineered system remains largely based on the application of heuristics of what should be done but without quantitative support. One of the key issues in establishing such support is the need to establish a pragmatic means to measure the technical integrity of a complex engineered system. Technical integrity of a complex engineered system is not measurable directly and no means for measurement via other indirect means has previously been determined.

Technical integrity

In the Australian Defence Force (ADF), technical integrity is the term used to encompass the concepts of a system's fitness for service, safety and compliance with regulations for environmental protection, i.e. its compliance with technical standards and the minimisation of unintended consequences. Technical integrity has various different meanings and has synonyms and related concepts occurring in other civilian and military system contexts. (Edwards 2011) has surveyed those various meanings and related concepts and identified a recurrence of concepts including safety, compliance, dependability, effectiveness, quality, maturity and reliability as the technical objectives of regulation as well as business objectives like maintenance of asset value and return on investment.

At a generalised level, technical integrity of a complex engineered system may be defined as compliance to technical standards and the minimisation of unintended consequences.

The Consensual Assessment Technique

Teresa Amabile (Amabile 1996) has developed and validated a consensual technique for creativity assessment – termed the Consensual Assessment Technique (CAT) – which establishes the creativity embodied in a sample of products as judged by a range of judges. What Amabile has established, over a large number of studies, is that creativity can be reliably judged by appropriate judges (persons who are familiar with the domain, but not necessarily experts in the domain). Despite the fact that

‘creativity may be something difficult to describe, but it is still relatively easy for them [i.e., the judges] to identify [creativity] with a good degree of reliability’ (Amabile 1996).

The main features of the CAT include:

- Judges need to be familiar enough with the domain to have some implicit criteria for ‘technical goodness’.
- Judges make their assessment independently.
- Judges are asked to make assessments of dimensions other than creativity of the product (which allows the researcher to check for confounding effects).
- Ratings of products are made relative to one another rather than against an absolute standard.
- Judges consider the products in a randomized order.
- Judges ratings are analysed for interjudge reliability (equivalent to construct validity).
- Specific objectively discernible characteristics of the products are recorded so correlation of these with the judgments may be assessed.

Amabile reports that the CAT has also been applied outside visual and verbal artistic domains, to creativity in problem-solving tasks including computer-programming, high-technology services and design of structures with good reliability results.

While technical integrity of a complex engineered system and the creativity embodied in an artwork or other produced item are not direct analogues, there are enough similar characteristics of the problem space to suggest investigation into whether the CAT technique can be successfully applied to establishing the technical integrity of a complex engineered system by a group of appropriate judges.

Similarities of the problem space include:

- Both technical integrity and creativity are difficult to define precisely.
- They are not directly measurable.
- There is no complete nor agreed set of measurable components for either.
- Both represent a quality of a product/system that people familiar with the domain are readily willing to subjectively judge in a relative sense (e.g. artwork A exhibits higher creativity than artwork B; system 1 has higher technical integrity than system 2).

Research objective

The overall research objective is to determine an approach to practically measure the technical integrity of a complex engineered system, with the primary purpose being to assess the effectiveness of technical regulation.

To this end, CAT has been identified as a candidate measurement approach for the measurement of technical integrity. The research objectives of this paper therefore are to:

- Assess the reliability of CAT for technical integrity measurement,
- Assess the correlation of other subjectively evaluated characteristics of complex engineered systems with technical integrity that may indicate a *scale*[1] for technical integrity measurement.

METHOD

Strategy

CAT involves having three aspects developed or available. There are *judgments dimensions* that the

1 A measurement instrument that is a collection of items combined into a composite score and intended to reveal levels of theoretical variables not readily observable by direct means (DeVellis 2011, 11).

judges will be asked to rank the *products* on, as well as *objective measures* that may be determined for the *products.*

Products

The products in this investigation are complex engineered systems. Ideally, it would be desirable to focus on very large, very complex systems, but it is considered not tractable, for an initial investigation at least, to document the relevant characteristics of a number of very large systems and brief these to the judges in a time effective manner or in enough depth for them to reasonably make judgements.

Rather an approach of defining a limited number of contrived products was pursued. A set of naval/maritime radar systems was selected based on a consideration that these systems were reasonably complex, but their main function, operation and construction would be readily understood by most experienced system practitioners. The products for this study are three descriptions (akin to specification/product data sheets) of fictional but realistic radar systems that allow for the variation of characteristics of the systems which are expected to influence technical integrity and other judgement dimensions. No *a priori* expectation for the absolute or relative technical integrity of these three example systems was established.

The three products are included as attachments to this paper and are identified as:

- System A: an older technology, full military qualified navigation radar system with secondary air search capabilities.
- System B: a new technology, lightweight, commercially qualified navigation radar system.
- System C: a recent technology, part-military qualified navigation and surface search radar system.

Judgement dimensions

An essential element of the CAT procedure is to ask for judgements using 'non-definitional definitions' (Amabile 1996, 74). As an example, for technical integrity it is asked "Using your own subjective definition of technical integrity rate the degree to which each system exhibits technical integrity relative to the other example systems".

The following dimensions were evaluated by the judges for each system:

1 technical integrity,

2 impact to the natural environment,

3 inadvertent impact to other systems,

4 fit for service,

5 capability,

6 effectiveness,

7 efficiency,

8 construction quality,

9 operationally flexible,

10 durability,

11 survivability,

12 dependability,

13 reliability,

14 supportability,

15 maintainability,

16 specification/standard compliant,

17 safety,

18 usability,

19 security,

20 technical complexity, and

21 high technology.

On each dimension, the judge is requested to rate each system on a 16 point scale. The scale is to be interpreted as equi-spaced and the ratings for each system are enforced to be different in order to be to extract definite ranking data. The judge is to rate all products (systems) on one dimension before moving on to consider other dimensions. This was designed to assist in maintaining consistency in the judge's subjective criteria for a particular dimension and allows separability of the different dimensions. The order in which the dimensions were presented to the judge is randomised to minimise the chance of the order of evaluation unintentionally biasing the results.

Objective Measures

By way of example, in one of Amabile's studies, a paper mosaic product, she collected objective data about the number of pieces of mosaic used, the number of colours used, and the number of shapes used. These were shown to be correlated significantly with the judges rating of creativity.

Identifying objective measures for each of the systems and then correlating this with results obtained from the judgement dimensions will allow correlation analysis that may lead to an objective indicator of technical integrity or one or more of the other dimensions. The objective measures for each system are included in the attachments.

Survey sample

Part of the allure of CAT is that it does not appear that the level of the judge's expertise matters very much nor do they require training in the technique (Amabile 1996, 60, 72). Appropriate judges need only to be persons who are familiar with the domain, but not necessarily experts in the domain of the products/systems being considered. The identification of target respondents is best described as 'purposive sampling' (Tashakkori and Teddlie 1998, 76). Survey respondents were requested to participate based on their familiarity with the definition, design, integration, verification, operation and/or support of complex engineered systems. Demographic data collected allowed examination of variations in judgements against criteria that may be proxies for expertise.

The judgments were obtained via an online survey ensuring independence of the judges (under CAT the judges are not meant to collaborate) and allowing them to make their judgments as their time commitments allowed. The survey responses were collected in September through October 2012. 128 requests were made, with 54 of those commencing the survey and 43 completing the survey, representing a 34% completion rate.

Demographics of respondents

This section is summary of the demographic data provided by the respondents.

There was a preponderance of male respondents in the group (91%). This is not particularly unrepresentative in the engineering of defence systems. 95% of the respondents had a university level education. This may be compared to the general Australian population where 24% people aged between 15 and 64 years hold a bachelors level degree or higher (Australian Bureau of Statistics 2011) and demonstrates that the respondents as a group are significantly more educated than the general population as expected in the engineering and usage of complex systems. A majority of respondents (60%) identified their main technical education as being in electrical/electronic engineering.

The group exhibited significant professional experience (most with greater than 10 years as shown in Figure 1) and 56% with formal engineering delegations in ADF regulatory frameworks (refer Figure 3) representing an experienced and responsible group of professionals.

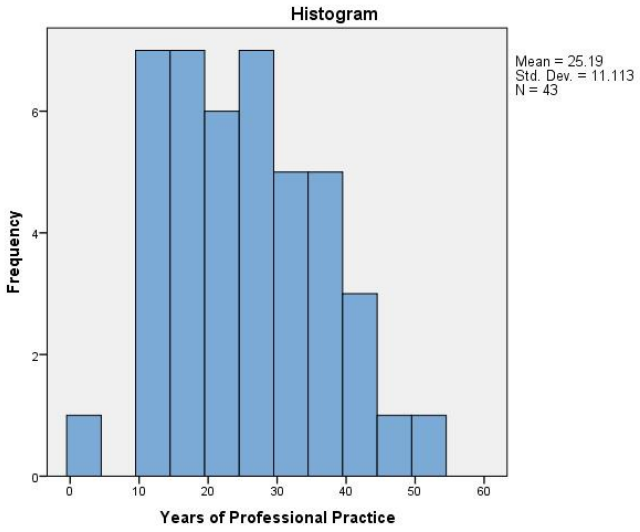


Figure 1. Histogram of the years of professional practice of the respondents.

Figure 2 shows the system application domain where the respondents claimed their main experience. A majority of respondents had their main experience in the application domains of Maritime Defence Systems and Aerospace Systems. It is possible that this reflects the researcher's own professional network from which most respondents were invited and also that the area of technical regulation and technical integrity is of high importance and is currently in active debate amid regulatory reform initiatives within the Australian Air Force and Navy particularly. The relatively small number of respondents, about 4.7%, in Commercial ICT may limit any generalisations that can be made across non-military application domains.

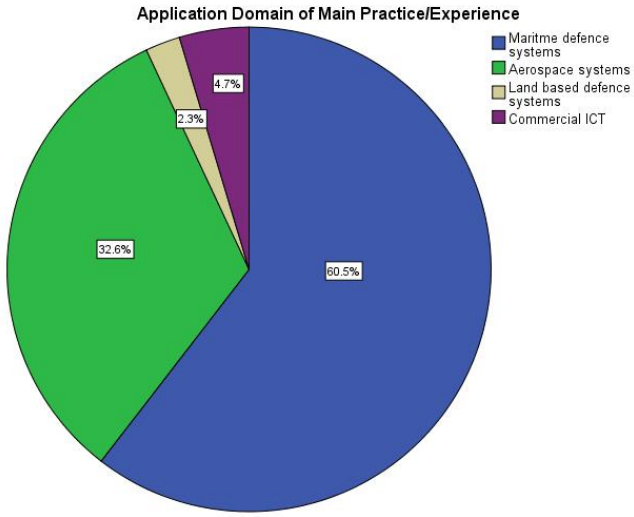


Figure 2. Pie chart of the system application domain where the respondent has most experience.

Twenty four (55.8%) of the respondents indicated that they hold or have held formal delegations of Engineering Authority in organisations governed by ADF technical regulations or other regulatory systems. The delegations held are shown in Figure 3. This demonstrated that many of the respondents represent a group of very experienced engineers and that they are, or have previously been, entrusted to make significant engineering decisions for the ADF. There is a bias in representation toward Naval Technical Regulatory System (NTRS) delegations (62.5% of all indicated delegations) that is consistent with the application domain of most experience (60.5% of respondents with main experience in maritime defence systems).

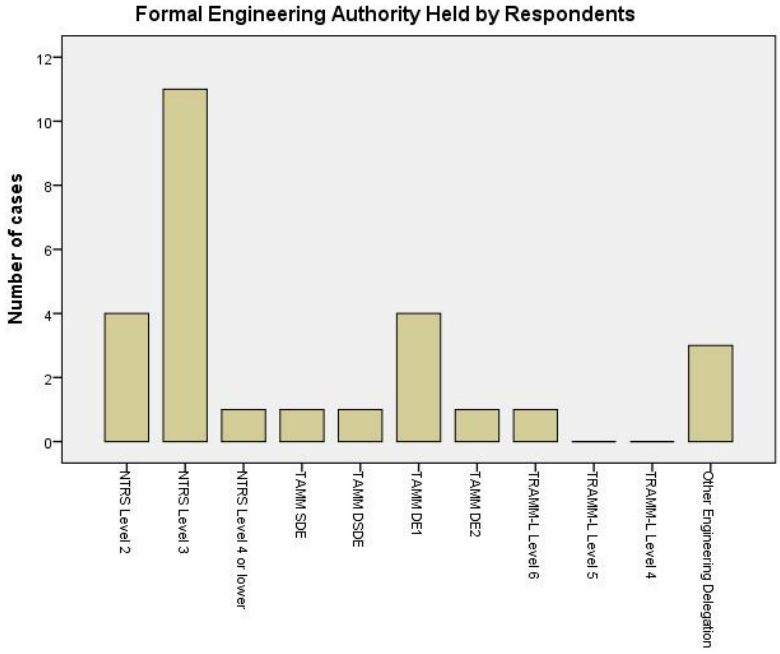


Figure 3. Bar chart of the engineering delegation held by respondents.

Figure 4 demonstrates that many of the group's experience with complex engineered systems was at what would be considered to be the higher levels of system abstraction: capability definition, system specification and design, system integration and test and project management. The relatively low numbers of respondents who cited experience with system operation, support and sustainment may cause a bias toward judgments being made with a development life-cycle perspective away from an in-service perspective.

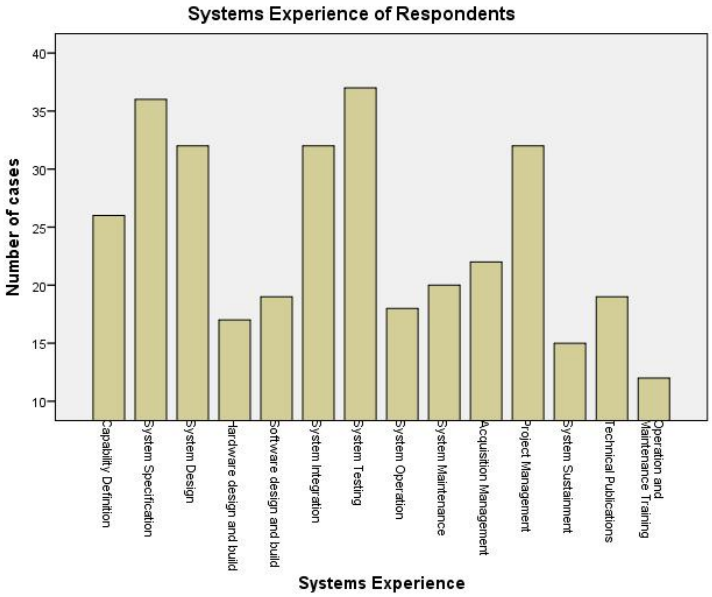


Figure 4. Systems experience of the respondents.

Analysis procedure

Ratings data was captured directly in an online, web-based survey instrument implemented in

SurveyMonkey. Data was transferred directly to commercial statistical software (SPSS v20) for analysis.

Non-parametric statistical analysis procedures are used since the rating scale is not intended to be interpreted as a ration scale. Estimates of mean, variance and standard deviation are therefore not mathematically valid. Central tendency of the ratings data is described by the median and the variability by range and interquartile range. Wilcoxon signed rank tests are used to compare the difference in medians and correlations are assessed using Spearman's rho (r_S).

RESULTS

Technical integrity ratings

The frequency of ratings from the judges for Systems A, B and C are shown in Figure 5, Figure 6 and Figure 7 respectively. Normal distribution curves are included for visualisation only. Figure 8 summarises the descriptive statistics for the same ratings. Figure 9 shows the median, range and inter-quartile range in a box-plot format for visualisation of the results.

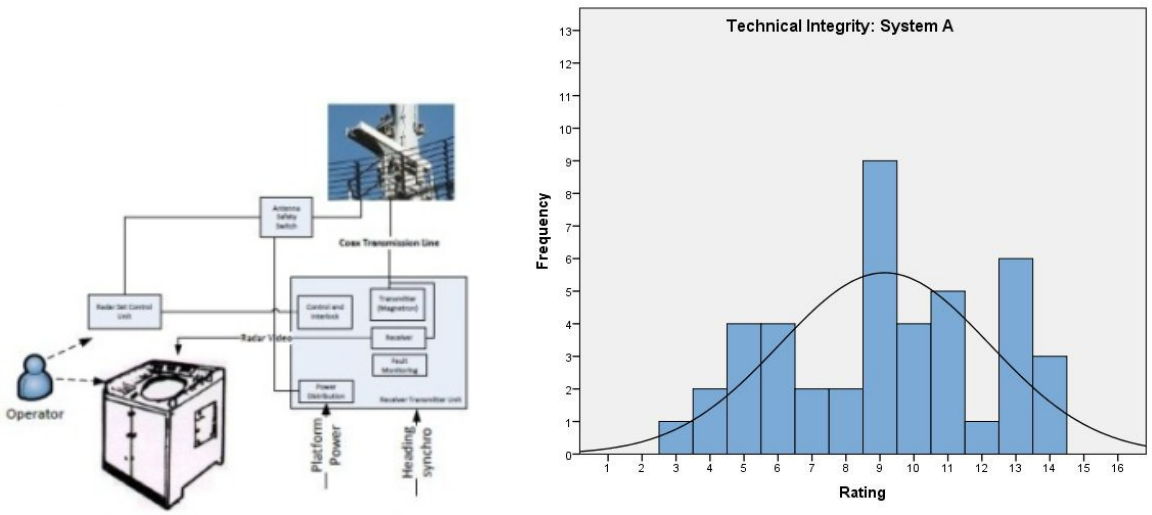


Figure 5. Ratings of technical integrity for System A.

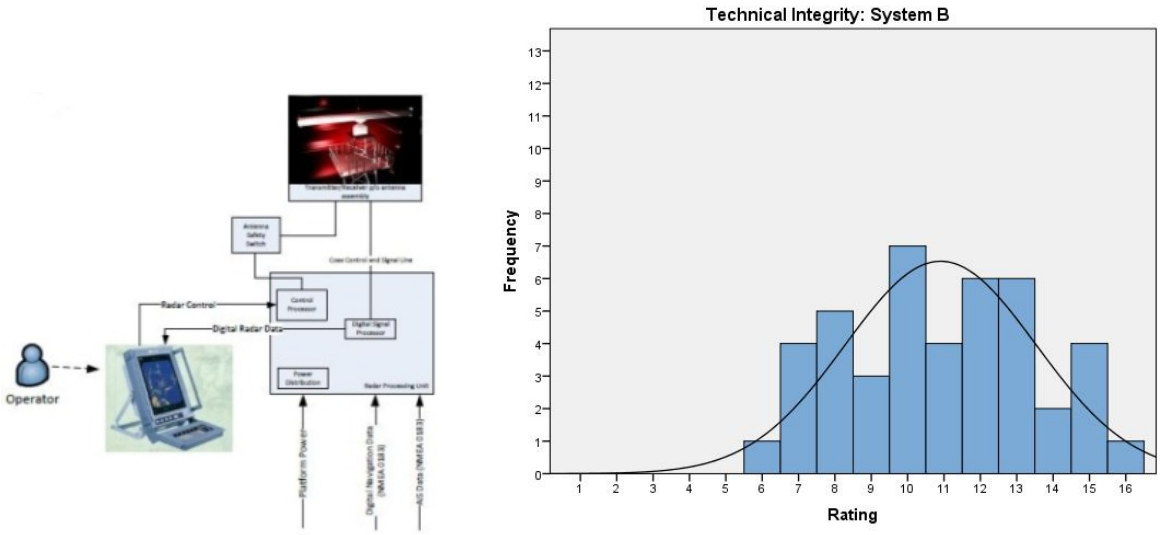


Figure 6. Ratings of technical integrity for System B.

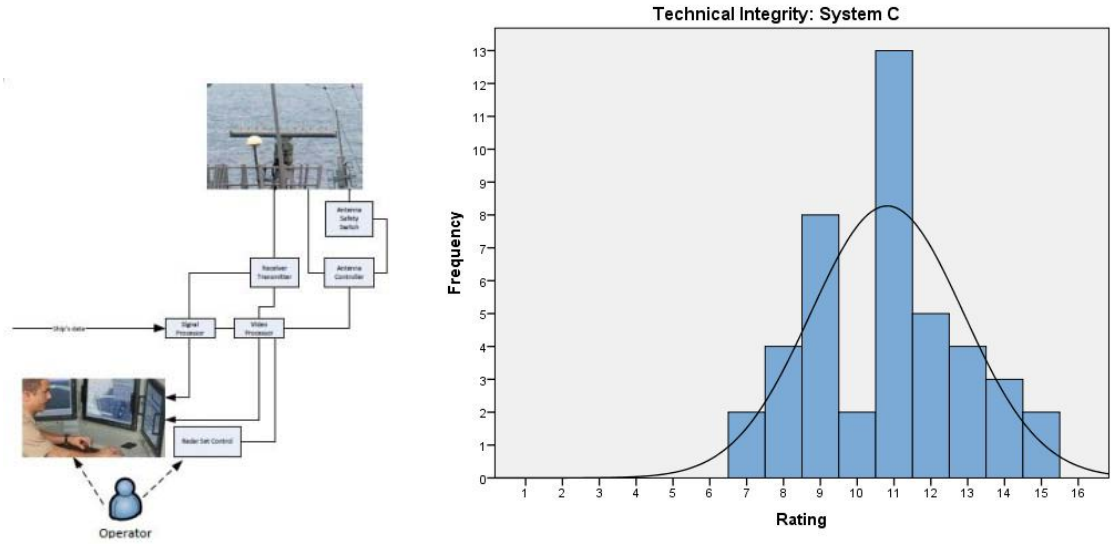


Figure 7. Ratings of technical integrity for System C.

Technical Integrity:		System A	System B	System C
N	Valid	43	43	43
	Missing	0	0	0
Median		**9**	**11**	**11**
Range		11	10	8
Minimum		3	6	7
Maximum		14	16	15
Interquartile range		**5**	**4**	**3**
Percentiles	25	6	9	9
	50	9	1	11
	75	11	13	12

Figure 8. Descriptive statistics for technical integrity ratings.

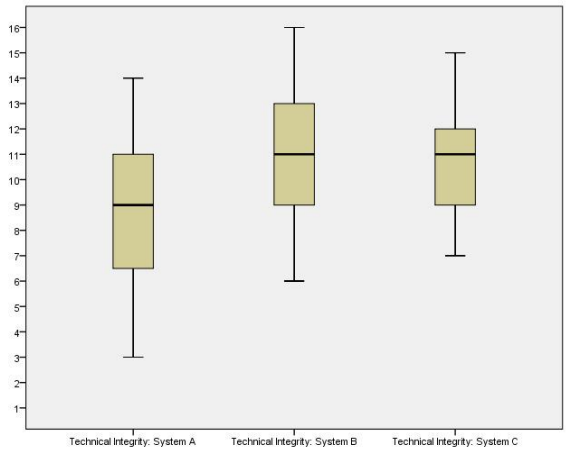


Figure 9. Medians, IQ range and range of technical integrity ratings.

Wilcoxon signed rank tests conducted at a significance level of 0.05 established that the median

ratings for System A and System B are significantly different, the median ratings of System A and System C are different and the median ratings of System B and System C are not significantly different. It is concluded from this that the consensus of judges is that the technical integrity of System B and System C are equivalent and exhibit higher technical integrity that that of System A. It is also noted that the variability as estimated by interquartile range is least for System C and most for System A.

Reliability

Interjudge reliability

(Amabile 1996, 68) utilises 3 techniques to calculate interjudge reliability of subjective assessments. The second uses the Spearman-Brown prediction formula:

$$reliability = \frac{n.r}{1 + (n-1).r}$$

Where *n* is the number of judges and *r* is the mean interjudge correlation.

This is consistent with the method described by (Wuensch 2007) in using the SPSS intraclass correlation coefficient for *r*. SPSS provides the 'single measure intraclass correlation' which is 'an index of the reliability of the ratings for a typical, single judge' and an 'average measure intraclass correlation' which is 'the reliability for all the judges averaged together'.

The intraclass correlation coefficients produced from SPSS are provided at Figure 10:

Intraclass Correlation Coefficient

	Intraclass Correlation[b]	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.113[a]	.021	.850	8.683	2	84	.000
Average Measures	.845[c]	.476	.996	8.683	2	84	.000

Two-way mixed effects model where people effects are random and measures effects are fixed.

a. The estimator is the same, whether the interaction effect is present or not.

b. Type A intraclass correlation coefficients using an absolute agreement definition.

c. This estimate is computed assuming the interaction effect is absent, because it is not estimable otherwise.

Figure 10. Intraclass correlation coefficients.

Using these intraclass correlation coefficients and *n*=43, the reliability equation yields:

- *reliability* = 0.996 (for an average of judges)
- *reliability* = 0.846 (for a typical judge)

Amabile also reports using the Cronbach's alpha from SPSS for a measure of interjudge reliability and claims it gives very similar results. The SPSS provided value for Cronbach's alpha is shown in Figure 11.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.885	.855	43

Figure 11. Cronbach's alpha for technical integrity ratings.

Characteristics of ratings

Pattern of judge's ratings

The chart below shows a scattergram of the mean value of each of the judge's three technical integrity

ratings (for systems A, B and C) versus the range of the same ratings. Each point on the scattergram represents the mean and range of the each judge's 3 ratings of technical integrity. The range of average values is from 7.0 to 14.7 indicating some judges rate the systems much more highly than others. It is also observed that the more optimistic raters (higher mean rating) tend to utilise a lower range of ratings (i.e. all 3 systems are relatively highly rated). This may just be an artefact of range compression at the top end of the scale.

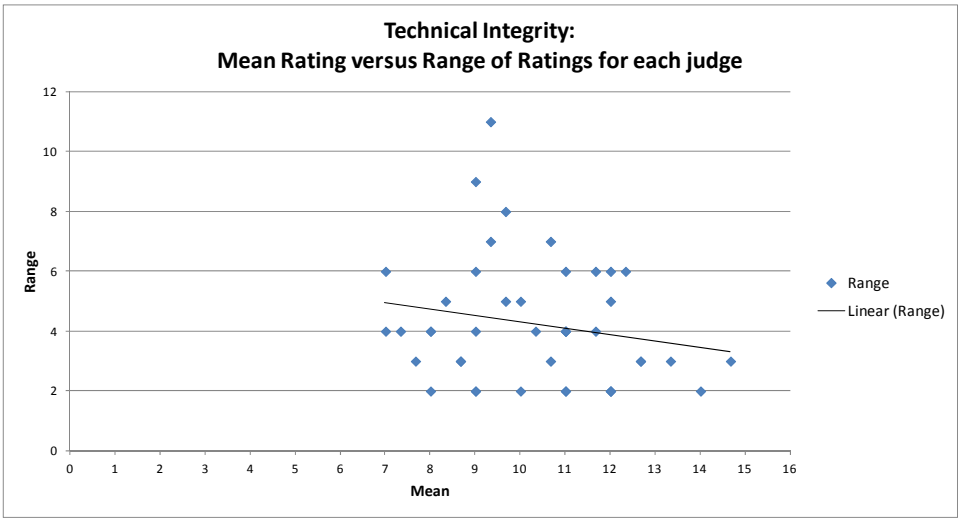


Figure 12. Scattergram of average of 3 ratings versus range of 3 ratings.

Judgment variation across demographic categories

In order to see if there is any correlation with any of the demographic data to explain observed ratings, a nonparametric correlation using Spearman's rho was conducted.

The mean of the three technical integrity ratings per judge was compared via nonparametric correlation to the demographic data comprising: highest level of technical education. Gender, age category, years of professional practice, application domain of main practice/experience, current employment sector, sector of most experience, ADF or other engineering delegation and the 14 experience categories. No statistically significant correlations exist. This is consistent with other CAT studies that indicate that judgments from experts and non-experts can be equally reliable (Amabile 1996, 72-73).

Correlation with other judgment dimensions

The judge's ratings for all 3 systems across all 20 dimensions were analysed for correlation to the ratings of technical integrity. Using Spearman's rho as the appropriate non-parametric statistic for correlation, the results are as shown in Figure 13. All correlations are statistically significant at the 0.01 level (2-tailed) with the single exception of Environmental Impact which is significant at the 0.05 level (2-tailed).

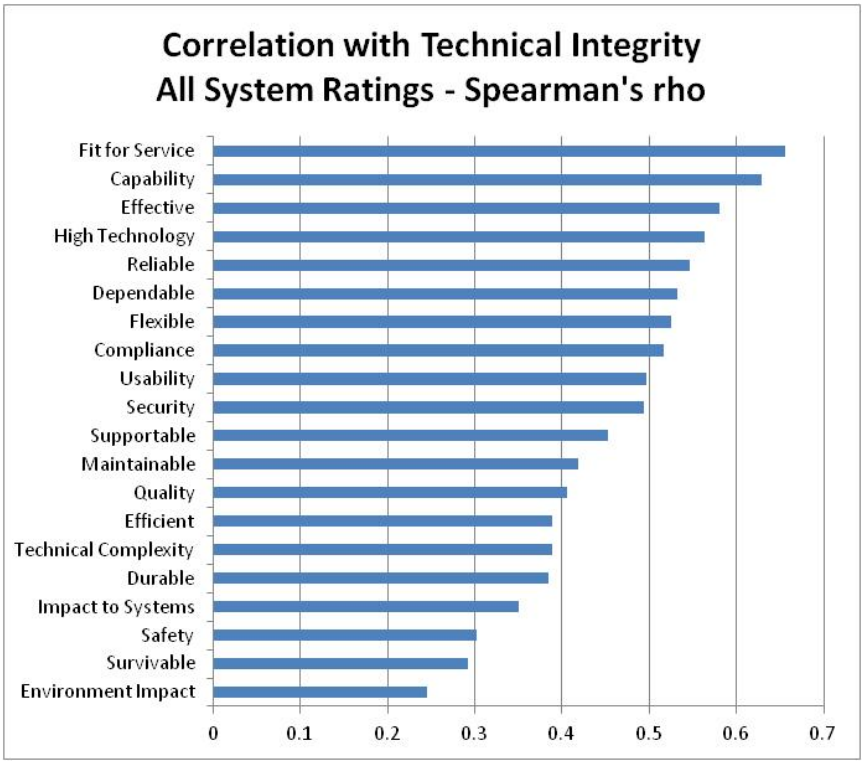


Figure 13. Correlation of judgment dimensions with technical integrity.

The strongest correlation is for Fit for Service at $r_S = 0.656$. A scatter plot with a linear regression line to assist visualisation of the relationship[2] is shown in Figure 14 (i). The weakest correlation is for Impact to the Natural Environment at $r_S = 0.245$. The correlation of Safety is also, surprisingly, weak at $r_S = 0.302$ as shown at in Figure 14 (ii).

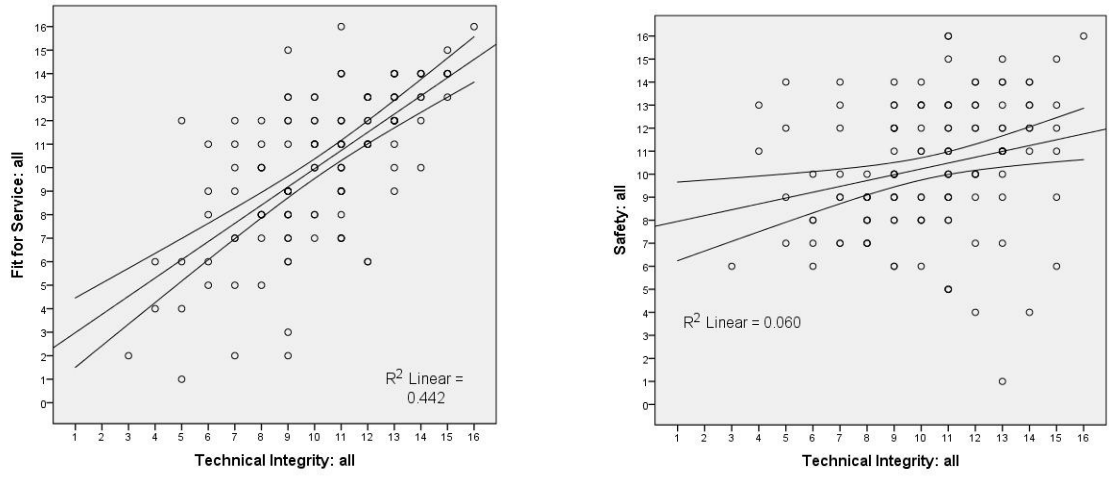


Figure 14. Scattergram of ratings for (i) Fitness for Service versus technical integrity ratings; (ii) Safety versus technical integrity ratings

2 Note – Spearman's rho does not presume or predict a linear correlation. It predicts the extent of a monotonic relationship.

DISCUSSION

Reliability

The reliability of a measurement instrument is an assessment as to the degree to which measured values are correlated with the true value of the property. In many cases, including technical integrity, where the value of the property cannot be directly observed or estimated, reliability assessment focuses on the correlation of observed scores on parallel applications of the instrument. That is, does the instrument objectively provide the same value when used by different people and/or at different times[3]?

(Amabile 1996, 69-71) reports the reliability results of studies using CAT across three domains with a range of Cronbach's alpha values as follows:

- Verbal creativity (e.g., stories and prose) – 0.67 to 0.91.
- Artistic creativity (e.g., drawings, paintings, sketches) – 0.53 to 0.92.
- Problem-solving creativity (e.g., computer programs, high-technology services, structures) – 0.61 to 0.87.

In this study, a value of alpha of 0.865 was obtained. This compares well with the higher end of the range of values reported by these other studies. An alpha of greater than 0.70 is generally considered to be an acceptable reliability for a measurement instrument for most social science research. The indirect measurement of technical integrity via expert judgement is reliable for this type of system at least.

Although this study was limited in scope to a particular type of engineered system (maritime radars), there is support for the assessment of technical integrity via CAT.

Correlation of judgment dimensions

The technical integrity judgements are positively correlated with all other judgment dimensions, however, there are no strong correlations[4].

The largest correlations are with Fitness for Service ($r_S = 0.656$) and Capability ($r_S = 0.629$) with other moderate correlations[5] being effectiveness, reliability, dependability, flexibility and compliance. Correlation of these judgment dimensions with technical integrity is as expected. 'High technology' also shows moderate correlation with technical integrity in this study. It is not expected that the level of technology employed in a system should be directly related to the technical integrity of the system and it would not have been surprising if this showed little or no correlation. Equally however, it may be readily construed that newer technology systems may often be considered to be technically "better".

The judgment dimensions of safety, impact to the natural environment, and survivability all show a weak correlation[6] with technical integrity ratings. The results for safety and impact to the natural environment are not as expected given that these judgment dimensions directly correspond to the elements of technical integrity explicitly defined in ADF technical regulations.

The other judgment dimensions exhibit a low correlation with technical integrity. The result that 12 of the 20 judgement dimensions have weak to low correlation with technical integrity suggests that inclusion of these dimensions as variables in a scale for technical integrity may not be necessary.

CONCLUSION

The CAT has been applied to a set of three contrived, example radar systems. Results indicate that

---

3 The reader is referred to (Allen and Yen 1979, Ch 4) and (DeVellis 2011, Ch 3) for a detailed description of measurement reliability.

4 A correlation is generally interpreted to be 'strong' when Spearman's rho is greater than 0.7.

5 A correlation is generally interpreted to be 'moderate' when Spearman's rho is in the range 0.5 to 0.7.

6 A correlation is generally interpreted to be 'weak' when Spearman's rho is in the range 0.16 to 0.3.

indirect estimation of the technical integrity of these systems is reliable (Cronbach's alpha = 0.885) and compares favourably with results obtained in other applications of the CAT. This suggests that indirect estimation of technical integrity of a complex engineered system via a consensus of judges can be a feasible and pragmatic measurement instrument.

The correlation of technical integrity ratings with judgement dimensions expected to be strongly correlated with technical integrity has provided some unexpected results. In particular, it would be expected that the ratings for safety and impact to the natural environment would correlate with technical integrity given the explicit inclusion of those dimensions in the ADF definition of technical integrity and the familiarity of most of the judges with ADF technical regulation.

The validity of generalising these conclusions is limited by the constrained example used in this study. Future research is intended to examine the correlation of the ratings of each judgment dimension with objective data about the example systems and the development of a scale from these directly measurable indicators.

REFERENCES

Allen, M.J., and W.M. Yen. 1979. *Introduction to measurement theory*: Brooks/Cole Monterey, CA.

Amabile, Teresa M. 1996. Consensual Assessment Technique. In *Creativity in Context*. Boulder, CO, USA: Westview Press.

Australian Bureau of Statistics. 2011. 6227.0 - Education and Work, Australia, May 2011 Australian Bureau of Statistics.

DeVellis, R.F. 2011. *Scale development: Theory and applications*. Vol. 26: Sage Publications, Incorporated.

Edwards, Michael. 2011. What is Technical Integrity and How to Measure It? In *INCOSE International Symposium 2011*. Denver, CO, USA.

Tashakkori, A., and C. Teddlie. 1998. *Mixed methodology: Combining qualitative and quantitative approaches*. Vol. 46: Sage Publications, Incorporated.

Wuensch, Karl L. 2012. *Inter-Rater Agreement*. East Carolina University 2007 [cited 12 October 2012 2012]. Available from core.ecu.edu/psyc/wuenschk/docs30/InterRater.doc.

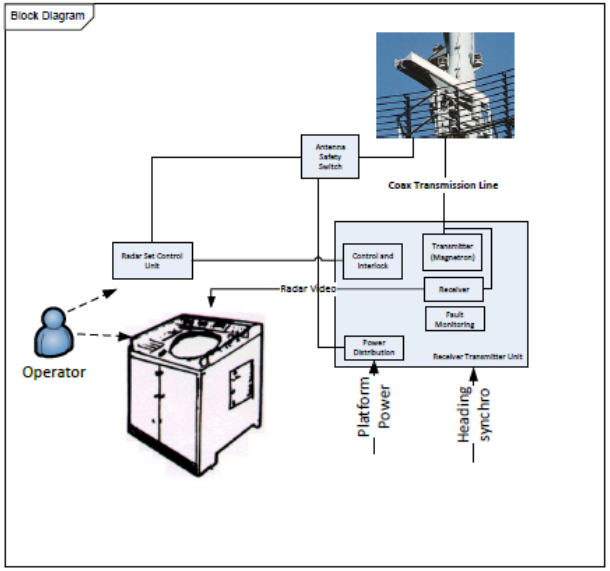
BIOGRAPHY

Michael Edwards holds a Bachelor of Engineering (Honours) and a Master of Engineering Science (Communications) from the University of NSW and a Master of Business Administration (Technology Management) from Deakin University. He is an Engineering Fellow with Raytheon Australia and is a part-time research student at the Defence and Systems Institute of the University of South Australia. He is a Fellow of the Institution of Engineers, Australia, and a Chartered Professional Engineer, a member of the Systems Engineering Society of Australia and a Member of INCOSE. His professional and academic interests build on an extensive experience base in the practice of engineering management and systems engineering largely in the domain of complex naval systems including applications in surface combatant and submarine combat systems, mine warfare systems, hydrographic systems and naval aviation systems.

ATTACHMENTS: CAT PRODUCTS – SYSTEM A, B AND C - THE 3 CONTRIVED MARITIME NAVIGATION RADAR SYSTEMS.

Radar A

Block Diagram



Capabilities & Performance

Primary Mission: detect ships, coastlines and hazards to navigation
Secondary Mission: low altitude air search
Instrumented Range: 190 km long pulse; 60 km short pulse
Range Accuracy: 22m, 198 m
Bearing Accuracy: 1.5°
Data rate: 16 rpm – 3.75sec
Auto Track Capacity: Nil; aids for manual tracking.

Interface to External Environment

RF Frequency: I-Band (X-band) tunable over 9.05 – 10.0 GHz
RF Bandwidth: 1.2 MHz long pulse; 10 MHz short puls
RF Power: 130-160kW peak
RF Waveform: 1µs pulse, 0.12 µs pulse
PRF: 750pps; 2250 pps

Military Threat Resistance

Unintentional jamming resistance: tunable frequency; polarisation
Intentional jamming resistance: polarisation
Counter Detection: N/A
Selective Emitter Identification: N/A

Shock and blast resistance: MIL-S-901

Technology

Antenna: two waveguide slotted arrays mounted back-to-back, one linearly horizontal polarised, one circularly polarised.
Transmission Line: coaxial
Transmitter: Magnetron
Receiver: Low noise RF front-end; sensitivity time control
Signal Processor: fast time constant filtering
Control: sector radiate control
Display: monochrome CRT; analog vide sweep

Supportability Features

Redundancy features: 5 display outputs
Maintenance Concept: on board repair by technicians
BIT Features: some, details not known
MTBF: 1200 hours
MTTR: 20 minutes

Provenance and Usage

Contracts: Originally contracted for in 1971.
Year first deployed: 1972
Number of units in service: >140
Platform classes: Frigates, destroyers, cruisers, minesweepers
Manufacturer Quality System: ISO 9001

Hazardous Materials:

Cadmium plated connectors used on all cable terminations.
No other hazmats disclosed.

Safety Features:

Antenna Safety Switch – when activated prevents rotation and radiation from the antenna by isolating power distribution to antenna assembly. Provides indication to operator.
Cabinet interlock for electrical safety.

Security Features:

Physical security: only as provided by installation
Computer security: no general purpose computer

Specification/Standard Compliance

Specified Requirements	System A
Radar Performance compliant with IMO Resolution A.477 (XII) when installed on platform	✓
Shall operate in I-Band (3cm)	✓
Range performance to IMO Resolution A.477 (XII)	✓
Detect 60 m high coastline at 26 nautical miles	✓
Detect large surface ship at 15 nautical miles	✓
Detect small surface ship at 10 nautical miles	✓
Detect navigational buoy at 4 nautical miles	✓
Range scales to 50 nautical miles	✓
Adjust pulse width with range scale in use (auto or manual)	✓
Display range rings	✗
Range discrimination of two small target seperated by 25 yards	✓
Radar shall process GPS inputs of position, couse and speed over ground	✗
Radar shall incorporate an ARPA in accordance with IMO resolution A.823	✗
Ability to store and recall up to 20 electronic maps	✗
Provide blanking signal to platform electronic support equipment.	✓
Antenna rotation rate selectable in the range 24 to 40 rpm	✗
The transceiver shall be an integral part of the antenna system	✗
Power input shall be to MIL-STD-1399 Section 300A.	✓
Key lockable Safe to Transmit and Safe to rotate switches to be provided.	i
EMI/EMC to MIL-STD-461D.	i
Alerts shall be audible/visible to the operator if there is a loss of signal at any interface.	i
Extensive BIT shall allow maintainer to isolate fault to a LRU.	i
The radar shall allow the operator to choose between SI and imperial units for display.	✗
Compass input required for azimuth stabilisation.	✓
EM Log input for sea stabilisation.	✓

✗ Not compliant to requirement
i Partially compliant to requirement
✓ Compliant to requirement

Build Standard and Environmental Resistance

Temperature:
(Receiver/Transmitter Unit and Radar Set Control)
Non-Operating: -62°C to +75°C
Operating: OIC to +50°C
Antenna Pedestal: -28°C to +65°C

Humidity:
Up to 95% at +60°C (MIL-E-16400)

Enclosure:
Receiver/Transmitter Unit and Radar Set Control 45°, drip-proof.
Safety Switch, splash-proof.
Antenna Pedestal, submergible to 3 feet.

Vibration, and Inclination:
MIL-E-16400

Salt Spray:
MIL-E-16400

Heat Test:
MIL-E-16400

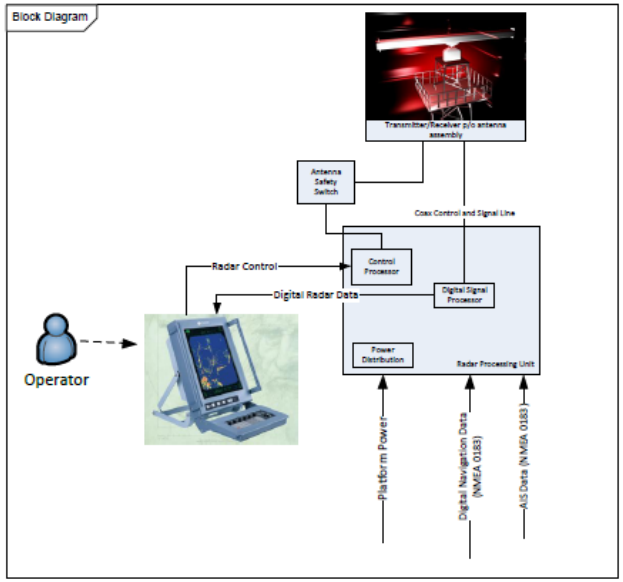
Platform Service Requirements

Power Type: 115 Vac/60Hz/1Φ
Power Draw: 1100W
Cooling Load to Air/Water: 100W
Data Inputs: synchro bearing

Weight:
Display: 140kg
Receiver/Transmitter: 260 kg
Antenna Group: 131 kg

Radar B

Block Diagram



Capabilities & Performance

Primary Mission: detect ships, coastlines and hazards to navigation
Instrumented Range: 200 km long pulse
Range Accuracy: 1% of target range
Bearing Accuracy: 1.2°
Data rate: 60 rpm – 1 sec
Auto Track Capacity: 20 auto tracks

Interface to External Environment

RF Frequency: I-Band (X-band) tunable over 9.05 – 10.0 GHz
RF Bandwidth: 10 MHz
RF Power: 100kW peak
RF Waveform: 0.1µs pulse
PRF: variable law selected range scale

Military Threat Resistance

Unintentional jamming resistance: matched filtering provides immunity form most interfering signals
Intentional jamming resistance: N/A
Counter Detection: N/A
Selective Emitter Identification: N/A

Shock and blast resistance: Nil

Technology

Antenna: circularly polarised waveguide slotted array
Transmitter: Solid state tranmitter integrated into antenna pedestal
Receiver: Low noise, solid state RF front-end integrated into antenna pedestal
Signal Processor: digital signal processor implementing coherent matched filter
Control: sector radiate control; variable PRF set auto with selected range scale
Display: high resolution LCD colour; electronic chart underlay; integration with AIS tracks

Supportability Features

Redundancy features: digital displayand track outputs can be provided to other display systems
Maintenance Concept: repair by return to OEM
BIT Features: full go/no go test results available to operator
MTBF: 4000 hours
MTTR: 30 minutes

Provenance and Usage

Contracts: Originally contracted for in 2001.
Year first deployed: 2003
Number of units in service: >30
Platform classes: Naval patrol boats, customs vessels, motor yachts
Manufacturer Quality System: ISO 9001

Hazardous Materials:

No hazmats disclosed.

Safety Features:

Antenna Safety Switch – when activated prevents rotation and radiation from the antenna by isolating power distribution to antenna assembly. Provides indication to operator.

Security Features:

Physical security: only as provided by installation
Computer security: password security enforced on display and processing unit

Specification/Standard Compliance

Specified Requirements	System B
Radar Performance compliant with IMO Resolution A.477 (XII) when installed on platform	✓
Shall operate in I-Band (3cm)	✓
Range performance to IMO Resolution A.477 (XII)	✓
Detect 60 m high coastline at 26 nautical miles	✓
Detect large surface ship at 15 nautical miles	✓
Detect small surface ship at 10 nautical miles	✓
Detect navigational buoy at 4 nautical miles	✓
Range scales to 50 nautical miles	✓
Adjust pulse width with range scale in use (auto or manual)	✓
Display range rings	✓
Range discrimination of two small target seperated by 25 yards	✓
Radar shall process GPS inputs of position, couse and speed over ground	✓
Radar shall incorporate an ARPA in accordance with IMO resolution A.823	✓
Ability to store and recall up to 20 electronic maps	✓
Provide blanking signal to platform electronic support equipment.	✓
Antenna rotation rate selectable in the range 24 to 40 rpm	✗
The transceiver shall be an integral part of the antenna system	✓
Power input shall be to MIL-STD-1399 Section 300A.	?
Key lockable Safe to Transmit and Safe to rotate switches to be provided.	✗
EMI/EMC to MIL-STD-461D.	✗
Alerts shall be audible/visible to the operator if there is a loss of signal at any interface.	✓
Extensive BIT shall allow maintainer to isolate fault to a LRU.	?
The radar shall allow the operator to choose between SI and imperial units for display.	✗
Compass input required for azimuth stabilisation.	✓
EM Log input for sea stabilisation.	✓

✗ Not compliant to requirement
i Partially compliant to requirement
✓ Compliant to requirement

Build Standard and Environmental Resistance

Meets the requirements of:
IEC 60945 Maritime Navigation and Radiocommunication Equipment and Systems – General Requirements – Methods of Testing and Required Test Results

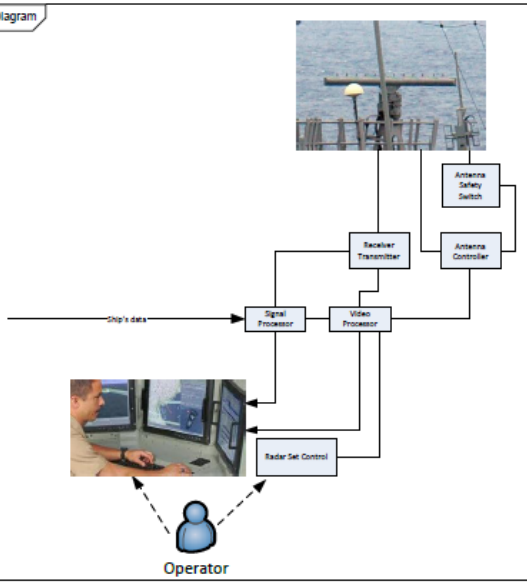
Platform Service Requirements

Power Type: 115 Vac/60Hz/1Φ
Power Draw: 1000W
Cooling Load to Air/Water:
Data Inputs: digital – NMEA 0183 format

Weight:
Display: 20 kg
Processor Unit: 40 kg
Antenna Group: 90 kg

Radar C

Block Diagram



Capabilities & Performance

Primary Mission: surface and air surveillance
Secondary Mission: navigation
Instrumented Range: 56 nm
Range Accuracy: 30m, 75 m, 300m
Bearing Accuracy: 1.5°
Data rate: 2 or 4 sec
Auto Track Capacity: 128, auto track initiation

Interface to External Environment

RF Frequency: C-Band 5.5 GHz
RF Power: 280kW peak
RF Waveform: 0.1µs pulse, 0.25 µs pulse, 1 µs pulse
PRF: 750, 1200, 2400

Military Threat Resistance

Unintentional jamming resistance: coherent EMI suppression
Intentional jamming resistance: jam strobe detection, sector CFAR
Counter Detection: none
Selective Emitter Identification: none

Shock and blast resistance: MIL-S-901D Grade A

Technology

Antenna: slotted array with built in IFF antenna
Transmission Line: coaxial
Transmitter: Coaxial Magnetron
Receiver: Low noise RF front-end; sensitivity time control
Signal Processor: sector by sector CFAR, coherent on receive Digital Moving Target Indicatot
Control: Auto Frequency Control, sector radiate control
Display: colour LCD display console; raw video sweeps with auto-detect track symbology

Supportability Features

Redundancy features: multiple display and track data outputs
Maintenance Concept: on board repair by technicians
BIT Features: extensive to replaceable part level
MTBF: 3000 hours
MTTR: < 30 minutes based on field data

Provenance and Usage

Contracts: For current variant let in 2000.
Year first deployed: 2003
Number of units in service: >50
Platform classes: Frigates, destroyers, cruisers,
Manufacturer Quality System: ISO 9001

Hazardous Materials:

Cadmium plated connectors used on all cable terminations.
Zinc chromate primers used on antenna group.
No other hazmats disclosed.

Safety Features:

Antenna Safety Switch – when activated prevents rotation and radiation from the antenna by isolating power distribution to antenna assembly. Provides indication to operator.
Cabinet interlock for electrical safety.

Security Features:

Physical security: only as provided by installation
Computer security: maintainer must use password to access computer elements of system.

Specification/Standard Compliance

Specified Requirements	System C
Radar Performance compliant with IMO Resolution A.477 (XII) when installed on platform	✓
Shall operate in I-Band (3cm)	✗
Range performance to IMO Resolution A.477 (XII)	✓
Detect 60 m high coastline at 26 nautical miles	✓
Detect large surface ship at 15 nautical miles	✓
Detect small surface ship at 10 nautical miles	✓
Detect navigational buoy at 4 nautical miles	?
Range scales to 50 nautical miles	✓
Adjust pulse width with range scale in use (auto or manual)	✓
Display range rings	✓
Range discrimination of two small target seperated by 25 yards	?
Radar shall process GPS inputs of position, couse and speed over ground	✓
Radar shall incorporate an ARPA in accordance with IMO resolution A.823	?
Ability to store and recall up to 20 electronic maps	✓
Provide blanking signal to platform electronic support equipment.	✓
Antenna rotation rate selectable in the range 24 to 40 rpm	?
The transceiver shall be an integral part of the antenna system	✗
Power input shall be to MIL-STD-1399 Section 300A.	✓
Key lockable Safe to Transmit and Safe to rotate switches to be provided.	✗
EMI/EMC to MIL-STD-461D.	✓
Alerts shall be audible/visible to the operator if there is a loss of signal at any interface.	✓
Extensive BIT shall allow maintainer to isolate fault to a LRU.	✓
The radar shall allow the operator to choose between SI and imperial units for display.	✗
Compass input required for azimuth stabilisation.	✓
EM Log input for sea stabilisation.	✓

✗ Not compliant to requirement
! Partially compliant to requirement
✓ Compliant to requirement

Build Standard and Environmental Resistance

Operating Temperature: -10°C to +50°C

Humidity:
Up to 95% at +45°C (MIL-STD-810)

Enclosure:
Inboard equipment: drip-proof to 15°
Antenna Pedestal: qualified for driving rain

Vibration:
MIL-STD-167-1

Platform Service Requirements

Power Type: 115 Vac/60Hz/1Φ
Power Draw: 2000W
Cooling Load to Air/Water: 1800W
Data Inputs: synchro and digital navigation data form ship (heading, speed, position)
Weight:
Display: 200 kg
Receiver/Transmitter: 20 kg
Antenna Group: 180 kg