Cranfield
UNIVERSITY

KAREN P. LANE

THE DEVELOPMENT OF USABILITY SCALES FOR FLIGHT MANAGEMENT SYSTEMS

COLLEGE OF AERONAUTICS

HUMAN FACTORS GROUP

PHD THESIS

ProQuest Number: 10832354

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.



ProQuest 10832354

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by Cranfield University.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Cranfield UNIVERSITY

COLLEGE OF AERONAUTICS

HUMAN FACTORS GROUP

PHD THESIS

ACADEMIC YEAR 2001-02

KAREN P. LANE

THE DEVELOPMENT OF USABILITY SCALES FOR FLIGHT MANAGEMENT SYSTEMS (VOLUME ONE)

SUPERVISOR: D. HARRIS

2001-02

THIS THESIS IS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

© Cranfield University 2002. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

ABSTRACT

Usability is a multidimensional concept, and although a number of techniques have been developed for assessing usability, none have been developed specifically for the civil flight deck or its components. The requirement for Human Factors tools to support civil flight deck certification have been well documented (e.g. FAA, 1996). This thesis reports on a research programme that resulted in the development of a series of scales to assess the usability of Flight Management System (FMS) Control and Display Units (CDUs). Since the goal of this research was to produce scales that could be used by certification authorities and FMS manufacturers, demonstrating that the scales had appropriate psychometric properties was critical. A rigorous and systematic approach to the construction of the usability scales was therefore adopted.

The research was conducted in three studies. The first study involved the development of a 49-item questionnaire that was distributed to airline pilots. A series of analyses on the data collected demonstrated that a reliable and valid usability factor structure was produced. The second study focused on the production of usability scales and examined the impact of flight phase on usability ratings. Analyses demonstrated that the factor structure produced in the first study was reliable, and was valid for data collected for the cruise and descent and approach phases of flight. A series of internally consistent usability scales was therefore produced which demonstrated evidence of discriminant, face, content, and construct validity.

The third study involved a content analysis on a large volume of participant comments derived during studies one and two, and examined the usability problems associated with current FMSs and suggested improvements for future FMSs. The content analysis produced evidence of construct validity for the usability scales developed, and permitted recommendations for FMS designers and certification authorities to be made.

ACKNOWLEDGEMENTS

I wish to express my gratitude to a number of people without whom this research could not have been undertaken. Firstly I must thank Dr. Don Harris for his unending patience, encouragement and support in supervising this thesis. I am also very grateful to all the other Human Factors staff at Cranfield University for their constructive criticism, advice and guidance throughout this research.

My thanks go out to the Human Factors team at Aerosystems International for providing the guidance and support necessary to conduct this research part-time. I would also like to thank the many airlines and airline pilots that participated in these studies. Finally, thanks to my family and friends for providing support through the more difficult times.

TABLE OF CONTENTS

1 INTRODUCTION 1

1.1 OVERVIEW 1
1.2 BACKGROUND 2
1.3 AVIATION AUTOMATION 5
1.3.1 Problems Associated with Aviation Automation 7
1.3.2 Human-Centered Aviation Automation 10
1.3.3 Aircraft Design Philosophies 10
1.3.4 The Modern Flight Deck 12
1.4 FLIGHT MANAGEMENT SYSTEMS 14
1.4.1 Introduction 14
1.4.2 Flight Management System Functions 14
1.4.3 FMS CDU 15
1.4.4 Deficiencies in Current CDUs 17
1.5 HUMAN FACTORS FOR FLIGHT DECK CERTIFICATION 19
1.5.1 Review of FAR/JAR 25 Regulations 19
1.5.2 Critical analysis of the role of usability in the Certification process 21
1.6 HUMAN-COMPUTER INTERACTION AND THE FMS 22
1.7 USABILITY 23
1.7.1 An Overview of Usability 24
1.7.2 Defining Usability 26
1.8 COMPONENTS OF USABILITY 31
1.8.1 Learnability 31
1.8.2 Visual Clarity 31
1.8.3 Consistency & Compatibility 32
1.8.4 Error Prevention & Correction 33
1.8.5 Appropriate Functionality 34
1.8.6 Effectiveness 35
1.8.7 Efficiency 35
1.8.8 Feedback 36
1.8.9 Explicitness 37
1.8.10 Flexibility 37
1.8.11 Workload 38
1.8.12 'Help' Facilities 38
1.8.13 Attitude 39
1.8.14 Summary of Usability Components 40
1.9 USABILITY AND THE FMS 42
1.10 BENEFITS OF ASSESSING USABILITY 51
1.11 METHODS FOR ASSESSING USABILITY 52
1.11.1 Subjective Evaluation Methods 52
1.11.2 Objective Evaluation Methods 53
1.11.3 Expert Evaluation Methods 54
1.11.4 Experimental Evaluation Methods 55
1.11.5 Selecting a Usability Assessment Method 55
1.11.6 Methodological Constraints and Considerations 57
1.11.7 Justification for the approach 59
1.12 SCALE PROPERTIES 60
1.12.1 Reliability 60
1.12.2 Validity 61
1.12.3 Properties of Usability Scales for Certification 62
1.12.4 Summary of potential benefits in a scale-based approach to usability in certification 63

1.12.5 Summary of potential weaknesses in a scale-based approach to usability in certification *64*
1.13 STUDY AIMS 66

2 STUDY ONE INTRODUCTION 68

2.1 PURPOSE 68
2.2 METHODOLOGICAL CONSIDERATIONS 69
2.2.1 Deriving Knowledge from Line Pilots *69*
2.2.2 Use of a Self-Completion Questionnaires *70*
2.2.3 Methodological Constraints *71*
2.2.4 Considerations of Rank and Culture *71*

3 STUDY ONE METHOD 73

3.1 OVERVIEW OF METHOD ONE PHASES 73
3.1.1 Phase One – Elicitation of Usability Issues from the Literature *73*
3.1.2 Phase Two - Familiarisation with the FMS *73*
3.1.2.1. Operator Manuals 74
3.1.2.2. CDU Part-Task Simulator 74
3.1.3 Phase Three - Flight Deck Visits *75*
3.1.4 Phase Four - Semi-Structured Interviews *75*
3.1.4.1. Methodological Considerations 75
3.1.4.2. Development of Interview schedule 76
3.1.4.3. Participants 76
3.1.4.4. Sampling 76
3.1.4.5. Procedure 76
3.1.5 Phase Five - Questionnaire Production *77*
3.1.5.1. Methodological Considerations 77
3.1.5.2. Derivation of Questionnaire Items 78
3.1.5.3. Pre-Pilot Questionnaire 79
3.1.5.4. Pilot Study 80
3.1.6 Study One Final Questionnaire *81*
3.1.6.1. Target Population 82
3.1.6.2. Response Rate Considerations 82
3.1.7 Questionnaire Distribution *83*

4 STUDY ONE RESULTS 84

4.1 OVERVIEW 84
4.1.1 Sample Description *84*
4.1.2 Treatment of data *88*
4.2 DESCRIPTIVE ANALYSIS 88
4.3 PRINCIPAL COMPONENTS ANALYSIS 91
4.3.1 Assumptions in PCA *91*
4.3.2 Extraction of Factor Structure *94*
4.3.2.1. Derivation of Factor Labels 96
4.4 DISCRIMINANT FUNCTION ANALYSES 102
4.4.1 Assumptions *102*
4.4.2 Discriminant Validity Evidence - DFA Examining Impact of Aircraft and FMS Manufacturer *103*
4.4.3 Discriminant Validity Evidence - DFA Examining Impact of FMS Manufacturer *108*
4.4.4 Discriminant Validity Evidence - DFA Examining Impact of Aircraft Manufacturer *111*
4.4.5 Discriminant Validity Evidence - DFA Examining Impact of Participants' Culture on FMS Usability Ratings *115*
4.4.6 Discriminant Validity Evidence - DFA Examining Impact of Pilot Rank on FMS Usability Ratings *117*
4.5 RELIABILITY ANALYSIS 120

5 STUDY ONE DISCUSSION 125

5.1 INTRODUCTION 125

5.1.1 Interpretation of Descriptive Analyses *126*
5.1.2 Interpretation of Factor Analysis *126*
5.2 RELIABILITY AND VALIDITY EVIDENCE 137
5.3 CONCLUSIONS 139

6 STUDY TWO INTRODUCTION 140

6.1 PURPOSE 140

7 STUDY TWO METHOD 142

7.1 OVERVIEW OF METHOD TWO PHASES 142
7.2 QUESTIONNAIRE MODIFICATIONS 142
7.3 METHODOLOGICAL CONSIDERATIONS 145
7.4 PILOT STUDY 145
7.5 RECRUITING THE PARTICIPANT SAMPLE 145
7.6 STUDY TWO FINAL QUESTIONNAIRE 146
7.6.1 Distribution Mechanism *146*

8 STUDY TWO RESULTS 147

8.1 OVERVIEW 147
8.2 SAMPLE DESCRIPTION 147
8.2.1 Response Rate *147*
8.2.2 Respondents *147*
8.3 TREATMENT OF DATA 151
8.4 DESCRIPTIVE ANALYSIS 151
8.5 ANALYSIS OF USABILITY RATINGS BY FLIGHT PHASE 153
8.6 CONFIRMATORY FACTOR ANALYSIS 154
8.6.1 Initial CFA Model *156*
8.6.2 Final Revised CFA Model *167*
8.6.3 CFA Model Validation for Cruise Phase *175*
8.6.4 CFA Model Validation for Descent & Approach Phase *182*
8.6.5 CFA Summary *189*
8.7 PRODUCTION OF SCALES 190
8.8 DISCRIMINANT FUNCTION ANALYSIS 192
8.8.1 Discriminant Validity Evidence - DFA Examining Impact of FMS and Aircraft Manufacturer *192*
8.8.2 Discriminant Validity Evidence - DFA Examining Impact of FMS Manufacturer *197*
8.8.3 Discriminant Validity Evidence - DFA Examining Impact of Aircraft Manufacturer *200*
8.8.4 Discriminant Validity Evidence - DFA Examining Impact of Pilot Rank *203*
8.8.5 Discriminant Validity Evidence - DFA Examining Impact of Flight Phase *205*
8.9 FURTHER ANALYSIS OF RELATIONSHIP BETWEEN PILOTS' EXPERIENCE AND USABILITY RATINGS 207
8.10 MULTIPLE REGRESSION ON FMS SATISFACTION RATINGS 208

9 STUDY TWO DISCUSSION 210

9.1 OVERVIEW OF STUDY TWO FINDINGS 210
9.2 INTERPRETATION OF DESCRIPTIVE ANALYSES 211
9.3 INTERPRETATION OF CONFIRMATORY FACTOR ANALYSES 211
9.4 RELIABILITY AND VALIDITY EVIDENCE 212
9.5 INTERPRETATION OF ANALYSIS ON USERS' EXPERIENCE 218
9.6 LIMITATIONS OF THE USABILITY SCALES DEVELOPED 219
9.7 FMS CDU SATISFACTION 221
9.8 CONCLUSIONS 222

10 STUDY THREE INTRODUCTION 223

10.1 PURPOSE 223
10.2 QUALITATIVE RESEARCH APPROACH 223
10.2.1 Methodological Considerations *224*

11 STUDY THREE METHOD 226

11.1 STUDY AIM 226
11.2 PROCEDURE 226
11.3 PARTICIPANTS 230

12 STUDY THREE RESULTS 232

12.1 OVERVIEW 232
12.2 CATEGORIES OF REPORTED FMS PROBLEMS & SUGGESTED IMPROVEMENTS 235
12.3 PROCESSOR SPEED 238
12.3.1 Processor Speed Problems 238
12.3.2 Processor Speed Suggested Improvements 238
12.3.3 Processor Speed Discussion 239
12.4 CDU ARCHITECTURE & LOGIC 240
12.4.1 CDU Architecture & Logic Problems 240
12.4.2 CDU Architecture & Logic Suggested Improvements 241
12.4.3 CDU Architecture & Logic Discussion 241
12.5 APPROPRIATE FUNCTIONALITY 242
12.5.1 Functional Limitations Problems 243
12.5.2 Additional Functionality Suggested Improvements 243
12.5.3 Appropriate Functionality Discussion 244
12.6 CONTROL DEVICES 245
12.6.1 Control Device Problems 245
12.6.2 Control Devices Suggested Improvements 246
12.6.3 Control Devices Discussion 246
12.7 ACCURACY OF PARAMETERS 247
12.7.1 Accuracy of Parameters Problems 247
12.7.2 Accuracy of Parameters Suggested Improvements 247
12.7.3 Accuracy of Parameters Discussion 248
12.8 FEEDBACK, ERROR PREVENTION & CORRECTION 249
12.8.1 Feedback, Error Prevention & Correction Problems 249
12.8.2 Feedback, Error Prevention & Correction Suggested Improvements 250
12.8.3 Feedback, Error Prevention & Correction Discussion 251
12.9 DISPLAY QUALITY & INFORMATION PRESENTATION 253
12.9.1 Display Quality & Information Presentation Problems 253
12.9.2 Display Quality & Information Presentation Suggested Improvements 253
12.9.3 Display Quality & Information Presentation Discussion 254
12.10 DATABASE / MEMORY SIZE 256
12.10.1 Database / Memory Size Problems 256
12.10.2 Database / Memory Size Suggested Improvements 256
12.10.3 Database / Memory Size Discussion 257
12.11 USER GUIDANCE & SUPPORT 259
12.11.1 User Guidance & Support Problems 259
12.11.2 User Guidance & Support Improvements 259
12.11.3 User Guidance and Support Discussion 260
12.12 CONSISTENCY AND COMPATIBILITY 261
12.12.1 Consistency and Compatibility Problems 261
12.12.2 Consistency and Compatibility Suggested Improvements 262
12.12.3 Consistency and Compatibility Discussion 262
12.13 TRAINING 264
12.13.1 Training Problems 264
12.13.2 Suggested Training Improvements 264
12.13.3 Training Discussion 265
12.14 INTUITIVENESS 266
12.14.1 Intuitiveness Problems 266
12.14.2 Intuitiveness Suggested Improvements 266
12.14.3 Intuitiveness Discussion 267
12.15 WORKLOAD 268

12.15.1 Workload Problems *268*
12.15.2 Workload Discussion *268*
12.16 OTHER CATEGORIES OF SUGGESTED IMPROVEMENTS REPORTED BY RESPONDENTS 270
12.16.1 Improved "What if?" Capability *270*
12.16.2 Improved "What If?" Capability Discussion *270*
12.16.3 Improved Barriers to Erroneous Operation *271*
12.16.4 Improved Barriers to Erroneous Operation Discussion *271*
12.16.5 Language Improvements *272*
12.16.6 Language Discussion *272*
12.17 SPECIFIC RECOMMENDATIONS FOR FMS DESIGNERS & CERTIFICATION AUTHORITIES 273
12.18 SUMMARY 276

13 CONCLUSIONS & RECOMMENDATIONS 277

13.1 CONCLUSIONS 277
13.1.1 Study One Conclusions *278*
13.1.2 Study Two Conclusions *281*
13.1.3 Study Three Conclusions *284*
13.2 OVERALL CONCLUSIONS 293
13.3 RESEARCH CONTRIBUTIONS 294
13.4 FUTURE DEVELOPMENTS 297

14 REFERENCES 299

VOLUME TWO – APPENDICES 313

TABLE OF FIGURES

Figure 1-1 - The effects of increasing automation on the pilot-aircraft relationship 6
Figure 1-2 - Boeing 747-400 Flight Deck 12
Figure 1-3 - Generic Glass Cockpit 13
Figure 1-4 - Boeing 737 FMS CDU 16
Figure 1-5 - Pre-Flight Process of making CDU inputs 18
Figure 1-6 - Causal Framework of Usability 26
Figure 1-7 - Usability Framework from ISO 9241 Part 11 p.3 30
Figure 1-8 - Conceptual Model of Usability 40
Figure 1-9 - Heuristic Chart for Selection of Usability Methods and Techniques 57
Figure 4-1 - Histogram of Pilots' Total Hours 85
Figure 4-2 - Histogram of Pilots' Hours on Current Aircraft Type 85
Figure 8-1 - Histogram of Pilots' Total Hours 148
Figure 8-2 - Histogram of Pilots' Hours on Current Aircraft Type 149
Figure 8-3 - Initial CFA Structural Diagram 155
Figure 8-4 - Final CFA Structural Diagram 167

TABLE OF TABLES

Table 1-1 - Characteristics of Usability Assessment Methods 56
Table 4-1 - Summary Participants' Age, Hours on Type & Total Hours 84
Table 4-2 - Summary of Participants' Rank 86
Table 4-3 - Summary of Airline Participation 86
Table 4-4 - Summary of Current Aircraft Type and Series 87
Table 4-5 - Summary of Reported FMS Manufacturer 87
Table 4-6 - Summary of Real FMS Manufacturer 88
Table 4-7 - Means & Standard Deviations for Questionnaire Items 89
Table 4-8 - KMO Test of Sampling Adequacy 92
Table 4-9 - Individual KMOs 93
Table 4-10 - Principal Components Analysis 95
Table 4-11 - Factor Labelling 96
Table 4-12 - Factor Loadings for Components 99
Table 4-13 - Questionnaire Items Not Loading On Any Component 101
Table 4-14 - Canonical Discriminate Function Analysis Summary Statistics 104
Table 4-15 - Functions at Group Centroids 104
Table 4-16 - Wilks' Lambda, Univariate F Ratios with 3,217 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients. 105
Table 4-17 - Group Statistics 106
Table 4-18 - Usability Ratings Ranked by Aircraft and FMS Manufacturer 107
Table 4-19 - Classification Matrix for Discriminant Function Analysis Predicting Group Members from Aircraft and FMS Manufacturer 107
Table 4-20 - Canonical Discriminate Function Analysis Summary Statistics 108
Table 4-21 - Wilks' Lambda, Univariate F Ratios with 1,210 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients. 109

Table 4-22 - Group Statistics 109
Table 4-23 - Usability Ratings Ranked by FMS Manufacturer 110
Table 4-24 - Classification Matrix for DFA Predicting Group Members from FMS Manufacturer.. 110
Table 4-25 - Canonical Discriminate Function Analysis Summary Statistics 111
Table 4-26 - Functions at Group Centroids 112
Table 4-27 - Wilks' Lambda, Univariate F Ratios with 2,218 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients. 112
Table 4-28 - Group Statistics 113
Table 4-29 - Usability Ratings Ranked by Aircraft Manufacturer 114
Table 4-30 - Classification Matrix for DFA Predicting Group Members from Aircraft Manufacturer114
Table 4-31 - Canonical Discriminate Function Analysis Summary Statistics 115
Table 4-32 - Functions at Group Centroids 115
Table 4-33 - Wilks' Lambda, Univariate F Ratios with 1,218 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients 116
Table 4-34 - Group Statistics 116
Table 4-35 - Classification Matrix for DFA Predicting Group Members from Participants' Culture 117
Table 4-36 - Canonical Discriminate Function Analysis Summary Statistics 118
Table 4-37 - Wilks' Lambda, Univariate F Ratios with 1,212 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients. 118
Table 4-38 - Group Statistics 119
Table 4-39 - Classification Matrix for DFA Predicting Group Members from Rank 119
Table 4-40 - Reliability Coefficient - Structure & Sequence 120
Table 4-41 - Reliability Coefficient - Appropriate Functionality 121
Table 4-42 - Reliability Coefficient - Feedback 121
Table 4-43 - Reliability Coefficient - Predictability 122
Table 4-44 - Reliability Coefficient - Information Presentation 122
Table 4-45 - Reliability Coefficient - Consistency 123
Table 4-46 - Reliability Coefficient - Flexibility & Control 123
Table 4-47 - Reliability Coefficient - Error Prevention & Correction 123
Table 4-48 - Reliability Coefficient - Processor Speed 124
Table 4-49 - Reliability Coefficient - Workload 124
Table 7-1 - Questionnaire Items Removed from Subsequent Study 142
Table 7-2 - Final Questionnaire Items 144
Table 8-1 - Summary Participants' Age, Hours on Type & Total Hours 148
Table 8-2 - Summary of Participants' Rank 149
Table 8-3 - Summary of Participation and Response Rate by Airline 150
Table 8-4 - Summary of Current Aircraft Type and Series 150
Table 8-5 - Summary of Reported FMS Type Currently in Use 150
Table 8-6 - Summary of Actual FMS Type Currently in Use 151
Table 8-7 - Means & Standard Deviations for Questionnaire Items 152
Table 8-8 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Structure & Sequence 157
Table 8-9 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Appropriate Functionality 158
Table 8-10 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Feedback 158

Table 8-11 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Predictability 159
Table 8-12 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Information Presentation 159
Table 8-13 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Flexibility & Control 160
Table 8-14 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Error Prevention & Correction 160
Table 8-15 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Processor Speed 161
Table 8-16 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Workload 161
Table 8-17 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Consistency 162
Table 8-18 - Squared Multiple Correlations 163
Table 8-19 - Summary of Models Initial CFA 165
Table 8-20 - Indicators Eliminated in Final Model 166
Table 8-21 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Structure & Sequence 168
Table 8-22 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Appropriate Functionality 168
Table 8-23 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Feedback 169
Table 8-24 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Predictability 169
Table 8-25 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Information Presentation 170
Table 8-26 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Error Prevention & Correction 170
Table 8-27 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Processor Speed 171
Table 8-28 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Workload 171
Table 8-29 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Consistency 172
Table 8-30 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Flexibility for Navigation 172
Table 8-31 - Squared Multiple Correlations 173
Table 8-32 - Summary of Models (Final Model) 174
Table 8-33 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Structure & Sequence 175
Table 8-34 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Appropriate Functionality 176
Table 8-35 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Feedback 176
Table 8-36 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Predictability 177
Table 8-37 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Information Presentation 177

Table 8-38 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Error Prevention & Correction 178
Table 8-39 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Processor Speed 178
Table 8-40 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Workload 178
Table 8-41 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Consistency 179
Table 8-42 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Flexibility for Navigation 179
Table 8-43 - Squared Multiple Correlations (Cruise) 180
Table 8-44 - Summary of Models (Cruise) 181
Table 8-45 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Structure & Sequence 182
Table 8-46 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Appropriate Functionality 183
Table 8-47 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Feedback 183
Table 8-48 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Predictability 184
Table 8-49 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Information Presentation 184
Table 8-50 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Error Prevention & Correction 185
Table 8-51 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Processor Speed 185
Table 8-52 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Workload 186
Table 8-53 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Consistency 186
Table 8-54 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Flexibility for Navigation 187
Table 8-55 - Squared Multiple Correlations (Descent & Approach) 188
Table 8-56 - Summary of Models (Descent & Approach) 189
Table 8-57 - Reliability Coefficients 190
Table 8-58 - Pearson Correlation Coefficients between the Usability Scales Produced 191
Table 8-59 - Canonical Discriminant Function Analysis Summary Statistics 193
Table 8-60 - Functions at Group Centroids 193
Table 8-61 - Wilks' Lambda, Univariate F Ratios with 2,328 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients 194
Table 8-62 - Group Statistics 194
Table 8-63 - Usability Ratings Ranked by Aircraft and FMS Manufacturer 195
Table 8-64 - Classification Matrix for DFA Predicting Group Members from Aircraft and FMS Manufacturer 196
Table 8-65 - Comparison of Usability Rating Rank Orders by Aircraft and FMS Manufacturer 197
Table 8-66 - Canonical Discriminant Function Analysis Summary Statistics 197
Table 8-67 - Wilks' Lambda, Univariate F Ratios with 1,329 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients. 198
Table 8-68 - Group Statistics 198

Table 8-69 - Usability Ratings Ranked by FMS Manufacturer 199
Table 8-70 - Classification Matrix for DFA Predicting Group Members from type rating on Aircraft with Smiths or Honeywell FMSs 199
Table 8-71 - Comparison of Usability Rank Orders by FMS Manufacturer 199
Table 8-72 - Canonical Discriminant Function Analysis Summary Statistics 200
Table 8-73 - Wilks' Lambda, Univariate F Ratios with 1,329 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients 200
Table 8-74 - Group Statistics 201
Table 8-75 - Usability Ratings Aircraft Manufacturer 201
Table 8-76 - Classification Matrix for DFA Predicting Group Members from Boeing/Airbus type-rating 202
Table 8-77 - Comparison of Usability Rank Orders by Aircraft Manufacturer 202
Table 8-78 - Canonical Discriminant Function Analysis Summary Statistics 203
Table 8-79 - Wilks' Lambda, Univariate F Ratios with 1,310 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients 203
Table 8-80 - Group Statistics 204
Table 8-81 - Classification Matrix for DFA Predicting Group Members from Rank 204
Table 8-82 - Canonical Discriminant Function Analysis Summary Statistics 205
Table 8-83 - Wilks' Lambda, Univariate F Ratios with 1,330 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients. 205
Table 8-84 - Group Statistics 206
Table 8-85 - Classification Matrix for DFA Predicting Group Members from Flight Phase Rated .. 206
Table 8-86 - Summary Model 209
Table 8-87 - ANOVA for Satisfaction 209
Table 8-88 - Group Statistics 209
Table 9-1 - Comparison of Discriminant Validity Evidence from Studies One and Two 217
Table 11-1 - Qualitative Categories of Problems with current FMSs 228
Table 11-2 - Qualitative Categories for Suggested Improvements to FMSs 229
Table 11-3 - Sample Description 230
Table 11-4 - Number of Comments by Participants' Current Type-rating 231
Table 12-1 - Qualitative Categories of Problems with current FMSs 233
Table 12-2 - Qualitative Categories of Suggested Improvements to FMSs 234
Table 12-3 - Problems with Current FMSs 236
Table 12-4 - Improvements Sought to Future FMSs 237
Table 13-1 - Psychometric Properties of Usability Factors 281
Table 13-2 - Psychometric Properties of Usability Scales 284

Abbreviations

A/T	Auto-Throttle
AAIB	Air Accidents Investigation Branch
AC	Advisory Circulars
ACARS	Aviation Communication And Reporting System
ACBI	Aviation Computer-Based Instruction
AFDS	Autopilot/Flight Director System
AGFI	Adjusted Goodness-of-Fit Index
ALPA	AirLine Pilots Association
ANOVA	Analysis of Variance
ASRS	Aviation Safety Reporting System
ATC	Air Traffic Control
ATCO	Air Traffic Control Officer
ATM	Air Traffic Management
AWO	All Weather Operations
BASI	Bureau of Air Safety Investigation
CAA	Civil Aviation Authority
CBT	Computer-Based Training
CDU	Control & Display Unit
CMIN	Chi-Square Minimum
CFA	Confirmatory Factor Analysis
CNS	Communications Navigation Surveillance
CRT	Cathode Ray Tube
DOS	Disk Operating System
DF	Degrees of Freedom
DFA	Discriminant Function Analysis
DV	Dependent Variable
EHSI	Electronic Horizontal Situation Indicator
ETA	Estimated Time of Arrival
FAA	Federal Aviation Administration
FANS	Future Air Navigation System
FMAQ	Flight Management Attitude Questionnaire
FAR	Federal Aviation Regulations
FMS	Flight Management System
GFI	Goodness-of-Fit Index
GUI	Graphical User Interface
HCI	Human-Computer Interaction
ICAO	International Civil Aviation Organization
IRS	Inertial Reference System
ISO	International Standards Organisation
IT	Information Technology

IV	Independent Variable
JAA	Joint Airworthiness Authority
JAR	Joint Airworthiness Requirements
Kb	Kilobyte
KMO	Kaiser-Mayer-Olkin Measure
LNAV	Lateral Navigation
MANOVA	Multivariate Analysis of Variance
MCP	Mode Control Panel
Mb	Megabyte
MHz	MegaHertz
MUMMS	Measuring the Usability of Multimedia Systems
MUSiC	Metrics for Usability Standards in Computing
PC	Personal Computer
PCA	Principal Components Analysis
PFD	Primary Flight Display
PGFI	Parsimonious Goodness-of-Fit Index
PIP	Product Improvement Packages
PUTQ	Purdue Usability Testing Questionnaire
QUIS	Questionnaire for Interaction Satisfaction
SD	Standard Deviation
SEM	Structural Equation Modelling
SID	Standard Instrument Departures
SME	Subject Matter Expert
SOP	Standard Operating Procedure
SPSS	Statistical Package for the Social Sciences
STAR	Standard Arrival Route
SUMI	Software Usability Measurement Inventory
VCR	Video Cassette Recorder
VDT	Visual Display Terminal
VHF	Very High Frequency
VNAV	Vertical Navigation
WAMMI	Web Analysis MeasureMent Inventory
WWW	World Wide Web

1 INTRODUCTION

1.1 Overview

The research presented in this thesis is concerned with the development of a series of usability scales for Flight Management Systems (FMSs) that could be used as part of future flight deck design and certification processes.

Evidence in the literature suggests that usability is a quality that does not exist in any real or absolute sense, but is a multidimensional concept concerned with how people perform a task, with a particular product and in a particular context (Brooke, 1996). The usability of a system must be viewed in terms of its context of use, and its "*appropriateness to that context*" (Brooke, 1996, p.189). A bank Automated Teller Machine (ATM) might demonstrate good usability properties by being: easy to use on first try, giving the user good feedback on the amount of money they are about to withdraw, and providing the user with an opportunity to cancel their transaction. While these properties might create an interface that demonstrates excellent usability characteristics when used in the evening or in conditions of dim light, it might be completely unusable on a bright sunny day when nothing on the screen can be read. The usability attributes of a product may be a coherent property of a system but they may only be revealed through operation in a particular context.

This research did not seek to propose that usability scales should be used as a sole method of designing for, or certificating, the usability of FMS CDUs. Rather, it was asserted that a series of reliable and valid usability scales designed specifically for a flight deck component such as the FMS, and for completion by real end users, would be a useful adjunct to the performance, mental workload and effectiveness data frequently elicited during system development.

In order to put the literature reviewed in this section into context, it is necessary to consider what a usability questionnaire for certification should represent. Since usability is a multidimensional concept, a usability questionnaire should be capable of eliciting a multidimensional profile of usability for the system under assessment. For a usability questionnaire to be appropriate for different FMSs and aircraft it would need to be able to access the generic usability characteristics associated with FMSs in a reliable and valid manner. If generic, stable usability factors could be elicited using a questionnaire, then it could be possible to generate usability scales on which comparisons between different products or different versions of the same product could be made. Reliable and valid usability scales would permit the introduction of a methodology into the certification process that would allow more extensive coverage of the end user population. This research therefore sought to elicit the generic stable usability attributes associated with FMSs, if any existed, and to develop usability scales based on those generic and stable properties.

Although, a psychometrically valid and reliable series of usability scales would allow quantitative comparisons to be made between different FMSs, it would not permit the elicitation of detail surrounding specific usability problems associated with those products. For this reason, it was determined that the usability questionnaire developed within this thesis should also permit the collection of qualitative data concerning problems associated with, and improvements sought to, FMS CDUs.

1.2 Background

Aircraft accident rates for the world-wide commercial jet fleet remained at a consistent two to three accidents per million departures over the period 1970 to 1996 (Boeing, 1997). Over the same time period flight decks have been altered greatly, with older mechanical instruments being replaced by electronic displays and computers that can control almost every aspect of a flight (Billings, 1997).

The prime purpose of these technological advances has been to increase fuel economy, reduce flight crew workload and increase safety. Advanced, highly automated aircraft can be more productive, reliable and safer than their predecessors when managed appropriately (Wiener, 1988). The introduction of this new automation onboard modern flight decks has not, however, been without problems and its application has also had unintended and unwanted effects (Scerbo, 1996). The result has been that while some pilot errors have been reduced or eliminated other new opportunities for different types of error have been introduced (Woods, Johannesen, Cook & Sarter, 1994). This is aptly illustrated in the following scenario described by Sweet (1995) p.31.

"After years of using simple word-processing software, you are switching to a windows system, and while you see some advantages, all those boxes seem confusing. Today as you are finishing a long document that you have not saved, filed, or even bothered to name, since you are unclear how to do that, you brush a key and the screen goes blank. In a panic, unaware that you have moved to an alternate window and can just as easily move back, you turn off the computer, thinking that maybe your file will return when you reboot. In your haste, you ignore a warning and then realise, even before your computer comes back on, that now your file really is lost. Darn! Nearly a day's work down the tubes.

Now put yourself in the "glass cockpit" of a modern transport aircraft, where, after years of flying planes with those old-fashioned electromechanical gauges and a simple autopilot, you confront a flight deck not only equipped with electronic displays that consolidate information from many sources, but also with a complex array of automated systems. These show your route on an easy-to read map, keep you on your flight plan, warn you of unseen dangers, get you into position to land and even – in a crisis - actually land the plane for you. This time as you are taking the plane down, you do not realise that because of a certain "mode" or computer function you have selected, the autopilot is simultaneously trying to abort the landing. The plane suddenly noses sharply up and stalls just a few hundred feet off the ground. Darn! Plane crashes with dozens or hundred of fatalities, including you!"

The reality of such a situation occurred on April 26, 1994, when an Airbus A300-600 operated by China Airlines crashed at Nagoya, Japan killing 264 passengers and flight crewmembers. Investigations following the accident indicated that conflicting actions taken by the flight crew and the aircraft's autopilot contributed to the crash. This crash provided a clear example of how the interface between flight crew and automation could adversely affect flight safety. Other incidents and accidents before and since have indicated that a breakdown in the flight crew/automation interface exists irrespective of aircraft type, manufacturer, operator and geographical region (FAA, 1996).

Although it is estimated that between 65% and 80% of air transport accidents are attributable in whole or part to human error, in general, little is known about the problems associated with pilot-automation interactions (Billings, 1997). The rapid growth of air travel predicted for the next two decades has made it necessary to develop a still safer, more efficient and productive aviation system than exists now and improvements in the pilot-automation interface clearly may have a role in achieving this (Graeber & Mumaw, 1999).

As a result of a number of incidents and accidents that pointed towards difficulties in flight crew interactions with flight deck automation, the Federal Aviation Administration's (FAA) Transport Airplane Directorate commissioned a study to evaluate the flight crew/flight deck automation interfaces across a range of current aircraft. The findings of the research published in the document, *The Interfaces Between Flightcrews and Modern Flight Deck Systems* (FAA, 1996), highlighted a number of vulnerabilities in modern flight decks including flight crew management of automation and situation awareness. In examining why these vulnerabilities existed the research identified a number of interrelated deficiencies in the current aviation system including insufficient criteria, methods, and tools for design, training and evaluation to aid the goal of achieving more human-centred aviation automation. The FAA Human Factors Team noted that *"Existing methods, data, and tools are inadequate to evaluate and resolve many of the important human performance issues"* (p.3), and subsequently made the recommendation that *"new tools and methods need to be developed and existing ones improved to accompany the process improvements"* (p.4).

Both the European Joint Airworthiness Authorities (JAA) and the US Federal Aviation Administration (FAA) have proposed the introduction of Human Factors Certification Criteria for civil aircraft flight decks. While the regulatory authorities have supported the philosophy behind the human factors certification process, no tools or methods have been proposed as techniques for evaluating the flight deck from a human factors perspective.

On the 8th of January 1989 a British Midland Airways B-737-400 that had left Heathrow airport bound for Belfast crashed at Kegworth, Leicestershire. The accident was later found to have been caused by the operating crew shutting down the number two engine after a fan blade had fractured in the number one engine. The aircraft remained in the air for nineteen minutes and forty seconds after the initial engine problem was diagnosed without the pilots noticing that the wrong engine had been shut down. One of the factors which prevented the pilots noticing the incorrect response they had made, was the difficulties they encountered trying to reprogram the aircraft's core

computer the Flight Management System (FMS). Valuable time was lost due to the difficulties interacting with the FMS and the aircraft struck a field adjacent to the eastern embankment of the M1 motorway. Forty-seven passengers lost their lives, and of the other 79 occupants, 74 suffered serious injury (AAIB, 1990).

It is believed that a similar failure to interact successfully with automation resulted in an Air Inter Airbus A320 crashing into mountainous terrain at Strasbourg. On this occasion the aircrew intended to make an automatic approach using a flight path angle of -3.3° from the final approach fix; this manoeuvre would have placed them at the correct point for a visual descent when they reached the minimum descent altitude. The pilots, however, appear to have executed the approach in heading/vertical speed mode instead of track/flight path angle mode (Pedrali, Cojazzi & Cacciabue, 1995). The result of this mode error was that the aircraft made a vertical descent at a rate of 3,300 feet per minute (Flight Path Angle of approximately -7°), until they crashed several miles short of the runway. Changing between the heading/vertical speed and track/flight path angle modes was by means of a push button on the Flight Control Unit, but inadequate feedback resulted in the pilots being unaware of their mode error. The manufacturers of this aircraft modified later series of the aircraft such that four digits were displayed on the speed/flightpath angle display when the aircraft was in vertical speed mode; while only displaying two digits when the aircraft was in flight path angle modes (Sparaco, 1994).

Flight Management Systems are one of the core systems of flight deck automation and can control flight from take-off to landing (Williams, 1995). FMSs can compute the most efficient flight profile and perform all the complex navigation calculations that flight crew would otherwise have to do. Functions that can be performed by a Flight Management System include control of the aircraft's flight path and speed, fuel-burn data, determining waypoints, course intercepts, holding patterns, optimum holding speeds, altitude crossing restrictions and estimated time of arrival. Research (for example, Sarter & Woods, 1992) has demonstrated that even with considerable experience pilots are still surprised by the behaviour of this piece of automation on board their aircraft and do not understand all the FMS modes and features. In fact it has been stated in the literature that pilots typically utilise only 20% of the Flight Management System's capabilities (Sherman, Helmreich & Merritt, 1997).

This research is focused on one aspect of Human-Computer Interaction (HCI), usability. The aim of the research to be described in this thesis was to develop a reliable and valid measure for assessing the usability of the control and display unit (CDU) of this core piece of flight deck automation, the FMS.

The literature reviewed in the remainder of this chapter presents a background to the research by introducing the areas of usability and flight deck automation. The importance of usability is discussed from both a human-centred and cost perspective. Current methods of assessing usability are presented followed by a discussion of the issues that make them unsuitable for use within the flight deck environment. Modern FMSs are described and their role on the modern flight deck is reviewed, with emphasis being placed on the context in which the FMS is used. The environment in which FMSs

are interacted with is particularly pertinent since the flight deck is a dynamic environment both internally and externally, a key point which makes the systems to be evaluated within this study distinctly different from desktop computers. An overview of the process of, and human factors issues associated with, flight deck certification is presented followed by a discussion of what usability for certification represents and a justification for the approach adopted in this research.

While a number of studies have examined attitudes to automation (e.g. Helmreich, Merritt, Sherman, Gregorich & Wiener, 1993), no known studies have either systematically evaluated the usability of FMS CDUs, or developed scales for the assessment of FMS CDU usability. The areas addressed in the following pages demonstrate the requirement for human factors tools to assess elements of flight deck automation and particularly the usability of flight deck automation, a requirement which this study will go some way towards fulfilling.

1.3 Aviation Automation

Automation is concerned with the allocation of functions to machines rather than humans (Kantowitz & Campbell, 1996). Aviation automation, therefore, *"consists of machines on the commercial transport aircraft flight deck which perform functions otherwise performed by pilots"* (Funk, Lyall & Riley, 1996, p.2). At the outset, aviation automation was designed to enhance and supplement human capabilities, to assist pilots perform tasks that were difficult or not possible to perform manually. Later the purpose of aviation automation was to reduce the operator workload associated with controlling the aircraft. Tasks such as managing the flight, navigating the route, and communicating with Air Traffic Control (ATC) were left to the flight crew. It is only since the 1970's digital revolution that tasks other than flight control have been greatly effected by automation (Wiener, 1988). One of the associated outcomes of this revolution has been to distance the pilot from the control of the modern flight deck (Endsley, 1996).

Figure 1-1 from Billings (1997) p.36 shows very effectively how aircrew have been moved to the periphery of control on the aircraft they operate.

The aircraft and flight deck design process is an evolutionary one that McClumpha & Rudisill (2000) claimed was worrying as it appeared to have evolved without adherence to any rigorous human factors science base. Norman (1991) similarly argued that cockpit procedures, flight instruments and regulation *are "guided more by instincts, anecdotes, and reactions to individual incidents rather than by systematic, scientific analysis of the issues"* (p.4). This implies a lack of compliance with basic human factors principles. Mosier & Skitka pointed out in 1996 that although the desirable effects of automation such as increased fuel efficiency or ability to operate in poor weather had been well documented, the negative or erosive effects had only begun to be defined.

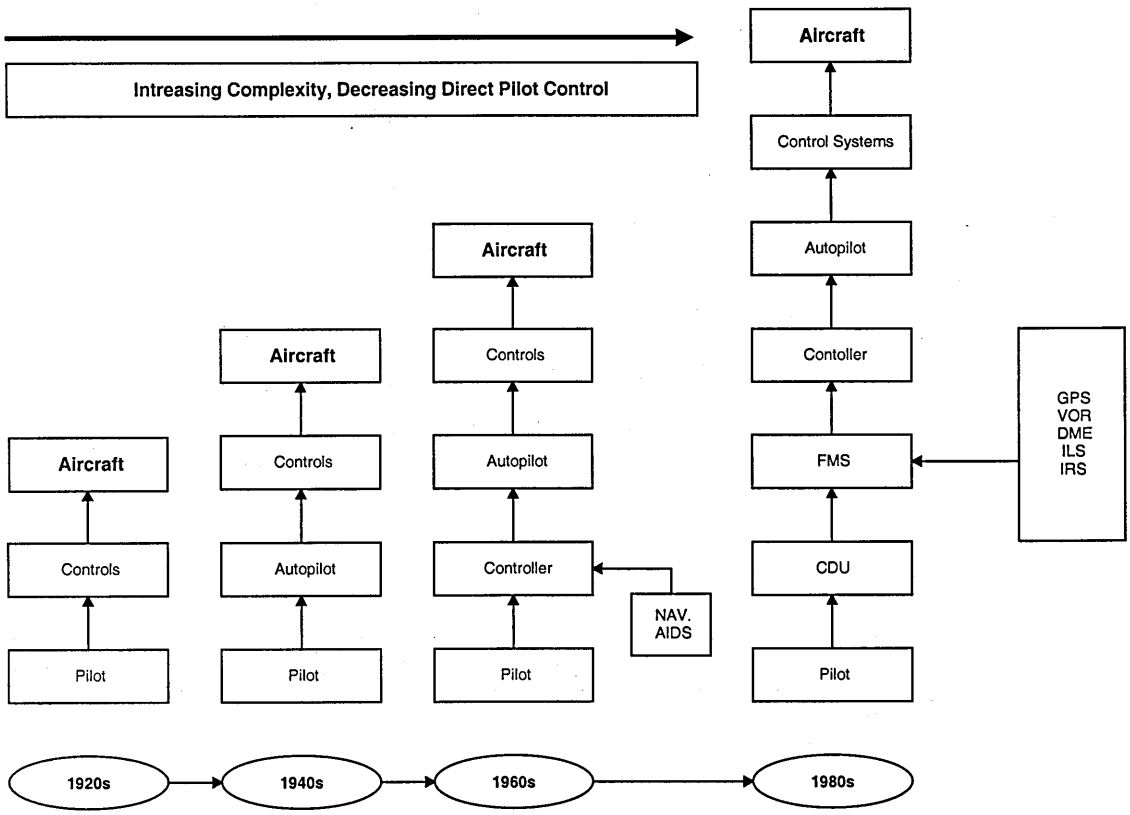


Figure 1-1 - The effects of increasing automation on the pilot-aircraft relationship (Billings, 1997, p.36)

Since highly automated aircraft can be more productive, reliable and can increase safety, flight deck automation has largely been well received by the aviation industry as a whole. Nevertheless, pilots, scientists and aviation safety experts have repeatedly made statements of concern about the impact of flight deck automation (e.g. Wiener, 1989; Billings, 1991). Among the prime concerns which have been voiced are that pilots may place too much reliance on the automation, that they may lose their manual flying skills and that flight deck interfaces may be poorly designed (Endsley & Kiris, 1995). Funk, Lyall and Riley (1996) reported that until their research, "*no comprehensive list of perceived flight deck human factors problems and concerns existed*" (p.2). After reviewing 960 documents including accident and incident reports the authors proposed that human factors concerns were related to the use of automation being justified by economy not safety; automation being poorly designed and the use of automation leading to flight deck problems. While this taxonomy did put some structure on the human factors issues associated with flight deck automation, it was very high level and related to the flight deck as a whole rather than any of its specific components.

1.3.1 Problems Associated with Aviation Automation

Very complex automation is frequently perceived as having a mind of its own, especially when something unexpected happens. Sarter & Woods (1992) referred to these unexpected happenings as "surprises" and highlighted how such occurrences could erode human trust in automation. As frequently happens with new automation, aviation automation has shaped the behaviour of pilots and not always in ways flight deck designers had intended. This has been partly due to the introduction of flight deck automation without "*an explicit guiding philosophy about its usability*" (Tenney, Rogers & Pew, 1998, p.103).

Woods (1996) described four major categories of problems that have been associated with automation-related aircraft incidents and accidents - complexity, coupling, autonomy and inadequate feedback. Woods argued that the complexity associated with flight deck automation made it difficult for pilots to always understand the ways in which the automation functioned and referred to it as the "*Apparent Simplicity, Real Complexity*" of aviation automation (p.3). The complexity associated with aviation automation is known to be particularly problematic when the use of rarely used functions is required.

Coupling is concerned with the relationship between automation functions. These interdependencies are frequently not obvious and the associated intricacies are not always detailed in flight deck manuals. Woods (1996) argued that while coupling could increase efficiency by reducing transfer costs, it also meant that problems had greater and more complex effects, effects that could propagate quickly. The following quote from the Batelle Aviation Safety Reporting System (ASRS) database illustrates the coupling effect - "*I failed to realise that the altitude restrictions are not in effect during a speed mode descent*".

Autonomy refers to incidences where automation performs functions that have not been requested by the operator, incidences where the automation appears to have a mind of its own. In this case confusion frequently occurs, as the pilot rapidly has to decide whether the function performed by the automation is sensible. An important characteristic of automation is whether the device involved only responds to user inputs or whether it can partially act autonomously. The FMS, the core system of flight deck automation that is the focus of this study, is a multiple-dynamic system which alters in response to operator inputs, situation factors and system factors.

For operator inputs, the pilots can select target values and activate corresponding modes and the FMS will respond to achieve the operator specified goals. If the pilot puts the aircraft into climb mode and requests a target altitude of flight level 310 (or 31,000 feet) the aircraft will respond to situation factors by automatically changing to altitude hold mode upon reaching the target altitude. System factors are built-in protections that help prevent the aircraft being operated in an unsafe configuration (Chidester, 1999). If the pilot puts the aircraft into an unsafe mode, the aircraft will automatically change the configuration into what it deems to be a safe configuration (Sarter & Woods, 1992).

Finally, inadequate feedback is caused by automation failing to communicate effectively with the operator. The problems related to inadequate feedback have been frequently cited in the aviation automation literature (e.g. Eldredge, Mangold & Dodd, 1992), and are associated with the questions *"what is it (the automation) doing now and why?" and "what will it do next?"* (Sarter & Woods, 1992, p.316/317).

Funk, Lyall & Riley (1996) in their taxonomy of flight deck automation problems and concerns found the pilot/automation interface to be one class of concern. The concerns which were particularly related to the FMS CDU, were:

- The control of automation being poorly designed.
- Flight deck interfaces being poorly designed.
- Data access being difficult.
- Data entry errors on the keyboard occurring.
- Displays being poorly designed.
- Data entry formats being inflexible.
- Data input prompts being poor.
- Inadvertent autopilot disengagement being too easy.
- Auditory displays being poorly designed.

The concerns raised by Funk et al (1996) and Sarter & Woods (1992) reflect both the "Gulf of Execution" and the "Gulf of Evaluation" described by Norman (1986). The "Gulf of Execution" described by Norman referred to the gap between a user knowing what they want to have happen and knowing what is required to bring that about. The term "Gulf of Evaluation", Norman argued, was the gap that existed between a user knowing their intention and knowing the effects of their actions. These terms are best illustrated through example. If a pilot cannot work out how to program a specific function into the FMS CDU, there is a Gulf of Execution. If the FMS has no method of informing the pilot that a programming error has been made, a Gulf of Evaluation exists. The gulf of execution therefore involves forming the intention, specifying the action sequences and execution, while the gulf of evaluation is concerned with perceiving the state of the world, interpretation and evaluating the outcome.

Norman argued that system designers could help to bridge the gulfs of evaluation and execution through presenting users with good conceptual models, visibility, feedback and mappings. Sarter & Woods (1992) demonstrated through experiment that FMSs failed to assist users in creating good conceptual models and argued that pilots' mental models were incomplete or inaccurate. Norman argued that systems could be made more usable by assisting users to bridge both the Gulf of Execution and the Gulf of Evaluation through:

1. Reducing the Gulf of Execution by specifying correction actions.

2. Reducing the Gulf of Evaluation by interpreting and evaluating the state of system and outcomes.

3. Providing good visibility of both alternatives for actions (Gulf of Execution) and the current system state (Gulf of Evaluation).

4. Providing good informative feedback about the results of actions.

5. Reducing the Gulf of Execution through helping users to learn.

6. Providing good mappings and logical relationships such that it is possible to determine the relationships between inputs and their effects.

Norman argued that adhering to these principles could also assist users by lowering their cognitive demands and therefore reducing the likelihood of error.

The FAA in 1996 issued a series of recommendations that they argued, should be adhered to, and would improve safety. The principles they identified which were most pertinent to FMS CDUs included:

- Minimise human error.
- Increase error tolerance.
- Avoid excess complexity as perceived by the user.
- Increase system observability.
- Evaluate new technology or operational changes for their effect on human performance.

It should be noted at this point that the principles outlined here reflect the types of factors which influence human-machine interaction as will be discussed in the subsequent section. A more in-depth discussion of concerns directly related to the FMS CDU pilot / automation interface is presented in Section 1.9.

1.3.2 Human-Centered Aviation Automation

In the flight deck operating environment, humans have ultimate control and should therefore be the primary focus of research attention. The concept of situation awareness has evolved from incidents and accidents in which pilots became preoccupied with automation detail at the expense of maintaining a bigger picture of the flight deck status. The automation present in today's aircraft is capable of limiting the authority of the pilot. The Airbus A-320, for example, is fitted with envelope protection that limits human authority by not permitting the pilot to fly the aircraft in a manner that would damage the aircraft structurally. Douglas (1990) outlined a design philosophy which argued *"if it is technically and economically feasible to automate a function, automate it"* (p.3). The issues arising from studies of situation awareness are not concerned with what functionality can be automated, but rather what should be automated. Pilots in the newer generation of aircraft have become managers of automation. The issues that have arisen from situation awareness research have focused on whether pilots have a sufficiently active role in flying the aircraft, or whether they are largely *"out of the loop"*, as monitors of anomalies (Endsley & Kiris, 1995, p.381). The problem arises when one is suddenly required to change from being a passive system monitor, to an active problem solver. This can be a difficult transition to make.

Kelly, Graeber & Fadden (1992) listed the pilot's role, limitations and needs based on the design philosophy for the most modern Boeing aircraft, the B-777. The pilot's roles and responsibilities stated both that aircrew flying the aircraft had ultimate authority for the safe operation of that aircraft, and that all decision making made on the flight deck was on a goal hierarchy. Pilot limitations meant that all aircrew were expected to recognise human performance limitations. Further, since pilots have different performance capabilities, these differences needed to be accommodated into the flight deck design. Flight deck design, Kelly et al stated, should create neither work overload nor underload; should incorporate error tolerance and avoidance techniques to improve safety; and should allow for crew operations based on training, past practices and intuitive operations. Finally, the pilot's needs were that within a comfortable working environment the automation should aid them, and the goal hierarchy used to perform their role.

1.3.3 Aircraft Design Philosophies

Human Factors has over the past fifteen years become a well established discipline in the aerospace industry, and most organisations in the aviation industry publicly support concepts such as human-centred design. Courteney (1997), however, reported that in practice human factors still seemed to be a *"peripheral concern"* to most aircraft manufacturers (p.111). In order to increase the human factors effort applied to flight deck design, manufacturers would be required to bear the cost themselves, while in the short-term at least, aircraft operators, trainers, insurers and maintainers would reap the benefits. In the military context, the same organisation normally pays for all aspects of the aviation system including design, development, training and ATC, and thus reaps the benefits of having good human factors processes. Courteney (1997) argued that this

explained why in the military world, the application of human factors had progressed at a rate largely not seen in the civil aviation world.

Two of the major manufacturers of large civil transport aircraft are The Boeing Company and Airbus Industrie. Their design philosophies differ significantly. All Boeing civil transport aircraft have, for example, chosen to retain conventional large controls columns even in their first commercial fly by wire aircraft the B-777. Conversely, Airbus Industrie have since 1988 with the advent of the A320 chosen to use a side-stick. The most modern civil aircraft are the Boeing 777 and the Airbus A330. While the Airbus A330 is as close in design concept as is feasible to the older A320 and A340, the latest Boeing aircraft is not a derivative of any previous aircraft although it has commonalties with the older B-747-400.

Aircraft manufacturers do not work completely independently, and their products are directed if not dictated by both certification and consumer requirements (Courteney, 1999). Many airlines resist flight deck changes due to the complexities that changes enforce on the way airlines operate. For an airline to introduce a radically different aircraft into their existing fleet, the pilots require significant re-training both in the simulator and on the line. While these pilots are re-trained, the airlines will incur not only these re-training costs, but also the costs of paying other pilots to fly the routes that the pilots being re-trained would otherwise have flown. Further, since pilots may only be type-rated on one aircraft at a time, any pilots type-rated on the new aircraft are not legally permitted to fly any other type of aircraft. If a pilot from one fleet falls ill, it is not possible to use a pilot from another fleet in their place. Rostering difficulties are therefore created for the airline, since some pilots will be permitted to fly one aircraft type and not the other. In an increasingly competitive market such as the airline industry, radical changes are not well received. Hence, from the manufacturer's perspective there is little incentive to alter flight decks or to accommodate many of the usability principles that will be outlined in Section 1.8. If radical change means airlines resist buying the product it does not become cost effective to manufacture that product.

The other constraint under which aircraft manufacturers operate is from the regulatory authorities. Every new component, fitted to every new or existing aircraft must be certificated as being fit to fly (Singer, 1999). Regulatory authorities such as the Civil Aviation Authority (CAA) and Federal Aviation Administration (FAA) oversee the certification procedure. Making changes to an aircraft is a further risk for the manufacturers since new components, or indeed a new aircraft, may be designed or manufactured but for some unforeseen reason may be deemed by the certification authorities as being unfit for entry into service. For reasons outlined in this section, an "if it isn't broke, don't fix it" philosophy prevails across the civil aviation industry.

1.3.4 The Modern Flight Deck

The term "modern" flight deck is now used to refer to aircraft that are predominantly of the "glass cockpit" variety in which the pilot's displays are made up of a series of Cathode Ray Tubes (CRTs) (James, McClumpha, Green, Wilson & Belyavin, 1991).

Figure 1-2 shows the glass cockpit of the Boeing 747-400. Today's flight decks are constituted of many systems and subsystems that are supplied by any number of avionics manufacturers. In assessing the usability of any one of these components it is not possible to isolate that component from the flight deck in which it resides. The aircraft in which it resides will therefore impact upon any usability assessment of a flight deck component.



Figure 1-2 - Boeing 747-400 Flight Deck
(Boeing Commercial Airplane Group, 2002)

Some of today's civil transport aircraft which have a glass cockpit include the Boeing B717, 737-300-900 series, B-757, B-767, B-747-400, B-777, Airbus Industrie A300-600, A310, A320, A321, A330, A340, the Fokker F-100 and the McDonald-Douglas (Boeing) MD-11.

The generic glass cockpit in Figure 1-3 shows the location of the FMS CDUs within the flight deck (Billings, 1997, p.89).

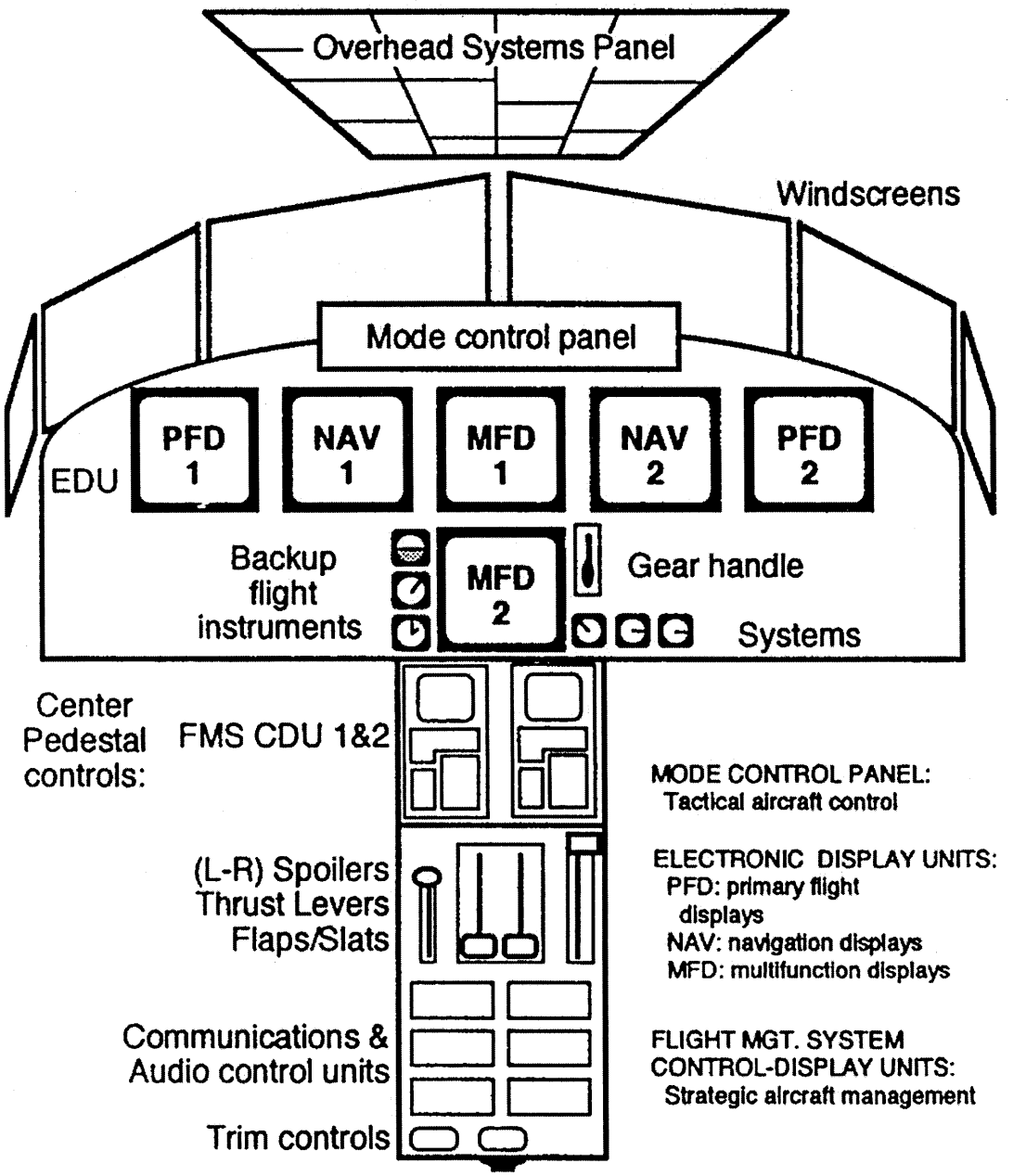


Figure 1-3 - Generic Glass Cockpit
(Billings, 1997, p.89)

1.4 Flight Management Systems

1.4.1 Introduction

The most influential change in the modern flight deck has been the introduction of the Flight Management System. The first truly automatic Flight Management System with external position sensors was introduced to the aviation industry in the early 1970s by Honeywell, although FMSs as used today have been in service for well over a decade. The modern FMS was originally designed to be the main means of navigation on board the flight deck, irrespective of conditions (Billings, 1997). FMS functionality has, however, been extended significantly over time as will be discussed in Section 1.4.2.

1.4.2 Flight Management System Functions

Bulfer (1996) compared a Flight Management System as being similar to an electronic flight deck Operations Manual. This is a simplification because in fact the FMS is an integrated inertial navigation, performance and fuel management system controlled by two independent Control & Display Units (CDUs) on the centre console of the flight deck (Stewart, 2000). FMSs provide pilots with a tool that performs both navigational and performance computations. The data stored in an FMS database includes accurate aircraft drag and engine model data, maximum altitudes, and maximum and minimum speeds. Computations related to lateral navigation are based on a stored or manually entered flight plan and include items such as courses to be flown, Estimated Times of Arrival (ETA) and distances to go. For vertical navigation, the FMS can compute items such as fuel burn data, optimum speeds and altitudes based on an aircraft's gross weight, cost index, predicted winds at cruise altitudes, and specific ATC constraints (Bulfer & Gifford, 1997).

The performance database of the FMS contains the average model of the aircraft and its engines. The average aircraft model includes high-speed drag polars, a buffet limit envelope, certificated operating limits, speed and altitude of engine out parameters. The performance model calculations which can be performed by the FMS include the cost optimum speed schedules used in climb, cruise and descent, as well as fuel flow, thrust, engine limits, target values and corrections for the effects of air-conditioning and anti-icing bleeds. Such performance computations constitute an essential difference between FMSs and conventional area navigation systems.

The FMS can also provide control and guidance commands, which can be coupled to the Autopilot/Flight Director System (AFDS), and Autothrottle (A/T). This allows integrated operation with automatic lateral and vertical navigation. In this way, the FMS relieves the crew from many routine system integration tasks and allows them to concentrate on the management of the flight.

Autopilots provide a middle level of automation between hand-flying an aircraft and using the FMS to control the complete flight from take-off to landing. The autopilot is

controlled through a Mode Control Panel (MCP) on which the pilot may specify a target airspeed, heading, altitude and vertical speed. With the autopilot and autothrottles engaged, if a pilot specifies a target altitude and selects an appropriate mode such as Vertical Navigation (VNAV), the autopilot will manoeuvre the aircraft pitch and power setting to achieve the target values. The autopilot is very limited in functionality when compared to the FMS since it cannot perform computations as the FMS does.

The manner in which an FMS is used is at the discretion of the pilots in that the crew can select to use any degree of automation (Billings, 1997). The crew may, for example, choose to use the FMS Control and Display Unit (CDU) only for reference during the flight, while flying the sector manually or through the autopilot. Alternatively, the crew may choose to select full FMS operation which would permit completely automatic flight path guidance and performance control.

FMS databases can contain most of the information that pilots would traditionally have accessed by using navigation charts. The databases contained within the FMS are tailored specifically for particular airlines, and therefore can contain many of the routes that an airline will fly. The type of data stored on an FMS will include navigation points (typically 32,000), airway route structure, location of VHF navigation aids, airports, runways, geographical reference points, standard instrument departures, standard arrival routes, approaches and company routes. Additionally, the operating pilots can typically input up to forty extra waypoints into the FMS.

1.4.3 FMS CDU

User interaction with all FMSs is through a Control and Display Unit (CDU) which consists of a display and keyboard combined into a single console located in the centre of the flight deck. There are typically two or three of these FMS CDUs on board modern flight decks, so each pilot has their own workspace of the CDU (Dekker and Orasanu, 1999). Data will typically, however, only be accepted from one of the CDUs at a time and although each FMS operates independently, they occasionally compare data in a resynchronization process. The primary FMS forwards all CDU inputs to the secondary FMS to ensure that the secondary FMS is synchronized with the primary FMS within 200 milliseconds of a CDU entry. An on-going synchronization allows each FMS to detect anomalies between itself and the other FMS for parameters such as flight plan and performance data. If the difference between the FMSs cannot be resolved during the synchronization process, the primary FMS takes over and the secondary FMS is conditionally failed. A single FMS operation message is then displayed to the pilots on the CDU scratchpad (Bulfer, 1996).

In many newer aircraft, tactical flight plan changes such as speed and heading, may also be made through the mode control panel. Figure 1-4 shows a Smiths FMS CDU (Schiratti, 1999).



Figure 1-4 - Boeing 737 FMS CDU
(Adapted from Schiratti, 1999)

The display on a CDU is a five-inch CRT display, which can display 14, 24 character lines. The display can be in colour, although the majority of passenger aircraft still have monochrome (light green on dark green) CDUs. The CDU contains line select keys on either side of the display, mode select keys, a numeric keypad and an alphanumeric keypad. Mode select keys can be used to access FMS function pages, while the alphanumeric and numeric keypads allow data to be entered on a scratchpad on the bottom of the display. Some newer CDU designs such as those featured on the Boeing 777 include cursor and touch-pad interaction devices, although as Singer (1999) pointed out, these are not as yet considered standard.

Since the FMS CDU display area is so limited, it has been necessary to split the information required to perform tasks across multiple pages. The title line on a CDU page indicates whether more information pertinent to the task being performed is contained in subsequent pages by indicating the current page number next to the total number of pages for that particular task. The first of three pages would thus be represented as "1/3" (Billings, 1997). The CDU typically contains upwards of fifty pages structured in a hierarchical tree, sometimes resulting in many pages having to be navigated through in order to perform a single function (Romahn & Schafer, 1997). Foreman (1996) reported that on one aircraft it could take in excess of 43 keystrokes to change from an approach on one runway to another runway. A further complication of these data structures is that they vary greatly across aircraft type and avionics manufacturer. Billings (1997) reported that even for the most proficient of aircrew, the reprogramming of the FMS through the CDU could require significant attentional demands since changes most commonly occur during departure and approach, periods when the pilot's attention to the outside world is most essential.

Figure 1-5 from Bulfer (1996) p.1.2-1 shows the typical CDU inputs that should be undertaken to programme the FMS CDU of a Boeing 737-300 during the pre-flight process. Although there are many steps in the pre-flight process shown in Figure 1-5, even this illustration assumes that a company route is already pre-programmed for selection. Many airlines do not have all their routes pre-programmed resulting in pilots additionally having to input their route manually.

1.4.4 Deficiencies in Current CDUs

Singer (1999) outlined some of the deficiencies in current FMS CDU design. The first deficiency highlighted by Singer was that although FMS CDU designs are typical across manufacturers, unlike QWERTY keyboards, no standard exists for the positioning of CDU letter keys, and each function key has a different purpose in each design. This leaves pilots vulnerable to negative transfer when use of a different system is required, as occurs when a new type-rating is undertaken.

Since the CDU display area is only five inches and normally displays fourteen lines of text, several menu layers are required in order for the CDU to contain all the data required by modern flight decks. Each CDU function may therefore have five or more pages of information. Singer (1999) reported that since the positioning of these menu pages is not logical to the pilot, the search for information is frequently conducted by a process of trial and error, resulting in the pilot navigating through ten to fifteen menus in order to find the required information. The outcome of such a process is increased workload and frustration (Molich & Nielsen, 1990).

Further, while some CDUs require critical flight changes such as flight plan alterations to be viewed and then accepted or cancelled prior to execution, others initialise a change immediately. Although, inputs to the CDU on a flight deck are prone to turbulence, parallax error and procedural errors, many current FMSs provide inadequate feedback, and do not always incorporate a review function to allow critical changes to be reviewed. Similarly, it is not standard for a CDU to have the "undo" type function

which a typical Personal Computer (PC) user would expect to find. It is possible, for example, to delete the entire flight plan by applying "Delete" to the wrong line select key and not be able to undo it.

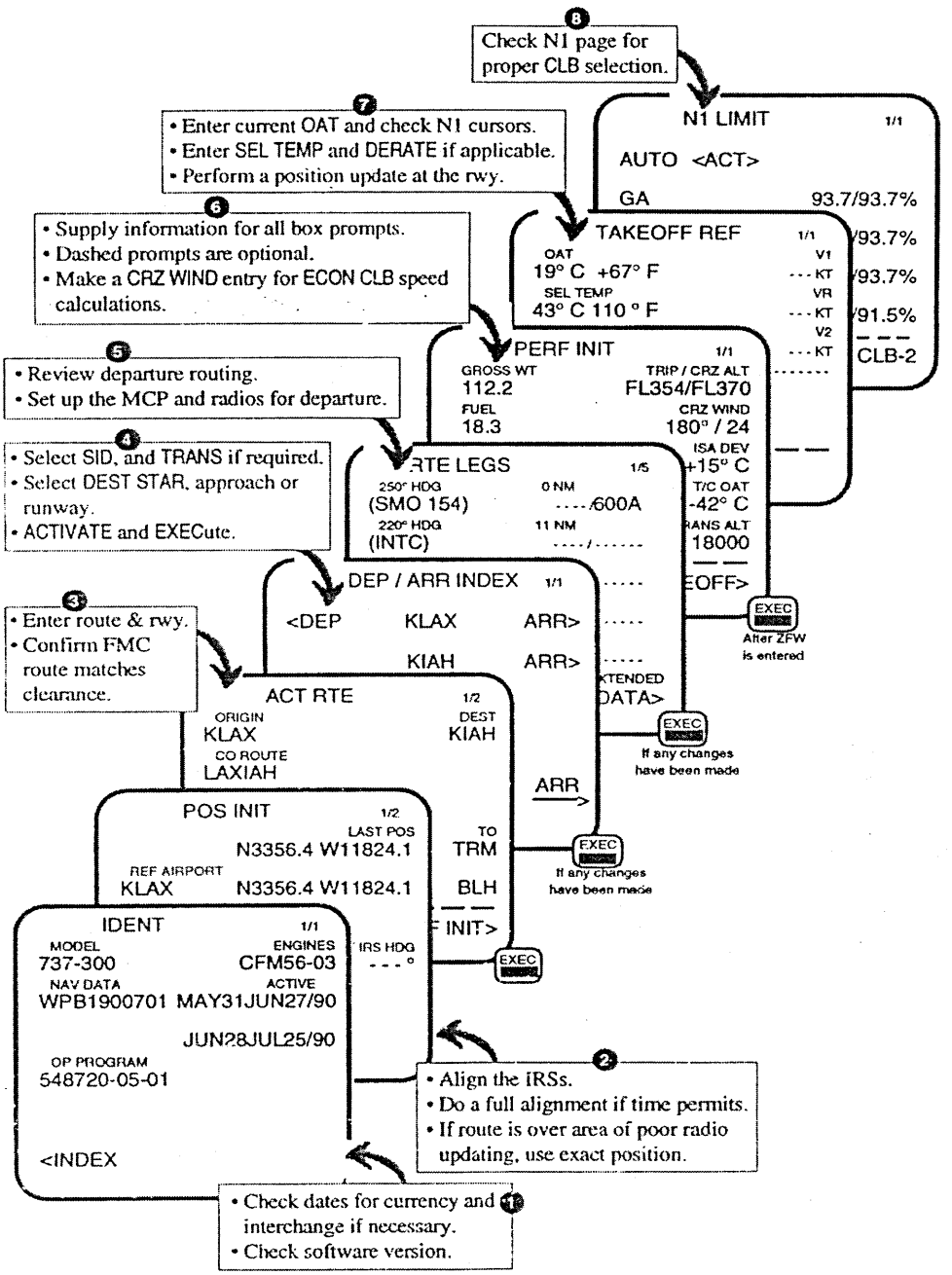


Figure 1-5 - Pre-Flight Process of making CDU inputs (Bulfer, 1996, p.1.2-1)

1.5 Human Factors for Flight Deck Certification

The purpose of certification is to guarantee that an aircraft fits the legal flight-safety requirements when flown "*by qualified standard pilots, as representative as possible of end users*" (Amalberti & Wibaux, 2000, p.361). Regulatory authorities such as the Joint Airworthiness Authorities (JAA), Federal Aviation Administration (FAA) and Civil Aviation Authority (CAA) are responsible for the airworthiness certification of all aircraft made by manufacturers (FAR Part 25), or operated by companies, based in Europe or the US (FAR Part 121 or JAA All Weather Operations (JAA-AWO) Part 1). The regulatory authorities define a number of requirements and criteria which aircraft must achieve in order to meet the minimum acceptable standard for certification.

There are those who would argue that since the safety of any system cannot be ensured, certification is costly, ineffective and should not be bothered with (Endsley, 1993). Endsley went on to argue that while these concerns were valid, the only thing worse that certificating an aircraft was not certificating an aircraft. The political and economic realities of the aviation industry dictate that little attention will be paid to human factors unless it is mandated, and unless the sales and acceptability of the product are dependent on it. While the inclusion of human factors and usability in the certification process does not guarantee a truly safe system, it does bring it closer than would just admitting defeat over the inadequacies of human factors as a science (Endsley, 1993).

While some Advisory Circulars (AC) do provide guidance on the application of human factors to FMSs, they are not mandatory and do not constitute regulation (e.g. FAA AC 25-11 (Electronic Displays); FAA AC 25-15 (Flight Management Systems)). Furthermore, even the applicable ACs do not present guidance on usability.

In the following subsection a critical analysis is made of existing FAR/JAR 25 guidelines and this is followed by a discussion of what usability for certification represents.

1.5.1 Review of FAR/JAR 25 Regulations

For large passenger aircraft the design requirements to meet correct standards, system safety levels and minimum levels of control are specified in FAA's Federal Aviation Regulations Part 25 (FAR 25, 1998) and by the JAA's Joint Airworthiness Requirements (JAR 25, 1994). While the technical requirements in these mandatory regulations are quite detailed, only the human factors requirements outlined in this section are mandatory when putting an FMS CDU through the certification process.

FAR/JAR 25.671(a) requires that controls and control systems must operate with the ease, smoothness and positiveness appropriate to its functions. While these terms were relevant to older flight decks controlled by levers and switches, they are clearly inadequate in the domain of glass CDU displays and keyboards.

The second pertinent regulation FAR/JAR 25.771(a) requires each pilot compartment and its equipment to allow the minimum flight crew to perform their duties without unreasonable concentration or fatigue. It does not state what unreasonable concentration is or how mental fatigue should be measured.

FAR/JAR 25.777(a) requires cockpit controls to be located such that they provide convenient operation and to prevent confusion and inadvertent operation. This guideline is again useful when examining older flight decks when it may have been possible to test whether a lever could inadvertently be operated or whether a switch could be reached. It would be difficult, if not impossible, to test the FMS CDU in this manner due to the infinite number of interaction possibilities (Singer, 1999). In, for example, conditions of turbulent flight simple errors are inevitable. Current regulations do not require FMS CDUs to have an undo button or to be fault tolerant by mandating two-step inputs such that inputs can be reviewed prior to their execution.

Finally, FAR/JAR 25.1301(b) requires installed equipment to be labelled such that its identification, function or operating limits can be identified. Unlike PCs for which international usability guidelines have been created (e.g. the International Standards Organisation's ISO 9241 Part 11: Guidance on Usability), there are no usability standards in existence for FMS CDU displays and controls.

While FAR/JAR guidelines were updated during the 1990's, they are in essence fifty-year-old documents and do not consider electronic flight deck instrumentation or FMSs. Singer (1999) in a review of the FAR/JAR 25 regulations concluded that *"existing requirements are general, vague and do not aid the manufacturer or authority evaluation groups in their decisions."* (p.93).

The design of an aircraft can affect human performance in terms of the likelihood of error, the ease of detection, the effect of an error, system awareness, situation awareness, cross-crew awareness, time pressure and workload (Woods & Courteney, 2001). The first step towards achieving any form of human factors certification will be the establishment of guidelines for human factors certification by multidisciplinary teams including human factors experts, designers, users, and regulatory authorities. The guidelines established will in turn refer to certification prerequisites such as goals, criteria, measures, methods, processes, standards, testers, test procedures, check and examination systems and penalties. This research attempts to feed into this process.

McClumpha & Rudisill (2000) argued that flight deck interfaces had evolved without being "*subjected to a rigorous and critical examination based on existing human factors and human-computer interaction methods, standards, or guidelines*" (p.338). They also cited the mix of avionics equipment as a source of problems for the civil airline industry, giving the example of a flight deck that had two different FMS keyboard layouts. McClumpha & Rudisill further asserted that the certification authority, in the quest for safe flight, should ensure conformance to good quality human engineering and human factors design principles as is currently required for the physical properties such as the structural integrity of an aircraft.

McClumpha & Rudisill (2000) outlined a number of organisational changes necessary for good human engineering to be adopted within the certification process. The first organisational change required was the recognition of the developments in human factors and HCI. A second proposed change was to involve certification and HCI support early in the flight deck development process before it was too late to do anything about poor interface design. The third change required was the involvement of line pilots in the development process, so that they could formally be involved in interface reviews. The final change recommended was the involvement of human factors experts in the certification or "signing off" of aircraft as being safe.

1.5.2 Critical analysis of the role of usability in the Certification process

While military customers specify the requirement for Human-Computer Interaction (HCI) prototyping and assessment by users to be conducted during the development lifecycle, there is no such requirement for civil transport aircraft. Usability in the certification process should be concerned with ensuring that equipment is designed around its intended end user, the tasks that end user will perform and the context in which the equipment will be used. Current civil flight deck certification is very equipment focused, and as outlined in the previous section, the certification authorities mandate no specific user-centred requirements. Although this is due to change to some extent under the JAA's interim policy document "Human Factors Aspects of Flight Deck Design", even the new policy does not consider usability (JAA, 2001). Any tools intended for use within the current certification requirements need to reflect elements of the equipment to be certificated rather than characteristics of the end user. It should be noted that this research is not promoting such an equipment-focused view as "best practice". On the contrary, whilst a user-centred approach would be much preferable, the current certification process does not accommodate such an approach and therefore for human factors tools to have any chance of being incorporated into the current certification process they need to be equipment oriented.

Stager (2000) argued that there are two significant challenges to the human factors certification of advanced systems. The first was the challenge associated with identifying and defining objective operational performance criteria, whilst the second challenge was that associated with articulating the most cost effective and appropriate measurement strategy for validation. Commercial flight decks are currently evaluated based on contractual requirements, time and scope constraints with a trend towards minimising risks to certification (Singer, 2001). Unlike military projects, commercial flight programmes typically allow only two years for design and one year for development and certification flights. The time constraints mean that "*very little experimentation is done for validation of new design features and approvals are made based on 'good engineering practices*'" (Singer, 2001, p.303). The commercial reality of civil aviation dictates that current human factors acceptability in the certification process is almost uniquely dependent on test pilot judgement. Their judgement is based on regulatory and company tests pilots' comparative experience following from actual or simulated flight exposure and their previous experience with existing designs (Paries,

2000). Other approaches to certification could include prototyping or the analysis of a body of evidence elicited during simulator or flight trials. While this approach to certification is preferable in many respects since it allows the inclusion of specific real world scenarios, it is extremely cost intensive and requires the product undergoing evaluation to be at an advanced stage in its development. Similarly, in terms of assessing usability during the FMS design and certification processes it would be preferable to do rigorous usability testing with large samples of line pilots who were as representative as possible of the ultimate end user. Cost factors also limit the involvement of real end users. The costs incurred by using operational line pilots means that even in major international aircraft development programmes, the number of end users involved is minimal. A further implication is that in most cases the modifications to system design required by certification are small. Amalberti & Wibaux (2000) reported that this was related to cost and noted that when problems were observed they tended to be resolved by putting effort into pilot training and regulation rather than design modifications.

1.6 Human-Computer Interaction and the FMS

Preece, Rogers, Sharp, Benyon, Holland & Carey (1994) defined Human-Computer Interaction as being about *"designing computer systems that support people so that they can carry out their activities productively and safely"* (p.1). Most of the early research concerned with how people interacted with computers was conducted within the domain of HCI and was focused on the design of desktop computers and the associated software. It is only in very recent years that this research has begun to extend to other environments, as computer interfaces have come away from the desktop and into the car for example.

Early in the 1970s researchers noted that many of the systems that seemed excellent from a functional perspective performed very poorly in the real world, a fact that still holds true today. Eason, Damodaran & Stewart (1975) highlighted the link between the poor performance of computers in the real world and the poor design of the interface with which users worked. Twenty-six years later computer users still suffer the effects of products designed by people who do not understand the user, the user's needs or the user's working environment. In an office environment the ill-effects of poorly designed interfaces can include stress, lower work rates, decreased job satisfaction and even higher absenteeism (Bevan & Macleod, 1994). In the flight deck environment the costs can be even greater.

Flight deck computers differ from the office based computers in the rate at which they have evolved. The first Flight Management Systems were introduced onto flight decks in the early eighties and have changed very little since (Foreman, 1996). When the Boeing 767 first came into service in 1982, for example, its FMS had a memory size of 100k. Now three memory sizes exist, 200k, 700k and 1Mb (10 times increase in size) and a sixteen-bit processor is common. Similarly FMSs have changed little in terms of their interface which remains a DOS-type interface.

Making a direct comparison with desktop computers over that same time period, memory capacity has changed dramatically from 640kb to 512Mb (820-fold increase in size), while typical processors have changed from 8-bit 4MHz to 32-bit 1500MHz (a 375-fold increase in speed). The desktop interface has seen a corresponding leap in appearance moving from the near obsolete DOS type interface to the Windows interface used world-wide today.

There are a number of reasons why FMSs have not evolved at a rate any way comparable to desktop computers. Airlines are slow to accept changes due to the huge costs involved in developing and running new training programmes, and of getting these training programmes approved by the regulatory authorities. Changes to flight deck equipment may create an additional problem for airlines in that if new equipment is added to one aircraft in a fleet, it has to be added to all the aircraft in that fleet in order to prevent commonality problems for pilots.

The severe consequences, which can ensue from any computer failure on the flight deck, have made manufacturers slow to make any unnecessary changes. Any new computer going onto a flight deck has to be extremely reliable and secure. Furthermore, because aircraft undergo rigorous testing by certification bodies, the development of any new equipment could serve to slow the process of developing a new aircraft or of having it certificated. Billings (1997) cites a personal communication from D.M. Fadden in discussing the evolution of Flight Management Systems stating that: "*Initial FMS designs were based on the notion of FMS use at high altitude in cruise flight. The success of the concept resulted in extension of the cruise concept into use throughout the flight, without redesign of the interfaces, a common problem with successful automation*" (p.116).

1.7 Usability

In this section some background is provided to the concept of usability. Issues surrounding the definition of the concept are discussed, and factors determining usability, as outlined in literature from the past two decades, are outlined.

The concept of usability in human factors is concerned with how people work with a product. Increasingly, the importance of usability is being recognised across a range of industries. Manufacturers are now beginning to acknowledge that product users no longer consider usability to be a bonus but rather a requirement (Jordan, 1998). It is important at this point to distinguish usability from the concepts of functionality and utility. Usability is concerned with how people perform a task, with a particular product and in a particular context. Testing usability is about making sure people can find and work with the product's functions to meet their needs. Functionality refers to what a product can do. Building functionality into a product does not mean that people will be able to use it (Shneiderman, 1992). Testing functionality is concerned with ensuring that the product operates according to its specification. Utility is concerned with the usefulness of a product (Nielsen, 1993).

While a number of techniques have been developed over the past fifteen years for assessing usability, none have been developed for, or applied specifically to, the flight deck or its components. For a number of reasons, including, the complex context in which flight crews operate, the usability of flight deck elements is distinct from other professional applications such as desktop computers and their related software, or home appliances such as VCRs. It is important, for example, that the automation on board an aircraft is dynamic since changes frequently need to be made very quickly (Foreman, 1996), something which is not so important when operating the aforementioned desktop computers or consumer products.

1.7.1 An Overview of Usability

Usability is concerned with "*the match between a product and the tasks for which it is intended, the users for which it is intended, and its context of use*" (Jordan, Thomas & McClelland, 1996, p.238). Designing a usable system requires knowledge of who the system's end user will be, what it is that the system will be used for, the work context and environment in which it will be used, and what is technically and logistically feasible. The fundamental tenet of usability is that a product should be easy to use. Dumas & Redish (1993) asserted that usability, like functionality, is an attribute of every product.

Evidence in the literature suggests that numerous qualities may contribute to a system's usability and the attributes that constitute usability may not be equally important for different systems. Stanton & Baber (1996) concluded that it was possible to argue the relative merits of different usability components, but suggested that "*this rarely becomes more than an exercise in semantics*" (p. 41). Considering a tourist information kiosk that users only interact with on a one-off basis, its most important usability attribute might be its learnability. Such interfaces need to have a zero learning time so that people can interact with them successfully on their first try. Regardless of how consistent the interface is in terms of its structure, if it is too complex to learn users are unlikely to persevere long enough to elicit the information they require. For a system that is used everyday, such as a software package, it might not be its learnability that is most critical but rather its flexibility to accommodate novice and expert users (Nielsen, 1993).

Most of the early work on usability was concerned with the design of computers and software within the context of Human-Computer Interaction (HCI). Usability, however, has gained attention within academia for the past twenty years and is now receiving attention across a range of industries (Rubin, 1994). Stanton & Young (1999) argued that the concept of usability had become "*common parlance in design*", and terms such as "*user friendly*" and "*ergonomically designed*" had become an "*ad-man's by-line*" (p.411). Not only has usability become a crucial issue within product marketing and sales, with more and more products being advertised as being "user-friendly", it is also increasingly being seen as a component of industry's responsibility to its customers. Usability has become a critical component in the design of technical equipment and designers are being forced to consider the needs and limitations of those who will be using their products. This is largely due to the growing user population who choose to

use an increasing number of sophisticated products, from complex Information Technology (IT) systems to everyday household items. Booth (1989) drew attention to the fact that manufacturers had continually focused on increasing the number of functions that their products could perform, while users became increasingly frustrated, annoyed and confused by not being able to operate products they had purchased.

Jordan (1998a) pointed out that users are becoming increasingly sophisticated in their product expectations and are no longer willing to pay the price of struggling with a product in return for technological wizardry. Many people are, for example, familiar with the frustration associated with programming a VCR or dealing with a central heating system that turns on and off when least expected. However, usability can also be associated with more safety critical appliances, as in the case of an emergency product such as a fire extinguisher. The design of such a safety item must match user expectations so that a member of the public can use it on their first time of trying.

The user interface refers to the space through which interactions occur between a system and its user. The user interface consists of a display via which information is presented and an interaction device through which information can be entered, manipulated and controlled. The user interface plays a large part in determining the effectiveness by which a system works and it holds a vital role in determining how effective the human interaction with a system is. Poorly designed interfaces can seriously hamper a user's ability to use a system and can lead to "*confusion and frustration, difficulty in learning how to use the system, misunderstanding of what the system is doing and of what the user should do, errors, and difficulty in using the system to carry out tasks successfully*" (Ravden & Johnson, 1989, p.1). The essential point is that every user interface should meet the requirements of those who will carry out tasks with that system.

A review of the literature suggests that a gap may exist between what academia is providing and what industry finds helpful. Jordan, McClelland & Thomas (1996) suggested that a lack of understanding of industrial practices and pressures may be a contributory factor to this. The fundamental notion of usability is that a product should be easy to use. The task is to ensure that problems of usability do not arise. Upon reviewing the previous fifteen years of usability literature, Booth in 1989 asserted that the challenge could best be expressed in the statement: *"Today we are just as capable of producing an unusable product as we have always been"* (p.104). Any IT user familiar with many of today's software packages will know that the validity of this statement still holds true.

1.7.2 Defining Usability

While a vast amount has been written on the importance of the concept of usability, debates concerning the meaning of the term are on-going. Stanton & Baber (1996) asserted that the main problem with the term "usability" is that it means different things to different people. The difficulty of defining the term usability was outlined as far back as 1981 when Shackel argued that *"Everyone knows what usability means until its recognition as a criterion implies evaluation, which requires measurement and operational definition"* (Shackel, 1981, cited in Booth, 1989, p.110). Bevan (1995) concurred with this opinion asserting that "*considerable confusion*" (p.115) existed over the meaning of the term "Usability".

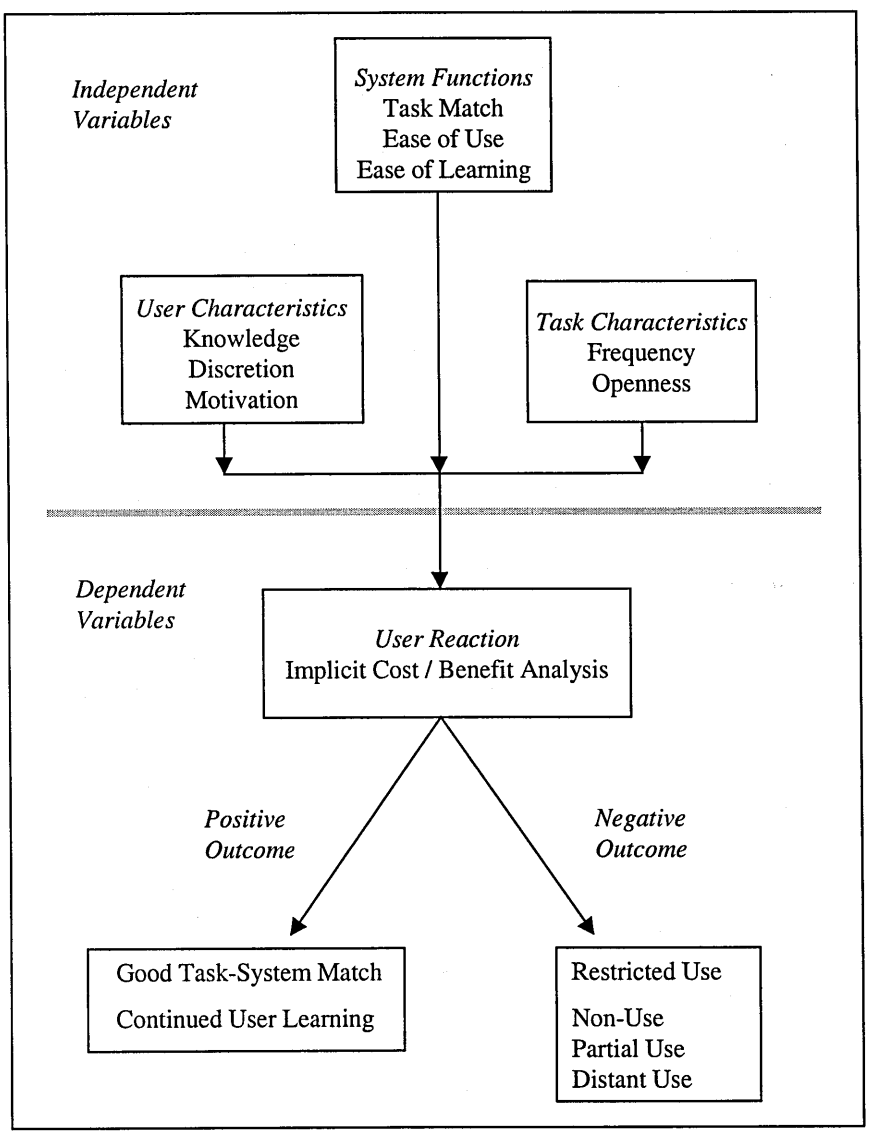


Figure 1-6 - Causal Framework of Usability (Eason, 1984, p.138)

Eason in 1984 highlighted the problem that the more multifunctional systems were becoming, the more difficulty users were having in using them successfully. Attention was also drawn to the danger of reaching the point of diminishing returns such that every new facility included in a multifunctional system could make it less usable. Through field study work, Eason demonstrated the existence of a "*partial use*" effect in which the full capabilities of flexible systems were not used (p.138). This meant that although designers were creating systems with greater functionality potential, much of this potential was unused. The concept of usability, as hypothesised by Eason, was constructed of three factors, system, task and user, which interacted with each other. Figure 1-6 shows Eason's (1984) causal framework of usability (p.138). According to this model, the principal input that could be manipulated was system function and this consisted of task match, ease of learning, and ease of use. Task match was concerned with the match or overlap between a user's requirements and the functions that a system provided. Ease of learning was concerned with the level of difficulty a user would encounter in using a system they were not familiar with, while ease of use was focused on the level of effort required by a user to use a system once they understood and could use its functions.

The user characteristics hypothesised to influence usability were knowledge, discretion and motivation. Knowledge was concerned with the information that the user chose to use in performing a task. The motivation variable was concerned with how motivated a user was to use the system and consequently how much effort they were willing to employ in overcoming difficulties. The final user characteristic that Eason argued would affect usability was discretion; which in this context referred to the user's ability to decide not to use some or all of a system.

The final set of inputs that Eason (1984) suggested would influence usability were the task characteristics of frequency and openness. Frequency was concerned with the number of times a user performed any particular task. The important point noted here was that tasks that were performed more frequently needed to be treated differently from those that were performed infrequently. So, for example, users may have required shorter methods of conducting frequently performed tasks while requiring guides to more infrequently performed tasks. The output of Eason's model was user reaction and the associated implicit cost-benefit analysis. Good task-system match and consequently continued user learning, it was hypothesised, would result in positive outcomes, while the restricted use of a given system on the other hand would result in negative outcomes. The restricted use element of this model is symptomatic of the type of problems associated with use of Flight Management Systems. It has been suggested that the lack of usability in FMS CDUs results in many pilots choosing to restrict their use below 10,000 feet (Hughes, 1995).

The crux of Eason's model was that the usability of a system was dependent on the interactions of users, tasks and systems. His subsequent attempt to define usability suggested that the major indicator of whether or not a system was usable was *"whether a system or facility was used"* (p.140). Eason's argument was that it was the act of choosing to use a system that defined the essence of usability.

The problem with Eason's definition and with many attempts at defining usability since, is that the definition is based on the premise of evaluating the system by implying that a system would have to be completely built and implemented in order to determine whether or not it was usable. While it was clear that Eason might have provided a global definition, it would clearly not be completely satisfactory from an industrial and cost perspective. Furthermore, as in the context of this research, interacting with a system is not always an optional activity, so whether or not the automation is used can not always be used as a measure of how usable the system is. It became evident at this point that the definition of usability hypothesised by Eason was not adequate.

In 1986 Shackel did succeed in suggesting an operational definition by integrating and developing the work of Eason (1984) and others by arguing that there were four rather than three factors which interacted to form usability and these were user, task, system and importantly, environment. The formalised definition of usability proposed by Shackel (1986) could be used to set a criterion for usability by assigning numerical values to set usability goals during the requirements specification part of the design cycle. Shackel argued that for a system to be usable the following goals would have to be achieved.

- Learnability: The system should allow learning to be achieved within some specified time from installation and start of user training based upon some specified amount of training and user support, and within some specified relearning time for intermittent users.
- Effectiveness: The required tasks should be performed at better than a specified level of performance, by a specified required percentage of a specified target range of users, and within some required proportion of the range of usage environments.
- Flexibility: The system should be adequately flexible to allow adaptation of some specified percentage variation in tasks and/or environments beyond those first specified.
- Attitude: The system should achieve acceptable levels of human cost in terms of tiredness, discomfort, frustration and personal effort.

Shackel's (1986) work, illustrated that, it was now possible to set a usability goal such that a user should be able to use the system within a given time period of beginning to use the system, assuming manuals were provided or training available. The important progress made by this approach, Shackel asserted, was that usability in this framework included elements which could all be measured.

Despite the interest that Shackel's definition has gained, it is not without limitations. One of the problems with the definition is that it does not specify the target values that can be used to distinguish between a "usable" and an "unusable" system. Other critics of Shackel's definition, including Booth (1989), have argued that, for example, "flexibility" is very difficult to specify, communicate and test in a real environment. A second criticism of Shackel's definition made by Booth was that it does not include usefulness i.e. the definition does not include the need for a product to help users achieve their goals. In reviewing Shackel's definition Booth (1989) argued that the term "flexibility" should be removed from the definition and "usefulness" included in its

place. A useful system, Booth (1989) argued was one *"that helps users to achieve their goals"* (p.112), the argument being that a system could meet the criteria set by Shackel's definition but might not be used because it did not match the user's goals.

Ben Shneiderman in 1987 similarly specified five measurable usability criteria; time to learn (learnability), speed of performance (efficiency), rate of errors by users, subjective satisfaction (attitude) and retention over time (memorability). These dimensions can be measured but are again prone to the same criticisms of Shackel's framework outlined previously.

Bevan, Kirakowski & Maissel in 1991 pointed out that the term usability was initially coined to replace the term "user friendly" which had developed vague connotations. They further argued that the term usability had, in the previous ten years, become equally devalued. Their taxonomy of how usability had been defined related to how usability could be measured and was therefore broken down into three categories:

- A product-oriented view, such that usability could be measured in terms of the ergonomic properties of the product.
- A user-oriented view, which focused on measuring the mental effort and attitude of the product users.
- A user-performance view, which focused on how the user uses a product; the emphasis being on ease of use.

More recently, the International Standards Organisation (ISO) proposed a contextually oriented definition of usability within the Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) Part 11: Guidance on Usability (1998), which appears to be paving the way toward a measure of usability that could be used during the product development process. The definition of usability proposed within this document stated that usability was *"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use"* (p.2). The model proposed under ISO 9241 Part 11 (p.3) is shown in Figure 1-7.

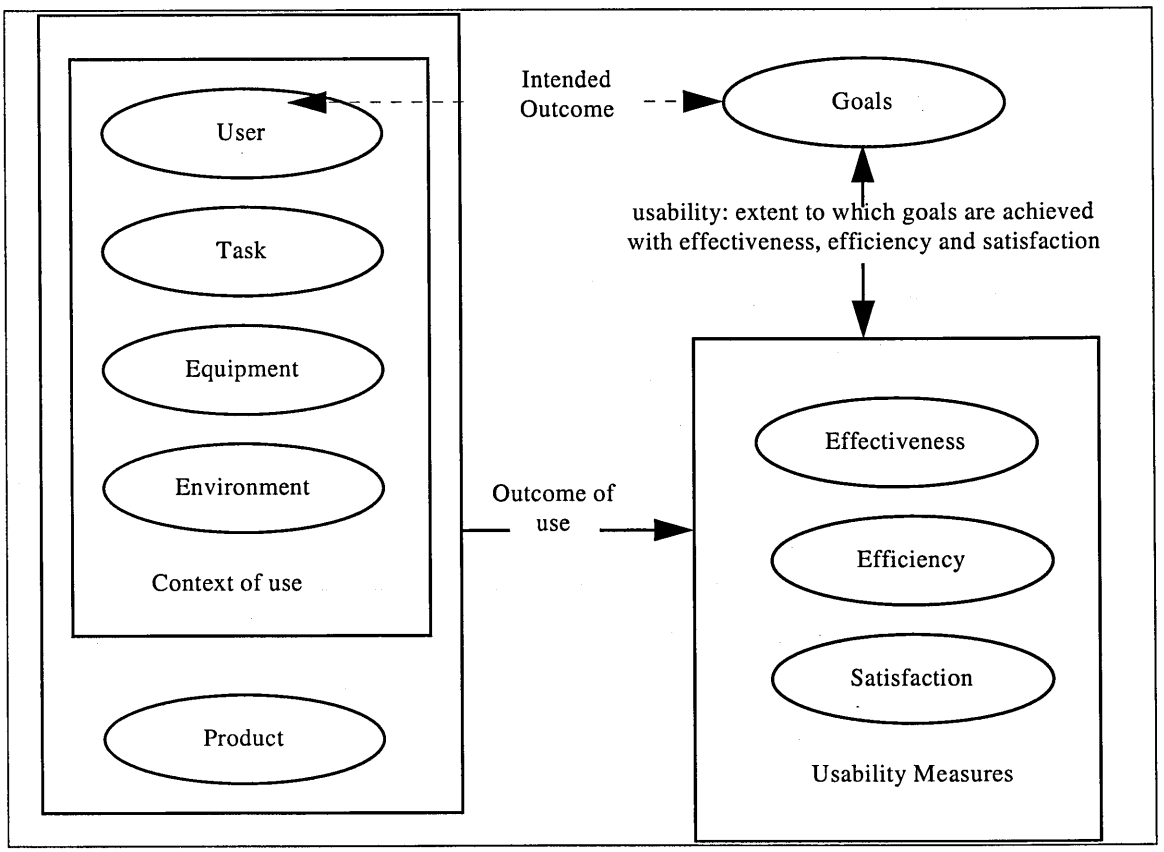


Figure 1-7 - Usability Framework from ISO 9241 Part 11 p.3

According to this definition and framework, in order to specify the usability of a product it is necessary to specify (a) a description of the intended goals (b) a description of the components of the context of use including users, tasks, equipment and environments and (c) target or actual values of effectiveness, efficiency and satisfaction for the intended contexts.

A criticism of this attempt at defining usability is that it relies on that which can be measured (Stanton & Baber, 1996). Stanton & Baber (1996) asserted that according to this definition *"usability is what usability evaluations do"* (p.41), and continued to argue that *"haste in producing a definition of usability should be chastened by rather more circumspect consideration about what is meant by usability"* (p.41).

Examining the issue from the perspective of usability problems, the picture is just as murky. Lavery, Cockton & Atkinson (1997) argued that the actual meaning of the term 'usability problem' had received *"little attention in the literature"* (p.251). Karat, Campbell & Fiegel (1992) argued that *"a usability problem was defined as anything that interfered with a user's ability to efficiently and effectively complete tasks"* (p.397). Both Desurvire (1994) and Dutt, Johnson & Johnson (1994), however, expressed doubt as to what constituted a usability problem.

1.8 Components of Usability

As previously stated there is little agreement about the definition of usability and the term usability has "*remained a fuzzy concept which has been difficult to evaluate and impossible to measure*" (Bevan & Macleod, 1994, p.132). Nielsen (1993) argued that "*Only by defining the abstract concept of "usability" in terms of these more precise and measurable components can we arrive at an engineering discipline where usability is not just argued about but is systematically approached, improved and evaluated*" (p.26). If as suggested by Stanton & Baber (1996) usability is more than an ephemeral concept, its components must be agreed upon. A review of the last two decades of research, however, indicates that numerous factors are associated with the concept of usability, and many will be described in the following subsections. A discussion of usability issues specifically related to the FMS CDU will be presented in Section 1.9.

1.8.1 Learnability

Learnability is concerned with ensuring that a system is easy to learn so that a user can rapidly start getting some work done with the system. Shackel (1986) asserted that learnability should be achieved within some specified time from installation and start of user training, based upon some specified amount of training and user support, and within some specified re-learning time for intermittent users.

Nielsen (1993) argued that learnability was perhaps the most fundamental usability component since most people's first experience with a system will be when they try to use it. Nielsen went on to outline how highly learnable systems have a learning curve which is steep in its first part, and allows users to achieve a reasonable level of usage within a short space of time. The exception to this, Nielsen argued, was in systems intended for one off usage such as museum information kiosks. These interfaces should have a zero learning time so that people can interact successfully with them at once. Stanton & Baber (1996) defined a learnable system as one which would allow users to reach "*acceptable performance levels within a specified time*" (p.40).

In terms of assessing usability, Nielsen (1993) pointed out that users do not take the time to learn an interface completely before they begin using it, and in fact users frequently starting using a system as soon as they have learned a part of it.

1.8.2 Visual Clarity

Ravden & Johnson (1989) in discussing visual clarity argued that information displayed on a screen should be clear, well organised, unambiguous and easy to read. Oppermann & Reiterer (1997) stated that visual clarity was typically associated with the attributes of "*clarity, discriminability, conciseness, consistency, detectability, legibility and comprehensibility*" (p.233).

Evidence from the literature, for example, Shneiderman (1992), suggested that for good visual clarity the following points were necessary:

- Screens should appear uncluttered.
- Users should be able to find the information they require and to see where information should be entered on screen easily and quickly.
- Important information on screen should be drawn to the user's attention.
- Users should be able to see and read information on the screen clearly and easily.

Much has been written over the past twenty years (e.g. Galitz, 1980) detailing how displays should be structured and organised so as to avoid unnecessary complexity and to group information in meaningful ways.

Jordan (1998) highlighted the role of visual clarity in creating a usable interface by asserting the importance of displaying information in a way that could be read quickly and easily and without causing confusion.

Issues for consideration when addressing visual clarity with on-screen interfaces have been examined by many authors (see Jordan, 1998; Nielsen, 1993) and include factors such as use of colour and the distance at which the operator will sit from the terminal being operated. When assessing how much information should be displayed on a single screen it might be considered that maximising the information on a single screen would mean that less depth in terms of menu structure would be required, while excessive information on a single screen may lead to the screen appearing cluttered. Having less information on a screen would mean that textual characters could be enlarged, thereby increasing the legibility of information provided, but this may contribute to a requirement to navigate through a number of menus in order to access the information required by the user.

1.8.3 Consistency & Compatibility

Consistency refers to the principle that when using a system, similar tasks should be performed in similar ways. Following a six-month period spent in the UK, Don Norman was triggered to write the book "The Psychology of Everyday Things" (1988) after becoming aware of numerous examples of design inconsistencies in his environment. One example he cited was that even water taps were not standardized, with hot water sometimes coming from the left tap, while at other times coming from the right tap.

Jordan (1998) cited an example of a word-processing package where the steps involved in making text bold or italic are consistent. To make text bold, the text to be formatted is highlighted, the "Format menu" is opened, and the "Bold" command is selected. Consistent with these steps, to italicise text, the text to be formatted is highlighted, the "Format menu" is opened, and the "Italic" command is selected. Inconsistencies, Jordan asserted were likely to lead to errors. Following the same example through, Jordan offered the example where the "Bold" command was in the "Format menu" but the

"Italic" command was not. A user who had learnt how to make text bold could well go to the "Format menu" and find that it was not possible to italicise text from there.

The issues raised for Consistency are also true when the compatibility between systems is examined. This component has sometimes been treated as a separate factor by for example, Ravden & Johnson (1989) under the term "*compatibility*" (p.32), but the underlying principles for compatibility between systems are similar to those within a system. Systems should look and operate in a manner that is compatible with user conventions and expectations.

Cars, for example, are designed by different manufacturers but in a compatible format to the extent that gear levers are usually placed in the centre console, and three pedals which have the same functions, are located in the same order. Should manufacturers not be compatible with other manufacturers by altering the layout of the clutch, brake and accelerator, drivers purchasing a new car would be more prone to making errors. The potential for mistakes would be evident in a stressful situation if a driver put their foot on the accelerator rather than the brake. Shneiderman (1992) argued that from a user perspective, slight differences among systems could be annoying and could lead to dangerous errors. Large differences between systems could burden users and could create the need for re-training.

1.8.4 Error Prevention & Correction

The focus of this usability component is that systems should be error tolerant and when users make errors, guidance should be provided on how to correct those errors. Nielsen (1993) asserted that errors that occur when using a system are critical to usability for two reasons. Firstly, they identify situations where users have trouble using a system. Secondly, they present an opportunity to assist the user in understanding the system better. Systems should correspondingly have good error recovery, or correction facilities. Nielsen cited examples of this, such as giving the user the ability to undo erroneous commands they have entered, and the opportunity to edit and re-enter previously used commands without having to enter them from scratch.

Shneiderman (1992) argued that users of interactive systems, such as word-processors or air traffic control systems, made errors far more frequently than one would expect. Using the example of a text editor, Shneiderman argued that novice users typically made mistakes in nineteen percent of their commands, while experts made mistakes in ten percent of their commands. Evidence from more demanding environments, Shneiderman argued, indicated that users made mistakes or used inefficient strategies in thirty-one percent of the tasks they were asked to do. Errors that occur in the desktop environment are expensive in terms of productivity, and are a source of frustration to users. In a safety critical environment they are additionally a safety risk.

Shneiderman (1992) suggested a number of guidelines for designing error messages into systems.

- Error messages should be specific and precise.
- Error messages should be positive.
- Error messages should tell the user what needs to be done to correct erroneous actions.
- Tying in with the consistency component, error messages should be consistent in the terminology and phraseology used.
- The phrasing of error messages should be user-centred.
- Error messages should use a consistent display format.
- The usability of error messages should be tested.
- The requirement for error messages should be reduced.

1.8.5 Appropriate Functionality

Ravden & Johnson (1989) asserted that appropriate functionality referred to the requirement for a usable system to meet the needs and requirements of users when carrying out tasks. They argued that if the system interface did not provide the information or options which user's required to carry out a task, it would lead to frustration in users and could impair their efficiency in carrying out the task. Appropriate functionality, has also been referred to as "task match" which was defined by Shackel (1986) as being the existence of an acceptable match between the functions provided by a product and the user's requirements. In considering the appropriate functionality for a system, the system's performance requirements must be determined with due consideration of the user and the system's context of use. In addition, function allocation which refers to the process of dividing activities between humans and machines (Fuld, 1997), can also impact on usability.

1.8.6 Effectiveness

Effectiveness was defined by ISO 9241 Part 11 as *"the accuracy and completeness with which users achieve specified goals"* (p.12). A specified proportion of users should achieve an acceptable level of performance, over a specified range of tasks in a specific range of environments. Jordan (1998) described effectiveness as *"the extent to which a goal, or task is achieved"* (p.18).

The most basic measure of product effectiveness for a task is whether the user can perform that task while using the product (Sutcliffe, 1988). Jordan (1998) used the example of a vacuum clearer to illustrate this dimension and argued that if the task the vacuum cleaner was designed for was to clean a carpet, then the vacuum cleaner would be effective if it could get the carpet clean. This is, however, a simplistic example since there are varying degree of effectiveness. In producing a document using a word-processor, a user might be required to perform a number of sub-tasks, and the system's effectiveness might be related to the number of sub-tasks that the user was able to perform successfully. A further determinant of effectiveness might be the quality of output produced by the system.

1.8.7 Efficiency

As previously discussed, Shackel (1986) described efficient systems as being those which permitted tasks to be performed at better than a specified level of performance, by a specified required percentage of a specified target range of users, and within some required proportion of the range of usage environments. Efficiency was defined in ISO 9241 Part 11 as *"The resources expended in relation to the accuracy and completeness with which users achieve goals"* (p.2). Effort could, for example, be measured as the time taken to complete a task. A further indicator of whether a system is efficient is the mental and physical workload incurred in performing a task. If a system is efficient then it does not cause high levels of workload for its users. Similarly high levels of productivity should be possible when using an efficient system.

Efficient properties are very evident to users of desktop applications in which numerous examples of efficiency reside. In word-processing packages, for example, the use of keyboard short cuts permits the user to perform numerous tasks without having to take their hands off the keyboard. For a beginner learning to use a word-processing package, it would be difficult to remember all the available short cuts, and drop-down menus provide a very explicit method of selecting commands such as "print" and "save". For the experienced user, however, the requirement to use such drop-down menus can become laborious and time consuming. Keyboard short cuts such as "Control-P" and "Control-S" provide a very efficient alternative. Similarly, the availability of cut, copy and paste commands in word-processing packages are efficient in that they eliminate the user's requirement to re-enter data that has previously been entered.

1.8.8 Feedback

The feedback principle is concerned with the fact that users interacting with a system should always be provided with informative feedback on their location within a system, the actions that have been taken and whether these actions have been successful (Jordan, 1998). Nielsen (1993) argued that a system should continuously update the user on what it is doing and on how the system is interpreting the inputs that the user has made. A user should not have to wait until they make an error to receive feedback, but rather should continuously receive feedback that their inputs have been positively received.

The most important consideration when giving feedback to a system user is that the feedback is meaningful. An example of informative feedback that most people are familiar with is the ring or engaged tone received when placing a telephone call. System feedback should not be expressed in either abstract or general terms but rather should re-state the information the user has put into the system in a format that tells the user what the system is doing with that information.

Nielsen (1993) described three types of persistence associated with feedback messages that were dependent on the context of the feedback message and its importance. Low persistence feedback was that associated with a particular action and which disappeared when that action had been completed and the feedback was no longer required. An example of such feedback in the desktop environment is the flashing "disk" icon that appears when a document is being saved. Once the document is saved, the icon stops flashing. Medium persistence feedback, Nielsen argued, required the user to acknowledge the feedback message. An example of medium persistence feedback occurs when a user prints a document and the PC re-routes the document to a different printer due to some problem with the printer selected. The feedback message would then require acknowledgement from the system user that the output would have to be collected from a different printer. The final type of feedback described by Nielsen (1993) was high persistence feedback, which occurred when the feedback message was so important that it remained a permanent feature of the interface. An example of high persistence feedback from the desktop PC environment is the indicator of how much free space remains on a computer's hard disk.

System response times should normally be as fast as possible, but not so fast as to make it difficult for the user to keep up with the feedback (Nielsen, 1993). Where a system cannot provide fairly immediate feedback, a "percentage done" indicator should be used to provide the user with feedback on how long the operation being performed will take to complete. In the desktop environment, the "egg-timer" symbol provides feedback that the computer is doing something, thereby preventing the user becoming frustrated by not knowing what the computer is doing.

Informative feedback is also required when systems fail. Nielsen (1993) reported that most systems are not designed to do so and simply stop responding to the user when they fail. Systems should rather be designed for graceful degradation, enabling them to provide some feedback to users when most of the system is down.

The interrelationship between the usability components is evident when the relationship between the feedback component and the error prevention and correction component are examined. If feedback messages are vague or imprecise, the user could, for example, continue to make inputs to the system while the previous input is still being processed resulting in the later inputs not being processed. Similarly, inadequate feedback could result in the user not knowing how to correct an erroneous input.

1.8.9 Explicitness

In order for a system to be explicit, it should be clear to the user how the system works and is structured (Marshall, Nelson & Gardiner, 1987). The interface should be transparent and its processes should be visible to the user (Mayhew, 1992). Norman (1992) used the term visibility to refer to the same concept.

Ravden & Johnson (1989) argued that if explicitness was achieved, systems users would be more likely to have an accurate representation of the interface they were working with. Jordan (1998) provided an example of explicitness in terms of how the doors in public buildings are designed. Prior to entering a building, individuals must determine if the door is opened through pushing or pulling. Well-designed doors, Jordan argued, should make it obvious which action is required. In the desktop environment, explicitness in menu-based systems is achieved through clearly naming command functions. The function to command the printing of a document is therefore called "print" and the saving of a document is controlled through use of a "save" command. Similarly, the representational properties of icons assist a user to determine the associated functionality. A picture of a printer is used to represent the "print" commands, while a picture of a disk is used to represent the "save" command.

Steps that can be undertaken to ensure that a system is explicit include providing informative feedback to the user as to where they are in the system. The usefulness of these measures has been seen in many hypermedia applications where a map of the system architecture is displayed somewhere on the screen so it is apparent to the user where they are within the system architecture, where they have come from and are going to.

The relationship between the usability components is evident when explicitness is examined, since other components such as visual clarity and consistency would improve the transparency of a system, thereby potentially reducing the requirement for constant feedback.

1.8.10 Flexibility

Lin, Choong & Salvendy (1997) referred to flexibility in terms of the interface being able to adapt to users' requirements. This adaptation could be made by either user customization or by the interface itself. Systems should be adaptable in terms of how information is presented and what can be done in order to suit the needs of all users, and to allow users to feel in control of the system they are interacting with.

Scapin and Bastien (1997) agreed that the rationale behind flexibility is that the more ways there are of doing a task, the more likely a user will find a way that suits their way of performing a task. They went on to argue that different procedures and commands should be available for a user to achieve a given goal. Some of the guidelines they suggested included:

- Provide users with some means of controlling their display configuration.
- Provide some means for users to alter the sequence of data entry, to their preferred method.
- For text formats that cannot be predicted in advance, allow users to specify and store the formats that might be needed.
- Allow users control over screen pages.

Shneiderman (1992) argued that flexibility, for example, in the Air Traffic Control context is desirable since controller's sometimes want to display arrival time, whilst at other times their principle concern is an aircraft's altitude. Systems should have the capability to adapt to user preferences. Shneiderman did, however, argue that flexibility should be used with caution since it goes against the consistency principle. Too much flexibility may only serve to confuse particularly the novice user and increase their workload.

1.8.11 Workload

Workload refers to the amount of effort involved in performing tasks with a system. Bevan (1995) argued that this dimension of usability was particularly important in safety critical applications. Workload incorporates both physical and mental aspects of a task, although only the mental or cognitive dimension of workload is normally discussed in the usability literature. Mental workload is concerned with the difference between the quantity of information processing resources required to perform a task in a particular situation and the quantity of such resources available to an individual at that time (Wickens, 1992). A task demanding too little effort can result in lower efficiency since it leads to boredom and a lack of vigilance, which results in lower effectiveness. Excessive mental workload can have negative influences on all aspects of information processing (BASI, 1998), and may cause information to be missed resulting in errors, thus also reducing effectiveness (ISO 9241 Part 11).

1.8.12 'Help' Facilities

Help Facilities, in terms of documentation (users manuals) and on-line help, are part of the user interface, and users can form impressions of how usable a system is by flicking through a manual or accessing its on-line help (Mayhew, 1992). Such "help" can therefore determine in large part both the perception and actual ease of learning and use of a system and, as Mayhew reported, will either degrade or enhance the overall usability of a system. Help facilities are useful for both new and experienced users, since all users tend to forget aspects of a system from time to time (Maguire, 1982).

Mayhew (1992) reported that user manuals were typically poorly designed, full of technical jargon, difficult to navigate through, difficult to read, and they *"simply do not support the natural learning and problem-solving processes of people"* (p.535). One common type of user manual is the user guide that is aimed at intermediate users who know some, but not all of a system's functionality. The type of user manual typically found on the flight deck, however, is the "Reference Manual". Reference manuals are typically aimed at experts and provide complete information on all system functions. Reference manuals should be organised to permit the rapid retrieval of specific information, and should contain complete information on all system functions.

User manuals should be written clearly, and as part of the user interface, should be aimed at the level of the end user. There is little point providing a user manual written by a computer programmer, for a computer programmer, to a user who has only used a computer a small number of times. Further, Mayhew (1992) reported that it was preferable to provide instructional text at a lower grade level than those who would use it were at. Contrary to popular belief, Mayhew argued, users did not find this condescending.

The alternative to paper-based help, is on-line help. Shneiderman (1992) argued that on-line help is probably most effective for knowledgeable users. Using on-line help, users of desktop systems typically select a help menu, and from there choose from a set of alphabetically arranged command names that have an associated paragraph or more of help. Unlike hard-copy documentation, on-line help is cheaper to update and potentially easier to navigate. The disadvantage of on-line help is that it may be more difficult to read due to screen resolution and page navigation is slower due to pages being small to fit on the screen. However, Ravden & Johnson asserted as far back as 1989 that it was then widely accepted that system users should have access to some form of on-line help.

1.8.13 Attitude

The attitude dimension of usability refers to the level of comfort felt by a user when using a product or as stated by Preece et al (1994) the *"attitude engendered in users by using a system"* (p.401). Shackel (1986) argued that for a system to be usable in terms of user attitude it should reach acceptable levels of human cost in terms of tiredness, discomfort, frustration and personal effort.

Attitude plays a significant role in the performance of interactive tasks and interactive systems can be designed to minimise anxiety, threat, boredom and apathy, and to maximise job satisfaction. Mayhew (1992) reported that emotional arousal affected performance with interactive systems in a U-shaped function. According to Mayhew, too little arousal impeded performance due to lack of attention, a moderate degree of arousal improved performance by increasing alertness, while a huge degree of arousal impeded performance perhaps because users found it difficult to concentrate. Rubin (1994) argued that users were more likely to perform well with a product that *"meets their needs and provides satisfaction than one that does not"* (p.19).

When Shackel recognised the importance of user attitude in 1986, the emphasis was on minimising the number of software bugs in a system that would cause the user's performance to be degraded as well as causing the user frustration. The importance of incorporating user values when designing systems has been recognised, and the emphasis is now on how to create a positive user attitude toward a particular system (Jordan, Thomas & McClelland, 1996).

1.8.14 Summary of Usability Components

From the literature it is evident that usability is dependent on a product being oriented around the user, the task to be performed with the product and the specific context or environment in which the product will be used. Figure 1-8 provides a conceptual overview of usability with examples of both system characteristics and associated effects on the system user.

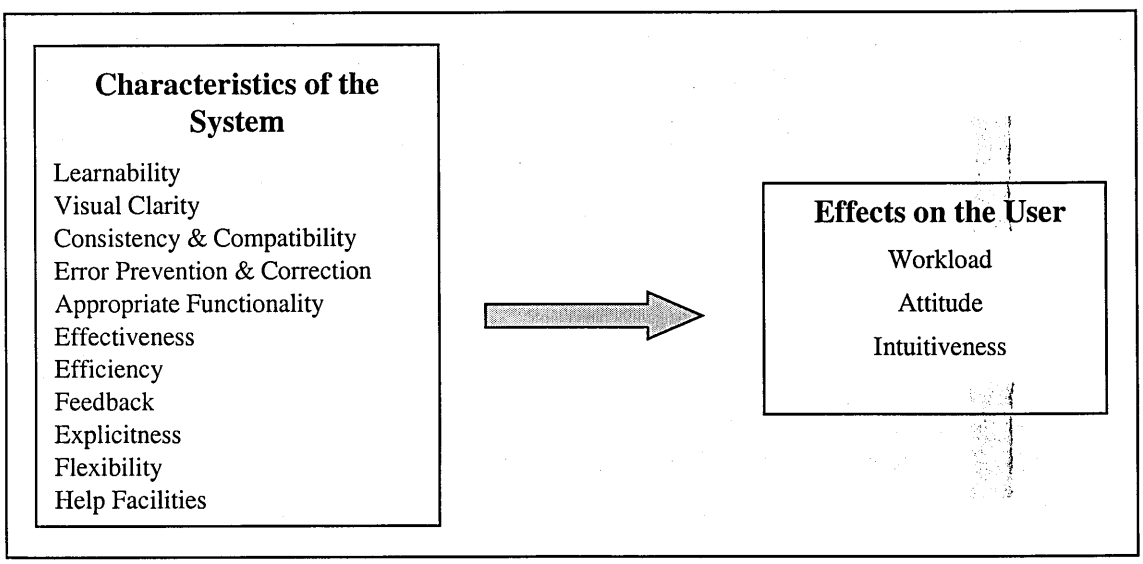


Figure 1-8 - Conceptual Model of Usability

Although there is no accepted definition of the hypothetical construct of usability, the components of usability that have been most frequently cited in the literature are summarised here:

1. Learnability: A system should be easy to learn so that the user can rapidly start getting some work done with the system.

2. Visual Clarity: Information on a screen should be displayed in a format that can be read easily and quickly without confusion.

3. Consistency & Compatibility: The way that a system looks and operates should always be consistent. The amount of information re-coding required in transferring between system interfaces should be minimised.

4. Error prevention & correction: Systems should be designed to minimise the possibility of error, with in-built facilities for detecting and handling those errors that do occur. Additionally, users should be able to check their inputs and to correct errors, or potential errors, before the inputs are processed.

5. Appropriate Functionality: A system should meet the needs and requirements of users when carrying out tasks.

6. Effectiveness: A specified proportion of users should achieve an acceptable level of performance, over a specified range of task in a specific range of environments.

7. Efficiency: The system should minimise resources expended in relation to the accuracy and completeness with which users achieve their goals.

8. Feedback: Users interacting with a system should always be provided with informative feedback on their location within a system, what actions they have taken, whether these actions have been successful and an indication of what actions should be taken next.

9. Explicitness: The way a system works and is structured should be intuitive to the user.

10. Flexibility: A system should be adaptable in terms of how information is presented and what users can do, to suit the needs of all users.

11. Workload: The mental and physical workload imposed by a system on a user should not be too high or too low.

12. Help Facilities: Users should be provided with an informative and easy to use help facility both within the system they are operating and in hard-copy form.

13. Attitude: Users should be subjectively satisfied when using the system.

It should again be emphasised when considering this list of components that most usability research to date has focused on office-based software, hardware and its associated environment. There are a number of key differences that make the assessment of Flight Management Systems fundamentally different to conventional systems such as personal computers. These key differences are:

- The flight deck is a highly complex environment.
- The flight deck is a safety critical environment.
- The flight deck is a dynamic environment both internally and externally, and is subject to the requirement for rapid change from sources such as the weather and Air Traffic Control. The flexibility of the FMS and the dynamism of the operating environment further impose cognitive demands on pilots. Pilots may have to intervene quickly to change parameters.

- FMSs are operated in a co-operative manner that is distinct from the individual use of a computer. Dourish & Bellotti (1992) noted that the flight deck is a synchronous system since both pilots are present in the same workspace, at the same time, and may both be programming the FMS. These authors noted that in asynchronous systems, information is shared explicitly through leaving messages for a subsequent user. However, in a synchronous system, such as the FMS, information about what is being worked on may be passively collected and distributed to the other user, and may as such create additional workload. Further, the real-time processing of information and the non-exclusive focus of the FMS, can further compound the user's workload.
- While not wishing to promote the view that an aircraft cannot be flown without use of an FMS, its use is an integral part of the job of flying modern aircraft. While aircrew may choose to fly the aircraft with differing levels of automation engaged, or even manually, FMS use is to some extent essential from an airline perspective since it increases flight efficiency. Jordan (1998) noted that in situations where interacting with high technology products is compulsory, it could be argued that usability is just as important as when its use is optional.

1.9 Usability and the FMS

The FMS CDU has the most complex interface on the flight deck, and is known to be the weakest link of the flight deck automation system in terms of high workload, dynamic situations in flight (FAA, 1996; Woods & Sarter, 1998). Sarter & Woods (1994) after a series studies which examined pilot interaction with flight deck automation discovered latent problems with pilot-FMS interaction that could effect even experienced pilots' performance in non-normal time-critical situations. They argued that "*The page architecture of the FMS CDU contains a huge amount of data that may be relevant at some point during the flight. Because only a very limited set of data can be presented on the CDU screen at any given time, pilots need to be able to navigate through the "hidden" data space*" (p.8).

Interacting with an FMS CDU requires close visual attention to the five-inch screen to ensure the accuracy of the alphanumeric data being entered. Pilots are thus required to interrogate a large database through a small keyhole. It is further possible with FMSs to make errors, which do not become obvious until much later. Route entry errors made manually during the pre-flight process outlined in Section 1.4 may, for example, not come to the pilot's attention until they are already in the climb or cruise phase of flight. Humans are known to be subject to making errors when manually entering alphanumeric data since numbers may be recalled incorrectly from memory, may be input incorrectly or may be misread prior to their execution. In addition, pilots may also have to remember data entry sequences since prompts indicting which data should be entered next are frequently unavailable. Billings (1997) asserted that "*In all FMSs, the complexity of the mode and display architecture poses substantial operational issues*" (p. 114).

Billings (1997) further reported that since the advent of radios that may only be tuned through CDU interactions, pilots have complained that data entry through the alphanumeric keypad is time consuming and requires more time "heads down" inside the cockpit than setting the rotary selector knobs in older aircraft. A further difficulty in interacting with FMS CDUs has been the lack of feedback provided when data is entered incorrectly. When data is entered in a format deemed unacceptable by the FMS, it is rejected with a message such as "invalid entry" without detailing why the entry is invalid, or what the correct format for entering the information is.

Avionics manufacturers have made some attempt to make FMSs CDUs more error tolerant by creating databases of frequently used routes, Standard Instrument Departures (SIDs) and Standard Arrival Routes (STARs). Little progress has been made, however, in providing information for corrective actions once errors have been made. Hughes (1995) reported that the senior vice president of Airbus Industrie Engineering argued that FMSs offered too many possibilities and were too complex, resulting in pilots relying on only about 20% of the software features. Sarter & Woods have conducted research examining automated cockpits since the late 1980s. Two of their studies in particular (1992 & 1994) have focused on FMS behaviours which have surprised pilots and caused them difficulties in interacting with the CDU.

In the first study, Sarter & Woods (1992) conducted a survey of Boeing 737-300 pilots prior to observing them in a one-hour session on a part-task training simulator. Twenty pilots participated in the study, 14 of whom were line pilots experienced on this type of aircraft, the remaining six pilots were newly trained on this aircraft type and were about to complete their line training. The study findings indicated that there were deficiencies in the pilots' mental models of the functional structure of how the aircraft's FMS operated in specific situations, such as an aborted take-off. The pilots did not understand how the automation operated in certain circumstances, although these problems were primarily associated with non-routine, time-critical situations. Further, research has indicated that pilots have significant problems with communicating and acknowledging changes in the FMS during all phases of flight. Hines (1997) reported that pilots further may not share the same mental model of the current and future actions of the FMS.

More pertinent to this thesis was that Sarter & Woods' (1994) research went on to examine pilots' knowledge of how the CDU pages were structured. The participants were asked to find single engine capabilities, wind data for fixes in flight, available fuel, and other information. Some of the pilot participants reported that they were unaware of how the FMS would behave in certain situations, and also that they found anticipating what would happen next to be difficult. The B-737-300 research further illustrated that only one of the 20 pilots could predict the whole sequence of mode indications during a take-off roll to permit them to stay a step ahead of the FMS. A subsequent study conducted by Sarter & Woods (1995) in an A320 full motion simulator found the same types of problems associated with that aircraft type.

Many of the most serious problems associated with interacting with FMS CDUs are related to the fact that FMSs were initially designed as a navigation aid for use at high altitude in the cruise phase of flight, when workload was low. The functionality of the

initial FMSs was so well received that their use was extended through the entire flight, from the pre-flight process right through to the descent and approach phases of flight, without any interface re-design. Dyck, Abbott & Wise (1993) reported that pilots have difficulty programming information into the CDU, and have difficulty understanding what the outcome of that programming will be. Pilots also reported disliking the inconsistencies in the layout of the alphanumeric keypad, and noted that a high degree of communication and co-ordination is required in use of the FMS.

In 1992, Dodd, Eldredge & Mangold produced a report for the FAA which examined FMS related incidents, and found that *"the FMS does not "lead" the pilot in terms of the expected series of steps that must be performed to accomplish the expected goal or end result. Furthermore, the placement of the various information sources that provide feedback to the pilot, has not been optimized, and requires significant visual and cognitive workload to obtain and understand the necessary information."* (p.vii).

Dodd et al (1992) further found that the FMS CDU was not optimally organised from a human-computer interaction perspective in terms of the screen presentation or the organization of that information. This team of researchers recommended that further research be conducted into FMS CDUs to ensure that:

- Both retrofit and next generation FMS equipment rectified FMS design problems which contributed to pilot error.
- Information was presented in an accurate and understandable form in terms of what the FMS was doing or was expected to do.
- Crews could make modifications to the flight plan in efficient, safe and predictable ways.

Billings (1997) concluded that while research was underway to improve cumbersome alphanumeric data entry, "*considerable attention*" to the CDU display was also warranted (p.117). Eldredge, Mangold & Dodd (1992) similarly reinforced the requirement to examine the FMS from a usability perspective arguing that: "*a tool (or tools) that would allow an overall assessment of how easy a specific FMS is to use would have great value, especially as a means for evaluating new FMSs when they are presented for certification*" (p. 6-2).

In order to develop a tool that would permit an overall assessment of how easy specific FMSs were to use, it would be necessary to consider which of the specific usability factors were applicable to the FMS. No usability components could be eliminated or viewed as less important a priori i.e. it was necessary to assume that all of the usability attributes outlined in the literature would be applicable to the FMS unless the research provided evidence to the contrary.

The following sub-sections consider the usability principles outlined in Section 1.8 in the context of the FMS CDU, which is the focus of this research.

Learnability

Learnability, for example, could raise some different issues for an FMS CDU than for a desktop application. Pilots normally learn to use the FMS CDU through an intensive training programme as part of an aircraft type-rating. The pilots are extremely motivated since they are not permitted to fly the aircraft until all parts of their training, including FMS training, have reached an adequate standard. Shackel's (1986) reference to intermittent users in his specification of learnability targets, for example, is not really applicable since pilots use their FMS on a regular basis. Nielsen's (1993) reference to users using a system as soon as they have learned a part of it should also be viewed differently for the FMS CDU, since a pilot will not be able to use an aircraft FMS as soon as a part has been learned.

Visual Clarity

The Visual Clarity aspect of usability is particularly pertinent when the size of the interface is limited by the space available, such as is the case with the FMS CDU on a flight deck. The available display area and the quantity of data to be displayed through that interface limit how the information will be structured in term of depth and breadth. Putting many functions on one page could cause clutter and increase the search time taken to find the required information but would reduce the depth required in the menu structure, and thus the opportunity to get lost within that menu structure (Jordan, 1998). It is then up to the designer to optimally structure the data in a form that optimises visual clarity.

The distance that the user will be from the FMS CDU must also be considered. It must also be considered that FMS CDUs are located in the centre of the flight deck and will therefore be viewed from either the right or left depending on which pilot it is being used by.

A further aspect of visual clarity that must be considered for the FMS CDU is that of labelling. The labels on controls and buttons, for example, must be large enough for the operator to see when working under extreme operating conditions such as under vibration in the flight deck environment.

Consistency & Compatibility

In discussing software interfaces for aviation systems, Williges, Williges & Elkerton (1987) discussed consistency in terms of the requirement to minimise dialogue differences across human-computer interfaces. Compatibility, they argued, referred to the requirement to minimise the amount of information re-coding that would be necessary in transferring between human-computer interfaces.

Consistency in the FMS CDU context might refer to similar menu information being located in the same place on every applicable CDU page. Similarly, one would expect FMS inputs to achieve the same outcome every time and to avoid the aircrew becoming frustrated the outcome should be achieved in a similar response time. It is anticipated that consistency and compatibility may be particularly pertinent for FMS users, since

within one fleet and within one airline, pilots are expected to use FMSs that are different both in speed of system response and in terms of CDU display layout. This is largely due to the way in which airlines operate. Aircraft are frequently bought by airlines from sources such as another airline rather than directly from manufacturers like Boeing and Airbus. Any modification or customisation made to the aircraft or FMS by the airline from which the aircraft was bought will result in differences in the fleet of the purchasing airline.

Similarly, in undertaking a new type-rating and moving from an aircraft made by one manufacturer to an aircraft made by another manufacturer, compatibility across aircraft types would make for a smoother transition for pilots, with less re-learning required. While a good degree of consistency has been maintained within some aircraft families like the Airbus A320/A321 and A330/A340, great differences exist between manufacturer philosophies, making the transition across fleets not as easy as it might otherwise be, and thus creating the opportunity for negative transfer of training.

In the Windows type environment, consistency is less of a problem as standards are evolving to which many packages adhere. In the FMS environment however, no standard exists in terms of the function any particular key performs. In the office environment QWERTY keyboards have become standard. For FMS CDUs however, there is no standard position for letter keys. The key that is used for navigation on one FMS CDU is used for flight planning on another CDU (Singer, 1999). This lack of consistency and compatibility is one of the major difficulties for pilots using the current generation of FMS CDUs.

Error Prevention & Correction

In discussing Error Prevention & Correction, Nielsen (1993) argued that errors with serious consequences could be reduced in frequency by asking users to confirm their actions. Although not present in many FMS CDUs, the "execute" key on the FMS provides a useful example of error prevention since the pilot is given the opportunity to review inputs prior to those inputs being processed by the FMS. Evidence from the literature suggests that recovery from errors when using the FMS CDU leaves much to be desired with error messages such as "Invalid Entry" providing the pilot with no guidance on how to correct their erroneous data inputs. Marshall, Nelson & Gardiner (1987) stated that in systems where there may be dangerous outcomes, tasks should not be completely automated and a disruptive element should be introduced to ensure that users think carefully about what they are doing.

Appropriate Functionality

The Appropriate Functionality requirement of a usable system refers to the requirement for the system to meet the requirements of its users in carrying out tasks. In the FMS CDU domain, this requirement would mean that in conducting a particular task the information required by users should be contained within the CDU page currently being used or in pages that can be accessed easily. Evidence from the literature (Singer, 1999) suggests that FMS CDUs are deficient in terms of Appropriate Functionality. It has

been suggested that FMSs may have excessive functionality, and in addition may have too many methods or modes of performing some tasks (Sarter & Woods, 1994).

Effectiveness

The ISO 9241 Part 11 definition of effectiveness refers *to "the accuracy and completeness with which user's achieve specified goals"* (p.2). Evidence of a lack of effectiveness in use of FMS CDUs, might be a reduced number of sub-tasks being performed using the CDU, or a lower level of flight deck automation being employed. The reduced effectiveness of the FMS CDU in the descent and approach phase of flight is evident from the fact that many pilots greatly restrict their usage of the FMS CDU below 10,000' by choosing to fly the aircraft manually or using the autopilot. There is thus clear evidence that the FMS CDU is not as effective as it might be (Hughes, 1995).

Efficiency

In a safety critical environment such as the flight deck, a trade-off exists between efficiency and error prevention. Extra steps involved in the performance of a task may reduce the number of operator errors by providing additional opportunities for the operator to notice input errors. The extra steps may, however, be tedious or may add additional workload and cause the user frustration as they attempt to perform a task as quickly as possible. FMSs differ from conventional PCs in the dynamic environment in which they are used. Reprogramming the FMS to make runway changes during the approach phase of flight, for example, is known to be an inefficient task that has the potential for errors and is a source of high workload for pilots (Endsley & Strauch, 1997). Increasing the number of steps required to perform a task may however only increase workload in an already busy phase of flight. The use of stored flight plans, for example, can be considered an efficient means of reducing data entry in flight in this dynamic environment. From a psychological perspective, however, there may be problems associated with the execution of pre-programmed plans in that, for example, errors may be more difficult to identify (Singer, 1999).

Feedback

Norman (1989) argued that a change in the type of system feedback or a complete loss of feedback was a problem associated with aviation automation in general. He went on to argue that "*Without appropriate feedback people were indeed out-of-the-loop. They may not know if their requests have been received, if the actions are being performed properly, or if problems are occurring*" (p.6). Norman (1989) and Endsley (1996) argued that this problem could be attributed to an erroneous belief by system designers that certain parameter information was no longer required by pilots once relevant functions had been automated.

The FMS CDU is again very different to a PC in that much of the feedback from inputs to the CDU is displayed outside the Control and Display Unit. For example, the pilot entered flight path on the CDU is continuously updated to reflect the current flight status, but this is displayed on the Horizontal Situation Indicator on the Primary Flight Display (PFD). Further, a typical feedback message from the CDU is "Invalid Entry", a

feedback message which neither assists the pilot in finding the correct input format nor informs the pilot of why the data entered are invalid (Sarter & Woods, 1992).

Explicitness

In discussing software interfaces for aviation systems, Williges, Williges & Fainter (1988) argued that it was important to "*assist the users in developing a conceptual representation of the structure of the systems so that they can navigate through the interface*" (p.466). This is particularly pertinent when the size of an interface is limited by the space available, such as is the case with the FMS CDU in a flight deck.

In order for a system's logic to be clear to users, design processes should aim to reflect users' mental model of the system. Evidence from the literature (e.g. Sarter & Woods, 1992) suggests that pilots' have gaps in their mental models of FMSs. Sarter & Woods (1994) suggested that that problems associated with these gaps occurred primarily in the context of non-routine, time-critical situations and recommended that pilots should be further supported in their mental model of the FMS.

Within the FMS CDU page structure, it is not always possible for the pilot to immediately determine how they got to the sub-menu they have navigated to, or how to go from the current page to another unrelated page. Providing an overview of the architecture would make it immediately explicit to the pilot where in the system architecture they currently were. Finally, it should be clear to the user how changes made in one part of a system affect other parts of the system. This latter component may be pertinent to FMS CDU usability since inconsistencies in data propagation through the CDU are known to be a source of "*surprises*" to pilots (Sarter & Woods, 1992, p.307).

Explicitness is particularly important to the type of system being evaluated within this thesis. Due to the small display space available on the CDU (typically 14 rows of text) and the large amount of navigation and performance data that needs to be available to the pilot, the menu structures of the FMS CDU have tended to require several layers. While there are usually only eight or ten top-level menus visible to the pilot, there can be up to five pages of functions buried underneath that top level. This has led to the recognition of the system architecture being narrow (on the top level) but deep (with many pages hidden underneath). Many pages of the CDU architecture may have to be navigated through, in order to access required information. It is even more critical for this type of system to have an interface that is transparent. Sarter & Woods (1994) argued that "*the opaque interface between pilots and automation makes it difficult for pilots to track the state and activity of the automation*" (p.27).

Flexibility

The flexibility dimension of usability discussed in Section 1.8 highlighted the need for user interfaces to be adaptable to individual users. FMS interfaces and the various autoflight functions available to pilots offer a high degree of flexibility in terms of the level of automation they choose to use. In light of the safety-critical environment in which the FMS CDU is operated, it might be that such individualisation or

customisation would be less desirable for flight deck components than it would be for PCs. This is again a particularly important consideration in light of the workload peaks and troughs known to be present in flying an aircraft. Further, since there are multiple methods to achieve the same goal when flying an aircraft, it is difficult to properly instigate the operating procedures of monitor / cross-monitor & challenge responses when there is more than one way of performing a task.

Sarter & Woods (1994) reported that flexibility itself could impose cognitive demands on pilots and discussed the trade-off between providing additional flexibility for users and imposing additional workload and argued that for some tasks on the flight decks there is an *"overwhelming number of possible methods to do the job"* (p.312). They concluded that deciding which of these methods was the most appropriate, for a particular task, and in a particular context, created additional mental workload for the pilots. Thus the level of flexibility incorporated into the FMS CDU raises different issues when compared to the considerations required for desktop computers.

Workload

Operator workload levels are particularly important in the flight deck environment where errors can threaten safety. A task demanding too little effort can result in lower efficiency since it leads to boredom and a lack of vigilance, which results in lower effectiveness. Excessive mental workload can have negative influences on all aspects of information processing (BASI, 1998), and may cause information to be missed resulting in errors, thus also reducing effectiveness (ISO 9241 Part 11). Automation may increase workload through being difficult to use, difficult to understand, and difficult to predict (Sarter & Woods, 1992). Parasuraman & Riley (1997) reported that one of the fundamental reasons for introducing automation into complex systems was to decrease the chance of human error by reducing the operator's mental workload. Research has, however, shown that this does not always occur (Sarter, Woods & Billings, 1997). Wiener (1989) coined the term *"Clumsy Automation"* to denote automation that reduced workload when it was already low, but required increased attention and interaction when workload was already high (p.6).

The type of high workload environment in which pilots interact with the FMS was evident from the Air Accidents Investigation Branch report into the Kegworth accident in Leicestershire. The crew had tried to reprogram the FMS for landing, while simultaneously informing their company of their problems and intentions, responding to ATC height and heading instructions, and obtaining weather information for East Midlands Airport (AAIB, 1990). Endsley & Strauch (1997) argued that pilots *"cannot be relied upon to integrate multiple, sometimes dissonant sets of information in time-critical, high stress situations"* (p.881).

Wickens (1992) argued that one of the ways in which automation increased workload was by increasing from one to three the number of decisions that an operator must take in making a diagnosis of a system malfunction. Parasuraman, Mouloua, Molloy & Hilburn (1996) illustrated how automation could increase workload using the example of an automated system that monitors an aircraft's doors during flight. Parasuraman et al argued that in the event of a failure, the flight crew would have to decide whether the

door was open, whether a failure had occurred with the automated monitor, or whether it was the display indicator of the automated system that had malfunctioned. Foreman (1996) concurred that the FMS had redistributed workload on the flight deck resulting in workload being greatly increased when changes occurred that required major FMS reprogramming. Wiener (1988) reported that flight deck workload must be reduced since this would achieve three important objectives.

- Pilots would be relieved of routine, manual tasks such as mental computation and would be free to supervise the flight and perform optimally in an emergency.
- At lower altitudes, and particularly in the terminal area, pilots should spend less time with their "head in the cockpit" and spend more time scanning for aircraft.
- Since two person crews now fly wide-body aircraft, a task that was traditionally performed by three person crews, the workload associated with that task must be reduced.

Workload is therefore potentially an important component of usability in the flight deck, although it has not been considered in depth by many who have written on subject of usability (e.g. Jordan, 1998; Shneiderman, 1992).

Help Facilities

While FMS manufacturers do offer the facility to have some basic help functionality incorporated into CDUs, most airlines do not purchase this capability. This may be due to the additional financial commitment involved in buying the additional software. Further, since the FMS CDU does not operate in a Windows type environment, accessing help pages would require pilots to navigate to a help page, and this action could potentially interrupt the pilot's primary task.

Since for the moment the reference manual is destined to be the most common "help facility" on-board the flight deck, the main considerations should be focused on ensuring that the technical language used is appropriate for the pilot, and is not written for the engineer or computer programmer. Further, the reference manual should be structured such that the information it contains is easily navigated through, read and learned (Mayhew, 1992).

Attitude

While it is known that usability problems exist in the use of FMS CDUs, pilot attitudes towards FMSs are generally favourable. A possible reason for this is the users' awareness of the additional workload that would be involved in flying passenger aircraft without this piece of equipment.

1.10 Benefits of Assessing Usability

Since manufacturers are now striving to produce usable products, the measurement of usability is becoming an important component of modern hardware and software product engineering. Lewis (1995), however, argued that the measurement of usability was "*particularly difficult because it is not a unidimensional product or user characteristic, but emerges as a multidimensional characteristic in the context of users performing tasks with a product in a specific environment*" (p.58).

There are many potential benefits to assessing usability such as those outlined by Bevan (1995).

- Making usability an objective criterion when specifying the requirements for a product creates an incentive to put resources into designing towards a usable end system.
- Measuring a product against usability objectives makes it possible to estimate how much additional work is required in order to meet the objectives.
- Benchmarks may be established such that comparison may be made between alternative configurations of a product e.g. between differing versions of a product.

Industry budgets for usability engineering are growing. Nielsen (1994) reported that in 1971, three percent of project funds was allocated to usability engineering. In 1993, a survey determined that the median share of budget costs allocated to usability engineering was six percent.

Companies benefit by selling more of their product, selling other products, enhancing the company's reputation, saving money on internal products, reducing support costs, reducing training costs, reducing the need for updates and maintenance releases and making documentation and training easier to develop (Dumas & Redish, 1993). Remington (1998), for example, cited a survey by a VCR manufacturer that found that 75% of consumers had never used the advanced VCR features. Difficult to use products increase manufacturer costs through training and subsequent support. Users benefit from interaction with a usable systems through a reduced learning curve, and being able to use more functionality with less effort. So while pilots, for example, do not buy the FMSs they work with, they can still incur a cost of employing increased effort when required to work with systems which lack a reasonable level of usability.

Where safety is concerned the costs associated with poor usability can be even greater. Taking the example of control rooms for nuclear power plants, the Three Mile Island accident in the USA was blamed on human error. Norman (1988), however, argued that major design faults caused the misdiagnosis to occur in the first place with one of the crucial instruments which had to be checked being located on the back of a control panel. On the flight deck, as previously stated, the costs associated with interacting with a system that is difficult to use can be much higher as evidenced from the numerous incidents and accidents caused, at least in part, by sub-optimal design.

1.11 Methods for Assessing Usability

Since the aim of this research was to produce a usability assessment tool, it was necessary to review the numerous methods of assessing usability that have been proposed in the literature, and to determine the most appropriate methodology for assessing the FMS CDU.

Two of the major considerations in determining which methodology is most appropriate in assessing a system's usability are (i) who does the assessment of the system; and (ii) where in the product's lifecycle the evaluation is to be conducted. Typically the assessment of a system's usability may be conducted by either a typical system end user or by a HCI expert. The points in a product's lifecycle that the assessment can take place might be at the concept, prototype or finished product stage.

There are four broad categories of usability assessment methods currently in use. These are subjective evaluation methods; objective evaluation methods; expert evaluation methods; and experimental evaluation methods (Oppermann & Reiterer, 1997). Each of these methods will be outlined briefly in the following sub-sections. The methods described here are not a complete list, but rather represent some of the more frequently used assessment techniques.

1.11.1 Subjective Evaluation Methods

Subjective evaluation methods are based directly on users' judgements. Real system end users are involved in the assessment but the evaluation is not conducted while the users are interacting with the system (Macleod, 1992). Frequently used subjective evaluation methods are questionnaires, interviews and focus groups.

The main advantage of questionnaires is that they permit the collection of opinions from real end users, rather than from theory or experts. Additionally, due to the relatively small cost involved in using and administering questionnaires, the technique permits opinions to be gathered from a large sample of users.

Some examples of existing usability questionnaires include Norman and Shneiderman's (1989) "Questionnaire for User Interaction Satisfaction (QUIS 5.0)" and the "Software Usability Measurement Inventory" developed within the Metrics for Usability Standards in Computing programme (MUSiC, 1992; Kirakowski & Corbett, 1993). The main disadvantage to using questionnaires as reported by Robson (1993), is that to be meaningful the questionnaire must be *"painfully constructed, with very clear and unambiguous instructions and careful wording of questions"* (p.243).

Interviews are another useful assessment method in that as before real end users, rather than HCI expert opinion is gained during the assessment. A further advantage is that interviews may be flexible and can therefore be shaped to the assessor's responses and be highly informative (Oppermann & Reiterer, 1997).

The major drawback to using this method is the time taken to conduct the interview and analyse its content. Conducting an interview takes a significant amount of planning and reasonable expertise on the part on the interviewer. The time and cost taken to conduct and analyse interviews therefore limits the number of participants that may be involved in the assessment.

The key distinguishing factor of focus group research from interviews is that it employs more than one user in a single assessment. The prime advantage of group discussions is that they help to determine the "*universality of comments*" (Shneiderman, 1992, p.484). However, the main disadvantage to conducting focus group research is the time and cost taken to conduct the interview and analyse the data. Although costs are likely to be lower for focus group research than for individual interviews, the analysis can still be very time consuming. A further disadvantage to group interviews is that it is sometimes difficult to pursue the views of any one participant (Oppermann & Reiterer, 1997).

1.11.2 Objective Evaluation Methods

Objective (or observational) evaluation methods are assessments that involve real end users interacting with working systems under the observation of experts. The systems used may be early prototypes or may be complete working systems. The observation method can range from very unstructured to very structured (Oppermann & Reiterer, 1997).

Using this assessment technique assessors are given a predetermined set of tasks to complete, and observations of the assessors' interactions with the system are made by experts, typically through direct observation, video recording, interaction monitoring or co-operative evaluation. Frequently the observer can view the system being used in its working context. For example, an observer could watch an assessor using a new word-processing tool in their office, and could therefore learn much about the problems the user encountered.

Video recording is conducted in the same manner as direct observation but the use of cameras permit the recording of the user's interactions and analyses of the video to be conducted separately. Interaction monitoring techniques use data loggers or capture tools to automatically record a user's interactions with the system being assessed. For example, the menu selections, mouse clicks or information displayed to a user could be recorded to determine difficulties encountered in finding a particular menu option.

Finally, co-operative evaluation extends the previously outlined observation methods through involving the user in the assessment process by asking about the problems encountered either at the point the problem occurs or retrospectively. Methods that fall into this category include retrospective testing and the "*Thinking Aloud*" method in which users perform a task while verbalising their thoughts (Nielsen, 1993, p.18).

The advantages to using objective evaluation methods are that they are user-based, have a broad scope, and can be used to examine many work tasks. The main disadvantages to using objective methods are that they can be difficult to use early in a system's development, and also may intrude on the user's performance of a task.

In traditional assessments that have examined the usability of desktop-based systems, the issues associated with intrusive techniques have been less consequential. However, when examining the usability of an FMS CDU, observational techniques which involve the presence of an expert evaluator could be intrusive, and airlines would be extremely reluctant to permit an evaluator to make video-recordings of a pilot's FMS interactions while flying a sector. Similarly, it is highly unlikely that the automatic logging of live data would be permitted by an airline even if this was physically possible.

1.11.3 Expert Evaluation Methods

Expert evaluation methods use an expert's knowledge of human factors and HCI to make judgements concerning a system's usability (Sweeney, Maguire & Shackel, 1993). These methods tend to be subjective in that they rely on the expert's judgement, but are objective in that if properly conducted, the assessment is based on a series of precise rules. Some examples of expert evaluation methods include usability inspections, specialist expert walkthroughs, cognitive walkthroughs, checklists and heuristics.

Usability Inspection Methods is the term used for a generic group of methods which involve evaluators inspecting and making judgements about particular dimensions of a user interface. The inspectors may be usability experts but they may also be designers with a particular expertise in the type of interface being developed (Nielsen, 1994a). Oppermann & Reiterer (1997) argued that *"The defining characteristic of usability inspection is the reliance on judgement as a source of evaluative feedback on specific elements of a user interface"* (p.241).

Specialist reports involve human factors experts using their expertise to assess how usable the system being evaluated might be for the end users. The limitation associated with this sort of method is that the assessment will only be as good as the assessor's experience. A more structured specialist report, such as an expert walkthrough in which the expert performs a sample of the tasks that the end user might perform, has increased validity, but again depends on the expert being able to predict what the end user might find difficult or easy.

Cognitive walkthroughs are again conducted by experts but these are theoretically founded on learning by exploration, and use mock-ups or prototypes, task scenarios and make assumptions about the end users and the context in which they would use the system explicit.

Usability checklists are usually formulated from HCI guidelines and they therefore assess how well a system's interface conforms to the guidelines from which the checklist was derived. Checklists are probably the most frequently used method of assessing a system's usability because they are perceived as being easy and quick to use. They do, however, require some expert knowledge both in applying and interpreting the guidelines. Guidelines are usually specific to a particular platform and are therefore useful when an assessment is being conducted to examine a system's conformance with a *"specific look and feel"* (Oppermann & Reiterer, 1997, p.241). A useful guide when

designing a Graphical User Interface (GUI), for example, is Microsoft's User Interface Style Guide (1995).

Heuristic evaluation, described by Nielsen (1993), is another method based on guidelines that can be used to assess the usability of an interface. Nielsen (1993) put forward some very broad characteristics of usable interfaces as a basis for systematic inspection. These heuristics stated that usable systems should have simple and natural dialogue, should speak the user's language, minimise user memory load, be consistent, provide feedback, have clearly marked exits, have short cuts, have good error messages, prevent errors and provide help and documentation. The goal of heuristic evaluation is to find the usability problems in an interface using these ten guidelines.

While there are many examples of useful guidelines to assist designers in producing usable products (e.g. Ravden & Johnson, 1989), these guidelines are imprecise and not universally applicable. They are instead a useful guide for designers which may be used by experts to make quick assessments (Bevan, 1995).

It is also important to note the distinction between questionnaires and checklists. Checklists are simply a list of statements of features that may or may not be desirable properties for a system to demonstrate. Checklists are not, for example, amenable to scaling and cannot be combined to form any measure of usability (Kirakowski, 2001).

1.11.4 Experimental Evaluation Methods

Experimental evaluation methods involve the design of controlled empirical experiments to test hypotheses about various aspects of an interface design. Many large organisations like IBM, for example, build usability laboratories in which prototypes under development may be empirically assessed. The advantages to using experimental studies are that rigorously designed and conducted user-based evaluations may be performed, and reliable and valid results can be produced (Oppermann & Reiterer, 1997). There are, however, disadvantages associated with such experiments in that they lose the environmental or contextual effects that impact upon a user when using an interface in the "real world". An FMS CDU, for example, which performed very well in the laboratory may be disastrous when used on the flight deck under conditions of vibration during turbulence, or when its use is only one of a number of tasks being conducted simultaneously. It is extremely difficult to design experiments which control for all the "real world" independent variables.

1.11.5 Selecting a Usability Assessment Method

Each of the four broad categories of usability assessment method has relative advantages and disadvantages as discussed. The most important characteristics of these methods along with their relative advantages and disadvantages were summarised by Oppermann & Reiterer (1997, p.243) and are displayed in Table 1-1.

Table 1-1 - Characteristics of Usability Assessment Methods (Oppermann & Reiterer, 1997, p.243)

Type of Method	Subjective	Objective	Expert	Experimental
Timing of use in product development	Prototype or later	Prototype or later	Any stage	Any stage
User-based	Yes	Yes	No	Yes
Scope of evaluation	Broad	Broad	Broad	Very Narrow
Type of data:				
Quantitative	Yes	Yes	No	Yes
Qualitative	Yes	Yes	Yes	Yes
Diagnostic	Yes	Yes	Yes	Narrowly
Expertise Required	Medium	Medium	High	High
Place	In-house	In-house or laboratory	In-house or at expert's workplace	Laboratory
Costs				
Time	Low/medium	Medium/high	Low	High
Money	Low/medium	Medium/high	low	High
Main Advantages	Quick	User-based	Low cost	User-based
	User based	Performance on work tasks	Quick	Rigorous
	User's view	Diagnostic	Diagnostic	Can produce results with reliability and validity
	Diagnostic	Broad scope	Broad scope	Narrow scope
	Broad scope	Brings designers and users together		High cost
Main Disadvantages	Less effective early in the design cycle	Less effective early in the design cycle	Not user based	Requires careful design by scientists
	Only retrospective views of usability	May be intrusive on user	Dependent on quality of expert	
			Questionable reliability	

Stanton & Baber (1996) produced an "*Heuristic Chart*" to assist in the selection of the most appropriate method for conducting usability evaluations in a given set of constraints (Stanton & Baber, 1996, p.46) and this is shown in Figure 1-9.

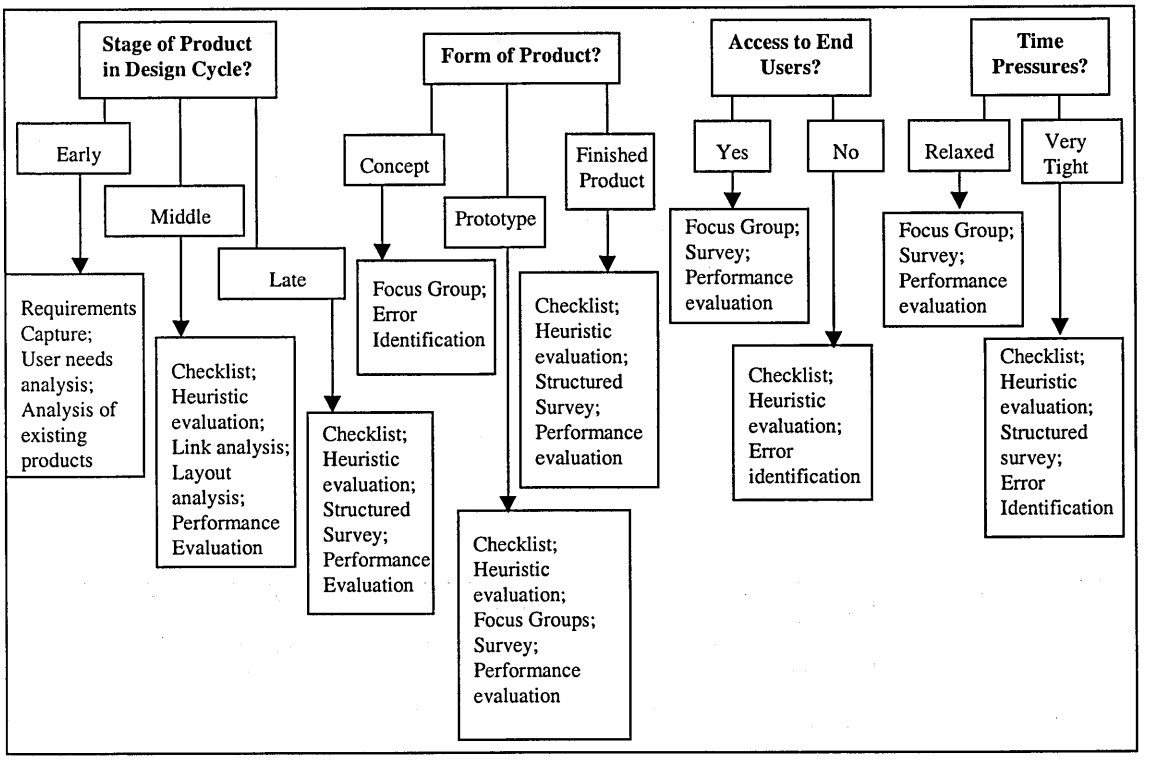


Figure 1-9 - Heuristic Chart for Selection of Usability Methods and Techniques (Stanton & Baber, 1996, p.46)

1.11.6 Methodological Constraints and Considerations

Recognising the limitations of the flight deck environment in which the FMS CDU is situated; that the product to be evaluated is finished in terms of its design lifecycle; the constraints under which airlines operate (security restrictions) and research cost considerations, many of the outlined methods would not be feasible. A subjective measure was the most realistic possibility.

Further, having reviewed the charts produced by Oppermann & Reiterer (1997) and Stanton & Baber (1996), the most appropriate assessment type for examining the usability of an FMS CDU was determined to be the subjective class of methods.

A review of existing usability checklists, guidelines and subjective measures revealed that most usability assessment tools had been derived for use with desktop computers. Even those usability scales that claimed to be generic were not found to be applicable to the flight deck domain for a number of reasons. The Software Usability Measurement Inventory (SUMI), developed by Kirakowski & Corbett (1993) within the Metrics for Usability Standards in Computing programme, for example, demonstrated good psychometric properties but was developed using office-based applications such as word-processing and spreadsheet packages. Of the fifty items in the questionnaire numerous were not directly applicable to the flight deck.

Examples included:

- "Getting data files in and out of the system is not easy."
- "The software has a very attractive presentation."
- "I would not like to use this software every day."
- "I would recommend this software to my colleagues."
- "It takes too long to learn the software commands."

Since SUMI was not sufficiently generic for use with even interactive desktop applications, its authors produced a separate usability measure for Measuring the Usability of Multimedia Systems (MUMMS). Similarly, it was found that SUMI was not directly applicable when examining the usability of World Wide Web (WWW) sites. In this case, research indicated that although parts of SUMI were applicable, a new sub-scale "excitement" was required. The researchers therefore developed a new usability tool the Web Analysis MeasureMent Inventory (WAMMI) (Kirakowski & Cierlik, 1998).

A review of the Purdue Usability Testing Questionnaire (PUTQ) developed by Lin, Choong & Salvendy (1997) indicated that it too was limited in the same dimensions as SUMI (Kirakowski & Corbett, 1993), in that it too contained items that were not pertinent to the FMS CDU. Such items included "Are users allowed to customise windows?"; "Is the control of cursor compatible with movement?" and "Are symbols for graphic data standard?".

While the usability checklist developed by Ravden & Johnson (1989) did contain many generic items that were suitable for adapting to the flight deck context, it also contained many items that were not pertinent to the study being undertaken. The checklist item, "Are graphical displays compatible with the user's view of what they are representing?", for example, was not applicable to the FMS CDU since the FMS CDU does not have a graphically-oriented interface. Further, since Ravden & Johnson had produced a checklist of considerations for designers rather than developing usability scales, no psychometric properties were associated with the checklist.

Using generic questionnaires would also have meant that the terms "flight deck" or "FMS" would not have been mentioned, and respondents would have faced questions concerned with, for example, the windows environment which were clearly not applicable to the FMS CDU being evaluated. Finally, many of the existing usability questionnaires were very long and would have appeared very unwieldy to participants. This may therefore have further threatened the response rate achieved by the questionnaire which would be so vital to a study of this nature (Robson, 1993).

Certification authorities such as the FAA have outlined the requirement for human factors tools which are both reliable and valid (FAA, 1996). Reliable tools indicate the extent to which "*individual differences in scores are attributable to "true" differences in the characteristics under consideration and the extent to which they are attributable to chance errors*" (Anastasi, 1988, p.109). The validity of a measure is concerned with "*what*" the test is measuring and how well it does so (Anastasi, 1988, p.139). The

concepts of reliability and validity are critical to this thesis and are discussed in more depth in Section 1.12. After reviewing other unsuitable usability tools and guidelines, it became evident that no reliable, valid tool existed which could be readily applied to FMS CDUs and that the usability scale methodology had the best characteristics of those potentially available. It is important to note that other methods were considered in determining which approach would be most appropriate for eliciting the stable properties of FMS usability for the purposes of certification. A summary of the other approaches considered has been included at Appendix A, along with an analysis of their applicability to this research.

1.11.7 Justification for the approach

The justification for the approach adopted in this thesis was founded on both published literature and a real world requirement for reliable and valid tools that could be used in the flight deck certification process. It is important to note that it was not the goal of this research to produce a tool that could be used as a standalone method of certificating FMSs from a usability perspective. The purpose of the research was rather to investigate whether such a tool could be useful in a multi-method approach to flight deck certification.

Research in the literature (Kirakowski, 2001) demonstrated that it was possible to develop a generic paper-based usability questionnaire for Personal Computers (SUMI), which was standardised against usability profiles of over 200 different types of applications (e.g. work processors, spreadsheets, graphics packages). Field trials conducted by the questionnaire developers and by several independent research organisations throughout Europe had shown the questionnaire to be reliable and valid (Porteous, Kirakowski & Corbett, 1993). The terminology used in the questionnaire in addition to differences between desktop interfaces and FMS interfaces meant that the tool could not be applied directly to FMSs. What was not known was whether a generic tool could be developed for FMSs or how useful such a tool would be to certification authorities and FMS manufacturers.

A further consideration in developing a methodology that could contribute to FMS development and certification processes was one of economic viability. The argument for significant financial investment in usability engineering can be a difficult one to make. This is largely owing to the benefits of increased usability frequently not being visible until after a product's release (Nielsen, 1993). It is true to say that there is at some level a contradiction throughout the whole aviation industry between that which is safe and that which is economical. Methods that are developed to feed into the certification process need to be economically viable. While large-scale user trials might allow more flexibility in terms of the information elicited during product development, the costs and organisational logistics associated with such trials can make them less attractive from an economic perspective. The usability scales that this thesis sought to develop could encourage manufacturers to consider investing their resources in usability by allowing the involvement of more end users. The proposed methodology could also be advantageous in simultaneously making it possible for manufacturers to focus their

attention on the problem areas identified by usability scale ratings when conducting expensive experimental trials.

1.12 Scale Properties

DeVellis (1991) defined scales as being "*measurements that are collections of items intended to reveal levels of theoretical variables, not readily observable by direct means*" (p.8). DeVellis argued that all scales "*were not created equally*" (p.9) and that it was necessary to contrast the properties of a scale and an index. Bollen (1989) referred to a scale as being that which contained items whose values were due to an underlying construct. In contrast, an index, for example, is constituted of cause indicators or items that determine the construct level. DeVellis (1991) argued that for a scale measuring an underlying construct such as optimism, responses to items are caused by that construct i.e. more optimism causes higher item scores. However, on a socio-economic status index, for example, based partly on education, having more education is not caused by higher socio-economic status, it influences it.

The underlying phenomenon, construct or latent variable being examined in this research is FMS CDU usability and it is implied that a certain relationship exists between the latent and the items that constitute it. The latent variable is viewed as a cause of the item's score. When a number of items caused by the same latent are examined, their relationships to each other may also be examined.

Reliability and validity are essential properties of measurement scales. In order to be considered useful the scales elicited from this series of studies would have to demonstrate that they hold the properties of reliability and validity.

1.12.1 Reliability

Reliability is a fundamental issue in psychological measurement. It is concerned with how much a variable influences a set of items (Ghiselli, Campbell & Zedeck, 1981). DeVellis (1991) defined scale reliability as the "*proportion of variance attributable to the true score of the latent variable*" (p.24). The concept of reliability is the basis for the calculation of the error of measurement, which refers to the range of fluctuation likely to occur in an individual score in a measure as a result of irrelevant factors or chance.

Test-retest reliability is one of the oldest methods of estimating the reliability of scale scores. It involves repeating a test or scale administration on a second occasion and establishing the correlation between the scores obtained by the same person on the two administrations. However, there are several reasons why test-retest correlations may be inaccurate. The attribute or characteristics being measured may have changed between the first test and the re-test. There may be carry-over effects between administrations such that when people are re-tested they may remember their original answers or the experience of taking the test itself could have changed a person's true score. In addition to the theoretical problems associated with the test-retest method, there are practical

limitations to this method of estimating reliability. Since two administrations are required, and since testing can be a time-consuming and expensive process, "*re-testing solely for the purpose of estimating reliability may be impractical*" (Murphy & Davidshofer, 1991, p.79). In applied research, such as that described in this thesis gaining access to the same participants within an external organization on two occasions is frequently impractical.

The type of reliability of most concern in scale development is the internal consistency within scale items. The internal consistency of a scale refers to the extent to which its items are intercorrelated. High inter-correlations imply that the items are measuring the same construct. High inter-correlations between scale items further imply a strong relationship between the scale items and the latent construct. The internal consistency of a scale is most usually measured by Cronbach's (1951) coefficient alpha ($\propto$).

1.12.2 Validity

Validity is concerned with whether a particular variable is the underlying cause of item covariation (DeVellis, 1991). Validity is inferred from the way in which a test or scale was constructed, its ability to predict events or its relationship to other measures of the construct being examined. Validity should therefore be built into a measure from the outset, rather than being limited to the final stages of test or scale development. Six types of validity are of prime concern in scale development. These are face validity, content validity, criterion validity, construct validity, convergent validity and discriminant validity.

Face validity refers to the extent to which a test appears superficially to measure that which it claims to measure (Anastasi, 1988). For face validity to exist a test should, for example, appear plausible to respondents and to untrained personnel who decide to use it. Face validity is not sufficient in itself, although it does make a test more acceptable to potential respondents (Robson, 1993). It should, however, never be regarded as a substitute for objectively determined validity. Anastasi (1988) argued that face validity could be improved by merely editing test items.

Content validation is concerned with determining whether the content of a test is representative of the behaviour or the domain to be measured (Anastasi, 1988). In order to achieve content validity the subject matter being examined should be fully described in advance, rather than after a test has been developed. For the type of scale development being undertaken within this research, content validity may be built into a test by choosing appropriate items, through systematic review of relevant literature, and through consultation with Subject Matter Experts (SMEs). Content validation should allow a representative sample of the domain skills, and ratings of the scales should be reasonably free from the impact of irrelevant variables (Anastasi, 1988). A scale is said to have content validity when the items it contains are a randomly chosen subset of the universe of possible appropriate items.

Criterion or predictive validity is more of a practical issue than a scientific issue because it is concerned with predicting a process rather than understanding it. Criterion validation procedures involve predicting an individual's performance in specified

activities. In order to achieve this, performance on a test, for example, is compared against a criterion, which is a direct and independent measure of what the test is designed to predict. The most useful measure of criterion validity, as reported by Ghiselli et al (1981) and DeVellis (1991), is the rate of correctly classifying the category of the criterion based on the predictor categories.

Construct validity refers to the extent to which a test measures a theoretical construct, the extent to which an assessment is measuring something meaningful. Construct validation requires the gradual *"accumulation of information from a variety of sources"* (Anastasi, 1988, p.153). The sources from which construct validity evidence is derived may include correlations with other tests, factor analysis and convergent and discriminant validation. Guion (1977) argued that since test creators always have a theory or construct that they are trying to assess, all test validation is necessarily construct validation.

Correlations with existing valid tests may be cited as evidence of construct validity, if the correlations are moderately high. If the correlations with other tests are too high, then the new test might be seen as unnecessary duplication. Low correlations on the other hand would not ensure validity either.

Factor analysis, which examines the interrelations in a set of data, is very useful for demonstrating construct validation (Anastasi, 1988). During the process of factor analysis an initial number of variables can be reduced to a smaller number of factors or common traits. Once the factors have been determined, the factorial composition of a test may be described and the factorial validity determined.

Anastasi (1988) argued that *"in order to demonstrate construct validity, it must be shown that a test correlates highly with other variables with which it should theoretically correlate, but also should not correlate significantly with variables from which it should differ"* (p.156). The latter two forms of evidence are known as convergent and discriminant validity. Campbell & Fiske (1959) called convergent validity the correlation of a measure with other variables with which it should theoretically correlate. In order to demonstrate discriminant validity, Anastasi (1988) argued, evidence should be provided to show that the measure discriminates between that which it should theoretically discriminate but does not discriminate between that which it should not discriminate.

1.12.3 Properties of Usability Scales for Certification

Usability scales for certification should represent a means of eliciting reliable and valid data from a wider sample of line pilots than would otherwise be possible. For usability scales to be appropriate for different FMSs and aircraft they would need to be able to access the generic usability characteristics associated with FMSs. If generic, stable usability factors could be elicited using a questionnaire, then it could be possible to make assessments of a system's usability in different operational contexts, for example, in different flight decks. This research therefore sought to elicit the generic stable

usability attributes associated with FMSs. It was an acknowledged real world limitation that it would not be possible to separate the FMS from its flight deck operating context.

Since usability is a multidimensional concept, usability scales should be capable of eliciting a multidimensional profile of usability for the system under assessment. In terms of developing the scales, all of the usability attributes cited in the literature would need to be assessed for their applicability and no usability component could be omitted a priori.

The most distinguishing point to note when developing usability scales specifically for certification, was that certification as it stands today is concerned with measuring aspects of the machine (in this case the FMS) and not the end user. This research did not seek to propose that a series of paper-based scales should be used as a sole method of certificating the usability of flight deck components. Rather, it was hypothesised that a series of usability scales designed specifically for a flight deck component such as the Flight Management System, and completed by line pilots, would be a useful adjunct to the performance, mental workload and effectiveness data frequently elicited during system development.

1.12.4 Summary of potential benefits in a scale-based approach to usability in certification

The implicit beliefs concerned with the application of psychometrically robust scales to the assessment of usability for flight deck design and certification are outlined in this section.

- The usability scale approach offered the best characteristics of the available methodologies as discussed in Section 1.11 and Appendix A.
- Psychometrically robust scales could be used to access the stable usability characteristics associated with FMSs and allow comparisons to be made across FMS types and variants. It would be important, therefore, that all potential FMS usability attributes be included during the questionnaire development process and that no usability dimension would be viewed as less important, or eliminated, a priori.
- The scale-based methodology could be less biased than other approaches since, for example, an interviewer would not be directing the questions or responses.
- Usability scales could be a pragmatic approach to accessing more line pilots. The methodology could provide a cost effective method of involving more typical end users in a certification process which currently lacks end user involvement and which is intended to be oriented around pilots who are "*as representative as possible of end users*" (Amalberti & Wibaux, 2000, p.361). Any methodology which would increase human factors involvement in the design, development and certification of flight deck components should be advantageous.

- Reliable and valid usability scales could be used to provide guidance as to where usability problems lie allowing more cost and time intensive usability methodologies to be focused on problem areas. Through identifying which attributes of a product are rated exceptionally it could be possible to interview and observe users to find out why they are rating particular aspects of the FMS exceptionally. This would additionally be useful for focusing system development and evaluation work. The questionnaire methodology could also assist in the identification of non-detected system weaknesses and in terms of giving a rationale to pilots' problems in relation to system design.
- Nielsen (1993) argued that questionnaires were probably "*the only usability method that makes extensive* (user population) *coverage feasible, with the ensuing possibility for discovering differences between various user categories*" (p. 211). The usability scale approach could permit the inclusion of a range of end users in terms of culture, rank, age and operational experience.
- A series of reliable and valid usability scales could make it possible to generate a product's profile of stable or general FMS Usability characteristics. If it was possible to elicit FMS usability factors and subsequently FMS usability scales – then a product usability profile could be determined.
- The scales proposed for development in this research could also be amenable to Item Consensual Analysis which could be used in, for example, the manufacturing and certification of FMSs to list scale items on which the FMS being rated was significantly better or worse than the standard of comparison.
- The questionnaire methodology could also be used to elicit qualitative information concerning problems encountered in operational use of FMS.

1.12.5 Summary of potential weaknesses in a scale-based approach to usability in certification

While it is clear that there are some advantages to using a scale-based approach to the assessment of usability, it is also acknowledged that there are also some disadvantages. The weaknesses associated with using a scale-based approach to usability in the certification process are discussed in this section.

Usability scales such as those proposed for development in this thesis are less flexible than, for example, experimental, interview or observational research approaches in terms of being able to elicit information concerning specific operator experiences and the environmental context of use. Since FMS-related accidents and incidents occur within a specific context of operation, this was an important element that had to be considered in determining the most appropriate approach to take for this study. It is a recognised limitation of the scale-based approach that specific flight contexts are not easily accessed.

While a usability questionnaire methodology could permit elicitation of qualitative information concerning problems encountered in operational use of FMS, it would not permit those problems to be probed further with the end user.

Although psychometrically robust scales could access issues relevant to aircraft accidents or incidents they would not, for example, necessarily be the most useful methodology in determining how usable a particular FMS was in time critical situations such as that which occurred prior to the Kegworth Accident (AAIB, 1990). Similarly, although accidents and incidents occur within an operational environment, usability scales are unlikely to be the optimal method of accessing the experiences of operators and the environmental context of use. Usability scales would only be useful to the extent that they could assist FMS manufacturers to focus their attention on aspects of the FMS that achieve particularly poor usability ratings.

Whilst a psychometrically robust questionnaire could be useful in identifying usability problems, it would not provide FMS manufacturers with design solutions. Qualitative methods such as interviewing typical end users, conducting focus group sessions with end users, and conducting iterative prototyping with alternative design configurations are more useful mechanisms for assisting with improved design solutions.

The use of psychometrically robust scales would not negate the necessity to conduct usability testing as a product is developed and integrated with other systems, although it could provide clues as to where usability problems lie. Where the research aim is to evaluate the overall usability of a product, then the data elicited by a questionnaire such as that proposed would need to be enhanced and complemented with performance, mental effort and effectiveness data (Kirakowski, 2001).

1.13 Study Aims

The literature outlined in this review has proposed that usability is not a unidimensional product or user characteristic, but rather is a multidimensional characteristic of the product with which a task is performed and the context in which that interaction occurs. Having identified the lack of end user involvement in the certification process, it was concluded that the proposed scale-based methodology was pragmatic in that it would allow greater end user involvement in a certification process that currently lacks end user involvement. In light of the fact that no existing measure of usability was suitable for evaluating the FMS CDU, the series of studies conducted in this research programme set about developing a series of usability scales for FMS CDUs.

The primary aim of the first study was to elicit the dimensions underlying FMS CDU usability. Using a new set of participants, the second study set out to verify the factor structure elicited during Study One and determine its validity and generalizability with the goal of developing usability scales. Since flight deck workload is known to vary across flight phase, typically being at its lowest during the cruise phase and at its highest during descent and approach to the terminal area, the second study also examined if the usability scales elicited were independent of workload changes. Finally, the purpose of the third study was to examine a large volume of qualitative data elicited from pilots during Studies One and Two in order to examine the problems associated with current FMS CDUs and to make a series of recommendations for the development of future FMS CDUs. The specific aims of each study are outlined in the following sub-sections.

Aim Study One: Elicit valid and internally consistent components of FMS CDU usability.

The first study was designed to:

1. Generate a pool of items concerning the usability of FMS CDUs through reviewing existing literature on the usability concerns surrounding Flight Management Systems and by conducting in-depth interviews with line pilots.

2. Produce a usability questionnaire that could be distributed internationally to a large sample of line pilots, across a range of fleets and FMS CDUs.

3. Conduct analyses on the data elicited with the goal of determining the factor structure of FMS CDU usability.

4. Examine the internal consistency and validity of the factor structure derived.

5. Generate qualitative data concerning the problems associated with current FMSs, and the improvements pilots would like to see in next generation FMSs.

Aim Study Two: To validate the usability factor structure derived and to produce a series of valid and internally consistent FMS CDU usability scales.

The second study was designed to:

1. Verify the usability factor structure elicited in Study One in order to demonstrate that it was replicable and generalizable.

2. Produce internally consistent and valid usability *scales* from the factors.

3. Demonstrate that the scales developed could discriminate between FMS CDU types and aircraft types (i.e. demonstrate discriminant validity).

4. Demonstrate that respondents' group membership could be predicted on the basis of their scale ratings (i.e. demonstrate criterion validity).

5. Demonstrate that participants' scale ratings were dependent on the FMS CDU being assessed and the flight deck in which it resided, and were independent of pilot workload, flight phase and user characteristics (i.e. demonstrate convergent validity).

6. Determine which of the usability scales were significant predictors of self-reported satisfaction in using the FMS CDU.

7. Generate qualitative data concerning the problems associated with current FMSs, and the improvements pilots would like to see in next generation FMSs.

Aim Study Three: To produce a framework of the problems associated with FMS CDUs from an end user perspective, to categorise the improvements that pilots would like to see to the CDUs on future flight decks, and to aid interpretation of the usability scales produced.

The third study was therefore designed to:

1. Analyse the participants' qualitative comments to produce a framework of the problems associated with FMS CDUs from the end user perspective.

2. Analyse the participants' qualitative comments to produce categories of improvements that pilots would like to see to the CDUs on future flight decks.

3. Provide evidence to aid interpretation of the usability scales developed during the first two studies and of their validity.

4. Generate recommendations from a human factors perspective which could be used by certification authorities and design teams generating future FMSs.

2 STUDY ONE INTRODUCTION

2.1 Purpose

The purpose of Study One was to elicit a reliable and valid factor structure of FMS CDU usability that could subsequently be used to develop usability scales. Although there is no universally accepted definition of the hypothetical construct of usability, the suggested constituents of usability derived from the literature were: Learnability; Visual Clarity; Consistency & Compatibility; Error Prevention & Correction; Appropriate Functionality; Effectiveness; Efficiency; Feedback; Explicitness; Flexibility; Workload; Help Facilities and User Attitude. A more detailed explanation of each component was included in Section 1.8. Existing measures of usability were largely developed with consideration of the desktop environment and the office context and were thus not applicable to the flight deck context. Further, many of the measures were too generic to assess the usability issues pertinent to the flight deck effectively.

As stated in Section 1.13, the objectives of the first study were to generate a pool of items concerning the usability of FMS CDUs through use of existing usability measures and interviews with line pilots, and to produce a usability questionnaire that could be distributed internationally to flight crew type-rated on a range of aircraft and FMS CDU types. Through a series of analyses, the product of this study was therefore intended to be a usability factor structure, which would be supplemented with qualitative data concerned with the problems associated with current FMSs and the improvements pilots would like to see to FMSs in the next generation of flight decks.

In the remainder to this chapter, the methodological considerations pertinent to the study to be described are introduced. In Section 2.2.1, attention is focused on the pertinent methodological issues which had to be considered when deriving knowledge from line pilots (the targeted participants for this research). In Section 2.2.2, methodological issues related to the use of self-completion questionnaires (the methodology to be adopted for this research) are discussed.

Since the usability of the FMS CDU is intrinsically linked to the aircraft in which it resides, it was hypothesised that if the usability factor structure elicited was valid, it should be possible to discriminate between aircraft manufacturers and FMS manufacturers. Since usability is "*not a unidimensional product or user characteristic, but emerges as a multidimensional characteristic in the context of users performing tasks with a product in a specific environment*" (Lewis, 1995, p.58), a number of hypotheses were made.

In order to demonstrate discriminant validity, ratings of the usability factors should be able to (i) discriminate between aircraft manufacturers (Boeing, Airbus and Fokker in the study to be described) and (ii) discriminate between the FMS CDU types that are fitted to these aircraft as standard (Honeywell and Smiths FMSs in the study to be

described). The real world constraints applicable to studies of FMSs are presented in Section 2.2.3.

Through recording the pilots' command rank (First Officer / Captain) and the country in which they operated, the study was also designed to show that these user characteristics would not significantly affect usability ratings. It was hypothesised that since the usability of the FMS CDU was an independent product of the FMS CDU and the aircraft in which the CDU resided, neither the rank nor culture of the participants should significantly affect ratings. Thus, the corollary of discriminant validity would be demonstrated – ratings of usability should not discriminate between those aspects which they should not discriminate. Considerations of rank and culture based on research literature are presented in Section 2.2.4.

2.2 Methodological Considerations

After consideration of a number of techniques that could be used to develop the usability scales, it was decided that a self-completion questionnaire technique would be most appropriate. This method would permit information to be derived from a large number of participants without the cost and time involved in conducting, for example, an observational or in-depth interview study. The methodological issues associated with requesting line pilots to complete self-completion questionnaires are outlined in this section. The specific methodological issues considered in developing the questionnaire are addressed subsequently in Sections 3.1.4 and 3.1.5, while the implicit beliefs concerned with the application of psychometric robust scales to usability were discussed in Section 1.12.

2.2.1 Deriving Knowledge from Line Pilots

Noyes, Starr & Frankish (1996) outlined the issues that should be considered in eliciting information from SMEs with specific emphasis on the type of SMEs, line pilots, who would be the ultimate participants in this study. The issues considered were:

Familiarity – Since use of the FMS CDU is an everyday activity for the participants, it could be difficult for them to make objective assessments about a system with which they interact so regularly.

Past Research – Historically line pilots and system designers have not communicated effectively, and pilots are therefore not used to being questioned about the equipment they use.

Loyalty to current systems – Line pilots can exhibit a loyalty towards their current aircraft due to the inherent risk associated with flying that (or any) aircraft. The important implication from this point is that pilots may be reluctant to question features of the flight deck in which they operate, and may justify design features regardless of their suitability for the task being undertaken.

Frequency of Participation – Flight Deck Crew, like those involved in many specialised occupations, are the subject of much research and therefore could be reluctant to participate in a study of this nature.

In order to minimise the impact of these factors, the benefits to the flight deck crew of participating in the research were emphasised throughout the study phases. The prime benefit to participants was that research of this nature could contribute to improving the usability, and thus safety, of next generation flight decks. Further, by distributing the questionnaires through the pilots' airlines rather than to the pilots directly, airline support for the research was made evident.

2.2.2 Use of a Self-Completion Questionnaires

Since the purpose of this study was to elicit an FMS CDU usability factor structure that could be used to develop usability scales, a large sample of participants was sought for inclusion in the study to maximise the generalizability of the data collected.

It is commonly recognised that experts vary widely in their opinions (Wyatt & Emerson, 1990). Using a small sample of end users in this study would have accentuated this problem. It was therefore decided that a self-completion questionnaire, distributed to a large sample of line pilots would be the main source of information for generating the usability scales. The initial knowledge elicitation for the item pool, however, would be derived from academic literature and semi-structured interviews with line pilots. Line pilots, as opposed to training or test pilots, were selected for use in this study as they are the pilots who are the typical end users of the FMS CDU, and would therefore be best able to comment on its usability from a typical end user perspective.

The ultimate objective of a standardized questionnaire is that it reduces variability due to measurement error, and therefore increases reliability. Self-completion questionnaires also provided the advantage of being the easiest, if not the only way of retrieving information about the beliefs of a large set of aircrew. Further, it was an efficient method of obtaining large amounts of data at relatively low cost and in a relatively short period of time. The anonymity associated with self-completion questionnaires allowed the respondents to give honest responses about what could be a sensitive subject.

As with any methodology, the associated disadvantages had to be considered. Of particular concern was that a low response rate might be achieved. Herberlein & Baumgartner (1978) reported that response rates of less than 20 percent were common when using postal questionnaires. A further disadvantage was that since the researcher would not have information about non-respondents, it would not be possible to determine how representative the sample of respondents was. The final concern was that if any ambiguities existed within the questionnaire, they might go undetected. It was important to note these potential disadvantages prior to developing a self-completion questionnaire to ensure that countermeasures could be taken at each step to minimise their impact on the research. Since research (e.g. Miller, 1991) has shown that the response rates gained from postal questionnaires are higher when the relevance of

the questionnaire is evident to participants, for example, meant that a process of ensuring that the questionnaire developed within this research was salient to respondents could be undertaken. The countermeasures taken are described in the subsequent method section.

2.2.3 Methodological Constraints

In order to elicit evidence of discriminant validity this study aimed to produce a series of usability scales that could discriminate between different aircraft and FMS types. It is important that the reader notes at this point that it was not possible to employ a fully crossed research design in the methodology to be described since it is not feasible to completely separate the FMS CDU from the aircraft in which it resides. This was highlighted by Wiener (1985) in conducting a study of a similar nature which stated that *"It is impossible to conduct an "aircraft independent" project of this sort*" (p.2). While it is possible to conduct studies on parts of the flight deck by taking them away from their operating environment, many of the contextual complexities that impact on FMS CDU usability would be lost in a research design of that type.

In examining the impact of FMS / aircraft type on usability ratings, allowances had to be made for the real world limitation that Smiths FMSs are only fitted to B-737 aircraft as standard, while Honeywell FMSs are fitted to all the other aircraft types incorporated in the study to be described. This is a real world limitation rather than an oversight on the part of the research design methodology.

2.2.4 Considerations of Rank and Culture

In order to generate further evidence of discriminant validity, this study sought to demonstrate that it would not be possible to discriminate between pilots' culture or rank based on their ratings of FMS usability. This would demonstrate discriminant validity since the ratings would not discriminate between that which they should not discriminate.

It was hypothesised that since it is well documented that both rank and culture influence pilots' attitudes to aviation automation, and since usability is not a user characteristic, pilots' ratings of the usability scales ultimately produced should not discriminate between pilots' culture (characterised by the pilot's country of base) or rank (Captain or First officer) as they would if attitudes to automation were being assessed. In order to examine the ratings of FMS usability by rank and national culture, the study to be described in the following section has included both First Officers and Captains from North America and Europe.

National culture was defined by Sherman & Helmreich (1995) as *"the way a group of people live their lives – the attitudes values and customs that reflect a group's approaches of surviving and thriving*" (p.683). Research has illustrated that both pilot rank and culture impact on pilot attitudes towards aviation automation. Helmreich & Merritt (1998) reported on a large-scale survey of international pilots using the Flight Management Attitudes Questionnaire (FMAQ: Helmreich, Merritt, Sherman, Gregorich & Wiener, 1993) which was designed to measure pilots' attitudes toward *"command,*

communication stress, rules, automation, organization climate and work values" (p.64). Their findings indicated that one of the areas in which national differences were most notable was in attitudes to automation.

Merritt & Helmreich (1995) similarly reported on pilot attitudes towards flight deck automation, and found that attitudes were determined by the pilot's national culture, the level of automation of the aircraft and the crew member's position (first officer or captain). Sherman & Helmreich (1995), found that national culture was the factor which had the largest impact on attitudes to automation and was responsible for up to *"ten times the amount of variability on the scales than were other tested effects*" (p.685).

3 STUDY ONE METHOD

3.1 Overview of Method One Phases

This study was undertaken in six phases:

1. Elicitation of usability components from literature.
2. Familiarisation with the FMS CDU using a part-task trainer.
3. Undertaking flight deck observational visits.
4. Conducting semi-structured interviews with typical end users.
5. Questionnaire production – including derivation of questionnaire items; formulation of initial questionnaire; pre-piloting the questionnaire face-to-face with end users; piloting the questionnaire with a larger international sample and distributing the final questionnaire derived from the study to a still larger sample.
6. Analysis of questionnaire data to determine the underlying factor structure of FMS CDU usability and deriving evidence of its reliability and validity.

Each of these study phases will be described in turn.

3.1.1 Phase One – Elicitation of Usability Issues from the Literature

In order to begin addressing the issue of usability on the flight deck, a thorough investigation of the literature was conducted to examine the concept of usability and to review existing measures of usability.

The literature search concluded that there was no universally accepted definition or measure of usability, however, a number of elements that constituted usability were cited repeatedly (see Section 1.8). The research associating these factors with usability was almost exclusively derived from PC-based environments. Whether these factors would be applicable to a complex, dynamic environment such as a flight deck was yet to be determined.

3.1.2 Phase Two - Familiarisation with the FMS

As the researcher was not a pilot, it was important to establish a common frame of reference between the researcher collecting the data and the pilots who would ultimately be completing the questionnaires. It was necessary to begin this study by examining how the FMS was operated and observing the environment in which the pilot has to

work. Three methods were used to gain familiarity with the FMS CDU. Reviewing FMS operator manuals provided a basic understanding of FMS functionality and an overview of the steps involved in programming the FMS through the CDU. A CDU part-task trainer provided the researcher with the opportunity to interact with the FMS CDU. Finally, flight deck visits provided the researcher with the opportunity to observe real time FMS operation and to appreciate the dynamic context in which the FMS is used.

3.1.2.1. Operator Manuals

Generic, as opposed to airline specific, operator manuals were used to gain an overview of the FMS and CDU functionality and to examine in detail the procedures involved in using the FMS (for example Bulfer, 1996; Bulfer & Gifford, 1997). Reviewing the operating manuals was also a useful method of assessing the CDU page architecture and the number of CDU pages that had to be navigated through to access necessary information or to go through, for example, the pre-flight process steps.

3.1.2.2. CDU Part-Task Simulator

A basic understanding of how to interact with the FMS and how it operated was gained through extended use of the UNS-1B part-task simulator and the UNS Training Software User's Guide. Aviation Computer Based Instruction (ACBI) Inc., part of the Universal Avionics System Corporation, produced the UNS-1B simulator used for this study phase.

The most basic of the features provided by the UNS-1B simulator was a tutorial that provided an introduction to FMS operating procedures. The program also provided realistic aircraft performance based on continual updates through the CDU such that flight progress could be viewed on an Electronic Horizontal Situation Indicator (EHSI); on a map or on a profile depicting the aircraft's progress by phase of flight.

Operation of the CDU interface on the simulator was through clicking the mouse over virtual CDU buttons rather than real buttons as used on flight deck CDUs. A "help" icon could be accessed at any time, and an on-line training manual was also available to assist the researcher in learning about CDU operation.

Extended use of the simulator and operators' training manuals permitted the researcher to become familiar with the steps involved in pilot activities such as FMS initialization, flight plan building, creation of pilot defined waypoints, inputting flight and departure plans, and creation of pilot defined approaches.

Although this could provide the researcher with an initial understanding of how the FMS functioned, it did not provide any information about the context in which aircrew have to operate this computer.

3.1.3 Phase Three - Flight Deck Visits

In order to achieve an understanding of the flight deck environment, a number of flight deck visits were undertaken. During the first flight deck visit, an experienced captain described the interactions being made with the FMS on an Airbus A300-600 while operating the aircraft en-route from Manchester to Verona. This permitted the researcher to observe how interactions with the FMS fitted in with the operation of other automation on board the aircraft.

The researcher was also permitted to ask questions of both the pilot in command and the non-handling pilot during periods of low workload. Particular emphasis was placed on the different flight phases during which the FMS was used, from operations on the ground in the terminal area through take-off, cruise, descent, approach and landing. The visit also provided the researcher with the opportunity to gain an appreciation of the dynamic environment in which flight deck crews operate and to observe the impact of flight phase on pilot workload.

Four subsequent flight deck visits were undertaken for purely observational purposes. The crews observed on these occasions were operating a range of FMS-equipped aircraft – the Boeing 767, Airbus A-320, Boeing 737-400, and Boeing 737-800.

3.1.4 Phase Four - Semi-Structured Interviews

The objective of this study phase was to generate items that could subsequently be used to build the questionnaire to assess FMS CDU usability. With a view to the future development of a questionnaire, it was decided to formalise discussions with pilots using an interview schedule. The structure of these interviews was modelled around the hypothesised usability components from the literature, although it was thought that these components would individually be more or less applicable to the flight deck operating environment.

3.1.4.1. Methodological Considerations

Semi-structured interviews provided the advantage of permitting the interviewee to discuss issues that the interviewer had not considered, and which could thus be lost in a fully structured interview. In the study to be described here, semi-structured interviews permitted the pilots to be free to discuss all factors relevant to FMS CDU usability without being constrained by those factors hypothesised from the literature.

The disadvantage of this method was that it was relatively time consuming and expensive, particularly since complete transcription of tape-recorded interviews was involved. Robson (1993), for example, stated that the transcription of a one-hour tape typically takes ten hours to transcribe fully. The quantity of time and cost involved in transcribing and analysing the data further reduces the number of participants that can be involved in a study of this type.

3.1.4.2. Development of Interview schedule

The purpose of this development phase was to assess the relevance of each of the usability dimensions elicited from the literature to the FMS and flight deck environment, and to assist in the development of an interview schedule that could be used with other line pilots. The researcher presented the study aim and rationale to an airline captain (with in excess of 8,000 hours flying experience) and to four experienced human factors researchers all of whom worked in the aviation domain. Based on the literature and comments from these SMEs, an interview schedule was developed.

The interview schedule was only intended to provide a focus for the interviewer and interviewee, and was not intended to dictate the interview direction. Participants were permitted to speak freely on the usability issues that they saw to be of concern, while the interview schedule allowed the researcher to examine the pertinence of the usability dimensions elicited from the literature. It should be noted that no usability component from the literature was omitted a priori.

The semi-structured interview schedule used by the interviewer is included at Appendix B.

3.1.4.3. Participants

A total of fourteen line pilots, from UK airlines were involved in the interview phase of the research. The initial participants for this phase of the study were recruited on a relatively ad-hoc basis through personal contacts. Each subsequent interviewee was asked if they could propose any colleagues who might be interested in participating.

3.1.4.4. Sampling

The sampling strategy could be considered snowball sampling since the interviewees were used as informants to identify other members of the target population. It was important at this phase of the research to ensure that information elicited was not limited by aircraft type (B-737, B-757/767, B-747, A320/A321/A330 and F-100), hours on type (range 500-4,000), airline (four UK airlines), total hours (1000–9000), age (range 27-45) or rank (First Officers and Captains).

3.1.4.5. Procedure

All participants were informed that they could either be interviewed at Cranfield University, at their airline reporting base, or at their home. All but one participant chose to be interviewed at their own home, while the final participant chose to be interviewed in an office at Cranfield University.

The interviews were taped with the participants' consent. Participants were informed that some of the recordings would be transcribed literally but that the purpose of the transcripts was only to ensure that technical detail could be verified and any

misunderstandings clarified. Anonymity and confidentiality were guaranteed to the participants, and they were informed that the tapes would be erased post analysis. It was important that participants understood the confidentiality associated with the interviews so that they were not inhibited about discussing the usability problems associated with the FMS on board their aircraft, and further so that they could be confident that their company would have no access to the information provided. The interview began with the interviewer introducing herself, the purpose of the study and its associated rationale from a script. The introduction was followed by the interviewer asking the interviewee a series of questions.

The interview schedule evolved as the researcher learned more about the issues of concern to the pilots. Questions asked about colour on the CDU, for example, were found to be not applicable to many interviewees since most interviewees were type-rated on aircraft with monochrome FMS CDUs.

The first seven interviews were tape-recorded and transcribed verbatim. After the first four interviews it became evident that the same usability issues were being identified repeatedly and no new issues were being elicited. After conducting seven interviews, the transcripts were reviewed and the usability issues identified were compared against the literature. Literature considered included comparisons with existing usability measures and checklists such as those produced by Ravden & Johnson (1989); Smith & Mosier (1986); Shneiderman (1992); Marshall, Nelson & Gardiner (1987) and Kirakowski et al (1993).

Further, interviews were conducted with a further seven pilots to confirm that all the major FMS usability issues had been identified. A training captain from North America, with in excess of 10,000 hours flying experience, assisted throughout the process to ensure that North American participants could easily comprehend the questionnaire. A summary of the usability issues raised during this interview phase has been included at Appendix C.

3.1.5 Phase Five - Questionnaire Production

The purpose of the questionnaire production phase was to produce a questionnaire that could be distributed to a large sample of pilots in both the UK and North America. It was important, therefore, to ensure that no colloquialisms were incorporated into the questionnaire and that input was derived during questionnaire development both from pilots in the UK and North America.

3.1.5.1. Methodological Considerations

The prime methodological consideration at this phase was selection of the most appropriate format for the scales.

Likert-type scaling was the selected rating method for the questionnaire, primarily because it was easy to construct, was easy for respondents to rate and is the most common scaling technique used in occupational research (Robson, 1993). Using Likert-

type scales, a statement is presented followed by response options that indicate varying degrees of agreement or disagreement with that statement (Oppenheim, 1966). Based on DeVellis's (1991) guidance, seven points scales with the anchors "strongly disagree" to "strongly agree" were used in this study, since they are very common in scale development, and are therefore easy for respondents to understand.

A disadvantage of using Likert-type scales is that habitual response style may bias the respondent to answers at either the extreme ends or the middle of the scale. However, it was decided that the advantages of using Likert-type scales, particularly its ease of use for respondents, far outweighed its disadvantages.

3.1.5.2. Derivation of Questionnaire Items

The process of questionnaire item writing has often been seen as more of an art than a science. Murphy & Davidshofer (1991) commented that this was a possible reason why so little attention has been given to item writing in comparison to other aspects of the scale development process like scale construction and response sets. The key points noted from the literature (e.g. Murphy & Davidshofer, 1991; Marshall, 1994) on item writing were issues of item length, and the presence of sexist, racist, or other offensive language. Murphy & Davidshofer (1991) argued that the best method for eliminating the problems associated with written items, was to test them out with the target population. This procedure was followed for the items derived for this study.

The researcher first reviewed the interview transcripts in order to collate all statements referring to usability dimensions. This process was then repeated by two aviation Human Factors experts (both of whom had in excess of ten years experience of applying human factors in the aviation domain), to ensure that all pertinent issues elicited during the interviews had been included. The usability statements were all placed on individual pieces of card such that an item pool could be physically generated. This procedure followed that outlined in DeVellis's (1991) guidelines for scale development. Initially, one hundred items were included in the item pool. Through an iterative process, and with two independent SMEs as verifiers, the initial item pool was reduced to forty-seven items, which were mapped onto the usability components identified from the literature. This was done by grouping similar statements into categories, and eliminating duplicated statements or statements which were too similar.

The items were then turned into statements that could be rated against seven point Likert-type scales that had the anchors "strongly disagree" and "strongly agree". The statements were counterbalanced in terms of whether they were positively or negatively phrased to prevent any response biases. The items were preceded by an introductory paragraph describing the purpose of the study and an instruction paragraph detailing how the questionnaire was to be completed.

A demographics section was added to the end of the initial questionnaire requesting each respondent's sex, age, rank, airline, aircraft fleet, other type-ratings held, hours on aircraft type, total hours, type of FMS on board their current aircraft (e.g. Honeywell 757-200 PIP), and whether they had a PC at home.

Space was provided at the end of the questionnaire to permit the pilots to add any extra comments they wanted to make concerning the usability of the FMS on board their aircraft, or improvements they would like to see in future FMSs.

3.1.5.3. Pre-Pilot Questionnaire

During the pre-pilot stage, eight UK line pilots completed the questionnaire on a "face-to-face" basis i.e. the participants completed the questionnaire in the presence of the researcher so that any difficulties encountered could be observed by the researcher and discussed with the participants. The pilots at this stage were again recruited on a word of mouth basis through contacts established during the interview phase of the research.

Many minor changes were made at this stage and the questionnaire went through a number of iterations largely concerned with item clarification.

The pre-pilot questionnaire has been included at Appendix D. Changes made to the questionnaire after the face-to-face testing included:

- The inclusion of an example of how the questionnaire items should be completed.
- The addition of a sentence to emphasise that the pilots were rating the "FMS Control and Display Unit (CDU) only, NOT the navigation display".
- The word "sex" being changed to "gender". The North American pilots recommended that "gender" would be a more appropriate term.
- Questionnaire item seven which stated, "The format and method of entering information is not consistent throughout the FMS", was altered to "The format and method of entering information is not consistent throughout the operation of the FMS".
- For clarification purposes, an example was added to questionnaire item 11 which then read – "Established conventions are followed for the format in which information is displayed *(e.g. layout of dates)*".
- Questionnaire item 17 which stated "When using the FMS I sometimes lose track of where I am in the system; of what I am doing or have done; or of what the other pilot is doing or has done" was separated into two items. The new items were "When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done" and "When using the FMS I sometimes lose track of what the other pilot is doing, or has done".
- Both the researcher and the participants deemed questionnaire item 33, "The FMS is not adequately flexible to meet Air Traffic Control requirements", to be outside the scope of the study since it reflected ATC limitations rather than FMS CDU properties. The item was therefore removed.
- Questionnaire item 34, "It is easy to return to the main menu from any part of the FMS architecture", was deemed to be too complex by all the participants and was altered to "It is easy to return to the main menu from any screen of the FMS".

- Questionnaire item number 36 that read "The FMS is adequately flexible for all navigation" was divided into two questions, one for vertical navigation, and one for lateral navigation.
- Since the term "knock-on", included in questionnaire item number 40, was deemed by the North American reviewers to be a "British vernacular" the term was removed. The item was altered from "The FMS is protected against possible knock-on effects of changes in one part of the system" to "Changes in one part of the system may produce subsequent adverse effects in other parts of the system".
- Since the pilots indicated that workload on the flight deck was both physical and mental questionnaire item number 46 "The FMS can increase physical workload at inappropriate times" was added to the questionnaire.

Two senior line pilots in North America again assisted throughout this development process to ensure all items in the questionnaire were of a similar meaning in North America. The changes detailed in this section resulted in the production of a 49-item questionnaire that was used in the pilot study described in the next section.

3.1.5.4. Pilot Study

Two hundred pilot questionnaires were distributed through four airlines across North America. The questionnaires were distributed through the senior pilots who were contacted through the American Airline Pilots Association (ALPA), and who agreed to distribute the questionnaires by leaving them in the crew rooms at their respective base airports.

The questionnaire used during the pilot phase of the research has been included in Appendix E.

In total 60 responses were returned. This constituted a 30% response rate. Based on these responses a number of minor changes were made to the questionnaire.

- Some questionnaire items, which were negatively phrased, created confusion, and were altered. For example, questionnaire item one, "The information in the FMS display (CDU) is not easy to see and read", was altered to "The information in the FMS display (CDU) is easy to see and read". Nine of the questionnaire items were changed in this manner.
- Questionnaire item 31 which read "The FMS processes information too slowly for other activities going on in the flight deck", was changed to "The FMS processes information too slowly for other activities in the flight deck".
- As only one space was provided for participants to input both their current aircraft type and series, some participants omitted the latter information. To eliminate this problem, separate spaces were provided for current aircraft type and aircraft series for the final questionnaire used in this study.

- Since the analysis illustrated that almost all participants in the study had a personal computer at home, it was determined that the item recording whether or not participants had a computer at home was not particularly useful. It had originally been included so that analysis could be conducted to determine if PC literacy had an effect on usability ratings. Since such an analysis could not be conducted, there was no longer any valid reason for including the item.
- Similarly it was decided that the item requesting other type-ratings previously held by participants did not yield information particularly relevant to the study. This item was also eliminated.

A series of descriptive analyses was conducted to examine the data to ensure, for example, a range of responses was being achieved on each item. The results indicated that a range of responses was achieved and no further problems were identified at this stage. A final questionnaire for the first study was produced from this pilot questionnaire. All comments elicited from the comment boxes at the end of the questionnaire were collated and categorised, although no new FMS usability concerns were raised at this stage.

3.1.6 Study One Final Questionnaire

The final questionnaire for Study One contained forty-nine items as shown in Appendix F.

The response format for the final questionnaire was identical to that used in the pre-pilot and pilot studies and involved rating each statement against a seven point Likert-type scale which had the anchors "strongly disagree" to "strongly agree". Statements were again counterbalanced in terms of whether they were positive or negative (e.g. "The FMS (CDU) is sometimes cluttered"; "The FMS sometimes makes unexpected actions"). A rating of '7' thus indicated the highest usability ratings for some questionnaire statements, while indicating the lowest usability rating on other usability statements. Consistent with the changes discussed in Section 3.1.5.4, even those statements which referred to negative aspects of FMS usability were positively phrased e.g. "The FMS sometimes makes unexpected actions".

An example of the response format is shown here:

1. The information in the FMS display (CDU) is easy to see and read.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

The items were combined with an introductory paragraph outlining the purpose of the study and an instruction paragraph detailing how the questionnaire was to be completed. Emphasis was placed on the fact that participants were rating the FMS Control and Display Unit (CDU) only, and not the navigation display.

A demographics section was added to the end of the final questionnaire. Each respondent was asked to provide their sex, age, airline, current aircraft fleet, current

aircraft series, hours on aircraft type, total hours, and type of FMS on board their current aircraft (e.g. Honeywell 757-200 PIP).

The final questionnaire was set on four A4 sides of paper, but printed double sided on an A3 page. The final questionnaire has been included at Appendix F, although it is presented on A4 paper for presentation reasons only.

The questionnaires were packed in envelopes each accompanied by a letter explaining the purpose of the research (see Appendix I) and a FREEPOST return envelope. A label marked "FMS Usability Questionnaire" was placed on the outside of each envelope to indicate the purpose of the questionnaire, and also to minimise the number of questionnaires being opened by airline personnel for whom the questionnaire was not applicable.

3.1.6.1. Target Population

The target population was airline line pilots, as opposed to test pilots, in the UK and North America who were currently type rated on FMS-equipped aircraft. Line pilots were targeted, as they are "typical" airline pilots and therefore would have a wide range of experience in terms of hours on their current aircraft type and total flying hours. A number of airlines were included to eliminate any effect of airline specific practices.

3.1.6.2. Response Rate Considerations

A number of steps were undertaken to maximise the response rate from the postal questionnaire.

Questionnaire Length: Robson (1993) reported that questionnaire length had a significant impact on response rate, with shorter questionnaires achieving better response rates. The questionnaire used in this study was not very long covering only four pages. Further, the questionnaire was printed on double-sided A3 paper rather than four sides of A4 paper to ensure it did not appear too long or unwieldy to complete.

Postage: A FREEPOST envelope was provided for return of the questionnaire.

Covering Letter: A covering letter tailored to the respondents' airline was included with each questionnaire to explain the purpose of the study.

Confidentiality and Anonymity: Robson (1993) reported that guaranteeing confidentiality and anonymity was important to securing a good response rate. The questionnaire introduction included an assurance of confidentiality. This was particularly important since the questionnaire was being distributed through the pilots' employers.

Applicability: Since research (e.g. Miller, 1991) has shown that the response rates gained from postal questionnaires are higher when the relevance of the questionnaire is evident to participants, both Human Factors experts and four airlines pilots in the UK

assisted in reviewing the questionnaire to ensure its salience to pilots. As in the pilot study, two senior aircrew members from North America assisted the researcher in ensuring that the questionnaire would be as applicable in North America as it would be in the UK. This process involved the North American SMEs reviewing the questionnaires to ensure that every questionnaire item was applicable to North American participants and that no UK colloquialisms had been included.

3.1.7 Questionnaire Distribution

A total of 1150 questionnaires were distributed throughout North America and the UK. Eight hundred questionnaires were distributed through two sources, one in the US and one in Canada. In North America, the questionnaires were distributed by sending the questionnaires to the senior ALPA pilots who placed them in the airport crew rooms. As in the pilot study, two senior aircrew members from North America assisted the researcher in ensuring that the questionnaire would be as applicable in North America as it would be in the UK. Two hundred and fifty questionnaires were distributed to one airline in the UK, while a further 100 questionnaires were distributed in a second UK airline. These were similarly placed in the crew rooms at the airlines' bases.

4 STUDY ONE RESULTS

4.1 Overview

This results section presents a description of the sample that participated in the study. This is followed by an explanation of how the data were treated. The descriptive analysis section presents the means and standard deviations (sd) of all questionnaire data. A Principal Components Analysis (PCA) is then described which explores the principal components underlying the questionnaire items, and their associated construct and content validity. A series of Discriminant Function Analyses (DFAs) are presented which examine the evidence of discriminant validity from the factors derived. Finally, a series of Cronbach's Alphas analyses are described which examine the internal consistency of the items within the factors from which the usability scales would ultimately be developed. For all the analysis tables in this section, highlighting of text has been used to show when a specified criterion value (e.g. $p<0.05$) has not been achieved.

4.1.1 Sample Description

This section outlines a description of the respondents and the response rate achieved. The overall response rate at the time the analyses were conducted (seven months after distribution) was 21.91%. Thirteen months after initial distribution responses were still being received, with some respondents from North America reporting difficulty in using the International FREEPOST return envelopes.

A total of 252 questionnaires were used in the final analyses. Nine (3.6%) respondents were female, while 231 (91.7%) respondents were male. Twelve respondents (4.7%) did not complete the gender part of the questionnaire. The mean and standard deviation of participants' ages, hours on type and total hours are shown in Table 4-1.

Table 4-1 - Summary Participants' Age, Hours on Type & Total Hours

	Min	*Max*	*Mean*	*sd*
Age	23	63	44.36	8.0
Hours on Type	50	16000	3033.23	2762.85
Total Hours	1200	30000	11706.62	5802.59

Figure 4-1 shows a histogram of pilots' total hours, while Figure 4-2 shows the distribution of pilots' hours on their current aircraft type.

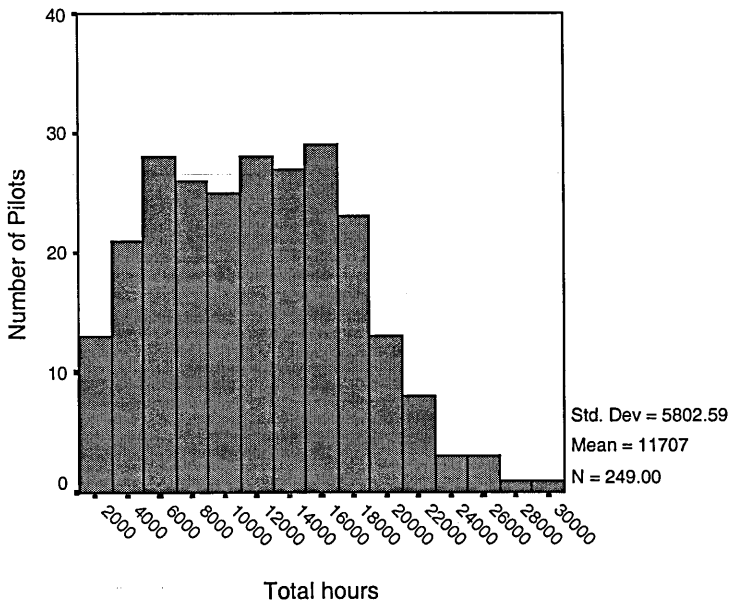


Figure 4-1 - Histogram of Pilots' Total Hours

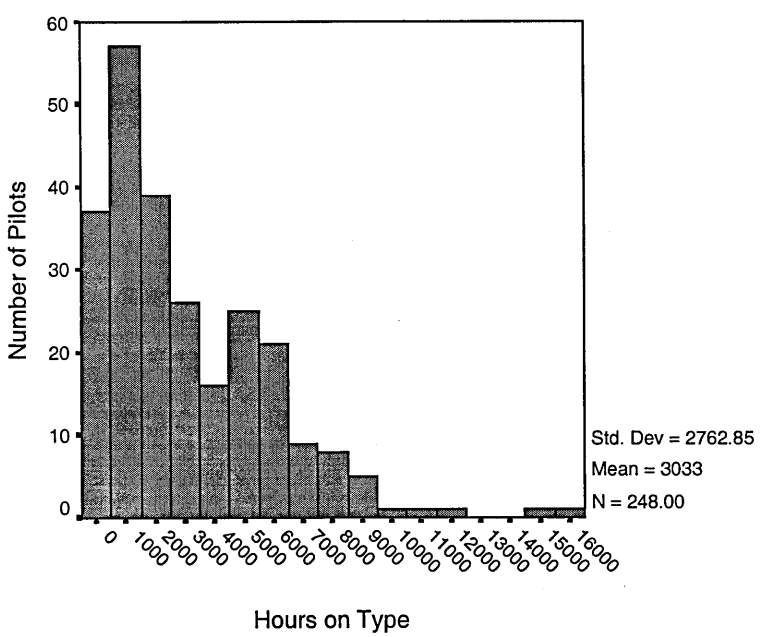


Figure 4-2 - Histogram of Pilots' Hours on Current Aircraft Type

The breakdown of participants' rank are shown in Table 4-2.

Table 4-2 - Summary of Participants' Rank

Rank	*Frequency*	*Percentage*
Captain	138	54.7
First Officer	105	41.7
Unspecified	9	3.6

The participants' airlines are shown in Table 4-3. It should be noted that there is typically an equal distribution of Captains and First Officers within an airline. Since 41.7% of respondents were First Officers this category of respondents was under represented. The names of the participant airlines have not been included since the airlines were guaranteed anonymity and confidentiality. Non-identifying terms such as, Airline 1 and Airline 2, have been used instead.

Table 4-3 - Summary of Airline Participation

Airline	*Frequency*	*Percentage*
N. American Airline 1	91	36.1
N. American Airline 2	25	9.9
N. American Other Airlines	16	6.4
UK Airline 1	86	34.1
UK Airline 2	34	13.5

The response rate for UK Airline 1 was 34.4%, while the response rate for UK Airline 2 was 34.0%. It was not possible to calculate response rates by airline for the North American airlines since the questionnaires were distributed through two Chief Pilots at their airline's reporting base, but responses were returned by participants from more than two airlines. It is likely that this is because more than one airline may use a crew room at any given airport. The overall response rate from North American airlines was 15.2%.

The aircraft on which the participants were type-rated is shown in Table 4-4. Participants type-rated on Boeing 757s and 767s have been combined since there is a common type-rating for these aircraft. The same is true of the Airbus A320/A321/A330 type-ratings which have also been combined.

Table 4-4 - Summary of Current Aircraft Type and Series

Aircraft Type	*Frequency*	*Percentage*	*Aircraft Series*	*Frequency*	*Percentage*
B-737	85	33.7	B-737-300/400/500	85	33.7
B-757/767	81	32.1	B-757/767-200	80	31.7
B-747	17	6.7	B-747-400	17	6.7
A320/A321/A330	47	18.8	A320/A321/A330	12	4.9
F-100	22	8.7	Series Not Specified	58	23.0

The type of FMS reported by participants as being the one on board the aircraft on which they were currently type-rated is described at Table 4-5.

Table 4-5 - Summary of Reported FMS Manufacturer

FMS Manufacturer	*Frequency*	*Percentage*
Honeywell	118	46.8
Smiths	22	8.7
Sperry	4	1.6
Collins	3	1.2
Unspecified	105	41.7

An examination of the reported FMS types revealed that many respondents were unable to specify the type of FMS on board their aircraft. A total of 41.7% of respondents did not complete this item. Evidence at the interview phase of this study, reinforced the validity of this observation, since many of the interviewees were unaware of the type of FMS on board their aircraft until they searched through their manuals during the interview. It was not a complete surprise that participants were unaware of the type of FMS they used, since in the day-to-day operation of the CDU, the name or type of FMS being used is largely irrelevant.

The actual FMS type on the aircraft on which the respondents were type-rated was generated from information made available by the participant airlines and from the aviation literature (Bulfer, 1996; Bulfer & Gifford, 1997). A summary of the actual FMS on board the aircraft on which the respondents were type-rated is shown on Table 4-6.

Table 4-6 - Summary of Real FMS Manufacturer

FMS Manufacturer	*Frequency*	*Percentage*
Honeywell	168	66.7
Smiths	84	33.3

4.1.2 Treatment of data

As stated in Section 3.1.5.2, in order to prevent response set effects the questionnaire was designed such that for some items a response of '7' indicated the usable end of the scale, while for other items '1' indicated greater usability. For ease of interpretation, any item for which '1' indicated the more usable end of the scale was inverted. In other words for all results appearing in this section, high scores indicate greater usability. Questionnaire items 2,9,10,14,15,17,18,31,41,43,44,46,47 and 48 were treated in this manner. All analyses were conducted using Statistical Package for the Social Sciences (SPSS) Version 8.0.

4.2 Descriptive Analysis

This section presents the descriptive analysis conducted on the data collected during Study One. The means and standard deviations for each questionnaire item are shown in Table 4-7.

Table 4-7 - Means & Standard Deviations for Questionnaire Items

Questionnaire Item	*Mean*	*sd*
1. The information in the FMS display (CDU) is easy to see and read.	5.06	1.40
2. The FMS display (CDU) is sometimes cluttered.	4.13	1.58
3. Important information is adequately highlighted on the screen.	4.15	1.57
4. Information is organised logically on the CDU screen (familiarity aside).	4.55	1.46
5. The way the FMS (CDU) look and works is consistent at all times.	4.33	1.63
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	5.38	1.31
7. The format and method of entering information is consistent throughout the operation of the FMS.	4.82	1.52
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	4.01	1.57
9. The FMS sometimes makes unexpected actions.	3.60	1.89
10. I prefer to stick to the FMS functions and facilities I know best.	3.12	1.63
11. Established conventions are followed for the format in which information is displayed (e.g. layout of dates).	4.92	1.27
12. The sequence of activities required in the use of the FMS is as I would expect.	4.61	1.51
13. The FMS works the way I think it should work.	4.15	1.62
14. The FMS buttons could be located more intuitively.	3.69	1.78
15. The feedback messages displayed by the FMS are frequently inadequate.	3.71	1.74
16. The FMS makes it clear what changes will occur as a result of an input being made.	3.60	1.71
17. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	5.02	1.70
18. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	3.93	1.89
19. The FMS clearly indicated which mode the system is currently in. It is informative and accurate.	4.36	1.64
20. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	2.82	1.56
21. It is clear to the user what should be done to correct any error that occurs.	2.52	1.40
22. The structure of the FMS is obvious to the user.	3.29	1.52
23. It is clear why the system is organised and structured as it is.	3.38	1.52
24. It is clear why the FMS screens are sequenced as they are.	3.89	1.58
25. The structure of the FMS is well organised from the user's point of view.	4.10	1.49
26. The FMS performs all the functions that I need it to.	3.84	1.77
27. FMS users can access all the information they require in all flight phases.	4.05	1.65
28. The FMS screens provide all the information which is relevant to the task.	4.17	1.53

Questionnaire Item	Mean	sd
29. The system allows users to do what they feel is necessary in order to carry out a task.	4.32	1.45
30. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	4.48	1.41
31. The FMS processes information too slowly for other activities in the flight deck.	3.09	1.92
32. FMS users have control over the order in which they request information or carry out a series of tasks.	4.37	1.52
33. The FMS allows the user to do the required task as efficiently as possible.	3.85	1.54
34. It is easy to return to the main menu from any screen of the FMS.	5.14	1.60
35. I feel in control of the FMS at all times.	4.88	1.62
36. The FMS is adequately flexible for all lateral navigation.	4.47	1.69
37. The FMS is adequately flexible for all vertical navigation.	3.58	1.76
38. It is always possible to check or correct inputs to the FMS before they are processed.	4.13	1.82
39. The FMS has facilities for detecting and handling input errors.	3.98	1.62
40. The FMS is protected against common trivial entry errors.	4.09	1.65
41. Changes in one part of the FMS may produce subsequent adverse effects in other parts of the system.	3.32	1.42
42. It was easy to learn how to use the FMS.	4.04	1.78
43. It is sometimes difficult to remember where information is located within the FMS architecture.	3.55	1.72
44. The FMS keyboard is difficult or awkward to use.	4.69	1.70
45. The FMS help facility (either on-line or manuals) is adequate.	2.76	1.59
46. It is irritating to have to input information at a rate slow enough for the FMS processor.	2.75	1.72
47. The FMS can increase mental workload at inappropriate times.	2.79	1.60
48. The FMS can increase physical workload at inappropriate times.	3.43	1.78
49. The FMS is pleasant to use; I like it.	5.27	1.33

From examining the means and standard deviations of the questionnaire items shown in Table 4-7, a number of ratings stood out as being particularly high (towards the usable end of the scale) and particularly low (towards the unusable end of the scale).

Items notably rated towards the usable end of the scale included were: "The information in the FMS display (CDU) is easy to see and read" *(mean=5.06, sd=1.40)*; "Abbreviations, acronyms & codes are used consistently throughout the FMS" *(mean=5.38, sd=1.31);* "When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done" *(mean 5.02, sd=1.7)*; "It is easy to return to the main menu from any screen of the FMS" *(mean=5.14, sd=1.60);* and "The FMS is pleasant to use; I like it" *(mean=5.27, sd=1.33).*

Correspondingly, some items were notably rated towards the unusable end of the scale and these were:

"FMS error messages explain clearly where errors are, what the errors are and why they have occurred" (*mean=2.82, sd=1.56*).

"It is clear to the user what should be done to correct any error that occurs" (*mean=2.52; sd=1.4*).

"The FMS help facility (either on-line or manuals) is adequate" (*mean=2.76; sd=1.59*).

"It is irritating to have to input information at a rate slow enough for the FMS processor" *(mean=2.75, sd=1.72).*

"The FMS can increase mental workload at inappropriate times" *(mean=2.79, sd=1.60).*

4.3 Principal Components Analysis

A PCA was used to elicit the structure underlying the questionnaire items by examining the matrix of correlations in search of groups of highly intercorrelated variables. The final questionnaire item; "The FMS is pleasant to use; I like it", was omitted from this phase of the analysis on the basis that it was taken as an overall assessment of satisfaction in using the FMS CDU. This item was instead used in a Multiple Regression analysis to determine which of the factors derived from the PCA were significant predictors of the FMS being easy to use and liked by the respondents. The Multiple Regression is described in Section 8.10.

Factor analysis refers to a group of techniques that examine the correlations between a set of variables to identify those groups of variables that are highly correlated or homogeneous. PCA is perhaps the most straightforward type of factor analysis both conceptually and mathematically. PCA derives underlying dimensions that "*when interpreted and understood, describe the data in a much smaller number of concepts than the original individual variables*" (Hair et al, 1998, p.90).

4.3.1 Assumptions in PCA

The following subsections outline the assumptions that should be considered in using any multivariate statistical technique. PCA assumes a normal distribution of data and therefore consideration is given to sample size, homoscedasticity and heteroscedasticity, multicollinearity and sampling adequacy. Any departures from normal distribution, homoscedasticity and linearity, are only applicable to the extent that they diminish observed correlations.

Sample Size

PCA assumes a normal distribution of data, and thus warrants large sample sizes. Tabachnick & Fidell (1983) recommend the general guideline, to include five times as many observations as there are variables to be analysed. Hair, Anderson, Tatham & Black (1998) recommend that the preferable sample size to use when conducting a PCA is 100 or larger. Since the sample size used in the analysis described here is greater than 100, and since there are at least five cases for each variable observed, an acceptable sample size may be assumed.

Homoscedasticity and Heteroscedasticity

Homoscedasticity refers to the assumption that variables demonstrate equal variance across the range of predictors. The assumption of heteroscedasticity follows from the assumption of multivariate normality. When the assumption of normality is met, the relationship between variables can be said to be homoscedastic.

Multicollinearity

Multicollinearity is defined by Hair et al (1998) as "*the extent to which a variable can be explained by the other variables in the analysis*" (p.89). In PCA, multicollinearity is necessary since the objective of PCA is to identify interrelationships between variables.

Sampling Adequacy

A further measure, which can be used to determine the appropriateness of factor analysis, is sampling adequacy. Sampling adequacy measures the degree of intercorrelation between variables. Sample adequacy is an index that ranges from zero to one, where one indicates a variable that is perfectly predicted without error by another variable (or variables). Sampling adequacy was examined in this study using the Kaiser-Mayer-Olkin Measure (KMO) as shown in Table 4-8. High KMOs demonstrate that there is lots of shared variance between a particular variable and the other variables, thus making it suitable for factor analysis. Kaiser (1974) suggested that 0.5 is an acceptable KMO for use in PCA. Hair et al (1998) described a measure of sampling adequacy above 0.8 as being "*meritorious*" (p.99), the overall KMO Test of Sampling Adequacy in this study was 0.907 and was thus well above the meritorious level.

Table 4-8 - KMO Test of Sampling Adequacy

Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.907

Since only three of the individual KMOs (6.3% of the items) shown in Table 4-9 failed to exceed the recommended value of 0.5 (Hair et al, 1998), evidence suggested that the correlations between pairs of items could be explained by other items. The values that failed to meet the required value have been highlighted in Table 4-9. The individual KMOs provided further evidence that PCA was an appropriate method of analysis.

Table 4-9 - Individual KMOs

Questionnaire Item	*Extraction*
1. The information in the FMS display (CDU) is easy to see and read.	0.497
2. The FMS display (CDU) is sometimes cluttered.	0.535
3. Important information is adequately highlighted on the screen.	0.577
4. Information is organised logically on the CDU screen (familiarity aside).	0.620
5. The way the FM (CDU) look and works is consistent at all times.	0.623
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	0.489
7. The format and method of entering information is consistent throughout the operation of the FMS.	0.598
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	0.627
9. The FMS sometimes makes unexpected actions.	0.630
10. I prefer to stick to the FMS functions and facilities I know best.	0.630
11. Established conventions are followed for the format in which information is displayed (e.g. layout of dates).	0.564
12. The sequence of activities required in the use of the FMS is as I would expect.	0.630
13. The FMS works the way I think it should work.	0.642
14. The FMS buttons could be located more intuitively.	0.575
15. The feedback messages displayed by the FMS are frequently inadequate.	0.568
16. The FMS makes it clear what changes will occur as a result of an input being made.	0.541
17. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	0.722
18. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	0.672
19. The FMS clearly indicated which mode the system is currently in. It is informative and accurate.	0.522
20. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	0.686
21. It is clear to the user what should be done to correct any error that occurs.	0.650
22. The structure of the FMS is obvious to the user.	0.701
23. It is clear why the system is organised and structured as it is.	0.742
24. It is clear why the FMS screens are sequenced as they are.	0.684
25. The structure of the FMS is well organised from the user's point of view.	0.751
26. The FMS performs all the functions that I need it to.	0.566
27. FMS users can access all the information they require in all flight phases.	0.753
28. The FMS screens provide all the information which is relevant to the task.	0.728
29. The system allows users to do what they feel is necessary in order to carry out a task.	0.690
30. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	0.594
31. The FMS processes information too slowly for other activities in the flight deck.	0.728

Questionnaire Item	Extraction
32. FMS users have control over the order in which they request information or carry out a series of tasks.	0.556
33. The FMS allows the user to do the required task as efficiently as possible.	0.577
34. It is easy to return to the main menu from any screen of the FMS.	0.570
35. I feel in control of the FMS at all times.	0.668
36. The FMS is adequately flexible for all lateral navigation.	0.633
37. The FMS is adequately flexible for all vertical navigation.	0.565
38. It is always possible to check or correct inputs to the FMS before they are processed.	0.540
39. The FMS has facilities for detecting and handling input errors.	0.690
40. The FMS is protected against common trivial entry errors.	0.560
41. Changes in one part of the FMS may produce subsequent adverse effects in other parts of the system.	0.580
42. It was easy to learn how to use the FMS.	0.511
43. It is sometimes difficult to remember where information is located within the FMS architecture.	0.535
44. The FMS keyboard is difficult or awkward to use.	0.543
45. The FMS help facility (either on-line or manuals) is adequate.	0.459
46. It is irritating to have to input information at a rate slow enough for the FMS processor.	0.731
47. The FMS can increase mental workload at inappropriate times.	0.684
48. The FMS can increase physical workload at inappropriate times.	0.729

Since the data met the necessary assumptions, a PCA was conducted and is reported in the subsequent section.

4.3.2 Extraction of Factor Structure

The purpose of performing the PCA was to form the basis for making scales that have construct validity. Factors were extracted from the PCA using the guidelines specified by Hair et al (1998), and extraction termination was based on Kaiser's criterion (Kaiser, 1974).

Eigenvalues

The rationale chosen for extracting factors with an eigenvalue greater than one was that any individual factor interpreted should account for the variance of at least a single variable. Eigenvalues of less than one for a factor indicate that it contains less variance than that of a variable from which it is composed. Thus, only factors which have eigenvalues greater than one are deemed useful.

Rotation Method

Unrotated factor solutions from PCA extract factors in order of their importance, the first factor accounting for the most variance, and each subsequent factor extracted accounting for less and less variance. The purpose of rotating the factor structure is to simplify the factor structure and create a more meaningful factor pattern through redistributing the variance from the first extracted factors to the later ones. Since the purpose of the analysis was to develop scales based on the factors elicited, ease of interpretation was of prime concern. An orthogonal, varimax method of rotation was applied to the data in this study to simplify the factors by maximising the variance of loadings within and across variables. The varimax rotation was chosen, as it tends to minimise the number of variables that load strongly on a factor. This type of rotation also equalises the proportion of variance explained by each factor.

Identification of Components

Hair et al (1998) suggested that those loadings greater than +/- 0.3 meet the minimum level for practical significance, loadings greater than +/- 0.4 are more important, while those greater than +/- 0.5 are considered practically significant. Further, based on criteria suggested by Hair et al, for the 0.01 significance level, and a sample size of 252, a minimum loading of 0.45 would be statistically significant.

The unrotated and rotated results of the PCA are presented in Table 4-10. The sum of the squared loadings of the items on a given factor indicate their relative importance.

Table 4-10 - Principal Components Analysis

	Initial Eigenvalues			*Rotation Sums of Squared Loadings*		
Component	*Eigenvalue*	*% of variance*	*Cumulative %*	*Eigenvalue*	*% of Variance*	*Cumulative %*
1	14.578	30.370	30.370	6.496	13.533	13.533
2	2.269	4.726	35.096	3.576	7.451	20.983
3	1.927	4.014	39.110	2.815	5.864	26.848
4	1.815	3.781	42.890	2.812	5.859	32.706
5	1.540	3.209	46.099	2.514	5.237	37.943
6	1.509	3.143	49.242	2.509	5.227	43.170
7	1.374	2.863	52.105	2.047	4.264	47.434
8	1.287	2.681	54.786	2.045	4.261	51.695
9	1.287	2.580	57.366	1.852	3.858	55.553
10	1.117	2.328	59.694	1.701	3.543	59.096
11	1.012	2.108	61.802	1.299	2.705	61.802

4.3.2.1. Derivation of Factor Labels

When the factor loading solution had been obtained, the researcher labelled the factors in order to give meaning to the pattern of factor loadings. A review of the literature was undertaken to consider other labels that might be pertinent and labels were selected based on their appropriateness for representing the underlying dimensions of a particular factor. The factor labels were selected on the basis of the content of the questionnaire items that loaded onto each factor, and the relationship between those items, rather than statistical similarities. Since items with higher loadings were more important, however, their content and meaning had a greater influence on the label selected for a factor (Hair et al, 1998). Table 4-11 shows the items that loaded onto each factor and the label that was selected as being most appropriate on the basis of the content of the applicable questionnaire items.

Table 4-11 - Factor Labelling

Factor Name	*Questionnaire Items*
Structure & Sequence	23. It is clear why the system is organised and structured as it is.
	22. The structure of the FMS is obvious to the user.
	25. The structure of the FMS is well organised from the user's point of view.
	24. It is clear why the FMS screens are sequenced as they are.
	12. The sequence of activities required in the use of the FMS is as I would expect.
	11. The FMS works the way I think it should work.
	42. It was easy to learn how to use the FMS.
	33. The FMS allows the users to do the required task as efficiently as possible.
	14. The FMS buttons/switches could be located more intuitively.
	30. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.
	44. The FMS keyboard is difficult or awkward to use.

Table 4-11 - Factor Labelling (Contd.)

Component	*Questionnaire Items*
Appropriate Functionality	27. FMS users can access all the information required during flight phases.
	28. The FMS screens provide all the information which is relevant to the task.
	26. The FMS performs all the functions that I need it to.
	29. The system allows users to do what they feel is necessary in order to carry out a task.

Component	*Questionnaire Items*
Feedback	20. FMS error messages explains clearly where errors are, what the errors are and why they have occurred.
	21. It is clear to the user what should be done to correct any errors that occur.
	15. The feedback messages displayed by the FMS are frequently inadequate.
	45. The FMS help facility (either on-line or manual) is adequate.
	41. Changes in one part of the system can have adverse effects on other parts of the system.

Component	*Questionnaire Items*
Predictability	17. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.
	18. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.
	9. The FMS sometimes makes unexpected actions.
	43. Difficult to remember where information is within architecture.

Component	*Questionnaire Items*
Information Presentation	6. Abbreviations, acronyms and codes are used consistently throughout the FMS.
	3. Important information is adequately highlighted on screen.
	7. The format and method of entering information is consistent throughout the operation of the FMS.
	4. Important information is organised logically on screen (familiarly aside).

Component	*Questionnaire Items*
Consistency	5. The way the FMS (CDU) looks and works is consistent at all times.
	8. The way the FMS responds to a particular action is consistent at all times.

Component	*Questionnaire Items*
Flexibility & Control	34. It is easy to return to the main menu from any screen of the FMS.
	35. I feel in control of the FMS at all times.
	37. The FMS is adequately flexible for all vertical navigation.

Table 4-11 - Factor Labelling (Contd.)

Component	*Questionnaire Items*
Error Prevention & Correction	38. The FMS has facilities for detecting and handling input errors.
	40. The FMS is protected against common trivial entry errors.
	16. The FMS makes it clear what changes will occur as a result of an input being made.

Component	*Questionnaire Items*
Processor Speed	46. It is irritating to have to input information at a rate slow enough for the FMS processor.
	31. The FMS processes information too slowly for other activities in the flight deck.

Component	*Questionnaire Items*
Workload	47. The FMS can increase mental workload at inappropriate times.
	48. The FMS can increase physical workload at inappropriate times.

Factor	*Questionnaire Items*
11	36. The FMS is adequately flexible for all lateral navigation.
	11. Established conventions are followed in the format in which information is displayed.

The final component extracted from the analysis contained two items that appeared unrelated in terms of content.

Since no meaningful relationship could be determined between the items loading on Factor 11, it remained undefined and was not given a label.

Table 4-12 shows the rotated component matrix using Varimax rotation with Kaiser Normalisation method and with listwise extraction for variables greater than 0.45.

Table 4-12 - Factor Loadings for Components

Factor Name	*Questionnaire Items*	*Loading*
Structure & Sequence	23. It is clear why the system is organised and structured as it is.	0.797
	22. The structure of the FMS is obvious to the user.	0.743
	25. The structure of the FMS is well organised from the user's point of view.	0.741
	24. It is clear why the FMS screens are sequenced as they are.	0.731
	12. The sequence of activities required in the use of the FMS is as I would expect.	0.696
	11. The FMS works the way I think it should work.	0.651
	42. It was easy to learn how to use the FMS.	0.535
	33. The FMS allows the users to do the required task as efficiently as possible.	0.533
	14. The FMS buttons/switches could be located more intuitively.	0.511
	30. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	0.479
	44. The FMS keyboard is difficult or awkward to use.	0.455

Component	*Questionnaire Items*	*Loading*
Appropriate Functionality	27. FMS users can access all the information required during flight phases.	0.822
	28. The FMS screens provide all the information which is relevant to the task.	0.728
	26. The FMS performs all the functions that I need it to.	0.684
	29. The system allows users to do what they feel is necessary in order to carry out a task.	0.654

Component	*Questionnaire Items*	*Loading*
Feedback	20. FMS error messages explains clearly where errors are, what the errors are and why they have occurred.	0.732
	21. It is clear to the user what should be done to correct any error that occurs.	0.594
	15. The feedback messages displayed by the FMS are frequently inadequate.	0.542
	45. The FMS help facility (either on-line or manuals) is adequate.	0.523
	41. Changes in one part of the system can have adverse effects on other parts of the system.	0.508

Component	*Questionnaire Items*	*Loading*
Predictability	17. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	0.771
	18. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	0.715
	9. The FMS sometimes makes unexpected actions.	0.492
	43. Difficult to remember where information is within architecture.	0.483

Component	*Questionnaire Items*	*Loading*
Information Presentation	6. Abbreviations, acronyms and codes are used consistently throughout the FMS.	0.647
	3. Important information is adequately highlighted on screen.	0.626
	7. The format and method of entering information is consistent throughout the operation of the FMS.	0.547
	4. Important information is organised logically on screen (familiarly aside).	0.531

Component	*Questionnaire Items*	*Loading*
Consistency	5. The way the FMS (CDU) looks and works is consistent at all times.	0.699
	8. The way the FMS responds to a particular action is consistent at all times.	0.647

Component	*Questionnaire Items*	*Loading*
Flexibility & Control	34. It is easy to return to the main menu from any screen of the FMS.	0.645
	35. I feel in control of the FMS at all times.	0.609
	37. The FMS is adequately flexible for all vertical navigation.	0.469

Component	*Questionnaire Items*	*Loading*
Error	38. The FMS has facilities for detecting and handling input errors.	0.754
Prevention	40. The FMS is protected against common trivial entry errors.	0.609
& Correction	16. The FMS makes it clear what changes will occur as a result of an input being made.	0.492

Component	*Questionnaire Items*	*Loading*
Processor Speed	46. It is irritating to have to input information at a rate slow enough for the FMS processor.	0.808
	31. The FMS processes information too slowly for other activities in the flight deck.	0.801

Component	Questionnaire Items	Loading
Workload	47. The FMS can increase mental workload at inappropriate times.	0.781
	48. The FMS can increase physical workload at inappropriate times.	0.671

Factor	Questionnaire Items	Loading
11	36. The FMS is adequately flexible for all lateral navigation.	0.707
	11. Established conventions are followed in the format in which information is displayed.	-0.466

The final component extracted from the analysis contained two items that appeared unrelated to each other in terms of content. This component was therefore eliminated from subsequent analyses. Where a questionnaire item loaded on two components, it was included in the previous table under the component on which it had the highest loading.

As stated in Section 4.3.2, based on criteria suggested by Hair et al, for the 0.01 significance level, and a sample size of 252, a minimum loading of 0.45 would be statistically significant. A number of items in the questionnaire did not achieve this minimum loading onto any of the components. These items have been listed in Table 4-13.

Table 4-13 - Questionnaire Items Not Loading On Any Component

Questionnaire Item
1. The information in the FMS display (CDU) is easy to see and read.
2. The FMS display (CDU) is sometimes cluttered.
19. The FMS clearly indicated which mode the system is currently in. It is informative and accurate.
32. FMS users have control over the order in which they request information or carry out a series of tasks.
37. The FMS is adequately flexible for all vertical navigation.
38. It is always possible to check or correct inputs to the FMS before they are processed.

4.4 Discriminant Function Analyses

The ultimate objective of this section was to demonstrate the validity of the ten components derived from the PCA.

DFA is a special case of multivariate analysis of variance (MANOVA) for independent groups' designs that allows a weighted linear composite score to be formed for the predictor variables. The weighted linear composite score can subsequently be used to assess the predictive ability of the solution using a simple classificatory procedure based on Bayesian statistics. Analysis of the components of the weighted linear combination produced allowed the relative contributions of the dependent variables to the canonical discriminating function to be assessed (Hair et al, 1998).

A number of DFAs are described in this section with the particular goal of demonstrating the discriminant validity associated with the factor structure derived. The DFAs were conducted using regression based factor scores derived from the PCA described in Section 4.3. Based on the knowledge that the FMS is intrinsically linked to the aircraft in which it resides a number of hypotheses were formulated. It was hypothesised that ratings on the usability factors would differ by aircraft manufacturer and FMS manufacturer if discriminant validity was to be demonstrated. Conversely, since usability is not a user characteristic but is rather a product of users performing specific tasks in a specific environment, it was hypothesised that ratings on the usability factors should not differ by participants' rank or culture. For the factors to be valid predictors of usability and not, for example, attitudes to automation or workload, the ratings should reflect aspects of the equipment and not user characteristics.

4.4.1 Assumptions

Similar to PCA, DFA assumes a normal distribution of data and therefore consideration is given to sample size, homoscedasticity and heteroscedasticity, and multicollinearity.

Sample Size

Since DFA assumes a normal distribution of data, it warrants large sample sizes. Tabachnick & Fidell (1983) recommend as a general guideline, to include five times as many observations as there are variables to be analysed. With reference to DFA, Hair et al (1998) specified that if the sample sizes were small and the covariances were unequal, then the statistical significance of the estimation process was adversely affected. In the data set analysed here, there are at least five cases for each variable observed.

Homoscedasticity and Heteroscedasticity

As discussed in the previous section, the assumption of heteroscedasticity follows from the assumption of normality. When the assumption of normality is met, the relationship between variables can be said to be homoscedastic. Heteroscedasticity is a product of either non-normality of one of the variables or due to one variable not being directly rated to another variable.

Multicollinearity

As is the case in any multivariate technique, there is an implicit assumption that relationships between variables are linear. In this type of analysis non-linear relationships are not identified in the discriminant function unless specific variable transformations are made to represent non-linear effects.

4.4.2 Discriminant Validity Evidence - DFA Examining Impact of Aircraft and FMS Manufacturer

The primary goal of this four-group DFA was to discriminate between FMS manufacturers (Honeywell and Smiths) and aircraft manufacturers (Boeing, Airbus and Fokker) on the basis of respondents' ratings of the usability of their FMS CDU. Since FMS type is intrinsically linked to the aircraft manufacturer's design philosophy, it was hypothesised that those pilots type-rated on, for example, a Boeing aircraft with a Smiths FMS would rate the usability scales differently from those type-rated on an Airbus aircraft with a Honeywell FMS. A further goal of this four-group DFA was therefore to determine on which of the predictor components the aircrew type-rated on:

(i) Boeing aircraft with Smiths FMSs (Boeing Smiths)

(ii) Boeing aircraft with Honeywell FMSs (Boeing Honeywells)

(iii) Airbus aircraft with Honeywell FMSs (Airbus Honeywells) or

(iv) Fokker aircraft with Honeywell FMSs (Fokker Honeywells)

were maximally different, and to predict group membership on the basis of those predictor variables used to create the dimension.

It is important to reiterate at this point that the FMS and aircraft are integrated and could not be separated in a research methodology. It was not possible, for example, to use a fully crossed design since neither Fokker nor Airbus aircraft have a Smiths FMS installed as standard.

Of the original 252 cases, 31 cases were eliminated from the analysis on the basis that they had at least one missing discriminating variable. Table 4-14 shows the three discriminant functions calculated. The two significant discriminant functions on Table 4-14 accounted for 35.0% and 15.2%, respectively, of the between group variability.

The non-significant function, which is highlighted in Table 4-14, accounted for 6.6% of the total sample variance.

Table 4-14 - Canonical Discriminate Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*
1	0.539	0.592
2	0.180	0.390
3	0.070	0.256

Function	*Wilks' Lambda*	*Chi²*	*df*	*p*
1 - 3	0.514	141.249	33	0.000
2 - 3	0.792	49.581	20	0.000
3	0.934	14.426	9	0.108

Table 4-15 - Functions at Group Centroids

	Function		
Manufacturer	*1*	*2*	*3*
Boeing Smiths	-0.539	-0.060	0.336
Boeing Honeywell	-0.348	0.045	-0.298
Airbus Honeywell	1.310	0.544	0.034
Fokker Honeywell	0.555	-1.252	0.098

When the group centroids for the canonical discriminant functions were analysed, it became apparent that for the first function, the solution was maximally discriminating between the usability ratings of pilots type-rated on Boeing aircraft with Smiths FMSs (group centroid = -0.539) and those type-rated on Airbus aircraft with Honeywell FMSs (group centroid = 1.310).

For the second function, the solution was maximally discriminating between the usability ratings of pilots type-rated on Airbus aircraft with Honeywell FMSs (group centroid = 0.544) and those type-rated on Fokker aircraft with Honeywell FMSs (group centroid = -1.252).

For the final non-significant function (Table 4-14), the solution was maximally discriminating between the usability ratings of pilots type-rated on Boeing aircraft with Smiths FMSs (group centroid = 0.336) and those type-rated on Boeing aircraft with Honeywell FMSs (group centroid = -0.298).

On the first discriminant function the four predictor components, **Structure & Sequence**, **Appropriate Functionality**, **Flexibility & Control**, and **Processor Speed** had loadings in excess of 0.45, while on the second discriminant function the predictor components **Predictability** and **Information Presentation** had loadings in excess of 0.45.

Table 4-16 - Wilks' Lambda, Univariate F Ratios with 3,217 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients.

Variable	*Wilks' Lambda*	*F Ratio*	*p*	*DFA Coefficient 1*	*DFA Coefficient 2*	*DFA Coefficient 3*
Structure & Sequence	0.906	7.474	0.000	-0.522	0.392	0.218
Appropriate Functionality	0.929	5.554	0.001	-0.532	-0.050	0.152
Feedback	0.966	2.511	0.060	0.310	-0.068	-0.404
Predictability	0.962	2.823	0.040	-0.094	-0.504	0.149
Information Presentation	0.947	4.071	0.008	0.252	-0.527	0.031
Consistency	0.967	2.445	0.065	0.101	0.435	-0.280
Flexibility & Control	0.923	6.007	0.001	0.543	0.161	-0.075
Error Prevention & Correction	0.981	1.407	0.242	-0.204	0.247	-0.148
Processor Speed	0.873	10.553	0.000	0.571	0.346	0.591
Workload	0.989	0.773	0.510	-0.074	-0.076	0.372

The group means and standard deviations for each component (calculated from regression-based factor scores) are shown in Table 4-17, and permit comparisons to be made across aircraft/FMS combination. It should be noted that higher scores on the regression-based PCA scores used in this DFA indicate greater usability.

Table 4-17 - Group Statistics

	Boeing Smiths[a]		*Boeing Honeywell*[b]		*Airbus Honeywell*[c]		*Fokker Honeywell*[d]	
Factor	*Mean*	*sd*	*Mean*	*sd*	*Mean*	*sd*	*Mean*	*sd*
Structure & Sequence	0.274	0.876	0.119	0.990	-0.419	1.086	-0.582	0.755
Appropriate Functionality	0.245	0.934	0.067	1.016	-0.502	1.017	-0.039	0.730
Feedback	-0.242	1.073	0.030	0.933	0.245	0.918	0.175	1.090
Predictability	0.073	0.937	-0.098	1.027	-0.158	1.071	0.559	0.739
Information Presentation	-0.091	0.956	-0.146	0.954	0.153	1.087	0.658	0.916
Consistency	-0.119	0.939	0.120	1.036	0.142	0.987	-0.461	0.956
Flexibility & Control	-0.222	0.837	-0.072	1.032	0.539	0.944	-0.101	1.138
Error Prevention & Correction	0.030	0.924	0.120	1.044	-0.161	0.963	-0.300	1.101
Processor Speed	-0.093	0.917	-0.236	0.908	0.67	1.121	-0.413	0.674
Workload	0.144	0.928	-0.093	1.073	-0.050	0.999	0.033	0.906

[a]n=69, [b]n=89, [c]n=44, [d]n=19

On the first function, Boeing aircraft with Smiths FMSs were rated more favourably than Boeing or Fokker aircraft with Honeywell FMSs on the **Structure & Sequence**, **Appropriate Functionality** and **Processor Speed** factors.

Boeing aircraft with Smiths FMSs were rated more favourably than Airbus aircraft with Honeywell FMSs on the **Structure & Sequence** and **Appropriate Functionality** factors.

Boeing aircraft with Honeywell FMSs were rated more favourably than Airbus aircraft with Honeywell FMSs on the **Structure & Sequence** and **Appropriate Functionality** components, and were also rated more favourably than Fokker aircraft with Honeywell FMSs on the **Structure & Sequence**, **Appropriate Functionality**, **Flexibility & Control** and **Processor Speed** components.

Airbus aircraft with Honeywell FMSs were rated more favourably on the **Flexibility & Control** and **Processor Speed** dimensions than either Boeing aircraft with Smiths or Honeywell FMSs or Fokker aircraft with Honeywells FMSs.

On the second function (Table 4-16), Fokker aircraft with Honeywell FMSs were rated more favourably than Boeing aircraft with Smiths FMSs or Boeing or Airbus aircraft with Honeywell FMSs on the **Predictability** Component. Boeing aircraft with Smiths FMSs were rated more favourably than Boeing aircraft with Honeywell FMSs, which were in turn rated more favourably than Airbus aircraft with Honeywell FMSs on the **Predictability** component (see Table 4-17).

Fokker aircraft with Honeywell FMSs were rated most favourably on the **Information Presentation** components, followed by Airbus aircraft with Honeywell FMSs, Boeing aircraft with Smiths FMSs and Boeing aircraft with Honeywell FMSs.

Although the third discriminant function was non-significant, it can be seen from Table 4-16 that it contained a predictor component with a loading in excess of 0.45, **Processor Speed**, which was examined for completeness. Airbus aircraft with Honeywell FMSs were rated most favourably on the Processor Speed dimension followed by Boeing aircraft with Smiths, Boeing aircraft with Honeywell FMSs and Fokker aircraft with Honeywell FMSs.

Table 4-18 summarises the findings by providing a rank ordered overview of usability ratings by FMS and aircraft type for DFA functions 1 and 2.

Table 4-18 - Usability Ratings Ranked by Aircraft and FMS Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)			
Factor	*Boeing Smiths*	*Boeing Honeywell*	*Airbus Honeywell*	*Fokker Honeywell*
Structure & Sequence	1	2	3	4
Appropriate Functionality	1	2	4	3
Predictability	2	3	4	1
Information Presentation	3	4	2	1
Flexibility and Control	4	2	1	3
Processor Speed	2	3	1	4

Table 4-19 shows the classification Matrix for the DFA predicting group membership from aircraft and FMS manufacturer. Overall, a total of 53.4% of cases were correctly classified on the basis of whether the pilot currently held a type-rating on a Boeing aircraft with a Smiths FMS, or on a Boeing, Airbus or Fokker aircraft with a Honeywell FMS. Table 4-19 shows the classification matrix for the DFA. Correct classification percentages are shown in parentheses.

Table 4-19 - Classification Matrix for Discriminant Function Analysis Predicting Group Members from Aircraft and FMS Manufacturer

	Predicated Group Membership				
Actual Group	*Boeing Smiths*	*Boeing Honeywell*	*Airbus Honeywell*	*Fokker Honeywell*	*Total*
Boeing Smiths	34 (49.3%)	21 (30.4%)	4 (5.8%)	10 (14.5%)	69
Boeing Honeywell	24 (27.0%)	38 (42.7%)	13 (14.6%)	14 (15.7%)	89
Airbus Honeywell	2 (4.5%)	5 (11.4%)	33 (75.0%)	4 (9.1%)	44
Fokker Honeywell	3 (15.8%)	1 (5.3%)	2 (10.5%)	13 (68.4%)	19

In order to assess the predictive power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. This measure compares the

number of correct classifications with the total sample size and the number of groups. The calculated value is then compared with a critical value (the chi-square value for 1 degree of freedom) at the desired confidence level. Press's Q in this classification matrix was 95.02, where the critical value at a significance level of 0.01 was 6.63. The classification rate in this analysis was therefore statistically better than chance at the $p<0.01$ level. It should however be noted that while the classification matrix was significant in statistical terms, no practical evidence of criterion validity was demonstrated since 46.6% of cases were incorrectly classified overall. Furthermore, since 30.4% of Smiths FMSs on Boeing aircraft were incorrectly classified as being Honeywell FMSs, evidence suggested that in practical terms it was not possible to predict FMS type accurately.

Section Summary

The results of this DFA supported the hypothesis that participants' ratings of the questionnaire items, would be significantly affected by both the type of aircraft and FMS on which the pilot was currently type-rated. Evidence of discriminant validity was therefore demonstrated. While the classification matrix predicting group membership from aircraft and FMS manufacturer was statistically significant ($p.<0.01$), no practical evidence of criterion validity was demonstrated since it was not possible to predict aircraft and FMS type accurately.

4.4.3 Discriminant Validity Evidence - DFA Examining Impact of FMS Manufacturer

The purpose of this section was to further examine the evidence of discriminant validity from the factors produced from the PCA. The analyses described in this section were conducted on the basis that if the factors from which scales were to be developed could discriminate between FMS types, then evidence of discriminant validity would be provided. An associated goal was therefore to assess on which of the predictor components aircrew type-rated on aircraft with Smiths FMSs and Honeywell FMSs were maximally different.

Of the original 252 cases, 31 were eliminated from the analysis on the basis that they had at least one missing discriminating variable.

Table 4-20 - Canonical Discriminate Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*	*Wilks' Lambda*	*Chi²*	*df*	*p*
1	0.159	0.371	0.863	31.531	11	.001

A single significant discriminant function was calculated as shown in Table 4-20. The function accounted for 13.8% of the between group variability. The group centroids for the canonical discriminant functions of the usability ratings were –0.589 for those type-rated on aircraft with Smiths FMSs and 0.268 for those type-rated aircraft with

Honeywell FMSs. As can be seen from Table 4-21 three predictor components, **Structure & Sequence**, **Appropriate Functionality** and **Feedback** had loadings in excess of 0.45 on the discriminating function.

Table 4-21 - Wilks' Lambda, Univariate F Ratios with 1,210 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients.

Component	*Wilks' Lambda*	*F Ratio*	*P*	*DFA Coefficient Function 1*
Structure & Sequence	0.966	7.755	0.006	-0.528
Appropriate Functionality	0.973	6.150	0.014	-0.474
Feedback	0.973	6.059	0.015	0.470
Predictability	0.998	0.537	0.464	-0.144
Information Presentation	0.996	0.827	0.364	0.178
Consistency	0.994	1.426	0.234	0.233
Flexibility & Control	0.978	5.037	0.026	0.431
Error Prevention & Correction	1.000	0.090	0.764	-0.059
Processor Speed	1.000	0.009	0.926	0.018
Workload	0.991	2.077	0.151	-0.280

As can be seen from the group means (calculated from regression-based factor scores) in Table 4-22, the usability ratings of Smiths FMSs were more favourable than those for the Honeywell FMSs on both the **Structure & Sequence** and **Appropriate Functionality** components. Honeywell FMSs were rated more favourably than Smiths FMSs on the **Feedback** dimension. The highlighted components reflect those which failed to meet the 0.45 criterion on Table 4-21 and therefore were not interpreted.

Table 4-22 - Group Statistics

	Honeywell FMS[a]		*Smiths FMS*[b]	
Factor	*Mean*	*sd*	*Mean*	*sd*
Structure & Sequence	-0.124	1.030	0.274	0.876
Appropriate Functionality	-0.111	1.012	0.244	0.934
Feedback	0.110	0.948	-0.243	1.073
Predictability	-0.032	1.029	0.073	0.937
Information Presentation	0.041	1.020	-0.091	0.956
Consistency	0.054	1.024	-0.119	0.939
Flexibility & Control	0.101	1.053	-0.222	0.837
Error Prevention & Correction	-0.014	1.035	0.030	0.924
Processor Speed	0.042	1.038	-0.093	0.917
Workload	-0.065	1.027	0.144	0.928

[a]n=152, [b]n=69

To help summarise these findings, Table 4-23 shows a rank ordered overview of the usability ratings by the respondents' FMS type for the DFA Function.

Table 4-23 - Usability Ratings Ranked by FMS Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)	
Factor	*Honeywell FMS*	*Smiths FMS*
Structure & Sequence	2	1
Appropriate Functionality	2	1
Feedback	1	2

Overall, a total of 68.3% of cases were correctly classified on the basis of whether the pilot currently held a type-rating on an aircraft with a Smiths or a Honeywell FMS. Table 4-24 shows the classification matrix for the DFA. Correct classification percentages are shown in parentheses.

Table 4-24 - Classification Matrix for DFA Predicting Group Members from FMS Manufacturer

	Predicated Group Membership		
Actual Group	*Smiths FMS*	*Honeywell FMS*	*Total*
Smiths FMS	50 (72.5%)	19 (27.5%)	69 (100%)
Honeywell FMS	51 (33.6%)	101 (66.4%)	152 (100%)

Press's Q in this classification matrix was 29.69, where the critical value at a significance level of 0.01 was 6.63. While the classification rate in this analysis was statistically better than chance at the $p<0.01$ level, it should be noted that over a third of Honeywell FMSs were incorrectly classified as being Smiths FMSs.

Section Summary

The results of this DFA supported the hypothesis that the factors derived from participants' questionnaire ratings were significantly different based on the type of FMS on which the pilot was currently type-rated. Further evidence of discriminant validity was therefore demonstrated. While evidence of criterion validity was demonstrated based on the classification matrix being statistically significant ($p<0.01$), over a third of Honeywell FMSs were incorrectly classified as Smiths FMSs indicating that no practical evidence of criterion validity was demonstrated.

4.4.4 Discriminant Validity Evidence - DFA Examining Impact of Aircraft Manufacturer

With the goal of demonstrating discriminant validity, the hypothesis tested in this DFA was that the factor scores derived would differ significantly based on the aircraft type on which the respondents were currently type-rated. If the factors from which scales were to be developed could discriminate between aircraft types, then evidence of discriminant validity would be demonstrated.

Since FMS type is intrinsically linked to the aircraft manufacturer's design philosophy, it was hypothesised that those pilots type-rated on, for example, a Boeing aircraft would rate usability differently from those type-rated on an Airbus aircraft irrespective of FMS type.

A further goal of this three-group DFA was to determine on which of the predictor components the aircrew type-rated on Boeing, Airbus or Fokker aircraft were maximally different, and to predict group membership on the basis of those predictor variables used to create the dimension.

Of the original 252 cases, 31 were eliminated from the analysis as they had at least one missing discriminating variable.

Table 4-25 - Canonical Discriminate Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*	*Wilks' Lambda*	*Chi²*	*df*	*p*
1	0.529	0.588	0.556	125.08	22	0.000
2	0.177	0.388	0.850	34.673	10	0.000

Two significant discriminant functions were calculated as shown in Table 4-25. The two functions together accounted for 34.6% of the between group variability, while the second function alone accounted for 15.1%.

When the group centroids for the canonical discriminant functions were analysed, it became apparent that for the first function, the solution was maximally discriminating between the usability ratings of Boeing (group centroid = -0.419) and Airbus type-rated pilots (group centroid = 1.401). The pilots' ratings of usability reflected the differences between the aircraft on which the pilots were type-rated.

For the second function, the solution was maximally discriminating between the usability ratings of Airbus (group centroid = 0.213) and Fokker type-rated pilots (group centroid = -1.354).

Table 4-26 - Functions at Group Centroids

	Function	
Manufacturer	*1*	*2*
Boeing	-0.419	0.104
Airbus	1.401	0.213
Fokker	0.244	-1.354

Table 4-27 - Wilks' Lambda, Univariate F Ratios with 2,218 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients.

Component	*Wilks' Lambda*	*F Ratio*	*p*	*DFA Coefficient Function 1*	*DFA Coefficient Function 2*
Structure & Sequence	0.911	10.704	0.000	-0.515	0.409
Appropriate Functionality	0.934	7.670	0.001	-0.520	-0.039
Feedback	0.980	2.270	0.106	0.286	-0.110
Predictability	0.968	3.647	0.028	-0.073	-0.486
Information Presentation	0.947	6.069	0.003	0.264	-0.517
Consistency	0.977	2.520	0.083	0.074	0.404
Flexibility & Control	0.927	8.552	0.000	0.533	0.158
Error Prevention & Correction	0.982	1.957	0.144	-0.218	0.227
Processor Speed	0.882	14.614	0.000	0.602	0.419
Workload	0.999	0.075	0.928	-0.048	-0.035

On the first discriminant function, the four predictor components **Structure & Sequence**, **Appropriate Functionality**, **Flexibility & Control** and **Processor Speed** had loadings in excess of 0.45 as shown on Table 4-27. The group means and standard deviations calculated from the regression-based factor scores show the relative usability ratings of the Boeing, Airbus, and Fokker aircrew on the first function. These are shown in Table 4-28. The highlighted rows reflect the components that did not achieve the 0.45 criterion in Table 4-27.

Table 4-28 - Group Statistics

	Boeing[a]		Airbus[b]		Fokker[c]	
Factor	*Mean*	*sd*	*Mean*	*sd*	*Mean*	*sd*
Structure & Sequence	0.187	0.942	-0.419	1.086	-0.581	0.755
Appropriate Functionality	0.145	0.982	-0.503	1.017	-0.039	0.730
Feedback	-0.089	1.002	0.245	0.918	0.175	1.090
Predictability	-0.023	0.980	-0.158	1.071	0.559	0.739
Information Presentation	-0.122	0.953	0.153	1.087	0.658	0.915
Consistency	0.016	0.999	0.142	0.987	-0.461	0.956
Flexibility & Control	-0.138	0.952	0.539	0.944	-0.101	1.138
Error Prevention & Correction	0.081	0.991	-0.161	0.963	-0.23	1.101
Processor Speed	-0.137	0.916	0.670	1.121	-0.412	0.674
Workload	0.010	1.016	-0.051	0.100	0.033	0.906

[a]n=158, [b]n=44, [c]n=19

The larger normalised factor scores shown on Table 4-28 reflect how pilots type-rated on Boeing aircraft rated the usability of their FMS more favourably than those pilots type-rated on Airbus aircraft on both the **Structure & Sequence** and **Appropriate Functionality** dimensions. The questionnaire ratings made by Boeing type-rated pilots were also more favourable than the Fokker type-rated pilots on the **Structure & Sequence**, **Appropriate Functionality** and **Processor Speed** components.

Airbus type-rated aircrew made more positive usability ratings than Boeing type-rated aircrew on the questionnaire items that constituted the **Flexibility & Control** and **Processor Speed** dimensions. Their ratings were also more positive than the Fokker type-rated aircrew on the questionnaire items that constituted the **Structure & Sequence**, **Flexibility & Control** and **Processor Speed** components.

Finally, Fokker type-rated aircrew rated the usability of their FMS more positively than Boeing type-rated aircrew on the **Flexibility & Control** dimension, and were also more positive than Airbus type-rated aircrew on the **Appropriate Functionality** dimension.

As can be seen from Table 4-27, two predictor components, **Predictability** and **Information Presentation**, had loadings in excess of 0.45 on the second discriminating function. A review of the group means on Table 4-28, shows that Fokker type-rated pilots made more favourable usability ratings than either Boeing or Airbus type-rated pilots on both the **Predictability** and **Information Presentation** components.

On the second function, Boeing type-rated pilots made more favourable ratings than Airbus pilots on the **Predictability** component, while Airbus type-rated pilots made more positive ratings than Boeing type-rated pilots on the **Information Presentation** component.

To help summarise these findings, Table 4-29 shows a rank ordered overview of the usability ratings aircraft type for DFA functions 1 and 2.

Table 4-29 - Usability Ratings Ranked by Aircraft Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)		
Factor	*Boeing*	*Airbus*	*Fokker*
Structure & Sequence	1	2	3
Appropriate Functionality	1	3	2
Predictability	2	3	1
Information Presentation	3	2	1
Flexibility & Control	3	1	2
Processor Speed	2	1	3

Table 4-30 shows the classification matrix for the DFA (percentages are given in parentheses). Overall, a total of 79.2% of cases were correctly classified on the basis of whether the pilot currently held a Boeing, Airbus or Fokker type-rating.

Table 4-30 - Classification Matrix for DFA Predicting Group Members from Aircraft Manufacturer

	Predicted Group Membership			
Actual Group	*Boeing*	*Airbus*	*Fokker*	*Total*
Boeing	149 (94.3%)	6 (3.8%)	3 (1.9%)	158
Airbus	20 (45.5%)	22 (50.0%)	2 (4.5%)	44
Fokker	13 (68.4%)	2 (10.5%)	4 (21.1%)	19

In order to determine the predictive power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. Press's Q in this classification matrix was 209.09, where the critical value at a significance level of 0.01 was 6.63. The classification rate in this analysis was therefore statistically better than chance at the $p<0.01$ level. While an overall total of 79.2% of cases were correctly classified on the basis of whether the pilot currently held a Boeing, Airbus or Fokker type-rating, it should be noted that the usability ratings of many pilots type-rated on Fokker aircraft (78.9%) and Airbus aircraft (50.0%) were incorrectly classified. Only 21.1% of participants type-rated on Fokker aircraft were correctly classified. No practical evidence of criterion validity was demonstrated since it was not possible to predict participants' current aircraft type accurately.

Section Summary

The results of this DFA supported the hypothesis that that the factors derived from participants' questionnaire ratings were significantly different based on the type of aircraft on which the pilot was currently type-rated. Evidence of discriminant validity

was therefore demonstrated. While the classification matrix predicting group membership from aircraft manufacturer was statistically significant ($p.<0.01$), no practical evidence of criterion validity was demonstrated since the classification rates were notably poor for both Airbus and Fokker aircraft.

4.4.5 Discriminant Validity Evidence - DFA Examining Impact of Participants' Culture on FMS Usability Ratings

Research has shown that pilot culture impacts upon pilot attitudes towards aviation automation (e.g. Helmreich & Merritt, 1998). In this study, however, it was hypothesised that a reliable and valid usability measure would not discriminate between pilots' culture (characterised by the country in which the pilot resided), since usability "*is not a.... user characteristic, but emerges as a multidimensional characteristic in the context of users performing tasks with a product in a specific environment*" (Lewis, 1995, p.58).

The primary goal of this two-group DFA was therefore to demonstrate that participants' ratings were truly of FMS usability, and that it would be the pilots' FMS and aircraft types, and not individual characteristics such as culture, which determined usability ratings. The DFA also permitted the researcher to examine on which of the predictor components the aircrew based in North America and Europe were maximally different, and to predict group membership on the basis of those predictor variables used to create the dimension.

Of the original 252 cases, 31 were eliminated from the analysis as they had at least one missing discriminating variable. As can be seen from Table 4-31, a single non-significant function was calculated which accounted for only 8.4% of the variance. The discriminant function shown on Table 4-31 has been highlighted to show that it was non-significant.

Table 4-31 - Canonical Discriminate Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*	*Wilks' Lambda*	*Chi²*	*df*	*p*
1	0.091	0.289	0.917	18.486	11	0.071

The functions at group centroids are shown in Table 4-32.

Table 4-32 - Functions at Group Centroids

	Function
Country	*1*
North America	-0.287
Europe	0.314

The Wilks' Lambda, Univariate F Ratios and Standardized Canonical Discriminant Functions are shown on Table 4-33.

Table 4-33 - Wilks' Lambda, Univariate F Ratios with 1,218 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients

Variable	*Wilks' Lambda*	*F Ratio*	*P*	*DFA Coefficient1*
Structure & Sequence	0.997	0.606	0.437	0.194
Appropriate Functionality	0.993	1.559	0.213	-0.303
Feedback	1.000	0.017	0.896	0.032
Predictability	0.973	5.981	0.015	-0.583
Information Presentation	1.000	0.091	0.763	-0.068
Consistency	0.999	0.256	0.613	0.116
Flexibility & Control	0.991	1.740	0.161	0.343
Error Prevention & Correction	0.990	2.257	0.134	0.367
Processor Speed	1.000	0.099	0.753	-0.076
Workload	0.976	5.282	0.022	0.552

For completeness, the means and standard deviations for each component are shown on Table 4-34.

Table 4-34 - Group Statistics

	North America[a]		*Europe*[b]	
Factor	*Mean*	*sd*	*Mean*	*sd*
Structure & Sequence	-0.054	1.010	0.051	0.996
Appropriate Functionality	-0.080	0.962	-0.090	1.041
Feedback	-0.008	1.010	0.01	0.999
Predictability	0.156	0.925	-0.171	1.058
Information Presentation	0.014	0.935	-0.027	1.069
Consistency	-0.026	0.915	0.042	1.085
Flexibility & Control	-0.092	1.020	0.098	0.982
Error Prevention & Correction	-0.100	0.915	0.103	1.023
Processor Speed	0.020	0.952	-0.023	1.058
Workload	-0.152	0.972	0.155	1.007

[a]n=115, [b]n=105, total n=220

The results of this DFA illustrated that the country in which pilots were based had no significant impact on ratings of FMS usability. Although the DFA was non-significant, the classification matrix has been included in Table 4-35 for completeness.

Table 4-35 - Classification Matrix for DFA Predicting Group Members from Participants' Culture

	Predicated Group Membership		
Actual Group	*North America*	*Europe*	*Total*
North America	84 (73.0%)	31 (27.0%)	115 (100%)
Europe	45 (42.9%)	60 (57.1%)	105 (100%)

Overall, a total of 65.5% of cases were correctly classified on the basis of whether the pilot was based in Europe or North America. Percentages of correct classification are shown in parentheses. In order to determine the predictive power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. Press's Q in this classification matrix was 21.02, where the critical value at a significance level of 0.01 was 6.63. The classification rate in this analysis was therefore statistically better than chance at the $p<0.01$ level. However, since more than a third of cases were incorrectly classified overall, and since more than 42.9% of pilots based in Europe were incorrectly classified as being based in North America, no evidence of criterion validity was demonstrated.

Section Summary

The results of this DFA supported the hypothesis that the factors derived, which were the basis for developing usability scales, were not significantly different based on participants' culture. This is further evidence of discriminant validity from the usability factors derived, since the factors were not discriminating between user characteristics, indicating that it was the usability of the participants' FMS and not personal characteristics that significantly influenced ratings. While the classification matrix predicting group membership from the country in which participants were based was statistically significant ($p.<0.01$), no practical evidence of criterion validity was demonstrated.

4.4.6 Discriminant Validity Evidence - DFA Examining Impact of Pilot Rank on FMS Usability Ratings

The primary goal of this two-group DFA was to determine if ratings on the factors developed would discriminate between the respondent pilots on the basis of their rank (First Officer or Captain). It was hypothesised that usability ratings would not be significantly different based on the pilots' rank, and that it would be the pilots' FMS and aircraft types, and not individual characteristics, that would determine usability ratings.

Of the original 252 cases, 31 were eliminated from the analysis as they had at least one missing discriminating variable. As can be seen from Table 4-36, a single non-significant function was calculated which only accounted for 6.0% of the variance. The discriminant function shown on Table 4-36 has been highlighted to show that it was non-significant.

Table 4-36 - Canonical Discriminate Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*	*Wilks' Lambda*	*Chi²*	*df*	*p*
1	0.063	0.244	0.94	12.678	11	0.315

The group centroid for First Officers was 0.288 while being –0.217 for Captains. The Wilks' Lambda, Univariate F Ratios and Standardized Canonical Discriminant Functions are shown on Table 4-37.

Table 4-37 - Wilks' Lambda, Univariate F Ratios with 1,212 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients.

Variable	*Wilks' Lambda*	*F Ratio*	*p*	*DFA Coefficient Function1*
Structure & Sequence	0.999	0.274	0.601	-0.125
Appropriate Functionality	0.967	7.154	0.008	0.762
Feedback	1.000	0.044	0.834	0.051
Predictability	1.000	0.013	0.911	-0.021
Information Presentation	0.995	0.997	0.319	0.300
Consistency	1.000	0.011	0.916	0.014
Flexibility & Control	0.998	0.334	0.564	-0.173
Error Prevention & Correction	0.992	1.799	0.181	0.403
Processor Speed	0.998	0.464	0.496	0.179
Workload	1.000	0.023	0.880	0.047

For completeness, the group means and standard deviations for each component are described in Table 4-38.

Table 4-38 - Group Statistics

	First Officer[a]		Captain[b]	
Factor	*Mean*	*sd*	*Mean*	*sd*
Structure & Sequence	-0.017	1.030	0.055	0.965
Appropriate Functionality	0.207	0.911	-0.156	1.029
Feedback	0.030	0.924	0.011	1.064
Predictability	-0.030	1.013	-0.014	0.993
Information Presentation	0.084	0.892	-0.052	1.050
Consistency	-0.020	1.014	-0.015	0.990
Flexibility & Control	-0.060	0.912	0.024	1.054
Error Prevention & Correction	0.103	1.039	-0.083	0.973
Processor Speed	0.060	1.067	-0.034	0.970
Workload	0.093	1.061	-0.016	0.953

[a]n=92, [b]n=122, total n=214

The results of this DFA illustrated that there was no significant difference in the ratings of FMS usability by whether the participant was a Captain or a First Officer. Although the DFA was non-significant, the classification matrix has been included in Table 4-39 for completeness. Overall, a total of 61.7% of cases were correctly classified on the basis of whether the pilot was a Captain or a First Officer. Percentages of correct classification are shown in parentheses.

Table 4-39 - Classification Matrix for DFA Predicting Group Members from Rank

	Predicated Group Membership		
Actual Group	*Captain*	*First Officer*	*Total*
Captain	95 (77.9%)	27 (22.1%)	122 (100%)
First Officer	55 (59.8%)	37 (40.2%)	92 (100%)

In order to determine the predictive power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. Press's Q in this classification matrix was 11.68, where the critical value at a significance level of 0.01 was 6.63. While the classification rate in this analysis was therefore statistically better than chance ($p<0.01$), pilot rank could not be predicted accurately and almost sixty percent of First Officers were incorrectly classified as Captains.

Section Summary

The results of this DFA supported the hypothesis that that the factors derived from participants' questionnaire ratings were not significantly different based on the pilot's rank. This was further evidence of discriminant validity since the factors were not discriminating between user characteristics, indicating that it was the usability of the

participants' FMS and not personal factors that significantly influenced ratings. While pilot rank could be predicted at a rate significantly better than chance ($p<0.01$), more than a third of cases were incorrectly classified overall, and almost 60% of First Officers were incorrectly classified as being Captains. From a practical perspective it was not possible to predict whether respondents were Captains or First Officers on the basis of their questionnaire ratings.

4.5 Reliability Analysis

In order to examine the internal consistency of the factors from which the scales were to be developed, reliability analyses in the form of Cronbach's Alphas were conducted.

The internal consistency of a factor refers to the extent to which its items are intercorrelated. High inter-correlations imply that the items are measuring the same construct. High inter-correlations between factor items further imply a strong relationship between the factor items and the latent construct. Internal consistency is most usually measured by Cronbach's coefficient alpha ($\propto$) (1951). Hair et al (1998) suggested that Cronbach's alphas with values in the range of 0.6 to 0.7 were within the lower limit of acceptability for scale construction. The results of the reliability analyses are shown in Tables 4-40 to 4-49.

Table 4-40 - Reliability Coefficient - Structure & Sequence

Dimension	*Reliability Coefficient*	*Number of Items*
Structure & Sequence	Alpha = 0.903	11

Questionnaire Items	*Alpha if Item Deleted*
21. It is clear why the system is organised and structured as it is.	0.887
20. The structure of the FMS is obvious to the user.	0.889
23. The structure of the FMS is well organised from the user's point of view.	0.885
22. It is clear why the FMS screens are sequenced as they are.	0.891
10. The sequence of activities required in the use of the FMS is as I would expect.	0.895
11. The FMS works the way I think it should work.	0.893
42. It was easy to learn how to use the FMS.	0.904
30. The FMS allows the users to do the required task as efficiently as possible.	0.897
12. The FMS buttons/switches could be located more intuitively.	0.897
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	0.896
38. The FMS keyboard is difficult or awkward to use.	0.901

The questionnaire item, "It was easy to learn how to use the FMS", seemed to be an inappropriate item in terms of content within the **Structure & Sequence** component.

The reliability analysis further demonstrated that the internal consistency would increase from 0.903 to 0.904 if this item was deleted. The item has not been included in the factors used in the subsequent study.

Table 4-41 - Reliability Coefficient - Appropriate Functionality

Dimension	*Reliability Coefficient*	*Number of Items*
Appropriate Functionality	Alpha = 0.823	4

Questionnaire Items	*Alpha if Item Deleted*
25. FMS users can access all the information required during flight phases.	0.757
26. The FMS screens provide all the information which is relevant to the task.	0.748
24. The FMS performs all the functions that I need it to.	0.816
27. The system allows users to do what they feel is necessary in order to carry out a task.	0.785

The Cronbach's alpha for the **Appropriate Functionality** component shown in Table 4-41 is well above the lower band of acceptable values for internal consistency of a factor.

Table 4-42 - Reliability Coefficient - Feedback

Dimension	*Reliability Coefficient*	*Number of Items*
Feedback	0.746	5

Questionnaire Items	*Alpha if Item Deleted*
18. FMS error messages explains clearly where errors are, what the errors are and why they have occurred.	0.664
19. It is clear to the user what should be done to correct any errors that occur.	0.667
13. The feedback messages displayed by the FMS are frequently inadequate.	0.701
39. The FMS help facility (either on-line or manuals) is adequate.	0.718
41. Changes in one part of the system can have adverse effects on other parts of the system.	0.747

The questionnaire item, "Changes in one part of the FMS may produce subsequent adverse effects on other parts of the system", did not seem immediately related in terms of content to the remainder of the items in the **Feedback** component. The PCA also demonstrated the item had the lowest loading on the **Feedback** component. Further, the internal consistency of the component demonstrated that the component reliability would increase from 0.746 to 0.747 if the item was deleted. This item was treated the same as the item "It was easy to learn how to use the FMS" on Table 4-40, and has therefore not been included in the subsequent study.

Table 4-43 - Reliability Coefficient - Predictability

Dimension	*Reliability Coefficient*	*Number of Items*
Predictability	0.736	4

Questionnaire Items	*Alpha if Item Deleted*
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	0.634
16 When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	0.626
9. The FMS sometimes makes unexpected actions.	0.695
43. Difficult to remember where information is within architecture.	0.739

The questionnaire item, "It is sometimes difficult to remember where information is located within the FMS architecture", did not seem to fit within the predictability component in terms of content. It contributed the lowest loading to the component (0.483), and the reliability analysis demonstrated that if the item were deleted from the component the internal consistency would increase from 0.736 to 0.739. The item has not been included in the subsequent study.

Table 4-44 - Reliability Coefficient - Information Presentation

Dimension	*Reliability Coefficient*	*Number of Items*
Information Presentation	0.709	4

Questionnaire Items	*Alpha if Item Deleted*
6. Abbreviations, acronyms and codes are used consistently throughout the FMS.	0.671
3. Important information is adequately highlighted on screen.	0.673
7. The format and method of entering information is consistent throughout the operation of the FMS.	0.629
4. Important information is organised logically on screen (familiarly aside).	0.606

The Cronbach's alpha for the **Information Presentation** component is well above the lower band of acceptable values for the internal consistency of a factor.

Table 4-45 - Reliability Coefficient - Consistency

Dimension	*Reliability Coefficient*	*Number of Items*
Consistency	0.723	2

Questionnaire Items
5. The way the FMS (CDU) looks and works is consistent at all times.
8. The way the FMS responds to a particular action is consistent at all times.

The Cronbach's alpha for the **Consistency** component described in Table 4-45 is well above the lower band of acceptable values for internal consistency of a factor. It should be noted that on this dimension no "Alpha if item deleted" has been included for the coefficient since it is not possible to remove one item when only two items are contained within a factor.

Table 4-46 - Reliability Coefficient - Flexibility & Control

Dimension	*Reliability Coefficient*	*Number of Items*
Flexibility & Control	0.635	3

Questionnaire Items	*Alpha if Item Deleted*
31. It is easy to return to the main menu from any screen of the FMS.	0.560
32. I feel in control of the FMS at all times.	0.551
34. The FMS is adequately flexible for all vertical navigation.	0.497

The Cronbach's alpha for the **Flexibility & Control** component described in Table 4-46 is lower than that which would be generally desirable, but is above the lower band of acceptable values for internal consistency of a factor (Hair et al, 1998).

Table 4-47 - Reliability Coefficient - Error Prevention & Correction

Dimension	*Reliability Coefficient*	*Number of Items*
Error Prevention & Correction	0.607	3

Questionnaire Items	*Alpha if Item Deleted*
36. The FMS has facilities for detecting and handling input errors.	0.411
37. The FMS is protected against common trivial entry errors.	0.507
14. The FMS makes it clear what changes will occur as a result of an input being made.	0.599

The Cronbach's alpha for the **Error Prevention & Correction** component described in Table 4-47 is lower than that which would be generally desirable, but is above the lower band of acceptable values for internal consistency of a factor (Hair et al, 1998).

Table 4-48 - Reliability Coefficient - Processor Speed

Dimension	*Reliability Coefficient*	*Number of Items*
Processor Speed	0.753	2

Questionnaire Items
40. It is irritating to have to input information at a rate slow enough for the FMS processor.
29. The FMS processes information too slowly for other activities in the flight deck.

The Cronbach's alpha for the **Processor Speed** component described in Table 4-48 is well above the lower band of acceptable values for internal consistency of a factor. Note that again no "Alpha if item deleted" has been included for this coefficient since it is not possible to remove one item when only two items are contained within the dimension.

Table 4-49 - Reliability Coefficient - Workload

Dimension	*Reliability Coefficient*	*Number of Items*
Workload	0.700	2

Questionnaire Items
41. The FMS can increase mental workload at inappropriate times.
42. The FMS can increase physical workload at inappropriate times.

The Cronbach's alpha for the **Workload** component described in Table 4-49 is above the lower band of acceptable values for internal consistency of a factor. Note that again no "alpha if item deleted" has been included for this coefficient either since it is not possible to remove one item when only two items are contained within the dimension.

Based on these analyses, it was determined that a number of the questionnaire items would not be included in the subsequent study. The questionnaire items loading on Component 11, "I prefer to stick to the functions and facilities I know best" and "Established conventions are followed for the format in which information is displayed", did not seem to be related to each other in terms of content and were therefore eliminated from the subsequent study.

5 STUDY ONE DISCUSSION

5.1 Introduction

The overall aim of this study, as set out Section 1.13, was to develop a series of reliable, valid factors that could be used to develop scales to assess the usability of FMS CDUs. The rationale for the development of these scales was that despite the vast quantity of literature that has been produced about the usability concept, little has been written concerning the usability of the civil flight deck. Furthermore, whilst a number of techniques had been developed for assessing the usability of computers, most had evolved around the desktop environment, in the office context, or around consumer products such as VCRs. Even those usability scales that claimed to be generic, such as the SUMI developed by Kirakowski et al (1993), were not applicable to the flight deck environment due to their office-based orientation (Reuzeau, 1997). Using generic questionnaires would have meant that the terms "flight deck" or "FMS" would not have been mentioned, and respondents would have faced questions concerned with, for example, the "windows environment" which were clearly not applicable to the FMS CDU being evaluated. Finally, many of the existing usability questionnaires were very long and may have appeared very unwieldy to participants. Since it has been reported that response rates gained from postal questionnaire are higher when the relevance of the questionnaire is evident to participants, a generic measure may have threatened the response rate achieved (Miller, 1991). Existing usability checklists such as that developed by Ravden & Johnson (1989) did contain items, which could be adapted for the flight deck. The checklists were however intended for use by interface designers and not end users, and also lacked the psychometric properties associated with the scales that this research intended to produce.

From the literature, it was hypothesised that since usability was "*not a unidimensional product or user characteristic, but emerges as a multidimensional characteristic in the context of users performing tasks with a product in a specific environment*" (Lewis, 1995, p.58), it would be influenced by the type of FMS and aircraft the respondent used. Conversely, it was hypothesised that if the ratings truly reflected the usability of the FMS, the pilots' rank or culture (characterised by country of residence) would have no significant impact on participants' usability ratings since these factors were user characteristics rather than properties of the FMS CDU being assessed.

In terms of generalizability, it was useful that respondents were of such varied experience. The sample was not restricted to pilots of very little or a very great deal of experience. It was also particularly useful that the sample was made up of respondents which were Captains / First Officers and from UK / North American airlines since analysis could be conducted to examine the hypothesis that the usability concept was independent of the assessors' characteristics. Evidence of discriminant validity would be provided if the factors derived could discriminate between that which they should be able to discriminate between (aircraft & FMS manufacturers), but did not discriminate between that which it should not discriminate (the personal attributes of the respondent

in terms of culture or rank). Similarly it was useful that respondents were type-rated on a range of aircraft, since analyses could be undertaken to determine if the usability factors could be used to distinguish between the ratings of participants type-rated on different aircraft and FMSs.

5.1.1 Interpretation of Descriptive Analyses

In reviewing the analyses it is important to consider the analyses both in absolute and relative terms. From the descriptive analyses (see Table 4-7) it is interesting to note that some items consistently received positive ratings. These included "The information in the FMS display (CDU) is easy to see and read"; "Abbreviations, acronyms & codes are used consistently throughout the FMS"; "When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done"; "It is clear to the user what should be done to correct any error that occurs"; "It is easy to return to the main menu from any screen of the FMS" and "The FMS is pleasant to use; I like it." It is not very surprising that pilots rated this latter item positively, since clearly life without an FMS would make the task of flying much more cumbersome.

Correspondingly, some items were notably rated towards the unusable end of the scale. These were: "FMS error messages explain clearly where errors are, what the errors are and why they have occurred"; "The FMS help facility (either on-line or manuals) is adequate"; "It is irritating to have to input information at a rate slow enough for the FMS processor" and "The FMS can increase mental workload at inappropriate times". It was interesting to note that some items were rated toward the unusable end of the scale. The first two of these items loaded onto the feedback component in the PCA, while the latter items loaded onto the processor speed and workload components respectively. This trend would be examined again in the subsequent study.

5.1.2 Interpretation of Factor Analysis

A PCA was conducted to explore the factor structure underlying the questionnaire items. The overall KMO test of sampling adequacy was very high and indicated that the data were suitable for analysis using the PCA technique. The factor structure elicited was subjected to a varimax rotation method and was shown to contain ten components. The components derived were labelled by establishing similarities with labels from the literature. For components that did not correspond to factors previously identified in the literature, labels were derived based on the content of the items that constituted the component. This will be expanded upon subsequently.

Structure & Sequence

The first factor of FMS CDU usability derived from the PCA was labelled **Structure & Sequence**, and was concerned with the FMS page architecture being structured such that it was clear how it should be operated. This factor accounted for the largest amount of variance of variance (30.4%). Williges, Williges & Fainter (1988) defined system structure in terms of it assisting the user to develop a conceptual representation that

could be used when navigating through the system interface. Shneiderman (1992) similarly referred to structure and sequence interface issues in terms of "Semantic Organisation" which he asserted referred to "*sensible, comprehensible, memorable, and convenient semantic organization relevant to the user's task*" (p.99). As most usability research has focused on PC-based environments, most of the **Structure & Sequence** type issues which have been discussed in the literature have focused on Windows interfaces and their associated menu hierarchies (e.g. Nielsen, 1993).

Eleven of the questionnaire items loaded onto this factor. The items, which loaded onto this component, were similar to those that constituted the Explicitness component referred to by Ravden & Johnson (1989), who argued that the way a system works and is structured should be evident (explicit) to the user.

Reviewing the items that loaded onto this factor, most were concerned with the FMS structure and architecture. Other items that loaded onto this factor referred to the FMS working the way participants thought it should, and that there was a clear logic between frequently used parts of the FMS and ease of access to those parts. The complete list of items, which loaded onto the Structure & Sequence factor, are shown in Table 4-12.

The **Structure & Sequence** factor structure was supported by the literature (e.g. Jordan, 1998), and further had a high internal consistency coefficient indicating that all the items in the factor were measuring the same construct. The item that looked inappropriate in terms of content was "It was easy to learn how to use the FMS". When the internal consistency of the component was examined it became clear that this was indeed an inappropriate item. The ratings of the remaining individual questionnaire items that loaded onto this factor were largely rated toward the poor usability end of rating scale (see Table 4-7).

Evidence from the interview transcripts suggested that **Structure & Sequence** is an appropriate construct of FMS CDU usability as can be seen from the following quotes:

"There are two start pages and very often if you want to enter fuel/information in the second of the two start pages you can't until the first start page has been filled out.....the first page is actually to do with things like zero fuel weight, fuel that has been loaded, and the trim position, for example, on the aircraft. And the second page is more, when you are aligning the IRSs [Inertial Reference Systems]. The first thing you want to do when you get on the aircraft is to align the IRSs because that gives you the flying display, the navigation display and brings up all the bits and pieces. The cockpit is now in an eye pleasing manner!! But you can't program the IRSs until you have the fuel information in but the fuel information, for example, zero fuel weights is probably the last piece of information you get before you close the doors for going because that is dependent on the number of passengers."

"...you probably have two or three steps to get to the information you want. You might have to go to one page, you then might have to give that particular page some information, having given it some information to get to the bit that you actually really want you then have to do a lateral selection. So there would be three steps there to get to the actual information you want".

"You can see how we can end up with ten or twelve button presses just to find a function".

The limited display area available in the FMS CDU, and the quantity of data it contains emphasises the "*Keyhole effect*" discussed by Woods (1995) and means that layers of data are unavoidable (p.167). The design aim should thus be to minimise the number of layers and to keep the pilot aware of the current position within the FMS structure at all times. On some FMS designs pilots have to navigate through up to five pages of menus, each page incorporating up to eight selectable functions with a multiple page menu for each. Singer (1999) argued that since the positioning of these functions is not logical, the pilot "*needs to search by means of 'trial and error' through up to 10-15 menus until finding the right function. Since no cues are available for this search the pilot has to rely on memory or perform the time consuming search*" (p.99). Furthermore as evident from the Kegworth accident discussed in Section 1.2, operators are not good at data entry and error identification in stressful, time-critical situations. In that scenario the pilots struggled to program the FMS while simultaneously informing their company of their problems and intentions, responding to ATC altitude and heading instructions, and obtaining weather information for East Midlands Airport (AAIB, 1990). Molich & Nielsen (1990) argued that such trial and error was a characteristic that increases workload and frustration and should therefore not be a feature of a modern interactive system.

As discussed in the literature review, authors such as Sarter & Woods (1992) have argued that a trade-off exists between providing pilots with flexibility within FMS page structures and sequences and imposing additional workload. The inability to enter zero fuel weight before aligning the IRSs does however provide a good example of the **Structure & Sequence** issues that are associated with the FMS CDU. One of the sources of **Structure & Sequence** problems may be related to the different way in which the programmers who design the databases, and the pilots who use them on a daily basis think both in terms of data flow and parameters required for the next task. Successful FMS designs need to be organised such that it is possible within each CDU page to select 'NEXT', 'PREVIOUS' and 'MAIN MENU' page options (Singer, 1999). Some FMS designs are now incorporating this fundamental usability requirement.

Appropriate Functionality

The second factor elicited from the PCA was termed **Appropriate Functionality**. Ravden & Johnson (1989) used the term **Appropriate Functionality** to define the need for a system to meet the requirements of users when carrying out tasks which corresponded to the items which loaded onto this factor. Johnson (1996) discussed the importance of prioritising functionality and information on the basis of, for example, how often particular functions are used or the comparative importance of different functions to the user. Eason (1984) similarly described a usability component, task match, which he described as the match between the functionality provided and the needs of the user. The items that loaded onto this factor, as shown on Table 4-12, were all clearly related to the issues proposed by these researchers.

The internal consistency of the **Appropriate Functionality** factor was well above the minimum recommended level for scale development indicating that all the items were highly related and were measuring the single construct referred to as **Appropriate Functionality**. Again, evidence from the interview transcripts suggested that this component was pertinent to the FMS CDU, as can be seen from the following extracts:

"The screen will only give you five waypoints and if you take the "Direct To" function, if the waypoint you want to go to is number six you then obviously have to scroll down or you have to physically spell it in here".

"What Air Traffic require of you is way beyond the capability of the FMS, to respond that quickly and to program that quickly".

The types of **Appropriate Functionality** problems reported by participants in this study have also been reported in the aviation automation literature. Singer (1999), for example, discussed how the 'UNDO' function which most PC users expect to be present in software packages is absent in many of today's FMS CDUs. In some FMSs, Singer argued, it was possible to select the wrong line select key and delete the entire flight plan without having any opportunity to undo or review that action. Some of the Appropriate Functionality issues presented may also have arisen from the use of "add-on" FMSs, which are missing features present in integrated systems due to older aircraft lacking modern displays and the fast avionics buses present in newer aircraft.

Feedback

Many researchers have discussed the **Feedback** dimension of usability and the problems associated with weak feedback are a well-known deficiency of human-computer interfaces. Feedback was defined by Jordan (1998) as being concerned with the fact that users interacting with a system should always be provided with informative feedback on their location within a system, what actions they have taken and whether these actions have been successful. The problems associated with inadequate feedback have also been cited frequently in the aviation automation literature e.g. Billings (1997). Sarter & Woods (1992) in discussing FMSs outlined how weak feedback on system status and behaviour was a common deficiency which led pilots to have difficulties "*finding, integrating and interpreting all the relevant data from*" from the CDU pages (p.317). An associated outcome, Sarter & Woods (1992) argued was that pilots could lose their situation awareness and be forced to ask *"What is it doing? Why did it do that? and What will it do next?"* (p. 317).

It is necessary for feedback to be informative. FMS CDUs can, for example, refuse to accept incorrectly formatted entries, but do not provide feedback as to why data entries are rejected. System response times should normally be as fast as possible, but not so fast as to make it difficult for the user to keep up with the feedback (Nielsen, 1993). Where a system cannot provide fairly immediate feedback, a "percentage done" indicator should be used to provide the user with feedback on how long the operation being performed will take to complete. Informative feedback is also required when systems fail. Nielsen (1993) reported that most systems are not designed to do so and

simply stop responding to the user when they fail. Singer (1999) also highlighted the importance of improved feedback for FMSs users.

The questionnaire items, which loaded onto this component, as shown on Table 4-12 were clearly related to the feedback issues discussed in the literature. The internal consistency of items was also well above the acceptable limit for scale development (See Table 4-42). Evidence from the interview transcripts also suggested that feedback was a notably deficient part of FMS CDU interaction. *"Invalid entry"* is a common feedback message from the CDU and as one interviewee reported *"it never tells you how to reformat the entry"*. This is also related to the **Error Prevention & Correction** dimension of usability that will be discussed subsequently.

The final questionnaire item which loaded on the **Feedback** component, "Changes in one part of the FMS may produce subsequent adverse effects in other parts of the system", did not seem to be so clearly related to **Feedback** in terms of content. The internal consistency analysis provided further evidence that reliability would be increased if this item was eliminated.

Predictability

The next component derived from the factor analysis could best be described as "**Predictability**". Billings (1997) argued that within the automation system the pilot "*must be able to predict how the aeroplane will be affected by that automation, not only at the time of selection but throughout the flight*" (p.241). Although, the term "**Predictability**" is not frequently referred to as a component of usability, it is a very important dimension in a dynamic system such as the flight deck (Sarter & Woods, 1992). Boy & Cacciabue (1997) defined predictability as "*the ability to anticipate consequences of actions on highly automated systems*" (p.70).

Ravden & Johnson in 1989 discussed a usability component with similar attributes under their checklist heading "System Usability Problems". They did not define the term system usability but rather suggested a list of system problems which system users could rate in categories of "no problems", "minor problems" or "major problems". Items they included were "*An inflexible, rigid system structure*" and "*Unexpected actions by the system*" (p.42/43). Sarter & Woods (1995) argued that FMS predictability problems were associated with pilots frequently asking the question the "what will it do next?". The problems, they argued were related to serious shortcomings in pilots' understanding of the FMS on board their aircraft. Courteney (1999), however, argued that the cliché "*what's it doing now?*" would in the future be replaced by "*'look what it's doing now', 'oh no, it's doing that again!' or even 'we'll have to work around what it's going to do next'*" (p.44). Courteney argued that these predictability questions were associated with flaws in the FMS rather than an incomplete understanding of the FMS by pilots.

The items included on the **Predictability** component from the PCA are shown on Table 4-12. Reviewing the components of FMS CDU usability elicited from this study, many of the components previously cited in the literature have been derived within the usability framework derived in this study. The predictability component, however, is

not frequently referred to in the literature as a dimension of usability. Javaux, Masson & De Keyser (2000) argued that predictability was an essential principal which any advanced aviation system should conform to. It may be that its importance within the FMS CDU usability framework is due to the safety critical nature and dynamic environment in which the FMS is operated.

The questionnaire item, "It is sometimes difficult to remember where information is located within the FMS architecture", did not seem to be related in terms of content to the other three items which loaded onto the **Predictability** component. The internal consistency analysis (Table 4-43) provided further evidence that the reliability of the component would be increased if this item was eliminated. The item was not included in the subsequent study.

Evidence derived from the interview transcripts supported the applicability of the **Predictability** component with comments from the participants including "You do sit there sometimes and think what is it doing?".

Consistency

Consistency was the next usability factor elicited from the analysis. The questionnaire items which loaded onto this factor are shown on Table 4-12, and the internal consistency coefficient shown on Table 4-45 demonstrated that the items loading onto this factors were all measuring the same construct. Evidence for the appropriateness of the **Consistency** component was derived from the interview transcripts where one pilot commented that "*As you can see, when we are setting up things are all over the place. Sometimes they are on the right of the screen, sometimes they are on the left. Certainly a common standard between the manufacturers would be good.*"

Consistency is concerned with the fact that when using a system it should look consistent, and similar tasks should be performed in similar ways (Jordan, 1998). Nielsen (1993) argued that **Consistency** was one of the most important usability characteristics and should be applied across all system media including the total user interface, the associated documentation and any on-line help. **Consistency** should be applied to successive releases of a product, and across entire product families. Such consistency across product ranges has been incorporated by Airbus Industrie, allowing pilots to undertake a common type-rating for a series of aircraft types.

Shneiderman (1992) argued that the consistency principle was frequently violated but the advantages to its adherence were large. The author went on to assert that "*Consistent sequences of actions should be required in similar situations; identical terminology should be used in prompts, menus, and help screens; and consistent commands should be employed throughout*" (p.72). Evidence from the aviation automation literature suggests that many FMS designs violate the **Consistency** principle. Singer (1999) argued that improved consistency was one area in which in which FMSs could benefit greatly from use of existing guidelines such as those identified by Molich & Nielsen (1990). User interface standards can play an important role in achieving consistency within and across products (Nielsen, 1993). However, no standard exists in positioning of FMS screens, line select keys, alphanumeric keys or

specific function keys (Singer, 1999). The findings of this research indicated that the **Consistency** dimension of usability is clearly an issue of importance in the FMS CDU context.

Error Prevention & Correction

The questionnaire items, which loaded onto the **Error Prevention & Correction** dimension, are shown in Table 4-12 and the internal consistency was at the acceptable level for scale development (see Table 4-47). As discussed in the literature review, Nielsen (1993) argued that errors that occur when using a system are critical to usability for two reasons. Firstly, they identify situations where users have trouble using a system. Secondly, they present an opportunity to assist the user understand the system better. Systems should correspondingly have good error recovery or correction, and users should be able to reverse their actions easily (Nielsen, 1993).

The **Error Prevention & Correction** dimension is distinct from the **Feedback** dimension of usability in that its focus is on preventing errors before they occur. The **Feedback** dimension is concerned with providing feedback which is not necessarily error related in that it might, for example, provide an indication that the FMS is processing information or that a resynching process is occurring between the FMSs.

Evidence in support of the applicability of this component can be gained from the following interview extract:

"Interviewer: Does the FMS minimise the possibility of you making an error?

Interviewee: No. Rubbish in, rubbish out. With our Jumbos one of our skippers has just resigned. He put the wrong departure into the FMS leaving Hong Kong.

Interviewer: Does the system have a mechanism for detecting and handling those errors that do occur?

Interviewee: Not really, no.

Interviewer: Does the FMS allow you to check your inputs and correct errors before those inputs are processed?

Interviewee: It will allow a certain amount of information to be displayed in the scratchpad, but if it can't be displayed in the scratch pad no it won't. The scratchpad is only one line anyway."

Another interviewee commented "The last gross error check that we tend to do is we check that the FMS route distance is about the same as our flight plan distance i.e. being within ten miles. I once found that it was out by a long way, by three or four hundred miles. Went through the whole route and could not find anything wrong with it. I re-entered the route and it came out correct. What was wrong with it I don't know."

Singer (1999) argued that the conditions under which the FMS is operated meant that errors and slips were common due to turbulence, parallax errors and procedural errors, and that it was essential for the FMS to be fault tolerant and for pilots to review inputs made to the CDU. Many FMS CDUs do not permit inputs made by one pilot to be cross-checked by the other pilot. This combined with the time pressure, the multiple simultaneous activities to be performed and the background of unambiguous normal practice under which pilots operate the FMS, makes it even more important that good error prevention and correction mechanisms are built into FMSs. Evidence from this study and the flight deck automation literature (e.g. Dekker & Orasanu, 1999) combine to suggest that well accepted Error Prevention & Correction guidelines such as those proposed by Nielsen (1993) and Shneiderman (1992) might be usefully considered by designers of future FMS CDUs.

Information Presentation

The next component derived from the analysis was **Information Presentation.** Johnson (1996) also referred to a similar component that was termed "*visibility*" (p.183). The questionnaire items that loaded onto this component were shown on Table 4-12, while the internal consistency coefficient, which was well within the range for scale development, was shown on Table 4-44.

Evidence from the interview transcripts indicated that FMS CDUs do not have all the attributes required for presenting information effectively as can be seen from the following quotes:

"You can't see it so well when the sun is directly on the screen".

"Sometimes the screen gets very dusty and there's a lot of static around it and you lose quite a lot of clarity that way".

"You can only get about half a dozen departures on a page, so you end up with ten pages of departures".

Billings (1997) noted that "*considerable attention*" (p.117) in terms of research and development was warranted to FMS CDU displays. The findings of this study support Billings assertion and indicate that **Information Presentation** is clearly a dimension of usability applicable to FMS CDUs.

Processor Speed

Early human factors work with computers found that system response time was an important determinant of both user acceptance and usability (Bailey, 1982). More recent literature has not however focused on processor speed. It is likely that this is largely due to the fact that so much of usability research has focused on the desktop environment, in which processor speeds have increased enormously over the past twenty years. Nielsen (1993) discussed processor speed in terms of response times arguing that system response times should "*be as fast as possible*" (p.135), but should not be so fast that the user cannot keep up with the system's responses. Miller (1968) argued that 0.1

seconds is about the limit for the user to feel a system is responding instantaneously, 1.0 seconds is about the limit for the user's flow of thought to remain uninterrupted, while 10 seconds is about the limit for maintaining a user's attention on the dialogue. These recommendations remain valid (e.g. Nielsen, 1993).

It is most probably due to the dynamic environment in which the FMS is operated that processor speed is so critical to the usability of FMSs. Significant time can be lost as pilots sometimes have to wait for their inputs to be processed before a second series of data entries can be made. In time critical situations, time spent trying to reprogram the FMS can have disastrous consequences such as those seen in the Kegworth accident (AAIB, 1990). The questionnaire items that loaded onto this component were shown at Table 4-12.

The problems associated with processor speed were discussed during the interview stage of this research with interviewees making comments such as:

"It is noticeably slow and it is frustrating".

"My biggest bug about the whole of the system is that the hardware itself is too slow. It is miles too slow".

"What I do find irritating is that it is too slow. There is no two ways about it. It is very slow. It gets a bit frustrating when the screen blanks, and when the screen does come back it comes back on the scratchpad saying "standby one". And it then might say "re-synching of FMS" before it does anything, which takes it twenty-five seconds. It is a pain in the neck".

As can be seen from Table 4-7, the means and standard deviations for items loading onto the processor speed component (questionnaire item 31 and 46) were well below the mid-point of the rating scale indicating that processor speed was rated as having poor usability. In light of the findings of this study and the discussion presented in this section, indications are that FMS **Processor Speed** is poor and has a negative effect on FMS CDU usability.

Flexibility & Control

Although **Flexibility & Control** could be considered as separate components, they are highly related to each other since the level of flexibility associated with using a system can determine the level of control the operator feels in using the system. As discussed in the literature review, Lin, Choong & Salvendy (1997) refer to flexibility in terms of the interface being able to adapt to users' requirements thus giving the users control. Systems should be adaptable in terms of how information is presented and what users can do in order to suit the needs of all users, and to allow them to feel in control of the system they are interacting with. Individually the components of flexibility and control have been referred to by many authors including Nielsen (1993) and Kirakowski (1996). In the flight deck domain, Javaux et al (2000) have argued that flexibility is generally appreciated by users although a trade-off exists between providing users with so much

flexibility that additional mental workload is created in remembering how to perform a task many different ways.

Evidence from the interviews supported the view that the FMS lacked flexibility and undermined control, particularly in the vertical navigation dimension. One interviewee commented:

"It just doesn't have the flexibility and control there. It's very good laterally. It gives you super accurate navigation but the vertical dimension usually doesn't work very well."

Three questionnaire items loaded onto the factor **Flexibility & Control** as shown on Table 4-12. Their internal consistency was at the acceptable level for scale construction and the mean ratings for the three items fell above the mid-point of possible ratings (see Table 4-46). Although **Flexibility & Control** is an important dimension of usability, its incorporation into a dynamic system such as the FMS, which is additionally a dynamic operating environment must be carefully planned.

Workload

Workload was the final usability component elicited from the Principal Components Analysis. This component refers to the amount of effort involved in performing tasks with a system and is a dimension of usability, which Bevan (1995) argued was particularly important in safety critical applications.

Two questionnaire items loaded onto the **Workload** factor as shown on Table 4-12. Good levels of internal consistency were again present in these questionnaire items. The mean usability ratings for both mental workload and physical workload were low, although ratings for mental workload were lower than those for physical workload.

Evidence from the literature suggests that workload is a very important issue on the flight deck, particularly since two-person crews now perform the work that was previously performed by three people. Wiener (1988) argued that workload reduction was a principle rationale for the introduction of flight deck automation since pilots prefer to have many routine and computational tasks performed by flight deck computers as it permits them to supervise their flights more effectively. A second rationale, outlined by Wiener (1988) was that within the terminal area, the use of flight deck automation would allow pilots to perform their tasks in a "heads up" manner, thereby allowing them to scan for other aircraft.

Kelly et al (1992) argued that flight deck design should create neither work overload nor work underload. Increases in flight deck automation have affected flight deck workload in a manner not envisaged prior to its implementation, as workload is frequently very low in the cruise phase of flight but can be very high in, for example, the descent and approach phase of flight. The physical dimension of workload is not often considered in discussions of usability, although this may be due to the fact that much of the usability research to date has focused on office-based applications.

Evidence from the interview transcripts again supported the importance of the workload dimension of usability for the FMS CDU:

"Coming into Heathrow, you will probably be told if you are holding or not ten miles from the fix. If you are doing three miles a minute, you have got roughly three minutes to go. How long does it take you to go to the hold page, get the hold down on the scratchpad, you bring your position down in there.....You put it in there and then it is wrong, then you have to change it again, so you are over the top of the hold by the start of it. So invariably what you end up doing is putting in the autopilot, the heading select and working it around its turn and that gives you an extra minute to get it sorted."

As highlighted in the literature review, accident reports such as the AAIB (1990) report into the Kegworth accident in Leicestershire, have also drawn attention to the high workload environments in which pilots reprogram the FMS. In the Kegworth example, the pilots programmed the FMS while simultaneously informing their company of their problems and intentions, responding to ATC altitude and heading instructions, and obtaining weather information for East Midlands Airport.

Overall Satisfaction

The satisfaction element of usability has been discussed in many papers including ISO 9241 Part 11 and Thomas (1996). This was defined by Preece et al (1994) as the positive attitude engendered in users by using a system. A single questionnaire item "The FMS is pleasant to use; I like it" was used as an overall measure of FMS usability and satisfaction in its use. While it is clear from the discussion in this section that many usability problems exist in use of FMSs, the general feeling towards their use is positive and it is clear that the modern pilot's task would be much more cumbersome in the absence of an FMS.

Although more complete analyses of this issue are described in the subsequent study, both the interview transcripts and the descriptive analyses (Table 4-7) indicated that although many usability problems exist in interacting with FMS CDUs, pilot satisfaction in use of the FMS was very favourable. There are a number of possible reasons for something that might seem contradictory. Many of the functions performed by the FMS are extremely useful and accurate and do take a significant workload from the pilots particularly during the cruise phase of flight. Life without the FMS would be more difficult for those operating modern two-crew flight decks. A further reason may be related to the fact that use of the FMS is an essential part of the pilot's job. As discussed by Noyes, Starr & Frankish (1996), line pilots can exhibit a loyalty towards their current aircraft due to the inherent risk associated with flying their (or any) aircraft. Pilots may therefore be reluctant to question features of the flight deck in which they operate, and may justify design features regardless of their suitability for the task being undertaken.

5.2 Reliability and Validity Evidence

The method of questionnaire construction and subsequent analysis of data adopted for this study were devised such that construct and content validity should be ensured.

Construct validity is the extent to which an assessment is measuring something meaningful. Guion (1977) argued that since test creators always have a theory or construct that they are trying to assess all test validation is necessarily construct validation. Murphy & Davidshofer (1991) defined the goal of construct validation as being "*to determine whether test scores provide a good measure of a specific construct*" (p.116).

In this study, therefore, it referred to the extent to which the usability factors, from which the scales would be developed, measured the theoretical construct of usability. Construct validity is concerned with how theoretically meaningful the factors from which the scales will be developed are. Ideally, evidence of construct validity would be achieved by comparing the usability factors developed in this study with existing measures of FMS CDU usability. Since no objective measure of FMS CDU usability exists, construct validity is difficult to establish.

If the underlying usability construct exists, it should exist regardless of participants' rank or culture. On the other hand in order to be useful it should be able to discriminate between FMSs and aircraft type (since FMS and aircraft type are intrinsically linked). Through comparing the factors elicited with comments from the interview transcripts and the literature, evidence towards construct validity was derived. The literature, for example, indicated that workload was a significant issue in interaction with the FMS CDU (Goteman, 1999). Ratings of the usability factors in this study supported the literature, since mean ratings of workload indicated that they were amongst the poorest of any of the components derived. Further evidence of construct validity was derived from the fact that the factors did not discriminate between pilots' usability ratings on the basis of their rank or culture. Both rank and culture are known to influence pilot attitudes to automation (see Section 2.2.4). Evidence in this study therefore indicated that the factors elicited were assessing usability and not attitudes to FMSs. Evidence toward construct validity was therefore demonstrated.

DeVellis (1991) defined reliability in the context of scale development as the *"proportion of variance attributable to the true score of the latent variable"* (p.24). The factors, from which the scales were to be developed, have successfully demonstrated internal consistency through achieving Cronbach's alpha scores, all of which were above the acceptable level for scale development. It has not been possible to demonstrate test-re-test reliability in this study since it was not viable to ask the same participants to complete the same questionnaire twice.

Face validity refers to the extent to which a test looks plausible to respondents. This was achieved in this study through involving both human factors and pilot SMEs throughout the questionnaire development process, and through extensive pre-testing and piloting of the questionnaire to be used for scale development. Although, face

validity was not sufficient in itself, it did ensure that the factors derived were more acceptable to potential respondents (Robson, 1993).

A factor is said to have content validity when the items it contains are a randomly chosen subset of the universe of possible appropriate items. The methodology adopted for scale development should ensure content validity. The questionnaire items in this study were developed through an iterative process of reviewing relevant academic literature, talking to line pilots and other human factors researchers, observing the line pilots interacting with the FMS, and pre-testing and piloting the items extensively with typical end users. This process ensured that the items elicited were appropriate and looked plausible to the line pilots who would ultimately receive the questionnaire. Since the factors from which scales will be developed should have content validity, they should almost certainly have face validity (Cook, 1998).

With the goal of demonstrating discriminant validity, a series of Discriminant Function Analyses (DFAs) were conducted. Since the usability of the FMS CDU is intrinsically linked to the aircraft in which it resides the analyses were designed to examine if the factors elicited could discriminate between aircraft and FMS types. It was hypothesised that if the factors could discriminate between these different aircraft and FMS CDU types, then evidence of discriminant validity would be demonstrated.

As discussed in Section 1.13, it was further hypothesised that user characteristics would not impact significantly on the factors derived since usability should be a product of the FMS CDU and its operating context not participants' culture or rank.

The series of DFAs described in Section 4.4 presented some evidence of both discriminant validity and criterion validity. Discriminant validity was demonstrated since ratings on the factors discriminated between that which they should discriminate but did not discriminate between that which they should not discriminate. The **Structure & Sequence**, **Appropriate Functionality**, **Information Presentation**, **Flexibility & Control** and **Processor Speed** components discriminated between aircraft manufacturers & FMSs manufacturers when these were examined together. The **Structure & Sequence, Appropriate Functionality and Feedback** components discriminated between Honeywell and Smiths FMSs. Finally, the **Structure & Sequence**, **Appropriate Functionality**, **Predictability**, **Information Presentation**, **Flexibility & Control** and **Processor Speed** components discriminated between Boeing, Airbus and Fokker aircraft. More components discriminated between aircraft manufacturers than discriminated between FMS manufacturers. This demonstrated the importance of the aircraft operating environment in which the FMS resides and the role of aircraft design philosophies discussed in Section 1.3.3.

The most useful measure of criterion validity, as reported by Ghiselli et al (1981) and DeVellis (1991) is the rate of correctly classifying the category of the criterion (e.g. a Smiths or Honeywell FMS), based on the predictor categories (the usability ratings). While the DFA classification matrices in Tables 4-19, 4-24, and 4-30 demonstrated that it was possible to predict participants' type rating at a rate significantly better than chance ($p<0.01$), no evidence of criterion validity was demonstrated at a practical level. Over a third of Honeywell FMSs, for example, were incorrectly classified as being

Smiths FMSs. Similarly, fifty percent of pilots type-rated on Airbus aircraft, were incorrectly classified, and only 21.1% of participants type-rated on Fokker aircraft were correctly classified. Furthermore, although it was not possible to discriminate between, or predict with accuracy, participants' rank or culture, the Press's Q statistic was still significant ($p<0.01$). It was therefore concluded that no practical evidence of criterion validity was demonstrated.

5.3 Conclusions

The production of a series of factors, from which scales to assess the usability of FMS CDUs could be developed, satisfied the aims of this study. The factors were shown to be internally consistent and evidence of face validity, construct validity and content validity was demonstrated. As hypothesised, the factors derived from participants' questionnaire ratings, did discriminate between aircraft manufacturers and FMS manufacturers. As was also hypothesised, the factors derived did not discriminate between participants' usability ratings based on their rank or culture. Evidence of discriminant validity was therefore demonstrated. While evidence of criterion validity was shown from a statistical perspective, no practical evidence of criterion validity was demonstrated. It was not possible to predict participants' current aircraft or FMS type accurately from their questionnaire ratings.

6 STUDY TWO INTRODUCTION

6.1 Purpose

The purpose of Study Two was to verify the structure of the usability factors elicited in Study One in order to demonstrate that they were replicable, and subsequently to produce reliable and valid usability *scales* rather than factors. Scales would provide the advantage of providing a simpler, more parsimonious view of FMS CDU usability than the factor structure did.

As stated in the literature review, FMSs were initially designed as a navigation tool for use solely in the cruise phase of flight when workload levels are low (Billings, 1997). Over time their functionality has been extended throughout all flight phases, from departure through to arrival, without any re-design of the CDU hardware. A further aim of this study was therefore to examine if differences in FMS usability existed with respect to flight phase. Aircraft manufacturers anticipated that the introduction of new technology would decrease flight deck workload (Billings, 1997). What was not anticipated was that the effects on workload would not be uniform. As pointed out by Wiener (1989), flight deck computers, such as the FMS, support pilots most in the low workload phases of flight, but are of little use when they are needed most, in the time critical phases of flight such as in descent and approach.

During the descent and approach phase of flight, many changes in system status can occur in a very short period. Flight deck crews are required to perform a number of concurrent tasks including interacting with Air Traffic Control and monitoring other aircraft. Further, as pointed out by Sarter, Woods & Billings (1997), interacting with the FMS CDU for even a simple task, such as changing a runway in this phase of flight can turn into a "*programming nightmare*" (p.1928). Workload can, therefore, be very high during this phase of flight. During the cruise phase of flight, pilots are not under the same level of time criticality when making their FMS inputs, their inputs are therefore less prone to error.

The approach taken for this study was to use the underlying factor structure of usability identified in Study One to construct a questionnaire that could be used to assess the FMS CDU usability across flight phase and aircraft type. If the scales developed were valid, the underlying factor structure would be constant for both the cruise and descent flight phases.

It was decided that questionnaires for Study Two should only be distributed to pilots in the UK and Ireland. This was due to the fact that the FMS CDU usability factor structure elicited in Study One had demonstrated that respondents' culture had no significant effect on ratings of usability. Further, the time taken to receive questionnaires back from the US during Study One had significantly delayed analysis.

The cruise and descent and approach phases of flight were chosen for investigation in this study since it has been suggested in the literature that these are phases of workload extremes. This investigation sought to demonstrate that the factor structure of FMS CDU usability developed in Study One would be valid for both the cruise and descent and approach phases of flight. If the factor structure was shown to be replicable and generalizable, then evidence of construct validity would be demonstrated. The investigation also sought to demonstrate that since usability was a product of the equipment used, and its operational environment (the flight deck), ratings of the usability scales ultimately produced should be independent of pilot workload and thus flight phase. Evidence of discriminant validity would be demonstrated if it was not possible to discriminate between pilots' ratings of FMS usability on the basis of flight phase.

To produce further evidence of discriminant validity and be consistent with analyses conducted during Study One, this study was also designed to examine if the scales produced could (i) discriminate between Boeing and Airbus aircraft and (ii) discriminate between the Honeywell and Smiths FMS CDUs that are fitted to these aircraft as standard. It was hypothesised that valid scales should be able to discriminate between these different aircraft and FMS CDUs.

The study was also designed to examine the impact of participants' rank on usability ratings. As in Study One, it was hypothesised that since the usability of the FMS CDU was an independent product of the FMS CDU and the aircraft in which the CDU resided, participants' rank should not significantly influence usability scale ratings.

7 STUDY TWO METHOD

7.1 Overview of Method Two Phases

The study was undertaken in three phases:

1. Modification of Questionnaire used in Study One based on analysis conducted in that study.

2. Production of two new questionnaires, one for the Descent & Approach phase of flight and one for the Cruise phase of flight – including formulation of questionnaires; pre-piloting the questionnaires face-to-face with airline pilots; piloting the questionnaires with airline pilots; and distributing the final questionnaires to a larger sample of airline pilots. An independent groups design methodology was thus made possible.

3. Analysis of questionnaire data with the goal of producing FMS usability scales and deriving evidence of reliability and validity.

Each of these phases will be described in turn.

7.2 Questionnaire Modifications

Based on the results of the analyses and the underlying factor structure identified in Study One, six items were deleted from the questionnaire used in Study One since they did not load onto any factor. The questionnaire items shown in Table 7-1 have therefore been eliminated from this study.

Table 7-1 - Questionnaire Items Removed from Subsequent Study

Questionnaire Item
10. I prefer to stick to the FMS functions and facilities I know best.
11. Established conventions are followed for the format in which information is displayed (familiarity aside).
32. FMS users have control over the order in which they request information, or carry out a series of activities.
41. Changes in one part of the FMS may produce subsequent adverse effects in other parts of the system.
42. It was easy to learn how to use the FMS.
43. It is sometimes difficult to remember where information is located within the FMS architecture.

The remaining forty-three items constituted the questionnaire utilised for Study Two. Although, some further questionnaire items did not load onto any component, their

importance which was highlighted both during the interview stage and from the literature, has meant that they have been retained in this study for further exploration. For example, it would not make logical sense to include the questionnaire item "The FMS is adequately flexible for all vertical navigation" without including the corresponding item for lateral navigation. The emphasis was to derive scales using the factor analysis as the basis for scale construction. It was not necessarily a basis in itself.

The introductory paragraphs and individual questionnaire items remained identical apart from the words *Cruise Phase* being used in the introductory paragraph of one set of questionnaires, and *Descent and Landing* being used in the other set. Questionnaire item 25, which contained the words "all flight phases", was altered so as to include the flight phase being assessed "Cruise" or "Descent and Landing" as appropriate.

The questionnaire items retained in this phase are shown in Table 7-2.

Table 7-2 - Final Questionnaire Items

Questionnaire Items
1. The information in the FMS display (CDU) is easy to see and read.
2. The FMS display (CDU) is sometimes cluttered.
3. Important information is adequately highlighted on the screen.
4. Information is organised logically on the CDU screen (familiarity aside).
5. The way the FMS (CDU) look and works is consistent at all times.
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.
7. The format and method of entering information is consistent throughout the operation of the FMS.
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.
9. The FMS sometimes makes unexpected actions.
10. The sequence of activities required in the use of the FMS is as I would expect.
11. The FMS works the way I think it should work.
12. The FMS buttons could be located more intuitively.
13. The feedback messages displayed by the FMS are frequently inadequate.
14. The FMS makes it clear what changes will occur as a result of an input being made.
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.
17. The FMS clearly indicated which mode the system is currently in. It is informative and accurate.
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.
19. It is clear to the user what should be done to correct any error that occurs.
20. The structure of the FMS is obvious to the user.
21. It is clear why the system is organised and structured as it is.
22. It is clear why the FMS screens are sequenced as they are.
23. The structure of the FMS is well organised from the user's point of view.
24. The FMS performs all the functions that I need it to.
25. FMS users can access all the information they require in this flight phase (Cruise or Descent & Approach).
26. The FMS screens provide all the information which is relevant to the task.
27. The system allows users to do what they feel is necessary in order to carry out a task.
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.
29. The FMS processes information too slowly for other activities in the flight deck.
30. The FMS allows the user to do the required task as efficiently as possible.
31. It is easy to return to the main menu from any screen of the FMS.
32. I feel in control of the FMS at all times.
33. The FMS is adequately flexible for all lateral navigation.
34. The FMS is adequately flexible for all vertical navigation.
35. It is always possible to check or correct inputs to the FMS before they are processed.

Questionnaire Items
36. The FMS has facilities for detecting and handling input errors.
37. The FMS is protected against common trivial entry errors.
38. The FMS keyboard is difficult or awkward to use.
39. The FMS help facility (either on-line or manuals) is adequate.
40. It is irritating to have to input information at a rate slow enough for the FMS processor.
41. The FMS can increase mental workload at inappropriate times.
42. The FMS can increase physical workload at inappropriate times.
43. The FMS is pleasant to use; I like it.

7.3 Methodological Considerations

All the methodological considerations outlined in Section 2.2 of Study One were also valid for this study.

7.4 Pilot Study

Since no new items were added to the questionnaire, only a small pilot study was conducted in which five pilot participants completed the questionnaire on a face-to-face basis. Ten further pilot participants recruited through personal contacts completed the questionnaire by mail.

The only change made to the questionnaire at this point was to change the wording used on the *"Descent & Landing"* questionnaires to *"Descent & Approach"* as recommended by the participants. This recommendation was made on the basis that the term *"Descent & Approach"* is more consistent with the terminology used in flight deck manuals.

7.5 Recruiting the Participant Sample

A search was undertaken in "Flight International" magazine's annual edition of "Airlines in Europe", to identify airlines that might participate in the research being undertaken (Flight International, 1998). The criteria considered in assessing whether an airline might be suitable was (1) that the airline had not participated in the previous phase of the study and (2) that the airline had an FMS-equipped fleet of aircraft. Seven potential participant airlines were identified, and a point of contact (the airline's Chief Pilot) was identified so that letters requesting participation could be sent to a named point of contact rather than just being addressed to the airline.

A letter outlining the purpose of this study was sent to each of seven airlines in the UK and Ireland inviting them to participate in the study (see Appendix G). Five of the airlines agreed to participate and the questionnaires were sent direct to the single point of contact (Chief Pilot or their nominated representative) within each airline.

Distribution was either through the questionnaires being placed in the crew room, or through the airline representative placing the questionnaires in each pilot's mail slot.

7.6 Study Two Final Questionnaire

The final forty-three item questionnaire was set on four A4 sides of paper, but printed double sided on an A3 page. For the purposes of distinguishing the questionnaires, those requiring ratings for the Cruise phase of flight were printed on green paper, while those for the Descent and Approach phase were printed on yellow paper. The final questionnaires have been included at Appendix H, although they are printed on A4 paper for presentation reasons only.

7.6.1 Distribution Mechanism

The questionnaires were packed in envelopes with a letter explaining the purpose of the research as shown at Appendix I. A FREEPOST return envelope accompanied each questionnaire and an "FMS Usability Questionnaire" label was placed on the front of each envelope to indicate the content of the envelope. It was anticipated that this would also minimise the number of questionnaires being opened by airline personnel for whom the questionnaire was not applicable. One thousand questionnaires were distributed through the five airlines in Ireland and the United Kingdom.

8 STUDY TWO RESULTS

8.1 Overview

This results section presents a description of the respondents that participated in the study followed by an explanation of how the data were treated. A descriptive analysis section presents the means and standard deviations of all questionnaire data. Confirmatory Factor Analysis was used to validate the structure of the usability components elicited from Study One using usability ratings from both the Cruise and Descent & Approach phases of flight. Using independent data sets from each flight phase, it was possible to examine if the factor structure on which the scales were to be based was replicable and generalizable. A series of DFAs are presented which examined the evidence towards discriminant validity through examining ratings of FMS CDU usability *scales* across aircraft type, FMS type, flight phase and pilot rank. Finally, a Multiple Regression is described that examined which of the usability scales were significant predictors of the questionnaire item "The FMS is pleasant to use; I like it". This latter questionnaire item was used as an overall measure of satisfaction in using the FMS CDU.

8.2 Sample Description

This section outlines a description of the respondents and the response rate achieved.

8.2.1 Response Rate

The overall response rate at the time the analyses were conducted was 35.6%. Considering this study offered no incentives, the response rate achieved was extremely good. A total of 356 questionnaires were used in the final analyses for this study. Of the 356 pilots that participated in the study, 179 pilots rated FMS usability for the Cruise phase of flight, while 177 pilots rated the FMS usability for the Descent and Approach phase.

8.2.2 Respondents

A total of one thousand questionnaires were sent to either the Chief Pilot or their nominated representative, of five UK and Irish Airlines. Fifteen respondents were female (4.2%), while 323 respondents (90.7%) were male. Eighteen respondents (5.1%) did not complete the gender part of the questionnaire. The participants' ages, hours on type and total hours are shown in Table 8-1.

Table 8-1 - Summary Participants' Age, Hours on Type & Total Hours

	Min	*Max*	*Mean*	*sd*
Age	22	62	42.7	9.0
Hours on Type	50	11500	2776.6	2326.7
Total Hours	400	22000	9491.8	5042.6

Figures 8-1 shows a histogram of pilots' total hours, while Figure 8-2 shows the distribution of pilots' hours on their current aircraft type.

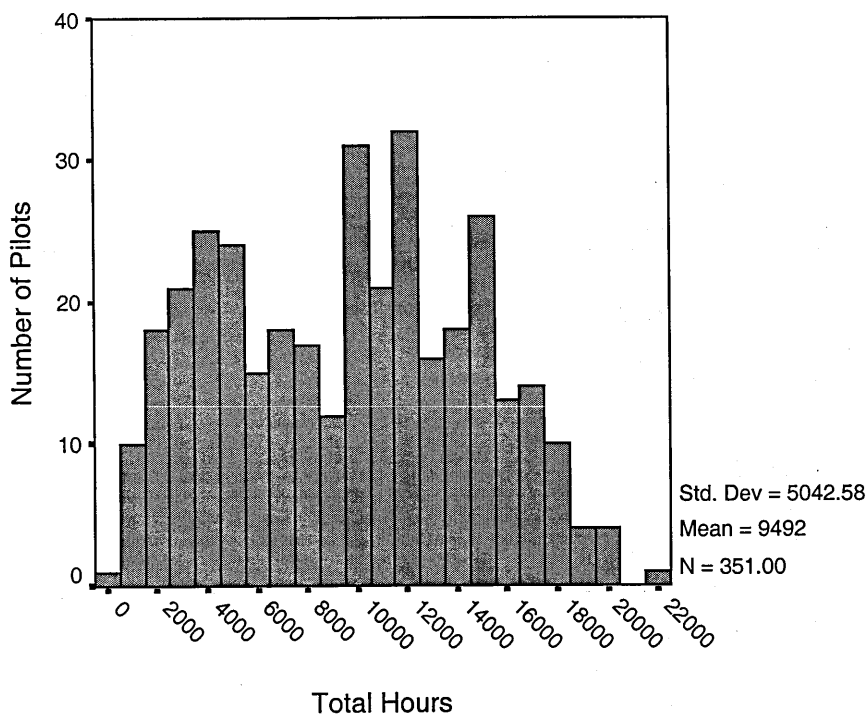


Figure 8-1 - Histogram of Pilots' Total Hours

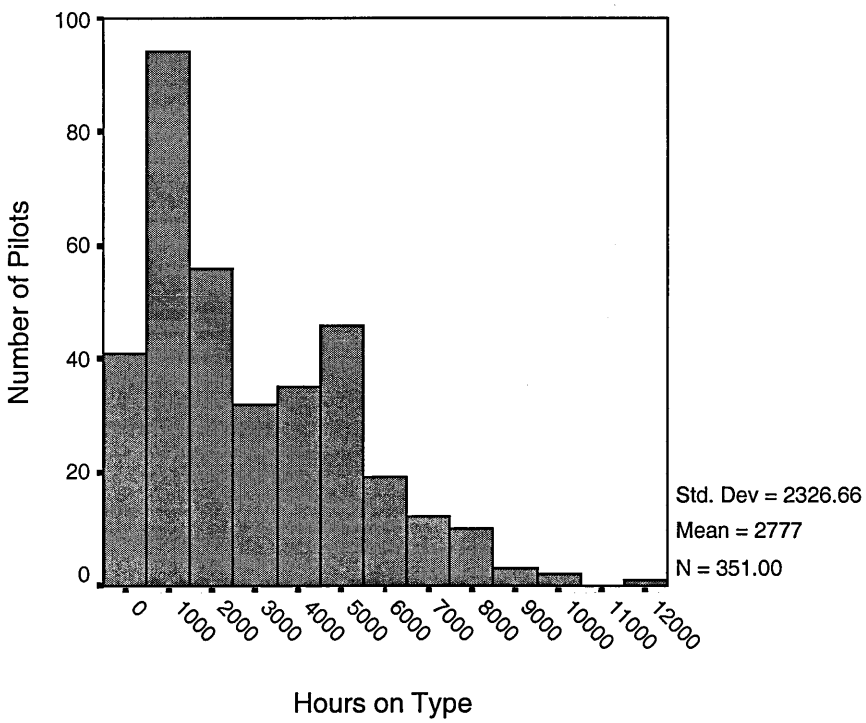


Figure 8-2 - Histogram of Pilots' Hours on Current Aircraft Type

The participants' ranks are shown in Table 8-2.

Table 8-2 - Summary of Participants' Rank

Participant Rank	*Frequency*	*Percentage*
Captain	216	60.7%
First Officer	120	33.7%
Second Officer	5	1.4%
Unspecified	15	4.2%

The participation and response rate by airline are shown in Table 8-3. The names of the participant airlines have not been included since the airlines were guaranteed anonymity and confidentiality. Non-identifying terms such as, Airline 1 and Airline 2, have been used instead.

Table 8-3 - Summary of Participation and Response Rate by Airline

Airline	*Frequency*	*Percentage*	*Response Rate*
Airline 1	101	28.4	50.5%
Airline 2	36	10.1	30.0%
Airline 3	51	14.3	25.5%
Airline 4	93	26.1	31.0%
Airline 5	73	20.5	40.6%
Unspecified	2	0.6	-

The aircraft type and series on which the participants were currently type-rated are shown in Table 8-4.

Table 8-4 - Summary of Current Aircraft Type and Series

Aircraft Type	*Frequency*	*Percentage*	*Aircraft Series*	*Frequency*	*Percentage*
B-757/767	177	49.8	B-757/767-200/300	172	48.4
B-737	94	26.4	B-737-300-500	94	26.4
A320/A321/ A330	41	11.5	A320/A321/A330	40	11.2
B-777	20	5.6	B-777-200	20	5.6
B-747	24	6.7	B-747-400	24	6.7
			Series Not Specified	6	1.7

The type of FMS participants reported to be on the aircraft on which they were currently type-rated is shown at Table 8-5.

Table 8-5 - Summary of Reported FMS Type Currently in Use

FMS Type	*Frequency*	*Percentage*
Honeywell	170	47.8
Smiths	30	8.4
Collins	3	0.8
Unspecified	153	43.0

An examination of the reported FMS types revealed that many respondents were unable to specify the type of FMS on board their aircraft. Many respondents (41.7%) did not complete this item. Evidence from the interview stage of Study One had suggested that this might be the case, since most of the interviewees had to interrogate their FMS manuals in order to find out their FMS type. The actual FMS type on board the aircraft on which respondents were currently type-rated was derived from information made available from participant airlines and from the aviation literature (e.g. Bulfer, 1996;

Bulfer & Gifford, 1997). Therefore, a summary of the respondents' actual FMS type is shown at Table 8-6.

Table 8-6 - Summary of Actual FMS Type Currently in Use

Aircraft Type	*Frequency*	*Percentage*
Honeywell	262	73.6
Smiths	93	26.1
Unknown	1	0.3

8.3 Treatment of data

As stated in Section 4.1.2, in order to prevent response set effects the questionnaire was designed such that for some items a response of '7' indicated the usable end of the scale, while for other items '1' indicated greater usability. For ease of interpretation, any item for which '1' indicated the more usable end of the scale was reversed (i.e. for all items appearing in this section, higher numbers indicate greater ratings of usability). Questionnaire items 2, 9, 12, 13, 15, 16, 29, 38, 40, 41 and 42 were treated in this manner. Since factor analysis is based on underlying correlations and covariances, the relative relationships in the underlying structure between items should remain constant regardless of whether they were rated higher or lower by flight phase (Hair et al, 1998). It was therefore possible to collapse the data sets across phase of flight, allowing subsequent analyses to be conducted on the integrated data collected across both flight phases. The data collected for the Cruise and Descent & Approach phases of flight were combined for all analyses reported for Study Two unless otherwise stated.

8.4 Descriptive Analysis

This section presents all the descriptive analysis conducted on the data collected during Study Two. The overall means and standard deviations for each questionnaire item are shown in Table 8-7 and include the data collected from both the Cruise and Descent & Approach phases of flight.

Table 8-7 - Means & Standard Deviations for Questionnaire Items

Questionnaire Item	*Mean*	*sd*
1. The information in the FMS display (CDU) is easy to see and read.	5.18	1.17
2. The FMS display (CDU) is sometimes cluttered.	4.26	1.56
3. Important information is adequately highlighted on the screen.	3.81	1.56
4. Information is organised logically on the CDU screen (familiarity aside).	4.35	1.40
5. The way the FMS (CDU) look and works is consistent at all times.	4.21	1.52
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	5.23	1.22
7. The format and method of entering information is consistent throughout the operation of the FMS.	4.65	1.52
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	4.01	1.56
9. The FMS sometimes makes unexpected actions.	3.83	1.79
10. The sequence of activities required in the use of the FMS is as I would expect.	4.57	1.38
11. The FMS works the way I think it should work.	4.36	1.45
12. The FMS buttons could be located more intuitively.	3.75	1.72
13. The feedback messages displayed by the FMS are frequently inadequate.	3.88	1.63
14. The FMS makes it clear what changes will occur as a result of an input being made.	3.49	1.68
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	4.89	1.66
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	3.71	1.81
17. The FMS clearly indicated which mode the system is currently in. It is informative and accurate.	4.36	1.52
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	2.94	1.49
19. It is clear to the user what should be done to correct any error that occurs.	2.72	1.41
20. The structure of the FMS is obvious to the user.	3.58	1.48
21. It is clear why the system is organised and structured as it is.	3.65	1.50
22. It is clear why the FMS screens are sequenced as they are.	4.29	1.42
23. The structure of the FMS is well organised from the user's point of view.	4.21	1.40
24. The FMS performs all the functions that I need it to.	3.97	1.71
25. FMS users can access all the information they require in this flight phase (Cruise or Descent & Approach).	4.72	1.56
26. The FMS screens provide all the information which is relevant to the task.	4.35	1.53
27. The system allows users to do what they feel is necessary in order to carry out a task.	4.51	1.36
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	4.41	1.39
29. The FMS processes information too slowly for other activities in the flight deck.	3.03	1.87
30. The FMS allows the user to do the required task as efficiently as possible.	3.86	1.48

Questionnaire Item	Mean	sd
31. It is easy to return to the main menu from any screen of the FMS.	5.12	1.61
32. I feel in control of the FMS at all times.	4.95	1.56
33. The FMS is adequately flexible for all lateral navigation.	4.51	1.60
34. The FMS is adequately flexible for all vertical navigation.	3.67	1.60
35. It is always possible to check or correct inputs to the FMS before they are processed.	4.30	1.70
36. The FMS has facilities for detecting and handling input errors.	3.82	1.59
37. The FMS is protected against common trivial entry errors.	3.79	1.72
38. The FMS keyboard is difficult or awkward to use.	4.43	1.73
39. The FMS help facility (either on-line or manuals) is adequate.	2.88	1.58
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	2.90	1.80
41. The FMS can increase mental workload at inappropriate times.	2.97	1.69
42. The FMS can increase physical workload at inappropriate times.	3.68	1.83
43. The FMS is pleasant to use; I like it.	5.00	1.40

From examining the means and standard deviations of the questionnaire items shown in Table 8-7, a number of ratings stood out as being particularly high (towards the usable end of the scale) and particularly low (towards the unusable end of the scale). Items notably rated towards the usable end of the scale were: "The information in the FMS display (CDU) is easy to see and read" *(mean=5.18, sd=1.17)*; "Abbreviations, acronyms & codes are used consistently throughout the FMS" *(mean=5.23, sd=1.22)*; "It is easy to return to the main menu from any screen of the FMS" *(mean=5.12, sd=1.61)* and "The FMS is pleasant to use; I like it" *(mean=5.00, sd = 1.40).*

Correspondingly, some items were notably rated towards the unusable end of the scale and these were: "FMS error messages explain clearly where errors are, what the errors are and why they have occurred" *(mean=2.94, sd=1.49)*; "It is clear to the user what should be done to correct any error that occurs" *(mean=2.72; sd=1.41)*; "The FMS help facility (either on-line or manuals) is adequate" *(mean=2.88; sd=1.58)*; "It is irritating to have to input information at a rate slow enough for the FMS processor" *(mean=2.90, sd=1.80)* and "The FMS can increase mental workload at inappropriate times" *(mean=2.97, sd=1.69).*

8.5 Analysis of Usability Ratings by Flight Phase

Further evidence for the appropriateness of collapsing the data collected for the Cruise and Descent & Approach flight phases into one data set was demonstrated from the results of t-tests shown in Appendix J. The independent t-tests examined individual questionnaire items to determine if significant differences existed between individual questionnaire items based on the phase of flight that was rated by the participants. As can be seen from Appendix J, results showed that no significant differences were found for forty of the forty-three individual questionnaire items. Significant differences were

found for questionnaire items 14, 26 & 41, each item being rated significantly more favourable for the Cruise phase of flight than for the Descent & Approach phase of flight. The questionnaire items in "Bold" typeface were those which were included in the final usability scales, while the highlighted rows indicate ratings that were not significantly different based on the flight phase rated ($p.<0.05$).

It is worth noting that all but two of the questionnaire items ("The way the FMS (CDU) look and works is consistent at all times" and "It is irritating to have to input information at a rate slow enough for the FMS processor") were rated more favourably for the Cruise phase of flight than for the Descent and Approach phase of flight.

8.6 Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) was used in this study to verify the factor structure elicited from Study One. Structural Equation Modelling (SEM) permits latent (hypothesised) and manifest (collected data) variables to be input to a single analysis, thereby allowing the hypothesised structure of FMS usability derived from Study One, to be validated against the manifest variables elicited during this study.

CFA is one of many SEM techniques which combine aspects of multiple regression and factor analysis to estimate a series of interrelated dependence relationships simultaneously (Hair et al, 1998). CFA is used to test a pre-specified hypothesised model i.e. it is a test to confirm a model specified prior to the analysis, as opposed to exploratory factor analysis which seeks to define possible relationships between variables.

The statistical package AMOS is one of the Structural Equation Modelling packages that allows CFA to be conducted. AMOS Version 4.01 was used to conduct the analyses described in this section.

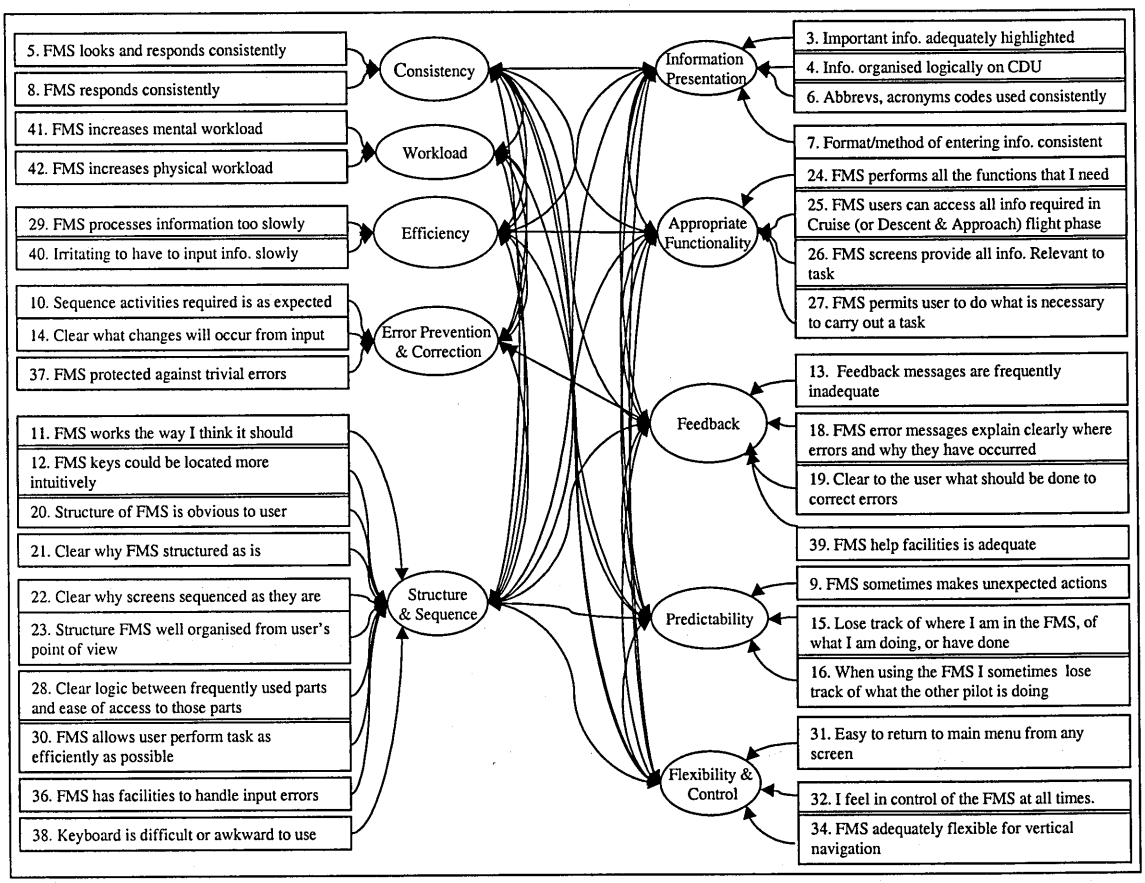


Figure 8-3 - Initial CFA Structural Diagram

In conducting the CFA a structure model was drawn showing the latent variables (the factors derived from Study One) and their indicators (the questionnaire items from which the factors were made) as shown in Figure 8-3. Arrows have been used to show which indicators (questionnaire items) loaded onto each latent variable (the factors derived from Study One).

The analyses conducted in Study One showed that some questionnaire items loaded onto more than one factor. Items that loaded onto more than one factor were included in this study on the factor on which they had the highest loading in Study One.

The questionnaire item, "The FMS sometimes makes unexpected actions", for example, loaded onto **Predictability** (0.492) and **Consistency** (0.487). Since the FMS making unexpected actions was more logically related to **Predictability** than **Consistency**, and since it also had a higher loading on that component, it was loaded onto the **Predictability** latent variable in this study. The Cronbach's alpha analysis conducted in Study One demonstrated that an acceptable level of internal consistency also existed when the questionnaire item was loaded on this component as was shown in Table 4-41.

Similarly, the questionnaire item, "Information is organised logically on the CDU screen (familiarity aside)" seemed more related to **Information Presentation** in terms of

content than **Structure & Sequence**. Since it also had a higher loading on that component (0.531 instead of 0.472) and since the Cronbach's alpha demonstrated an acceptable level of internal consistency (See Table 4-42), the item was loaded onto the **Information Presentation** latent variable in this CFA.

The structure model allowed the estimation of the strength of the relationship between the indicator variables and the latent factors based on a correlation or covariance matrix only.

Due to the fact that correlation matrices use a common range (are standardized), direct comparisons between the coefficients within a model are feasible. The coefficients from correlation matrices are always within the range –1.0 to +1.0. Missing data have a significant effect on the ability of SEM techniques to perform the estimation process. Therefore all cases containing missing data were deleted from the data set used in the CFA analyses. The remaining 320 complete cases were used in the calculation of the structural model.

In CFA, standardized regression weights indicate the actual size of a parameter, and range from a minimum of zero to a maximum of one. The standardized regression weights have equal variances and therefore closely approximate effect sizes. The higher the value the more importance it has in the causal relationships. The unstandardized coefficients are expressed in terms of the construct scale's variance. If the model is accepted, each of the constructs can be evaluated for statistical significance using the indicator loadings and by assessing any associated t-values which exceed the critical value.

8.6.1 Initial CFA Model

In order to conduct this CFA, the data sets from both Cruise and Descent and Approach flight phases were combined. The latent variables specified were the factors derived from the factor analysis described in Study One. The indicators were the questionnaire items, which loaded onto each factor in that analysis. Tables 8-8 to 8-17 show the unstandardized and standardized regression weights, the standard error and critical ratios elicited for each latent variable. It should be noted that one variable must have its loading set to 1.00 by the user on each latent variable. This marker variable has no critical ratio.

Table 8-8 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Structure & Sequence

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
23. The structure of the FMS is well organised from the user's point of view.	Structure & Sequence	1.000	0.828		
20. The structure of the FMS is obvious to the user.	Structure & Sequence	0.935	0.725	0.064	14.531
21. It is clear why the system is organised and structured as it is.	Structure & Sequence	0.964	0.741	0.064	14.986
22. It is clear why the FMS screens are sequenced as they are.	Structure & Sequence	0.859	0.705	0.061	13.995
10. The sequence of activities required in the use of the FMS is as I would expect.	Structure & Sequence	0.820	0.682	0.061	13.387
11. The FMS works the way I think it should work.	Structure & Sequence	0.914	0.729	0.062	14.639
30. The FMS allows the user to do the required task as efficiently as possible.	Structure & Sequence	0.817	0.628	0.068	12.065
12. The FMS buttons could be located more intuitively.	Structure & Sequence	0.777	0.516	0.082	9.520
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	Structure & Sequence	0.837	0.691	0.061	13.631
38. The FMS keyboard is difficult or awkward to use.	Structure & Sequence	0.823	0.545	0.081	10.166

As can be seen from Table 8-8, the critical ratios are all well above 1.96 indicating that the latent variable, **Structure & Sequence**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-9 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Appropriate Functionality

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
25. FMS users can access all the information they require in this flight phase (Cruise or Descent & Approach).	Appropriate Functionality	1.000	0.780		
26. The FMS screens provide all the information which is relevant to the task.	Appropriate Functionality	1.035	0.830	0.069	15.045
24. The FMS performs all the functions that I need it to.	Appropriate Functionality	0.946	0.672	0.079	11.960
27. The system allows users to do what they feel is necessary in order to carry out a task.	Appropriate Functionality	0.887	0.789	0.062	14.289

As can be seen from Table 8-9, the critical ratios are all well above 1.96 indicating that the latent variable, **Appropriate Functionality**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-10 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Feedback

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	Feedback	1.000	0.781		
19. It is clear to the user what should be done to correct any error that occurs.	Feedback	0.958	0.807	0.073	13.120
13. The feedback messages displayed by the FMS are frequently inadequate.	Feedback	0.789	0.571	0.083	9.552
39. The FMS help facility (either on-line or manuals) is adequate.	Feedback	0.568	0.432	0.079	7.159

As can be seen from Table 8-10, the critical ratios are all well above 1.96 indicating that the latent variable, **Feedback**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-11 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Predictability

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	Predictability	1.000	0.788		
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	Predictability	1.097	0.784	0.095	11.529
9. The FMS sometimes makes unexpected actions.	Predictability	0.499	0.366	0.085	5.900

As can be seen from Table 8-11, the critical ratios are all well above 1.96 indicating that the latent variable, **Predictability**, loaded significantly onto two of the indicators ($p<0.05$).

Table 8-12 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Information Presentation

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	Information Presentation	1.000	0.369		
3. Important information is adequately highlighted on the screen.	Information Presentation	1.745	0.508	0.318	5.486
7. The format and method of entering information is consistent throughout the operation of the FMS.	Information Presentation	2.005	0.587	0.346	5.796
4. Information is organised logically on the CDU screen (familiarity aside).	Information Presentation	1.903	0.610	0.324	5.870

As can be seen from Table 8-12, the critical ratios are all well above 1.96 indicating that the latent variable, **Information Presentation**, loaded significantly onto all of the possible indicators ($p<0.05$).

Table 8-13 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Flexibility & Control

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
34. The FMS is adequately flexible for all vertical navigation.	Flexibility & Control	1.528	0.568	0.254	6.019
32. I feel in control of the FMS at all times.	Flexibility & Control	1.595	0.612	0.258	6.184
31. It is easy to return to the main menu from any screen of the FMS.	Flexibility & Control	1.000	0.378		

As can be seen from Table 8-13, the critical ratios are all well above 1.96 indicating that the latent variable, **Flexibility & Control**, loaded significantly onto both of the possible indicators ($p<0.05$).

Table 8-14 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Error Prevention & Correction

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
14. The FMS makes it clear what changes will occur as a result of an input being made.	Error Prevention & Correction	0.713	0.490	0.105	6.822
37. The FMS is protected against common trivial entry errors.	Error Prevention & Correction	0.921	0.624	0.115	7.978
36. The FMS has facilities for detecting and handling input errors.	Error Prevention & Correction	1.000	0.737		

As can be seen from Table 8-14, the critical ratios are all well above 1.96 indicating that the latent variable, **Error Prevention & Correction**, loaded significantly onto both of the indicators ($p<0.05$).

Table 8-15 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Processor Speed

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
29. The FMS processes information too slowly for other activities in the flight deck.	Processor Speed	1.090	0.824	0.119	9.159
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	Processor Speed	1.000	0.782		

As can be seen from Table 8-15, the critical ratio is well above 1.96 indicating that the latent variable, **Processor Speed**, loaded significantly onto one of the indicators ($p<0.05$).

Table 8-16 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Workload

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
42. The FMS can increase physical workload at inappropriate times.	Workload	0.900	0.712	0.092	9.834
41. The FMS can increase mental workload at inappropriate times.	Workload	1.000	0.849		

As can be seen from Table 8-16, the critical ratio is well above 1.96 indicating that the latent variable, **Workload**, loaded significantly onto one of the indicators ($p<0.05$). Similarly, as shown on Table 8-17, the critical ratio for **Consistency** is well above 1.96 indicating that the latent variable loaded significantly onto one of the indicators ($p<0.05$).

Table 8-17 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Consistency

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	Consistency	1.009	0.721	0.092	11.011
5. The way the FMS (CDU) look and works is consistent at all times.	Consistency	1.000	0.727		

Table 8-18 shows the squared multiple correlations for each indicator, which describe the amount of variance in the indicators accounted for by the latent construct. Higher variance extracted values occur when the indicators are truly representative of the latent. Guidelines suggest that the variance extracted value for a construct should exceed 0.50 (Hair et al, 1998). Over fifty percent of the estimates (19 out of 37 items) failed to exceed the recommended value. Those estimates that fell below the recommended value have been highlighted in Table 8-18. None of the estimates exceeded the critical value on the **Information Presentation** and **Flexibility & Control** dimensions, while only one estimate achieved the criterion on the **Error Prevention & Correction** dimension.

Table 8-18 - Squared Multiple Correlations

Latent & Indicators	*Estimates*
Structure & Sequence	
23. The structure of the FMS is well organised from the user's point of view.	0.686
20. The structure of the FMS is obvious to the user.	0.525
21. It is clear why the system is organised and structured as it is.	0.550
22. It is clear why the FMS screens are sequenced as they are.	0.497
10. The sequence of activities required in the use of the FMS is as I would expect.	0.465
11. The FMS works the way I think it should work.	0.531
30. The FMS allows the user to do the required task as efficiently as possible.	0.395
12. The FMS buttons could be located more intuitively.	0.266
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	0.477
38. The FMS keyboard is difficult or awkward to use.	0.298
Appropriate Functionality	
25. FMS users can access all the information they require in this flight phase (Cruise or Descent & Approach).	0.608
26. The FMS screens provide all the information, which is relevant to the task.	0.690
24. The FMS performs all the functions that I need it to.	0.452
27. The system allows users to do what they feel is necessary in order to carry out a task.	0.623
Feedback	
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	0.610
19. It is clear to the user what should be done to correct any error that occurs.	0.651
13. The Feedback messages displayed by the FMS are frequently inadequate	0.326
39. The FMS help facility (either on-line or manuals) is adequate.	0.186
Predictability	
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	0.621
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	0.615
9. The FMS sometimes makes unexpected actions.	0.134

Latent & Indicators	*Estimates*
Information Presentation	
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	0.136
3. Important information is adequately highlighted on the screen.	0.258
7. The format and method of entering information is consistent throughout the operation of the FMS.	0.345
4. Information is organised logically on the CDU screen (familiarity aside).	0.372
Flexibility & Control	
34. The FMS is adequately flexible for all vertical navigation.	0.323
32. I feel in control of the FMS at all times.	0.375
31. It is easy to return to the main menu from any screen of the FMS.	0.143
Error Prevention & Correction	
14. The FMS makes it clear what changes will occur as a result of an input being made.	0.240
37. The FMS is protected against common trivial entry errors.	0.389
36. The FMS has facilities for detecting and handling input errors.	0.543
Processor Speed	
29. The FMS processes information too slowly for other activities in the flight deck.	0.680
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	0.612
Workload	
42. The FMS can increase physical workload at inappropriate times.	0.507
41. The FMS can increase mental workload at inappropriate times.	0.721
Consistency	
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	0.520
5. The way the FMS (CDU) look and works is consistent at all times.	0.529

Table 8-19 shows a summary of the model derived from the Initial CFA.

Table 8-19 - Summary of Models Initial CFA

Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*	*RMR*	*GFI*	*AGFI*	*PGFI*
Default	119	1365.23	584	0.000	2.338	0.192	0.807	0.768	0.671

The Number of Parameters derived in the model was 119. The Chi-Square Minimum (CMIN) was 1365.23, while the root mean square residual at 0.04 was within the acceptable range of 0.08 or less. The CMIN/DF result (2.338) shows the chi-square adjusted by the degrees of freedom. This measure is the normed chi-square, and is the ratio of the chi-square divided by the degrees of freedom. This measure can be used to assess the model fit, since (i) inappropriate models may be overfitted and therefore be over-reliant on chance, or (2) the model may not be truly representative of observed data and therefore needs improvement (Hair et al, 1998). Over-fitted data typically have normed chi-square values of less than one, while values greater than three indicate the model is not truly representative of the observed data (Jöreskog, 1970). The CMIN/DF value for the model described in this section (2.338) is well within the appropriate range of one to three.

The Goodness-of-Fit Index (GFI) is a measure by which the actual or observed input matrix (covariances or correlation) is predicted by the estimated model. GFI values range from zero to one, where zero indicates a poor overall fit and one indicates a perfect overall fit. The GFI value derived from this CFA was 0.807. The recommended level of acceptance for the GFI is 0.90, although Hair et al (1998) describe an example value of 0.865 as being at the "*marginal acceptance level*" (p.634). Since the GFI in this model fell outside the acceptable range, further analyses were undertaken to establish if a more parsimonious model could be achieved by removing offending estimates.

The Adjusted Goodness-of-Fit Index (AGFI) assesses the goodness of fit measures compared to a null model, the null model being hypothesised as a single-factor model with no measurement error. The recommended level of acceptance for an AGFI is 0.90, although Hair et al (1998) describe an example value of 0.810 as being of "*marginal*" acceptability (p.660). However, the AGFI derived in this model fell well below the acceptance level recommended, again suggesting that this may not be the most parsimonious solution.

The Parsimonious Goodness-of-Fit Index (PGFI) is a measure of overall goodness-of-fit representing the degree of model fit per estimated coefficient. This measure attempts to correct for any "overfitting" of the model and evaluates the parsimony of the model compared to the goodness of fit.

The summary model of this CFA, and particularly the GFIs, indicated that the data from Study Two were not a very good fit for the hypothesised model structure derived from Study One. A review of the high correlations between the latent variables further suggested that some of them could have been non-orthogonal. Finally, the high modification indices generated (see Appendix L), particularly within the **Structure &**

Sequence latent component, indicated that some of the indicators within that latent did not fit the model and were thus inappropriately inflating the Chi-Square.

Through iteratively modifying the model in line with suggestions from the modification indices a more parsimonious solution was produced. A number of poor indicators, particularly on the **Structure & Sequence** component in the original factor analysis seemed to be generating the problem. Once the poor indicators were eliminated the factor structure produced was much more parsimonious with the GFIs produced also being within the acceptable range. The indicators that were eliminated are shown on Table 8-20.

Table 8-20 - Indicators Eliminated in Final Model

Indicator
7. The format and method of entering information is consistent throughout the operation of the FMS.
9. The FMS sometimes makes unexpected actions.
11. The FMS works the way I think it should work.
12. The FMS buttons could be located more intuitively.
13. The feedback messages displayed by the FMS are frequently inadequate.
14. The FMS makes it clear what changes will occur as a result of an input being made.
20. The structure of the FMS is obvious to the user.
22. It is clear why the FMS screens are sequenced as they are.
27. The system allows users to do what they feel is necessary in order to carry out a task.
31. It is easy to return to the main menu from any screen of the FMS.
32. I feel in control of the FMS at all times.
38. The FMS keyboard is difficult or awkward to use.

The final CFA model generated is presented in Section 8.6.2.

8.6.2 Final Revised CFA Model

The revised CFA model showing the latent variables and their indicators (the questionnaire items which from which the factors were made) is shown in Figure 8-4.

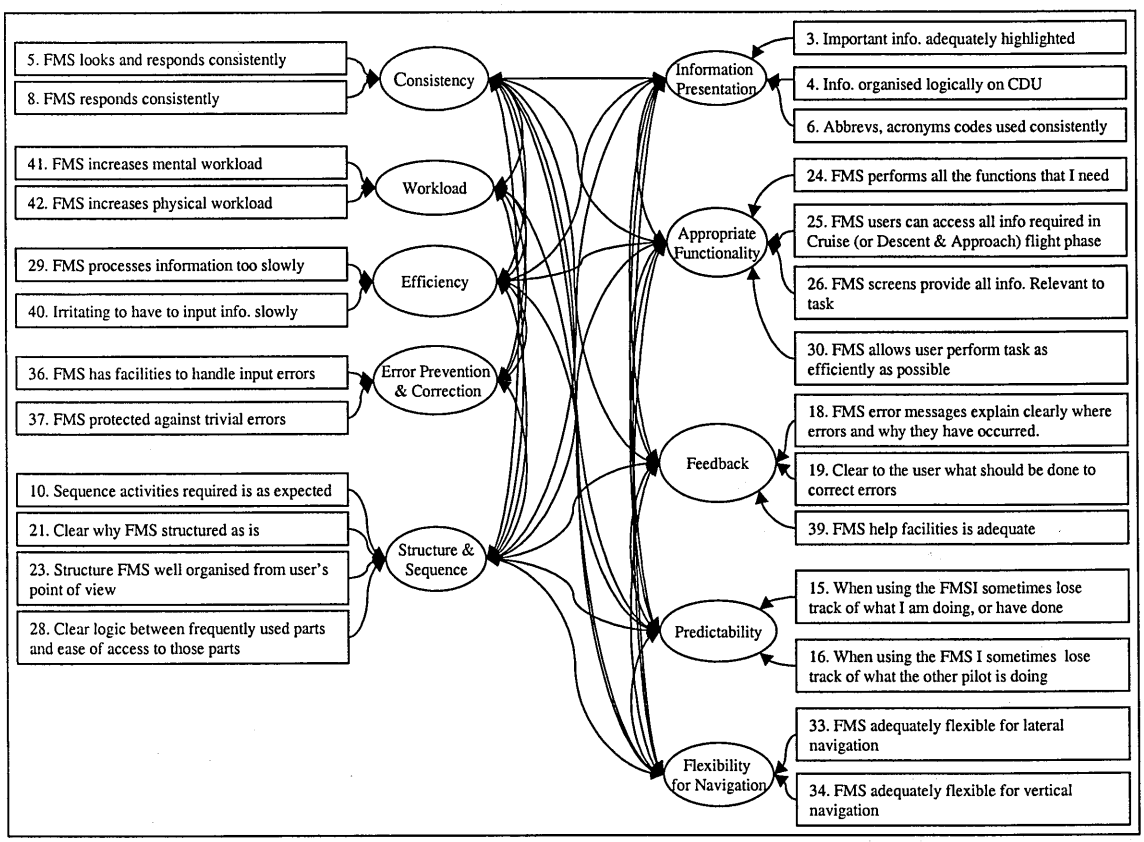


Figure 8-4 - Final CFA Structural Diagram

Table 8-21 to Table 8-30 show the unstandardized and standardized regression weights, the standard error and the critical ratios elicited for each latent variable. It should again be noted that one observed variable must have its loading set to 1.00 by the user on each latent variable. This marker variable has no critical ratio.

Table 8-21 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Structure & Sequence

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
23. The structure of the FMS is well organised from the user's point of view.	Structure & Sequence	1.000	0.811		
21. It is clear why the system is organised and structured as it is.	Structure & Sequence	0.928	0.698	0.072	12.919
10. The sequence of activities required in the use of the FMS is as I would expect.	Structure & Sequence	0.823	0.669	0.067	12.283
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	Structure & Sequence	0.894	0.722	0.066	13.450

As can be seen from Table 8-21, the critical ratios are all well above 1.96 indicating that the latent variable, **Structure & Sequence**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-22 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Appropriate Functionality

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
25. FMS users can access all the information they require in this flight phase (Cruise or Descent & Approach).	Appropriate Functionality	1.000	0.777		
26. The FMS screens provide all the information which is relevant to the task.	Appropriate Functionality	1.041	0.832	0.069	15.059
24. The FMS performs all the functions that I need it to.	Appropriate Functionality	0.963	0.682	0.079	12.130
30. The FMS allows the user to do the required task as efficiently as possible.	Appropriate Functionality	0.884	0.783	0.062	14.158

As can be seen from Table 8-22, the critical ratios are all well above 1.96 indicating that the latent variable, **Appropriate Functionality**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-23 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Feedback

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	Feedback	1.000	0.782		
19. It is clear to the user what should be done to correct any error that occurs.	Feedback	1.005	0.847	0.081	12.343
39. The FMS help facility (either on-line or manuals) is adequate.	Feedback	0.552	0.420	0.080	6.938

As can be seen from Table 8-23, the critical ratios are all well above 1.96 indicating that the latent variable, **Feedback**, loaded significantly onto both of the indicators ($p<0.05$).

Table 8-24 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Predictability

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	Predictability	1.000	0.824		
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	Predictability	1.025	0.767	0.099	10.313

As can be seen from Table 8-24, the critical ratio is well above 1.96 indicating that the latent variable, **Predictability**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-25 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Information Presentation

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	Information Presentation	1.000	0.328		
3. Important information is adequately highlighted on the screen.	Information Presentation	2.173	0.563	0.441	4.930
4. Information is organised logically on the CDU screen (familiarity aside).	Information Presentation	2.417	0.690	0.467	5.174

As can be seen from Table 8-25, the critical ratios are all well above 1.96 indicating that the latent variable, **Information Presentation**, loaded significantly onto both of the indicators ($p<0.05$).

Table 8-26 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Error Prevention & Correction

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
37. The FMS is protected against common trivial entry errors.	Error Prevention & Correction	0.706	0.600	0.122	5.768
36. The FMS has facilities for detecting and handling input errors.	Error Prevention & Correction	1.000	0.925		

As can be seen from Table 8-26, the critical ratio is well above 1.96 indicating that the latent variable, **Error Prevention & Correction**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-27 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Processor Speed

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
29. The FMS processes information too slowly for other activities in the flight deck.	Processor Speed	1.086	0.823	0.116	9.369
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	Processor Speed	1.000	0.784		

As can be seen from Table 8-27, the critical ratio is well above 1.96 indicating that the latent variable, **Processor Speed**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-28 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Workload

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
42. The FMS can increase physical workload at inappropriate times.	Workload	0.898	0.711	0.103	8.753
41. The FMS can increase mental workload at inappropriate times.	Workload	1.000	0.850		

As can be seen from Table 8-28, the critical ratio is well above 1.96 indicating that the latent variable, **Workload**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-29 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Consistency

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	Consistency	1.059	0.739	0.103	10.302
5. The way the FMS (CDU) look and works is consistent at all times.	Consistency	1.000	0.710		

As can be seen from Table 8-29, the critical ratio is well above 1.96 indicating that the latent variable, **Consistency**, loaded significantly onto one of the indicators ($p<0.05$). Similarly, as can be seen from Table 8-30, the critical ratio is well above 1.96 indicating that the latent variable, **Flexibility for Navigation**, loaded significantly onto one of the indicators ($p<0.05$). The factor label **Flexibility & Control** was replaced by the label **Flexibility for Navigation** since that label was more appropriate considering the content of the indicators that now loaded onto that latent variable. An assessment could not be made of the second observed variable for either **Consistency** or **Flexibility for Navigation** as they were set by the researcher as marker variables.

Table 8-30 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Flexibility for Navigation

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
34. The FMS is adequately flexible for all vertical navigation.	Flexibility for Navigation	0.961	0.696	0.092	10.490
33. The FMS is adequately flexible for all lateral navigation.	Flexibility for Navigation	1.000	0.722		

Table 8-31 shows the squared multiple correlations for each indicator, which illustrates the amount of variance in the indicators, accounted for by the latent construct. Over 65.4% of the estimates exceeded the 0.5 value recommended by Hair et al (1998). This was an improvement on the initial model in which only 48.6% of the estimates exceeded the recommended value (see Table 8-18). Those estimates that fell below the recommended value have been highlighted in Table 8-31.

Table 8-31 - Squared Multiple Correlations

Latent & Indicators	*Estimates*
Structure & Sequence	
23. The structure of the FMS is well organised from the user's point of view.	0.658
21. It is clear why the system is organised and structured as it is.	0.488
10. The sequence of activities required in the use of the FMS is as I would expect.	0.448
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	0.522
Appropriate Functionality	
25. FMS users can access all the information they require in this flight phase (Cruise or Descent & Approach).	0.604
26. The FMS screens provide all the information, which is relevant to the task.	0.692
24. The FMS performs all the functions that I need it to.	0.465
30. The FMS allows the user to do the required task as efficiently as possible.	0.614
Feedback	
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	0.611
19. It is clear to the user what should be done to correct any error that occurs.	0.717
39. The FMS help facility (either on-line or manuals) is adequate.	0.177
Predictability	
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	0.680
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	0.588
Information Presentation	
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	0.108
3. Important information is adequately highlighted on the screen.	0.317
4. Information is organised logically on the CDU screen (familiarity aside).	0.476
Error Prevention & Correction	
37. The FMS is protected against common trivial entry errors.	0.360
36. The FMS has facilities for detecting and handling input errors.	0.855
Processor Speed	
29. The FMS processes information too slowly for other activities in the flight deck.	0.677
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	0.615

Latent & Indicators	Estimates
Workload	
42. The FMS can increase physical workload at inappropriate times.	0.506
41. The FMS can increase mental workload at inappropriate times.	0.723
Consistency	
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	0.546
5. The way the FMS (CDU) look and works is consistent at all times.	0.504
Flexibility for Navigation	
34. The FMS is adequately flexible for all vertical navigation.	0.485
33. The FMS is adequately flexible for all lateral navigation.	0.521

Table 8-32 - Summary of Models (Final Model)

Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*	*RMR*	*GFI*	*AGFI*	*PGFI*
Default	97	345.577	254	0.000	1.361	0.103	0.927	0.899	0.671

The Number of Parameters derived in the model was 97. The Chi-Square Minimum was 345.577, with 254 degrees of freedom. The CMIN/DF was 1.361 and was therefore within the recommended range of 1 to 3. The root mean square residual 0.01 was within the acceptable range of 0.08 or less. The Goodness-of-Fit derived from this CFA was 0.927. The recommended level of acceptance for the GFI is 0.90. The GFI for this model fell well above the acceptable value, the AGFI derived in this model at 0.899 is therefore at the recommended level of acceptability. There is no recommended minimum level for PGFI (Hair et al, 1998). The summary model from this CFA and particularly the GFIs indicated that this model was a parsimonious solution.

As noted in Section 7.2, a number of questionnaire items which had been included in Study One were not included in the questionnaire used for Study Two. To provide further evidence for the appropriateness of their exclusion, these items were again tested in the CFA model described here. They did not, however, improve the Goodness of Fit indices.

8.6.3 CFA Model Validation for Cruise Phase

In order to evaluate the stability of the factor model and to examine its reliability and generalizability, the data set used during elicitation of the Initial and Final CFA structural models was separated by flight phase into two data sets. The factor structures from the Cruise phase and the Descent and Approach phase were thus examined separately. The hypothesis being tested in this section was that if the final CFA model described in Section 8.6.2 was valid it would be generalizable and would be a good fit for both of these data sets.

Table 8-33 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Structure & Sequence

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
23. The structure of the FMS is well organised from the user's point of view.	Structure & Sequence	1.000	0.822	-	-
21. It is clear why the system is organised and structured as it is.	Structure & Sequence	0.932	0.709	0.095	9.780
10. The sequence of activities required in the use of the FMS is as I would expect.	Structure & Sequence	0.777	0.626	0.093	8.402
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	Structure & Sequence	0.875	0.716	0.088	9.895

As can be seen from Table 8-33, the critical ratios are all well above 1.96 indicating that the latent variable, **Structure & Sequence**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-34 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Appropriate Functionality

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
25. FMS users can access all the information they require in this flight phase (Cruise or Descent & Approach).	Appropriate Functionality	1.000	0.781	-	-
26. The FMS screens provide all the information which is relevant to the task.	Appropriate Functionality	1.059	0.893	0.085	12.385
24. The FMS performs all the functions that I need it to.	Appropriate Functionality	0.948	0.699	0.101	9.354
30. The FMS allows the user to do the required task as efficiently as possible.	Appropriate Functionality	0.808	0.788	0.075	10.787

As can be seen from Table 8-34, the critical ratios are all well above 1.96 indicating that the latent variable, **Appropriate Functionality**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-35 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Feedback

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	Feedback	1.000	0.778		
19. It is clear to the user what should be done to correct any error that occurs.	Feedback	1.142	0.875	0.113	10.092
39. The FMS help facility (either on-line or manuals) is adequate.	Feedback	0.660	0.478	0.112	5.885

As can be seen from Table 8-35, the critical ratios are all well above 1.96 indicating that the latent variable, **Feedback**, loaded significantly onto both of the indicators ($p<0.05$).

Table 8-36 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Predictability

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	Predictability	1.000	0.893		
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	Predictability	0.901	0.702	0.124	7.253

As can be seen from Table 8-36, the critical ratio is well above 1.96 indicating that the latent variable, **Predictability**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-37 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Information Presentation

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	Information Presentation	1.000	0.366		
3. Important information is adequately highlighted on the screen.	Information Presentation	2.001	0.683	0.471	4.252
4. Information is organised logically on the CDU screen (familiarity aside).	Information Presentation	1.917	0.633	0.464	4.130

As can be seen from Table 8-37, the critical ratios are all well above 1.96 indicating that the latent variable, **Information Presentation**, loaded significantly onto both of the indicators ($p<0.05$).

Table 8-38 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Error Prevention & Correction

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
37. The FMS is protected against common trivial entry errors.	Error Prevention & Correction	0.824	0.647	0.161	5.119
36. The FMS has facilities for detecting and handling input errors.	Error Prevention & Correction	1.000	0.850		

As can be seen from Table 8-38, the critical ratio is well above 1.96 indicating that the latent variable, **Error Prevention & Correction**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-39 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Processor Speed

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
29. The FMS processes information too slowly for other activities in the flight deck.	Processor Speed	1.268	0.860	0.189	6.697
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	Processor Speed	1.000	0.730		

Similarly, as can be seen from Table 8-39, the critical ratio is well above 1.96 indicating that the latent variable, **Processor Speed**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-40 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Workload

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
42. The FMS can increase physical workload at inappropriate times.	Workload	0.825	0.696	0.142	5.814
41. The FMS can increase mental workload at inappropriate times.	Workload	1.000	0.864		

As can be seen from Table 8-40, the critical ratio is well above 1.96 indicating that the latent variable, **Workload**, loaded significantly onto one of the indicators ($p<0.05$). Again, an assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-41 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Consistency

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	Consistency	1.147	0.844	0.133	8.606
5. The way the FMS (CDU) look and works is consistent at all times.	Consistency	1.000	0.715		

As can be seen from Table 8-41, **Consistency**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-42 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Flexibility for Navigation

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
34. The FMS is adequately flexible for all vertical navigation.	Flexibility for Navigation	1.026	0.733	0.126	8.126
33. The FMS is adequately flexible for all lateral navigation.	Flexibility for Navigation	1.000	0.712		

As can be seen from Table 8-42, the critical ratio is well above 1.96 indicating that the latent variable, **Flexibility for Navigation**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-43 shows the squared multiple correlations for each indicator, which illustrates the amount of variance in the indicators, accounted for by the latent construct. Over 65% of the estimates exceeded the 0.5 value recommended by Hair et al (1998). Those estimates that fell below the recommended value have been highlighted in Table 8-43.

Table 8-43 - Squared Multiple Correlations (Cruise)

Latent & Indicators	*Estimates*
Structure & Sequence	
23. The structure of the FMS is well organised from the user's point of view.	0.675
21. It is clear why the system is organised and structured as it is.	0.503
10. The sequence of activities required in the use of the FMS is as I would expect.	0.392
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	0.512
Appropriate Functionality	
25. FMS users can access all the information they require in this flight phase (Cruise).	0.610
26. The FMS screens provide all the information which is relevant to the task.	0.798
24. The FMS performs all the functions that I need it to.	0.489
30. The FMS allows the user to do the required task as efficiently as possible.	0.621
Feedback	
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	0.605
19. It is clear to the user what should be done to correct any error that occurs.	0.766
39. The FMS help facility (either on-line or manuals) is adequate.	0.228
Predictability	
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	0.798
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	0.493
Information Presentation	
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	0.134
3. Important information is adequately highlighted on the screen.	0.466
4. Information is organised logically on the CDU screen (familiarity aside).	0.400
Error Prevention & Correction	
37. The FMS is protected against common trivial entry errors.	0.418
36. The FMS has facilities for detecting and handling input errors.	0.723
Processor Speed	
29. The FMS processes information too slowly for other activities in the flight deck.	0.740
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	0.533

Workload	
42. The FMS can increase physical workload at inappropriate times.	0.484
41. The FMS can increase mental workload at inappropriate times.	0.746
Consistency	
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	0.713
5. The way the FMS (CDU) look and works is consistent at all times.	0.511
Flexibility for Navigation	
34. The FMS is adequately flexible for all vertical navigation.	0.537
33. The FMS is adequately flexible for all lateral navigation.	0.507

Table 8-44 - Summary of Models (Cruise)

Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*	*RMR*	*GFI*	*AGFI*	*PGFI*
Default	96	339.20	255	0.000	1.330	0.134	0.871	0.822	0.632

The Number of Parameters derived in the model was 96. The Chi-Square Minimum was 339.20, with 255 degrees of freedom. The CMIN/DF of 1.330 was within the acceptable range of 1 to 3. The root mean square residual of 0.02 is below the critical value of 0.08. The Goodness-of-Fit derived from this CFA was 0.871 which was also above the level of acceptance. The recommended level of acceptance for AGFI is 0.90, although Hair et al (1998), described an example value of 0.810 as being of "*marginal*" acceptability (p.660). However, the AGFI derived in this model at 0.822 was just above the marginal level of acceptability. There is no recommended minimum level for PGFI (Hair et al, 1998).

The results of this CFA suggested that the Final Revised CFA model described in Section 8.6.2 was supported by the Cruise data. Evidence of the stability and generalizability of the final CFA model was therefore demonstrated.

8.6.4 CFA Model Validation for Descent & Approach Phase

A further CFA was conducted on the data elicited from ratings of the Descent and Approach phase. This examined whether the Final CFA model derived in Section 8.6.2 would fit this data set.

Table 8-45 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Structure & Sequence

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
23. The structure of the FMS is well organised from the user's point of view.	Structure & Sequence	1.000	0.808	-	-
21. It is clear why the system is organised and structured as it is.	Structure & Sequence	0.921	0.689	0.104	8.823
10. The sequence of activities required in the use of the FMS is as I would expect.	Structure & Sequence	0.898	0.737	0.094	9.558
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	Structure & Sequence	0.846	0.685	0.097	8.759

As can be seen from Table 8-45, the critical ratios are all well above 1.96 indicating that the latent variable, **Structure & Sequence**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-46 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Appropriate Functionality

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
25. FMS users can access all the information they require in this flight phase (Descent & Approach).	Appropriate Functionality	1.000	0.766		
26. The FMS screens provide all the information which is relevant to the task.	Appropriate Functionality	1.014	0.756	0.113	8.999
24. The FMS performs all the functions that I need it to.	Appropriate Functionality	0.982	0.652	0.127	7.744
30. The FMS allows the user to do the required task as efficiently as possible.	Appropriate Functionality	0.977	0.773	0.106	9.195

As can be seen from Table 8-46, the critical ratios are all well above 1.96 indicating that the latent variable, **Appropriate Functionality**, loaded significantly onto all of the indicators ($p<0.05$).

Table 8-47 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Feedback

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	Feedback	1.000	0.806		
19. It is clear to the user what should be done to correct any error that occurs.	Feedback	0.838	0.808	0.099	8.467
39. The FMS help facility (either on-line or manuals) is adequate.	Feedback	0.494	0.405	0.107	4.625

As can be seen from Table 8-47, the critical ratios are all well above 1.96 indicating that the latent variable, **Feedback**, loaded significantly onto both of the indicators ($p<0.05$).

Table 8-48 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Predictability

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	Predictability	1.000	0.770		
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	Predictability	1.117	0.818	0.149	7.517

As can be seen from Table 8-48, the critical ratio is well above 1.96 indicating that the latent variable, **Predictability**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-49 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Information Presentation

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	Information Presentation	1.000	0.275		
3. Important information is adequately highlighted on the screen.	Information Presentation	3.589	0.732	1.177	3.049
4. Information is organised logically on the CDU screen (familiarity aside).	Information Presentation	3.395	0.766	1.109	3.062

As can be seen from Table 8-49, the critical ratios are both above 1.96 indicating that the latent variable, **Information Presentation**, loaded significantly onto both of the indicators ($p<0.05$).

Table 8-50 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Error Prevention & Correction

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
37. The FMS is protected against common trivial entry errors.	Error Prevention & Correction	0.569	0.524	0.074	7.650
36. The FMS has facilities for detecting and handling input errors.	Error Prevention & Correction	1.000	1.000		

As can be seen from Table 8-50, the critical ratio is well above 1.96 indicating that the latent variable, **Error Prevention & Correction**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-51 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Processor Speed

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
29. The FMS processes information too slowly for other activities in the flight deck.	Processor Speed	0.865	0.754	0.128	6.735
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	Processor Speed	1.000	0.872		

As can be seen from Table 8-51, the critical ratio is well above 1.96 indicating that the latent variable, **Processor Speed**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-52 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Workload

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
42. The FMS can increase physical workload at inappropriate times.	Workload	0.973	0.720	0.135	7.230
41. The FMS can increase mental workload at inappropriate times.	Workload	1.000	0.839		

As can be seen from Table 8-52, the critical ratio is well above 1.96 indicating that the latent variable, **Workload**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-53 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Consistency

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	Consistency	0.984	0.589	0.170	5.783
5. The way the FMS (CDU) look and works is consistent at all times.	Consistency	1.000	0.664		

As can be seen from Table 8-53, the critical ratio is well above 1.96 indicating that the latent variable, **Consistency**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-54 - Unstandardized & Standardized Regression Weights, Standard Errors & Critical Ratios for Flexibility for Navigation

Indicator	*Latent*	*Unstandardized Regression Weight*	*Standardized Regression Weight*	*Standard Error*	*Critical Ratio*
34. The FMS is adequately flexible for all vertical navigation.	Flexibility for Navigation	0.812	0.618	0.121	6.688
33. The FMS is adequately flexible for all lateral navigation.	Flexibility for Navigation	1.000	0.754		

As can be seen from Table 8-54, the critical ratio is well above 1.96 indicating that the latent variable, **Flexibility for Navigation**, loaded significantly onto one of the indicators ($p<0.05$). An assessment could not be made of the second observed variable as it was set by the researcher as a marker variable.

Table 8-55 shows the squared multiple correlations for the Descent & Approach phase of flight. Guidelines suggest that the variance extracted value should exceed 0.50 for a construct (Hair et al, 1998). Over 65% percent of the estimates exceeded the 0.5 value recommended by Hair et al (1998). Those estimates that fell below the recommended value have been highlighted in Table 8-55.

Table 8-55 - Squared Multiple Correlations (Descent & Approach)

Latent & Indicators	*Estimates*
Structure & Sequence	
23. The structure of the FMS is well organised from the user's point of view.	0.653
21. It is clear why the system is organised and structured as it is.	0.475
10. The sequence of activities required in the use of the FMS is as I would expect.	0.543
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	0.469
Appropriate Functionality	
25. FMS users can access all the information they require in this flight phase (Descent & Approach).	0.587
26. The FMS screens provide all the information which is relevant to the task.	0.571
24. The FMS performs all the functions that I need it to.	0.425
30. The FMS allows the user to do the required task as efficiently as possible.	0.598
Feedback	
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	0.650
19. It is clear to the user what should be done to correct any error that occurs.	0.654
39. The FMS help facility (either on-line or manuals) is adequate.	0.164
Predictability	
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	0.593
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	0.669
Information Presentation	
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	0.075
3. Important information is adequately highlighted on the screen.	0.536
4. Information is organised logically on the CDU screen (familiarity aside).	0.586
Error Prevention & Correction	
37. The FMS is protected against common trivial entry errors.	0.274
36. The FMS has facilities for detecting and handling input errors.	1.000
Processor Speed	
29. The FMS processes information too slowly for other activities in the flight deck.	0.569
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	0.761

Latent & Indicators	Estimates
Workload	
42. The FMS can increase physical workload at inappropriate times.	0.519
Consistency	
41	0.704
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	
5. The way the FMS (CDU) look and works is consistent at all times.	
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	0.346
5. The way the FMS (CDU) look and works is consistent at all times.	0.440
Flexibility for Navigation	
34. The FMS is adequately flexible for all vertical navigation.	0.382
33. The FMS is adequately flexible for all lateral navigation.	0.569

Table 8-56 shows a summary of the model when applied to the Descent & Approach ratings.

Table 8-56 - Summary of Models (Descent & Approach)

Model	*NPAR*	*CMIN*	*DF*	*P*	*CMIN/DF*	*RMR*	*GFI*	*AGFI*	*PGFI*
Default	96	331.132	255	0.001	1.229	0.135	0.870	0.821	0.632

The Number of Parameters derived in the model was 96. The Chi-Square Minimum was 331.32, with 255 degrees of freedom. The CMIN/DF was 1.229 and was therefore within the acceptable range of 1 to 3. The root mean square residual 0.02 was below the critical value 0.08 or less. Since the GFI derived from this CFA was 0.870, the model fell above the critical value (Hair et al, 1998). However, the AGFI derived in this model at 0.821 was just above the marginal level of acceptability.

8.6.5 CFA Summary

Based on the results of this CFA it is suggested that the model is relatively stable. The hypothesis being tested in this section was that if the final CFA model described in Section 8.6.2 was valid it would be generalizable and would be a good fit for both the Cruise and Descent & Approach data sets when they were analysed independently. The CFA model was a good fit for both sets of data and evidence of the stability of the model was therefore demonstrated.

8.7 Production of Scales

Since the factor structure was shown to be replicable and stable in Studies One and Two, and for both the Cruise and Descent & Approach phases of flight, usability scales could be produced with confidence. Scales provide the advantage of producing a simplified and more parsimonious view of a multidimensional construct, such as usability, than is possible using a factor structure.

In order to examine the internal consistency of the scales produced, reliability analyses in the form of Cronbach's Alphas were conducted. The internal consistency coefficients were largely well above the acceptable level for scales as can been seen from Table 8-57 (Hair et al, 1998).

Table 8-57 - Reliability Coefficients

Scale	*Reliability Coefficient*	*Number of Items*
Structure & Sequence	Alpha = 0.81	4
Appropriate Functionality	Alpha = 0.85	4
Feedback	Alpha = 0.70	3
Predictability	Alpha = 0.77	2
Information Presentation	Alpha = 0.51	3
Error Prevention & Correction	Alpha = 0.71	2
Processor Speed	Alpha = 0.78	2
Workload	Alpha = 0.75	2
Consistency	Alpha = 0.69	2
Flexibility for Navigation	Alpha = 0.66	2

The scales used in the following series of analyses were created through summation of the respondent ratings on each questionnaire item within a scale, and subsequently dividing the total by the number of items in a scale. The correlation matrix at Table 8-58 shows the intercorrelations between all of the scales produced, permitting the shared variance between the scales to be examined. It should be noted that all of the correlations were significant ($p<0.05$) except for the correlation **between Processor Speed** and **Error Prevention & Correction**. Highlighting has been used to identify the non-significant correlation.

Table 8-58 - Pearson Correlation Coefficients between the Usability Scales Produced

Scales	*Appropriate Functionality*	*Feedback*	*Predictability*	*Information Presentation*	*Error Prevention & Correction*	*Processor Speed*	*Workload*	*Consistency*	*Flexibility for Navigation*
Structure & Sequence	0.471	0.489	0.466	0.605	0.236	0.256	0.354	0.570	0.492
Appropriate Functionality	---	0.379	0.234	0.371	0.210	0.327	0.253	0.413	0.609
Feedback	---	---	0.360	0.351	0.374	0.196	0.207	0.345	0.394
Predictability	---	---	---	0.329	0.231	0.181	0.351	0.334	0.310
Information Presentation	---	---	---	---	0.201	0.267	0.331	0.480	0.411
Error Prevention & Correction	---	---	---	---	---	0.102	0.105	0.145	0.199
Processor Speed	---	---	---	---	---	---	0.349	0.373	0.320
Workload	---	---	---	---	---	---	---	0.317	0.303
Consistency	---	---	---	---	---	---	---	---	0.396

Note: The correlation between Error Prevention & Correction and Processor Speed has been highlighted to show that it was not significant ($p<0.05$).

8.8 Discriminant Function Analysis

The purpose of the series of DFAs described in this section was to determine if the *scales* produced demonstrated discriminant validity as the factors had done in the previous study. It was also hypothesised that it would be possible to predict group membership (e.g. manufacturer of aircraft/FMS on which pilot was currently type-rated) on the basis of participants ratings of the ten usability scales. It should be noted that throughout this section highlighting has been used to show when a specified criterion value (e.g. $p<0.05$) has not been achieved.

8.8.1 Discriminant Validity Evidence - DFA Examining Impact of FMS and Aircraft Manufacturer

The primary goal of this three-group DFA was to discriminate between FMS manufacturers (Honeywell and Smiths) and aircraft manufacturers (Boeing and Airbus) on the basis of respondents' ratings of the usability scales.

A DFA permitted the researcher to determine on which of the predictor components, (irrespective of flight phase), the aircrew type-rated on Boeing aircraft with Smiths FMSs (Boeing Smiths), Boeing aircraft with Honeywell FMSs (Boeing Honeywells) and Airbus aircraft with Honeywell FMSs (Airbus Honeywells) were maximally different, and to predict group membership on the basis of those predictor variables used to create the scales.

Of the original 356 cases, 24 were eliminated from the analysis as they had at least one missing discriminating variable.

Table 8-59 shows that two discriminant functions were calculated, with a combined $\chi^2 = 59.494$, $df=20$, $p<0.01$. After removal of the first function, there was still some discriminating power although it was non-significant ($\chi^2 = 15.906$, $df=9$, $p=0.076$). The significant discriminant function accounted 12.7% of the total sample variance. The non-significant function, which is highlighted in Table 8-59, accounted for 4.7% of the total sample variance.

Table 8-59 - Canonical Discriminant Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*
1	0.145	0.356
2	0.049	0.217

Function	*Wilks' Lambda*	*Chi²*	*df*	*P*
1 – 2	0.832	59.494	20	0.000
2	0.953	15.606	9	0.076

When the group centroids for the canonical discriminant functions were analysed (Table 8-60), it became apparent that for the first function, the solution was maximally discriminating between the usability ratings of pilots type-rated on Boeing aircraft with Smiths FMSs (group centroid = -0.193) and those type-rated on Airbus aircraft with Honeywell FMSs (group centroid = 1.034).

For the second non-significant function, the solution was maximally discriminating between the usability ratings of those pilots type-rated on Boeing aircraft with Smiths FMSs (group centroid = 0.356) and those type-rated on Boeing aircraft with Honeywell FMSs (group centroid = -0.159).

Table 8-60 - Functions at Group Centroids

	Function	
Manufacturer	*1*	*2*
Boeing Smiths	-0.193	0.356
Boeing Honeywell	-0.115	-0.159
Airbus Honeywell	1.034	0.050

As shown on Table 8-61, on the first discriminant function, four predictor components had loadings in excess of 0.45. These were **Structure & Sequence, Information Presentation, Error Prevention & Correction** and **Processor Speed.** The non-significant function has been highlighted.

Table 8-61 - Wilks' Lambda, Univariate F Ratios with 2,328 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients

Variable	*Wilks' Lambda*	*F Ratio*	*p*	*DFA Coefficient 1*	*DFA Coefficient 2*
Structure & Sequence	0.999	0.114	0.892	-0.555	-0.434
Appropriate Functionality	0.985	2.441	0.089	-0.421	0.949
Feedback	0.976	4.092	0.018	0.395	-0.643
Predictability	0.992	1.354	0.260	0.184	0.085
Information Presentation	0.981	3.160	0.044	0.517	0.496
Error Prevention & Correction	0.965	5.884	0.003	0.470	0.206
Processor Speed	0.968	5.453	0.005	0.588	0.018
Workload	0.993	1.198	0.303	-0.022	0.311
Consistency	0.996	0.651	0.522	0.084	-0.160
Flexibility for Navigation	0.998	0.401	0.670	-0.353	-0.468

The group means and standard deviations (scale scores) show the relative usability ratings of the pilots type-rated on Boeing & Airbus aircraft with Smiths or Honeywell FMSs. These are shown on Table 8-62.

Table 8-62 - Group Statistics

	Boeing Smiths[a]		*Boeing Honeywell*[b]		*Airbus Honeywell*[c]	
Scale	*Mean*	*sd*	*Mean*	*sd*	*Mean*	*sd*
Structure & Sequence	4.16	1.06	4.22	1.14	4.17	1.27
Appropriate Functionality	4.63	1.12	4.31	1.32	4.19	1.38
Feedback	2.63	0.96	2.89	1.21	3.27	1.34
Predictability	4.24	1.62	4.27	1.54	4.69	1.49
Information Presentation	4.48	0.92	4.37	0.98	4.80	1.17
Error Prevention & Correction	3.73	1.31	3.71	1.47	4.55	1.48
Processor Speed	2.91	1.50	2.84	1.70	3.77	1.69
Workload	3.42	1.57	3.21	1.56	3.58	1.60
Consistency	4.04	1.31	4.07	1.40	4.32	1.13
Flexibility for Navigation	4.08	1.35	4.10	1.36	3.89	1.57

[a]n=86, [b]n=206, [c]n=39

Pilots type-rated on Boeing aircraft with Honeywell FMSs rated the usability of their FMS CDU more favourably than either those type-rated on Boeing aircraft with Smiths FMSs or Airbus aircraft with Honeywell FMSs on the **Structure & Sequence** scale. The respondents type-rated on Airbus aircraft with Honeywell FMSs made more favourable ratings than those type-rated on Boeing aircraft with Smiths FMSs.

Pilots type-rated on Airbus aircraft with Honeywell FMSs made more positive usability ratings than those type-rated on Boeing aircraft on the **Information Presentation, Error Prevention & Correction** and **Processor Speed** components. While those pilots type-rated on Boeing aircraft with Smiths FMSs made more positive usability ratings on the latter three components that did those type-rated on Boeing aircraft with Honeywell FMSs.

Table 8-61 shows that four predictor components had loadings in excess of 0.45 on the non-significant function. These were **Appropriate Functionality, Feedback, Information Presentation** and **Processor Speed** and are described here for information purposes only.

Pilots type-rated on Boeing aircraft with Smiths FMSs made more favourable usability ratings on the **Appropriate Functionality** dimension than those type-rated on Boeing aircraft with Honeywell FMSs or those type-rated on Airbus aircraft. However, those type-rated on Airbus aircraft gave more positive usability ratings than those type-rated on Boeing aircraft on the **Feedback, Information Presentation** and **Processor Speed** components. Those type-rated on Boeing aircraft with Smiths FMSs, however, made more positive usability ratings than those type-rated on Boeing aircraft with Honeywell FMSs.

Table 8-63 provides a rank ordered overview of usability ratings by FMS and aircraft type for the significant DFA function.

Table 8-63 - Usability Ratings Ranked by Aircraft and FMS Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)		
Scale	*Boeing Smiths*	*Boeing Honeywell*	*Airbus Honeywell*
Structure & Sequence	3	1	2
Information Presentation	2	3	1
Error Prevention & Correction	2	3	1
Processor Speed	2	3	1

Overall, a total of 63.7% of cases were correctly classified on the basis of whether the pilot's current type-rating was held on a Boeing aircraft with a Smiths FMS, or a Boeing or Airbus aircraft with a Honeywell FMS. Table 8-64 shows the classification matrix for the DFA. Percentages of correct classification are shown in parentheses.

Table 8-64 - Classification Matrix for DFA Predicting Group Members from Aircraft and FMS Manufacturer

Actual Group	*Predicated Group Membership*			
	Boeing Smiths	*Boeing Honeywell*	*Airbus Honeywell*	*Total*
Boeing Smiths	8 (9.3%)	74 (86.0%)	4 (4.7%)	86 (100.0%)
Boeing Honeywell	5 (2.4%)	196 (95.2%)	5 (2.4%)	206 (100.0%)
Airbus Honeywell	0 (0.0%)	32(82.1%)	7 (17.9%)	39 (100.0%)

In order to determine the discriminatory power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. Press's Q in this classification matrix is 137.77, where the critical value at a significance level of 0.01 is 6.63. The classification matrix is therefore statistically better than chance at the $p<0.01$ level. It should however be noted while the classification matrix was significant in statistical terms, no practical evidence of criterion validity was demonstrated. Over a third of cases (36.3%) were incorrectly classified overall and, for example, 86.0% of Smiths FMSs on Boeing aircraft were incorrectly classified as being Honeywell FMSs. Similarly 82.1% of Airbus aircraft with Honeywell FMSs were incorrectly classified as Boeing aircraft with Honeywell FMSs. No practical evidence of criterion validity was demonstrated since it was not possible to predict participants' current aircraft / FMS type accurately from their usability scale ratings.

To assist the reader in making comparisons between the rank order of usability ratings made by participants in Studies One and Two, a summary is shown on Table 8-65. It should be noted that there are only two aircraft manufacturers included for Study Two since no aircrew type-rated on Fokker aircraft participated in this study. The data from Study One is taken from Table 4-18 and uses the factors derived from the PCA in Study One.

Table 8-65 - Comparison of Usability Rating Rank Orders by Aircraft and FMS Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)						
	Study One				*Study Two*		
Scale	*Boeing Smiths*	*Boeing Honeywell*	*Airbus Honeywell*	*Fokker Honeywell*	*Boeing Smiths*	*Boeing Honeywell*	*Airbus Honeywell*
Structure & Sequence	1	2	3	4	3	1	2
Appropriate Functionality	1	2	4	3	---	---	---
Predictability	2	3	4	1	---	---	---
Information Presentation	3	4	2	1	2	3	1
Flexibility	4	2	1	3	---	---	---
Processor Speed	2	3	1	4	2	3	1
Error Prevention & Correction	---	---	---	---	2	3	1

8.8.2 Discriminant Validity Evidence - DFA Examining Impact of FMS Manufacturer

The primary goal of this two-group DFA was to discriminate between FMS manufacturers (Honeywell and Smiths) on the basis of respondents' ratings of their FMS CDU on the usability scales irrespective of flight phase. A DFA permitted the researcher to determine on which of the predictor components the aircrew type-rated on aircraft with Smiths FMSs and Honeywell FMSs were maximally different, and to predict group membership on the basis of those predictor variables used to create the dimension.

Of the original 331 cases, 25 cases were eliminated from the analysis as they had at least one missing discriminating variable. A single discriminant function (χ^2 = 18.108, df=10, p≤ .05) was calculated as shown in Table 8-66. The function accounted for 5.4% of the between group variability.

Table 8-66 - Canonical Discriminant Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*	*Wilks' Lambda*	*Chi²*	*df*	*p*
1	0.057	0.233	0.946	18.108	10	0.05

The group centroid for those pilots type-rated on aircraft with Honeywell FMSs was 0.142 while being –0.403 for those pilots type-rated on aircraft with Smiths FMSs.

As shown in Table 8-67, two of the predictor components had loadings in excess of 0.45 on the discriminating function.

Table 8-67 - Wilks' Lambda, Univariate F Ratios with 1,329 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients.

Component	*Wilks' Lambda*	*F Ratio*	*p*	*DFA Coefficient Function 1*
Structure & Sequence	1.000	0.147	0.701	0.141
Appropriate Functionality	0.986	4.615	0.032	-1.033
Feedback	0.986	4.678	0.031	0.752
Predictability	0.999	0.244	0.622	0.005
Information Presentation	1.000	0.101	0.751	-0.215
Error Prevention & Correction	0.999	0.406	0.525	0.025
Processor Speed	1.000	0.137	0.712	0.248
Workload	0.998	0.590	0.443	-0.287
Consistency	0.999	0.170	0.681	0.180
Flexibility for Navigation	1.000	0.009	0.926	0.261

The means and standard deviations for each component are shown in Table 8-68. As can be seen from the group means (scale scores) in Table 8-68, the usability ratings were more favourable for Smiths FMSs than for Honeywell FMSs on the **Appropriate Functionality** dimension. Conversely, Honeywell FMSs was rated more favourably on the **Feedback** dimension. The non-significant components ($p>0.05$) have been highlighted.

Table 8-68 - Group Statistics

	Honeywell[a]		*Smiths*[b]	
Scale	*Mean*	*sd*	*Mean*	*sd*
Structure & Sequence	4.21	1.16	4.16	1.06
Appropriate Functionality	4.29	1.32	4.63	1.12
Feedback	2.95	1.24	2.63	0.96
Predictability	4.33	1.53	4.24	1.62
Information Presentation	4.44	1.02	4.48	0.92
Error Prevention & Correction	3.84	1.50	3.73	1.31
Processor Speed	2.98	1.70	2.91	1.50
Workload	3.27	1.57	3.42	1.57
Consistency	4.11	1.36	4.04	1.31
Flexibility for Navigation	4.06	1.39	4.08	1.35

[a]n=245, [b]n=86

Table 8-69 provides a rank ordered overview of usability ratings by FMS and aircraft type for the significant DFA function.

Table 8-69 - Usability Ratings Ranked by FMS Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)	
Scale	*Honeywell*	*Smiths*
Appropriate Functionality	2	1
Feedback	1	2

Overall, a total of 73.4% of cases were correctly classified on the basis of whether pilots currently held a type-rating on an aircraft with a Smiths or a Honeywell FMS. Table 8-70 shows the classification matrix for the DFA. Percentages of correct classification are shown in parentheses.

Table 8-70 - Classification Matrix for DFA Predicting Group Members from type rating on Aircraft with Smiths or Honeywell FMSs

	Predicated Group Membership		
Actual Group	*Smiths*	*Honeywell*	*Total*
Smiths	3 (3.5%)	83 (96.5%)	86 (100%)
Honeywell	5 (2.0%)	240 (98.0%)	245 (100%)

In order to determine the discriminatory power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. Press's Q in this classification matrix is 72.58, where the critical value at a significance level of 0.01 is 6.63. While this indicates that the classification matrix is statistically better than chance at the $p<0.01$ level, the numbers indicate that almost all Smiths FMSs (96.5%) were incorrectly classified as being Honeywell FMSs. Therefore, no practical evidence of criterion validity was demonstrated.

To assist the reader in making comparisons between the rank order of usability ratings made by participants in Studies One and Two, a summary is shown on Table 8-71.

Table 8-71 - Comparison of Usability Rank Orders by FMS Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)			
	Study One		*Study Two*	
Scale	*Honeywell*	*Smiths*	*Honeywell*	*Smiths*
Structure & Sequence	2	1	---	---
Appropriate Functionality	2	1	2	1
Feedback	1	2	1	2

8.8.3 Discriminant Validity Evidence - DFA Examining Impact of Aircraft Manufacturer

The primary goal of this two-group DFA was to discriminate between aircraft manufacturers (Airbus or Boeing) on the basis of respondents' ratings of the usability of their FMS CDU. A DFA permitted the researcher to determine on which of the predictor components the aircrew type-rated on Boeing or Airbus aircraft were maximally different, and to predict group membership on the basis of those predictor variables used to create the dimension.

Of the original 331 cases, 25 were eliminated from the analysis as they had at least one missing discriminating variable. A single discriminant function was calculated (χ^2= 43.726, df=10, p≤ .05) as shown on Table 8-72. The function accounted for 12.6% of the between group variability.

Table 8-72 - Canonical Discriminant Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*	*Wilks' Lambda*	*Chi²*	*df*	*p*
1	0.144	0.355	0.874	43.726	10	0.000

The group centroid for those pilots type-rated on Boeing aircraft was -0.138 and was 1.037 for those pilots type-rated on Airbus aircraft. As shown in Table 8-73, four predictor components had loadings in excess of 0.45 on the discriminating function and these were the **Structure & Sequence**, **Information Presentation**, **Error Prevention & Correction** and **Processor Speed** scales.

Table 8-73 - Wilks' Lambda, Univariate F Ratios with 1,329 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients

Component	*Wilks' Lambda*	*F Ratio*	*p*	*DFA Coefficient Function 1*
Structure & Sequence	1.000	0.039	0.844	-0.578
Appropriate Functionality	0.997	0.949	0.331	-0.369
Feedback	0.984	5.208	0.023	0.359
Predictability	0.992	0.696	0.102	0.188
Information Presentation	0.983	5.597	0.019	0.545
Error Prevention & Correction	0.965	11.794	0.001	0.481
Processor Speed	0.968	10.817	0.001	0.588
Workload	0.996	1.311	0.253	-0.005
Consistency	0.996	1.277	0.259	0.075
Flexibility for Navigation	0.998	0.795	0.373	-0.378

The group means and standard deviations for each scale are shown in Table 8-74.

Table 8-74 - Group Statistics

	Boeing[a]		Airbus[b]	
Scale	*Mean*	*sd*	*Mean*	*sd*
Structure & Sequence	4.20	1.11	4.17	1.27
Appropriate Functionality	4.40	1.27	4.19	1.38
Feedback	2.81	1.15	3.27	1.34
Predictability	4.26	1.56	4.69	1.49
Information Presentation	4.40	0.96	4.80	1.17
Error Prevention & Correction	3.71	1.42	4.55	1.48
Processor Speed	2.86	1.62	3.77	1.69
Workload	3.27	1.56	3.58	1.60
Consistency	4.06	1.40	4.32	1.13
Flexibility for Navigation	4.09	1.35	3.88	1.57

[a]n=292, [b]n=39.

Boeing aircraft were given more favourable FMS usability ratings than Airbus aircraft on the **Structure & Sequence** dimension while Airbus aircraft were given more favourable ratings than Boeing aircraft on the **Information Presentation, Error Prevention & Correction** and **Processor Speed** scales ($p \leq .05$).

Table 8-75 provides a rank ordered overview of usability ratings by aircraft type for the significant DFA function.

Table 8-75 - Usability Ratings Aircraft Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)	
Scale	*Boeing*	*Airbus*
Structure & Sequence	1	2
Information Presentation	2	1
Error Prevention & Correction	2	1
Processor Speed	2	1

Overall, a total of 87.6% of cases were correctly classified on the basis of whether the pilot currently held a Boeing or Airbus type-rating. Table 8-76 shows the classification matrix for the DFA. Percentages of correct classification are shown in parentheses.

Table 8-76 - Classification Matrix for DFA Predicting Group Members from Boeing/Airbus type-rating

	Predicated Group Membership		
Actual Group	*Boeing*	*Airbus*	*Total*
Boeing	286 (97.9%)	6 (2.1%)	292 (100%)
Airbus	35 (89.7%)	4 (10.3%)	39 (100%)

In order to determine the discriminatory power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. Press's Q in this classification matrix is 187.31, where the critical value at a significance level of 0.01 is 6.63. While the classification matrix is therefore statistically better than chance at the $p<0.01$ level, a closer look at the matrix shows that no practical evidence of criterion validity has been demonstrated since almost all Airbus aircraft (89.7%) were incorrectly classified as Boeing aircraft.

To assist the reader in making comparisons between the rank order of usability ratings made by participants in Studies One and Two, a summary is shown on Table 8-77.

Table 8-77 - Comparison of Usability Rank Orders by Aircraft Manufacturer

	Rank of Usability Ratings (1 = Most Favourable Rating)				
	Study One			*Study Two*	
Scale	*Boeing*	*Airbus*	*Fokker*	*Boeing*	*Airbus*
Structure & Sequence	1	2	3	1	2
Appropriate Functionality	1	3	2	---	---
Predictability	2	3	1	---	---
Information Presentation	3	2	1	2	1
Flexibility	3	1	2	---	---
Processor Speed	2	1	3	2	1
Error Prevention & Correction	---	---	---	2	1

8.8.4 Discriminant Validity Evidence - DFA Examining Impact of Pilot Rank

The primary goal of this two-group DFA was to determine if ratings of usability on the scales would discriminate between the respondent pilots on the basis of their rank. It was anticipated that ratings on a usability measure would not be significantly influenced by whether the respondent was a Captain or a First Officer. A DFA permitted the researcher to determine on which of the predictor components the pilot's rank (First Officer or Captain) were maximally different, and to predict group membership on the basis of those predictor variables used to create the dimension.

Of the original 356 cases, 44 were eliminated from the analysis as they had at least one missing discriminating variable. As can be seen from Table 8-78, a single non-significant function was calculated ($\chi^2 = 17.85$, df=10, $p>0.05$).

Table 8-78 - Canonical Discriminant Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*	*Wilks' Lambda*	*Chi²*	*Df*	*p*
1	0.06	0.238	0.943	17.850	10	0.06

The group centroid for First Officers was 0.334 while being –0.179 for Captains. The Wilks' Lambda, Univariate F Ratios and Standardized Canonical Discriminant Function Coefficients are shown on Table 8-79.

Table 8-79 - Wilks' Lambda, Univariate F Ratios with 1,310 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients

Variable	*Wilks' Lambda*	*F Ratio*	*p*	*DFA Coefficient Function 1*
Structure & Sequence	0.995	1.52	0.219	0.330
Appropriate Functionality	0.997	0.854	0.356	0.128
Feedback	0.975	7.864	0.005	0.781
Predictability	0.999	0.296	0.587	-0.536
Information Presentation	0.999	0.238	0.626	-0.502
Error Prevention & Correction	0.995	1.667	0.198	0.149
Processor Speed	0.997	1.055	0.305	-0.463
Workload	0.998	0.738	0.391	0.400
Consistency	0.998	0.508	0.477	0.174
Flexibility for Navigation	1.000	0.148	0.701	-0.167

The means and standard deviations for each component are shown in Table 8-80 and are provided for information purposes only.

Table 8-80 - Group Statistics

	First Officer[a]		Captain[b]	
Scale	*Mean*	*Sd*	*Mean*	*sd*
Structure & Sequence	4.31	1.02	4.14	1.18
Appropriate Functionality	4.46	1.25	4.32	1.31
Feedback	3.13	1.20	2.74	1.16
Predictability	4.26	1.43	4.36	1.63
Information Presentation	4.42	0.96	4.48	1.00
Error Prevention & Correction	3.96	1.51	3.74	1.44
Processor Speed	2.83	1.55	3.03	1.71
Workload	3.39	1.60	3.23	1.56
Consistency	4.15	1.41	4.04	1.32
Flexibility for Navigation	4.11	1.32	4.05	1.41

[a]n=109, [b]n=203, total n=312

As hypothesised, the results of this DFA demonstrated that there was no significant difference in ratings of the FMS usability scales by Captains or First Officers. Although the DFA was non-significant, the classification matrix has been included in Table 8-81 for completeness. Overall, a total of 66.7% of cases were correctly classified on the basis of whether the pilot was a Captain or a First Officer. Percentages of correct classification are shown in parentheses.

Table 8-81 - Classification Matrix for DFA Predicting Group Members from Rank

	Predicated Group Membership		
Actual Group	*Captain*	*First Officer*	*Total*
Captain	190 (93.6%)	13 (6.4%)	203 (100%)
First Officer	91 (83.5%)	18 (16.5%)	109 (100%)

In order to determine the discriminatory power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. Press's Q for this classification matrix is 34.67, where the critical value at a significance level of 0.01 is 6.63. Although the classification matrix was significant in statistical terms ($p<0.01$), no practical evidence of criterion validity was demonstrated since 83.5% of First Officers were incorrectly classified as Captains.

8.8.5 Discriminant Validity Evidence - DFA Examining Impact of Flight Phase

The primary goal of this two-group DFA was to determine if ratings on the usability scales developed would discriminate between the Cruise and Descent & Approach phases of flight. As discussed in Section 6.1, it was hypothesised that it would not be possible to discriminate between flight phases on the basis of participants' usability ratings. A DFA permitted the researcher to determine on which of the predictor components the flight phases were maximally different, and to predict the flight phase rated on the basis of those predictor variables used to create the dimension.

Of the original 356 cases, 24 were eliminated from the analysis as they had at least one missing discriminating variable. As can be seen from Table 8-82 a single non-significant function was calculated which only accounted for 2.0% of the variance.

Table 8-82 - Canonical Discriminant Function Analysis Summary Statistics

Function	*Eigenvalue*	*Canonical Correlation*	*Wilks' Lambda*	*Chi^2*	*df*	*P*
1	0.021	0.143	0.979	6.757	10	0.748

The group centroid for the Cruise phase of flight was 0.142 while being –0.147 for the Descent & Approach phase of flight.

Table 8-83 - Wilks' Lambda, Univariate F Ratios with 1,330 Degrees of Freedom and Standardized Canonical Discriminant Function Coefficients.

Variable	*Wilks' Lambda*	*F Ratio*	*p*	*DFA Coefficient Function 1*
Structure & Sequence	0.998	0.668	0.414	0.194
Appropriate Functionality	0.993	2.457	0.118	0.690
Feedback	0.999	0.287	0.593	-0.019
Predictability	1.000	0.035	0.851	-0.429
Information Presentation	1.000	0.085	0.770	-0.224
Error Prevention & Correction	0.999	0.469	0.494	0.217
Processor Speed	0.999	0.273	0.602	-0.601
Workload	0.994	1.837	0.176	0.698
Consistency	0.999	0.219	0.640	0.030
Flexibility for Navigation	0.998	0.667	0.415	-0.043

The group means and standard deviations for each component are described in Table 8-84.

Table 8-84 - Group Statistics

	Cruise[a]		Descent & Approach[b]	
Scale	*Mean*	*Sd*	*Mean*	*Sd*
Structure & Sequence	4.25	1.13	4.15	1.13
Appropriate Functionality	4.48	1.34	4.26	1.21
Feedback	2.90	1.22	2.82	1.13
Predictability	4.29	1.54	4.32	1.57
Information Presentation	4.46	0.98	4.43	1.02
Error Prevention & Correction	3.87	1.44	3.75	1.47
Processor Speed	2.91	1.65	3.01	1.66
Workload	3.42	1.58	3.18	1.56
Consistency	4.12	1.48	4.05	1.19
Flexibility for Navigation	4.12	1.42	3.99	1.34

[a]n=169, [b]n=163, total n=332

The results of this DFA illustrated that there was no significant difference in the ratings of FMS usability by flight phase. An explanation for these results is included in the subsequent discussion section. Although the DFA was non-significant, the classification matrix has been included in Table 8-85 for completeness.

Table 8-85 - Classification Matrix for DFA Predicting Group Members from Flight Phase Rated

	Predicted Phase Rated		
Actual Phase Rated	*Cruise*	*Descent & Approach*	*Total*
Cruise	100 (59.2%)	69 (40.8%)	169 (100%)
Descent & Approach	80 (49.1%)	83 (50.9%)	163 (100%)

Overall, a total of 55.1% of cases were correctly classified on the basis of the flight phase rated. Percentages of correct classification are shown in parentheses. In order to determine the discriminatory power of the classification matrix, the matrix was compared with a chance model using Press's Q statistic. Press's Q for this classification matrix is 3.48, where the critical value at a significance level of 0.01 is 6.63. The classification matrix was therefore not significant ($p<0.01$).

8.9 Further analysis of relationship between pilots' experience and usability ratings

In order to examine the nature of the relationship between ratings of the usability scales and pilots' experience, a series of curvilinear regressions was performed. The purpose of the analyses was to examine if a point in terms of experience existed beyond which usability judgements stabilised. The series of curvilinear regression analyses were conducted between ratings of the usability scales as the dependent variables and both pilots' hours on type and pilots' total hours as independent variables.

Examining the relationship between pilots' total flying experience and their ratings of the usability scales, it was noted that a significant linear relationship existed between pilots' total hours and their ratings of the **Predictability** usability scale ($p<0.05$). More experienced pilots rated the **Predictability** usability scale more favourably, and ratings improved gradually from the midpoint of the usability scale as pilots' experience increased. The relationship, however, only accounted for 1.2% of the variance (see Appendix K). No significant F-values ($p.<0.05$) were established using non-linear regressions when modelled using a logarithmic, quadratic or a cubic relationship.

Significant F-values were established between pilots' total flying hours and their ratings of the **Error Prevention & Correction** scale using non-linear regressions when modelled using a quadratic and a cubic relationship ($p.<0.05$). The quadratic relationship accounted for 3.1% of the variance, while the cubic relationship accounted for 3.3% of the variance. Until approximately nine thousand hours, pilots' ratings of usability on the **Error Prevention & Correction** scale deteriorated from above the mid-point in the rating scale to below the midpoint in the rating scale. After approximately nine thousand hours, however, participants' ratings became more favourable again as they gained more experience eventually exceeding initial usability ratings (see Appendix K).

No significant F-values ($p.<0.05$) were established for the **Structure & Sequence, Appropriate Functionality, Feedback, Information Presentation, Error Prevention & Correction, Workload** and **Consistency** scales, using linear or non-linear regressions when modelled using a logarithmic, a quadratic or a cubic relationship. No consistent overall trend, either linear or non-linear, could be identified between participants' total number of flying hours and their ratings of the usability scales.

Examining the relationship between the number of hours pilots had flown on their current aircraft type and their ratings of the usability scales, it was noted that significant relationships existed between pilots' hours and their ratings of the **Predictability, Processor Speed and Flexibility for Navigation** usability scales ($p<0.05$).

Ratings of the **Predictability** scale were influenced by the number of hours pilots had flown on their current aircraft type, with more experienced pilots making more favourable usability ratings. Significant F-values ($p.<0.05$) were established using linear and non-linear regressions when modelled using a logarithmic, a quadratic and a

cubic relationship. The linear and non-linear regressions for **Predictability** accounted for between 2.0% and 2.4% of the variance (see Appendix K).

Ratings of the **Processor Speed** scale were also influenced by the number of hours pilots' had flown on their current aircraft type. Significant F-values ($p.<0.05$) were established using linear and non-linear regressions when modelled using a logarithmic, a quadratic and a cubic relationship (see Appendix K). For this scale, however, usability ratings of the **Processor Speed** scale deteriorated, as pilots became more experienced. There was some evidence from the regressions modelled using a quadratic and a cubic relationship that usability ratings improved again after approximately 8000 hours on type, although when modelled using a linear and a logarithmic relationship usability ratings continued to decrease with increased experience. The significant relationships accounted for between 3.6% and 4.5% of the variance (see Appendix K).

Finally, the analysis suggested that a significant linear relationship ($p.<0.05$), accounting for 1.2% of the variance, existed between the number of hours pilots had flown on their current aircraft, and their ratings of the **Flexibility for Navigation** scale. Usability scale ratings became more favourable as pilots gained more experience on their current aircraft type (see Appendix K).

No significant F-values ($p.<0.05$) were established for the **Structure & Sequence, Appropriate Functionality, Feedback, Information Presentation, Error Prevention & Correction, Workload** and **Consistency** scales, using linear or non-linear regressions when modelled using a logarithmic, a quadratic or a cubic relationship. No consistent overall trend, either linear or non-linear, could be identified between usability scale ratings and participants' hours on their current aircraft type.

8.10 Multiple Regression on FMS Satisfaction Ratings

The object of this analysis was to determine which of the scales developed predicted an overall assessment of usability. The Multiple Regression reported in this section was performed between the questionnaire item "The FMS is pleasant to use; I like it" as the dependent variable, and the scales confirmed by the CFA as the independent variables.

The purpose of multiple regression was to use the scales to predict a value on a single dependent variable. The technique assessed each Independent Variable (IV) as if it had entered the regression after all the other independent variables had been entered. Each IV, could then be evaluated in terms of what it added to the prediction of the Dependent Variable (DV), over and above the predictability afforded by all the other IVs (Tabachnick & Fidell, 1983).

The model summary in Table 8-86 shows the R, R^2, the Adjusted R^2 and the Standard Error of the Estimate values.

Table 8-86 - Summary Model

Model	*R*	*R Square*	*Adjusted R Square*	*Std Error of the Estimate*
1	0.675	0.456	0.439	1.05

The ANOVA results are shown in Table 8-87 and show that the R for regression was significantly different from zero (F=26.906, 10.331, p<0.01).

Table 8-87 - ANOVA for Satisfaction

Model	*Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*
Regression	295.48	10	29.548	26.906	0.000
Residual	352.52	321	1.098		
Total	648.00	331			

The standardized and unstandardized coefficients are shown in Table 8-88, and the non-significant (p>0.05) predictors have been highlighted.

Table 8-88 - Group Statistics

	Unstandardized Coefficients		*Standardized Coefficients*		
Scale	*B*	*Std Error*	*Beta*	*t*	*Sig.*
Constant	1.068	0.304		3.515	0.001
Structure & Sequence	0.456	0.078	0.368	5.826	0.000
Appropriate Functionality	0.136	0.060	0.124	2.260	0.025
Feedback	0.000	0.061	-0.023	-0.458	0.647
Predictability	0.154	0.044	0.171	3.505	0.001
Information Presentation	0.000	0.076	0.047	0.876	0.382
Error Prevention & Correction	0.000	0.043	-0.083	-1.852	0.065
Processor Speed	0.000	0.040	0.011	0.233	0.816
Workload	0.000	0.042	0.109	2.299	0.022
Consistency	0.000	0.056	0.000	-0.005	0.996
Flexibility for Navigation	0.124	0.056	0.123	2.197	0.029

From the results shown in Table 8-88, it can be seen that ratings on the usability scales **Structure & Sequence**, **Appropriate Functionality**, **Predictability**, **Workload** and **Flexibility for Navigation** were significant predictors of more positive ratings on the questionnaire item "The FMS is pleasant to use; I like it".

9 STUDY TWO DISCUSSION

9.1 Overview of Study Two Findings

The aim of this second study was to verify the factor structure produced during Study One and to develop a series of reliable and valid scales to assess the usability of FMS CDUs.

It was hypothesised that if the underlying factor structure developed during Study One was valid, it would underlie the new independent data sets collected during Study Two for the Cruise and Descent & Approach phases of flight. It was further hypothesised that if the scales developed were truly measuring usability, and not for example, workload, ratings of FMS usability would not be significantly affected by the flight phase participants were asked to rate. It was also hypothesised that the scales developed in this study should show the same discriminant validity as the factors had demonstrated in previous study. Analysis conducted to demonstrate discriminant validity should therefore show the same general pattern of results in discriminating between FMS manufacturers and aircraft manufacturers. Finally, a Multiple Regression analysis was presented that examined which of the scales produced were significant predictors of overall satisfaction in using the FMS CDU using the questionnaire item, "The FMS is pleasant to use; I like it", as a dependent variable.

The participants in this study varied widely in terms of age, hours on type and total hours. Two-thirds of respondents were Captains, while one-third of respondents were First Officers. Respondents were type-rated on a range of Boeing and Airbus aircraft, equipped with Smiths and Honeywell FMSs permitting the impact of these variables on the pilots' ratings of FMS usability to be examined. The diversity of respondents in terms of type-rating, age and experience increased the generalizability of the study findings. It was also useful that respondents were type-rated on a range of aircraft, since analyses could be undertaken to determine if ratings on the usability scales could be used to distinguish between participants type-rated on different aircraft and FMSs. This would provide evidence of discriminant validity.

No pilots type-rated on Fokker aircraft participated in this study, but this was not surprising since fewer airlines have Fokker aircraft than have Airbus or Boeing aircraft. It did, however, mean that Fokker aircraft could not be included in the analyses conducted within this study. Only participants from the UK and Ireland participated in this study but since the previous study had demonstrated no significant difference in the usability ratings of participants from the UK and North America, this was not considered to be a problem.

9.2 Interpretation of Descriptive Analyses

In reviewing the descriptive analyses it is important to consider the results of the analyses in both absolute and relative terms (See Table 8-7). The mid-point for the seven-point questionnaire items was four. It is interesting to note that only four of the questionnaire items were notably rated towards the usable end of the scales (mean greater than five). These items were "The information in the FMS display (CDU) is easy to see and read"; "Abbreviations, acronyms & codes are used consistently throughout the FMS", "It is easy to return to the main menu from any screen of the FMS" and "The FMS is pleasant to use; I like it." The highest mean usability rating was given to the questionnaire item, "Abbreviations, acronyms & codes are used consistently throughout the FMS", which achieved a mean rating of 5.23 out of a possible seven. The four questionnaire items notably rated towards the positive end of the scale in this study were also rated towards the usable end of the scale in Study One. This indicated that ratings on the questionnaire items were stable and consistent.

Correspondingly, some questionnaire items were notably rated towards the unusable end of the scale (mean less than three) and these were: "FMS error messages explain clearly where errors are, what the errors are and why they have occurred"; "It is clear to the user what should be done to correct any error that occurs"; "The FMS help facility (either on-line or manuals) is adequate"; "It is irritating to have to input information at a rate slow enough for the FMS processor" and "The FMS can increase mental workload at inappropriate times". The lowest mean usability rating was given to the questionnaire item, "It is clear to the user what should be done to correct any error that occurs", which achieved a mean rating of only 2.72 out of a possible seven (See Table 8-7). The five questionnaire items rated towards the unusable end of the scale in Study Two were identical to those items rated towards the unusable end of the rating scales in Study One. This again indicated that the questionnaire items from which the scales were constructed were stable and generalizable.

9.3 Interpretation of Confirmatory Factor Analyses

The Confirmatory Factor Analysis (CFA) confirmed that as in Study One, ten latent variables underpinned FMS CDU usability.

The initial CFA model examined the questionnaire items loading onto each of the latent variables elicited from Study One. Through iteratively modifying the initial model in line with suggestions from the modification indices a more parsimonious solution was produced. The result of this analysis was that a number of poor indicators, particularly on the **Structure & Sequence** component were eliminated, and a final CFA model of FMS CDU usability was produced. The initial usability components elicited – **Structure & Sequence**, **Information Presentation**, **Appropriate Functionality**, **Consistency**, **Processor Speed**, **Error Prevention & Correction**, **Feedback**, **Flexibility and Control** and **Workload** remained in the final model although the "**Flexibility and Control**" component was re-named "**Flexibility for Navigation**"

based on the items which loaded onto it. Overall, 26 of the questionnaire items loaded onto this final more parsimonious solution (See Section 8.6.2).

It was also demonstrated that the structure of FMS CDU usability derived was valid when applied to the Cruise and Descent and Approach phases of flight when data from these phases were examined individually. The Goodness-of-Fit indices for both phases supported the model derived, which suggested that the model was relatively stable and that the data fitted the hypothesised factor structure.

The evidence from the CFA suggested that stable usability factors had been developed and there was some evidence of reliability and validity for them. Since the factor structure was shown to be replicable and stable, usability scales could be produced with confidence. A series of FMS CDU usability scales was thus produced through summation of the respondent ratings on each questionnaire item within a scale, and subsequently dividing the total by the number of items in a scale.

9.4 Reliability and Validity Evidence

Face validity refers to the extent to which a test looks plausible to respondents. This was achieved in this study through involving both human factors and pilot SMEs throughout the scale development process, and through extensive pre-testing and piloting of the questionnaire used for scale development. Although, face validity was not sufficient in itself, it did ensure that the scales derived were more acceptable to potential respondents (Robson, 1993).

A scale is said to have content validity when the items it contains are a randomly chosen subset of the universe of possible appropriate items. The methodology adopted for scale development should ensure content validity. The scales derived in this research were constructed according to the procedure for scale development outlined by DeVellis (1991). An iterative process of reviewing relevant academic literature, talking to SMEs, observing the experts interacting with the FMS, and pre-testing and piloting the items extensively ensured that the items elicited were an appropriate subset of all possible items. Since the scales should have content validity, they almost certainly should have face validity (Cook, 1998).

DeVellis (1991) defined reliability in the context of scale development as the *"proportion of variance attributable to the true score of the latent variable"* (p.24). The usability scales produced have successfully demonstrated internal consistency through achieving acceptable Cronbach's alpha scores as shown at Table 8-57. It has not been possible to demonstrate test-retest reliability in this study since it was not viable to ask the same participants to complete the same questionnaire twice.

The most useful measure of criterion validity, as described by Ghiselli et al (1981) and DeVellis (1991) was the rate of correctly classifying the category of the criterion (e.g. a Smiths or Honeywell FMS), based on the predictor categories (the ratings on the usability scales). While the DFA classification matrices shown in Tables 8-64, 8-70 and 8-76 demonstrated that statistically it was possible to correctly predict participants' type-

rating at a rate significantly better than chance using Press's Q ($p.<0.01$), no practical evidence of criterion validity was demonstrated. More than a third of cases were incorrectly classified, when aircraft and FMS were predicted simultaneously, and 86% of Boeing aircraft with Smiths FMSs were incorrectly classified as Boeing aircraft with Honeywell FMSs. Although 73.4% of participants were correctly classified overall when FMS classifications were examined independently, 96.5% of Smiths FMSs were incorrectly classified as Honeywell. Similarly, when aircraft prediction rates were examined, 82.1% of Airbus aircraft were incorrectly classified as Boeing aircraft. While the classification matrices in Study Two were more accurate than the classification matrices in Study One, and while statistical evidence of criterion validity was demonstrated, no practical evidence of criterion validity was demonstrated. It was not possible to predict participants' current aircraft or FMS type accurately on the basis of their usability scale ratings.

Some of the classification problems associated with the DFAs could have been due to the weighting in the analyses being biased towards certain aircraft and FMS types. The majority of participants in this research were type-rated on Boeing aircraft and Honeywell FMSs. The classification matrices were also most successful at predicting Honeywell FMSs and Boeing aircraft. It is therefore possible that the weighted linear combinations from the DFAs and the associated classification matrices could have reflected this bias. In the DFA classification matrices, Airbus and Fokker aircraft were, for example, frequently incorrectly classified as being Boeing aircraft but this may have been at least partially due to the fact that the sample was biased towards Boeing aircraft. Of the pilots who participated in Study Two, 88.5% were type-rated on Boeing aircraft, while 73% were typed rated on Honeywell FMSs. The classification problems could also have been due to the sensitivity of the Press's Q statistic (Hair et al, 1998). Although it was not possible to discriminate between participants on the basis of their command rank (first officer or captain), or to predict participants' rank accurately, the Press's Q statistic was still significant ($p<0.01$). It was therefore concluded that the Press's Q statistic was very sensitive and that no practical evidence of criterion validity was demonstrated in this study.

With the goal of examining if the scales produced demonstrated discriminant validity, a series of Discriminant Function Analyses (DFAs) were conducted using the scales confirmed by the CFA. It was hypothesised that valid scales should be able to discriminate between these different aircraft and FMS manufacturers and would show the same general pattern of results in discriminating between FMS manufacturers and aircraft manufacturers as the factors had done in the previous study. Discriminant validity was demonstrated since ratings on the factors discriminated between that which they should discriminate but did not discriminate between that which they should not discriminate. In the following sections the discriminant validity evidence is discussed and where possible is put in the context of the research findings from Study One. The purpose of comparing the discriminant validity evidence between the scales produced in Study Two and factors produced in Study One, was to determine the consistency of the discriminant validity evidence, and further to examine if discriminating capability was lost in the transformation of factors into scales.

It was first demonstrated that the scales had discriminant validity since they could successfully discriminate between the two aircraft manufacturers and the two FMS CDU manufacturers on which the pilots were currently type-rated as hypothesised. Ratings of the **Structure & Sequence, Information Presentation, Error Prevention & Correction** and **Processor Speed** scales all discriminated between aircraft and FMS manufacturers.

Boeing aircraft with Honeywell FMSs were rated more favourably than Boeing aircraft with Smiths FMSs or Airbus aircraft with Honeywell FMSs on the **Structure & Sequence** scale, while Airbus aircraft with Honeywell FMSs were rated more favourable than Boeing aircraft with Smiths FMS. Usability ratings were, however, most favourable for Airbus aircraft with Honeywell FMSs, followed by Boeing aircraft with Smiths FMSs, then Boeing aircraft with Honeywell FMSs on the **Information Presentation, Error Prevention & Correction** and **Processor Speed** scales.

Comparing the findings with those from Study One, the **Structure & Sequence, Information Presentation**, and **Processor Speed** dimensions of usability had a significant effect on usability ratings in both studies. Some differences were, however, noted in the rank ordering of manufacturers on the **Structure & Sequence** dimension. On the **Structure & Sequence** factor in Study One, Boeing aircraft with Smiths FMSs achieved the highest usability ratings followed by Boeing aircraft with Honeywell FMSs then Airbus aircraft with Honeywell FMSs. In Study Two, however, Boeing aircraft with Honeywell FMSs achieved the highest mean usability rating followed by Airbus aircraft with Honeywell FMSs then Boeing aircraft with Smiths FMSs.

The rank ordering of FMS/aircraft manufacturers was identical when the **Information Presentation** factor in Study One was compared with the **Information Presentation** scale in Study Two. Airbus aircraft with Honeywell FMSs achieved the highest usability ratings followed by Boeing aircraft with Smiths FMSs then Boeing aircraft with Honeywell FMSs. It should be noted, however, that although Fokker aircraft with Honeywell FMSs were not included in the second study, they achieved the most favourable usability rating on the **Information Presentation** factor in Study One. Similarly, on the **Processor Speed** factor in Study One and the **Processor Speed** scale in Study Two Airbus aircraft with Honeywell FMSs achieved the highest usability ratings followed by Boeing aircraft Smiths FMSs then Boeing aircraft with Honeywell FMSs. An interesting observation from this analysis was that a smaller difference existed between the group centroids of Boeing aircraft with Smiths FMSs and Boeing aircraft with Honeywell FMSs, than was observed between the group centroids of Boeing aircraft with Smiths FMSs and Airbus aircraft with Honeywell FMSs (see Table 8-61). This may have suggested that the manufacturer's flight deck philosophy was interacting with the FMS usability assessment.

A second DFA analysis demonstrated that the type of FMS CDU on which the pilot rating the usability scales was currently type-rated, significantly affected their ratings of the usability scales. It was possible to discriminate between the manufacturers of respondents' FMS on the basis of their ratings of the **Appropriate Functionality** and **Feedback** scales. Honeywell FMSs were rated more favourably than Smiths FMSs on

the **Feedback** scale, while Smiths FMSs were rated more favourably than Honeywell FMSs on the **Appropriate Functionality** scale. These findings were consistent with the findings of Study One although the **Structure & Sequence** factor also discriminated between FMSs in Study One. The scales therefore demonstrated discriminant validity and since the scales showed the same general pattern of results as the factors had in Study One evidence of reliability and stability had also be demonstrated.

As hypothesised, the DFA analyses also demonstrated that the type of aircraft on which the pilot rating the usability scales was currently type-rated significantly affected ratings of the usability scales. It was possible to discriminate between aircraft types on the basis of pilots' ratings of the **Structure & Sequence, Information Presentation, Error Prevention & Correction** and **Processor Speed** scales. It was further possible to correctly classify 87.6% of respondents into their appropriate aircraft manufacturer groups, on the basis of their ratings of these scales.

Pilots type-rated on Boeing aircraft gave their FMS CDU more positive ratings on the **Structure & Sequence** scale than Airbus type-rated pilots did. Conversely, Airbus type-rated pilots rated their FMS CDU more favourably on the **Information Presentation, Processor Speed** and **Error Prevention & Correction** scales.

Comparing the findings with those in Study One, the **Structure & Sequence, Information Presentation**, and **Processor Speed** factors in that study also discriminated between aircraft manufacturers. The rank order of usability ratings by aircraft manufacturer on these components was identical in both studies as shown on (Table 8-77). Evidence of the reliability and validity of the scales developed was again demonstrated. Further, since it was still possible to discriminate between aircraft manufacturers on the basis of ratings of the scales it was evident that little discriminating capability had been lost in the creation of scales from the usability factors.

As hypothesised, it was demonstrated that it was not possible to discriminate between pilots' ratings of the FMS usability scales on the basis of their rank. The DFA analyses demonstrated that, as in Study One, rank did not significantly impact upon ratings of the usability scales. Therefore, evidence of discriminant validity was demonstrated since ratings on the usability scales did not discriminate between pilots' rank as hypothesised.

The purpose of the final DFA was to examine if it would be possible to discriminate between flight phases on the basis of pilots' usability scale scores. As hypothesised, it was demonstrated that it was not possible to discriminate between flight phases based on pilots' ratings of the FMS usability scales. A DFA analysis demonstrated that it was not possible to discriminate between the Cruise and Descent & Approach flight phases based on pilots' ratings of the usability scales. For the scales to be valid they should discriminate between those factors that they should discriminate between, but should not discriminate between factors that are unrelated to the scales' purpose. Usability scales, therefore, should distinguish between FMSs and the aircraft in which they reside and not phase of flight. If the ratings discriminated between flight phases they would be discriminating between factors which were unrelated to the scales purpose. The fact that the usability scale structure was valid and applicable to both the Cruise and Descent

& Approach flight phases was considered a positive finding in that it demonstrated the stability and generalizability of the scales derived. The t-test results shown at APPENDIX J also demonstrated that for forty of the forty-three individual questionnaire items, pilots' ratings were not significantly different based on the phase of flight rated. It was, however, noted that only two of the forty-three questionnaire items received more positive usability ratings for the Descent & Approach phase of flight than the Cruise phase of flight. So while no significant differences existed in ratings of the questionnaire items across the two flight phases examined, a consistent pattern of ratings did emerge. It was demonstrated that ratings of the usability scales were not a function of the flight of phase that was being rated by the participants, but rather were solely a rating of usability based on the equipment being rated and the operational environment in which that equipment resided. The scales were not, for example, measuring cognitive workload alone which is known to be higher in the Cruise phase of flight than in the Descent & Approach phase of flight (Sarter, Woods, Billings, 1997). In terms of certification process, however, questions can be raised as to the usefulness of a series of scales that cannot discriminate between flight phases. The usability problems associated with Flight Management Systems are accentuated in time critical situations, and it may be that a questionnaire administered in a non-operational environment may benefit from more detailed scenario information than was provided in this research.

A comparative summary of the discriminant validity evidence from Studies One and Two is presented at Table 9-1. The symbol '✓' has been used to identify the factors and scales on which ratings discriminated between aircraft and FMS manufacturers. Neither rank nor culture has been included in the summary since neither significantly influenced usability ratings. The pattern of ratings on the **Structure & Sequence**, **Feedback**, **Information Presentation** and **Processor speed** scales, for example, were almost identical to the factors derived in Study One in the way in which they discriminated between aircraft and FMS manufacturers. Since the analyses showed the same general pattern of results in discriminating between FMS and aircraft manufacturers in both Studies One and Two, it was demonstrated that little discriminant capability had been lost in the conversion of factors into scales. Evidence of the generalizability and reliability of the scales produced was therefore derived.

Table 9-1 - Comparison of Discriminant Validity Evidence from Studies One and Two

Study One	*Manufacturers*			*Study Two*	*Manufacturers*		
FACTOR	*Aircraft + FMS Manufacturer*	*FMS Manufacturer*	*Aircraft Manufacturer*	*SCALE*	*Aircraft + FMS Manufacturer*	*FMS Manufacturer*	*Aircraft Manufacturer*
Structure & Sequence	✓	✓	✓	Structure & Sequence	✓	✓	---
Appropriate Functionality	✓	✓	✓	Appropriate Functionality	---	---	✓
Feedback	---	---	✓	Feedback	---	---	✓
Predictability	✓	✓	---	Predictability	---	---	---
Information Presentation	✓	✓	---	Information Presentation	✓	✓	---
Consistency	---	---	---	Consistency	---	---	---
Flexibility & Control	✓	✓	---	Flexibility for Navigation	---	---	---
Error Prevention & Correction	---	---	---	Error Prevention & Correction	✓	✓	---
Processor Speed	✓	✓	---	Processor Speed	✓	✓	---
Workload	---	---	---	Workload	---	---	---

Construct validity is the extent to which an assessment is measuring something meaningful. Guion (1977) argued that since test creators always have a theory or construct that they are trying to assess, all test validation is necessarily construct validation. Murphy and Davidshofer (1991) defined the goal of construct validation as being "*to determine whether test scores provide a good measure of a specific construct*" (p.116). In this study, therefore, it referred to the extent to which the scales measured the theoretical construct of usability. Construct validity is concerned with how theoretically meaningful the scales are. Ideally, evidence of construct validity would be achieved by comparing the usability scales developed in this study with existing measures of FMS CDU usability. Since no objective measure of FMS CDU usability exists, construct validity is difficult to establish.

If the underlying usability construct exists, it should exist regardless of participants' rank or culture. On the other hand in order to be useful it should be able to discriminate between FMS and aircraft types (since FMS and aircraft type are intrinsically linked). Through comparing the scales developed with comments from the interview transcripts and the literature, evidence towards construct validity was derived. The literature, for example, indicated that workload was a significant issue in interaction with the FMS CDU (Goteman, 1999). Ratings of the usability scales in this study supported the literature, since mean ratings of workload indicated that they were amongst the poorest of any of the components derived. The usability scales were not, however, measuring workload since analyses of flight phase demonstrated that no significant difference existed in ratings of the scales in high and low workload phases of flight. Evidence toward construct validity was therefore provided. Further evidence of construct validity was derived from the fact that the scales did not discriminate between pilots' usability ratings on the basis of their rank. Pilot rank is known to influence pilot attitudes to automation (see Section 2.2.4). Evidence in this study therefore indicated that the scales elicited were assessing usability and not attitudes to FMSs. Further evidence toward construct validity was therefore demonstrated.

9.5 Interpretation of Analysis on Users' Experience

Analyses using linear and curvilinear regression demonstrated that no overall trend, either linear or non-linear, could be identified between usability scale ratings and participants' hours on their current aircraft type or their total flying hours (see Appendix K). While there was some evidence of a small amount of variance being accounted for, and while some significant trends were noted, the small amount of variance accounted for meant that the trends were practically non-significant.

Significant relationships were established between pilots' total hours and their ratings of the **Predictability** and **Error Prevention & Correction** scales. Similarly, it was noted that significant relationships existed between pilots' hours on their current aircraft type and their ratings of the **Predictability, Processor Speed** and **Flexibility for Navigation** usability scales ($p<0.05$). However, no overall trend was established between pilots' experience (in terms of hours on type or total hours) and their usability ratings.

This finding could be interpreted in one of two ways. It could be that the scales were sensitive to characteristics of the equipment (the aircraft & FMS) rather than characteristics of the end users such as their experience on type. Alternatively, it could be that the usability scales were only accessing the most obvious, high-level usability issues and were not sensitive to those usability concerns that only emerge through having greater experience on an aircraft.

It is clear that some relationships between human error and system design will only emerge in experienced pilots. The finding that no point could be identified beyond which usability scale ratings stabilised raised some concern in terms of how experienced pilots need to be in order to participate in the certification process. It could be considered a weakness of the usability scales approach that no deductions can be made on how experienced pilots should be in order to comment on FMS usability. Amalberti & Wibaux (2000) have discussed the question of how experienced pilots should be in order to be involved in the certification process and outlined how most pilots involved in the certification process have less than two hundred hours of experience on that aircraft. Experimental data (e.g. Sarter & Woods, 1992), however, suggest that pilot expertise for flying a glass cockpit aircraft can require more flying hours. Since behaviour changes and the nature of error change with experience, pilots involved in the certification process should represent the future range of company pilots from novice First Officers through to experienced Captains.

The usability issues raised in this study can be viewed as reflecting both the "Gulf of Execution" and "Gulf of Evaluation" described by Norman (1986). Poor usability ratings on the **Structure & Sequence**, **Appropriate Functionality**, **Information Presentation** scales reflected Norman's Gulf of Execution, whilst poor usability ratings on the **Feedback**, **Consistency**, **Predictability**, **Error Prevention & Correction**, and **Processor Speed** scales reflected a Gulf of Evaluation. Examining participants' ratings of the usability scales in Study Two across all aircraft and FMS types, it was a Gulf of Evaluation rather than a Gulf of Execution that was reflected in the ratings. The **Feedback**, **Processor Speed**, and **Error Prevention & Correction** scales all achieved mean ratings toward the poor usability end of the scale.

9.6 Limitations of the Usability Scales Developed

Ratings of the usability scales in this research produced a single value on each of the ten scales, calculated through summation of the scores on the individual questionnaire items and dividing the total by the number of questionnaire items loading onto that scale. In order to be useful in the certification process a "cut-score" in terms of acceptability would need to be identified for each scale. It is a recognised limitation of this research that a greater degree of validation of the usability scales with larger sample sizes would need to be conducted in order to identify what those cut-scores should be. It is also a recognised limitation of this research that no practical evidence of criterion validity could be demonstrated. While statistical evidence of criterion validity was presented, it was not possible to predict participants' current aircraft or FMS type accurately from their ratings of the usability scales.

A further limitation of this research was that it was not evident from the scale ratings why particular characteristics achieved poor ratings. It could have been due to participant's expectations of better usability based on prior exposure to other equipment (external experience), or it could have been due to respondents' personal experience of having particular problems with the specific FMS they were asked to rate.

Data from these scales should be enhanced and complemented with performance, mental workload and effectiveness data frequently elicited during system development and certification. Even if an FMS achieves an acceptable score on each of the usability scales, it is not going to be free from error in operation. A human end user will interact with the FMS once it is in-service, and it is unrealistic to expect totally error-free performance from a human. Making errors is part of the human condition and FMS designers can only be expected to minimise the opportunities for error. The use of psychometrically robust usability scales in a development or certification programme will not negate the necessity to conduct usability testing as a product is developed and integrated with other systems.

As discussed in Section 1.12.5, there are some weaknesses associated with a scale-based approach to usability in the certification process. The scales developed are less flexible than, for example, experimental, interview or observational research approaches in terms of being able to elicit information concerning specific operator experiences and the environmental context of use. Since FMS-related accidents and incidents occur within a specific context of operation, this was an important element that had to be considered in determining the most appropriate approach to take for this study. It is a recognised limitation of the questionnaire approach that specific flight contexts are not easily accessed. Context was only assessed in this study in terms of flight deck types and the phase of flight that participants were asked to rate the usability of their FMS for.

Similarly, although accidents and incidents occur within an operational environment the methodology developed in this research is unlikely to be the optimal method of accessing the experiences of operators and the environmental context of use. Usability scales are only useful to the extent that they could assist FMS manufacturers to focus their attention on aspects of the FMS that achieve particularly poor usability ratings. The incorporation of more detailed scenarios would be useful in future research since it would permit the examination of specific problems, at specific times or in specific situations such as using the FMS in the hold, on "go-arounds" or in making last minute runway changes.

As outlined in Section 1.2, the Kegworth accident was found to have been caused by the operating crew shutting down the number two engine after a fan blade had fractured in the number one engine. The subsequent accident investigation reported that that one of the factors which prevented the pilots noticing the incorrect response they had made, was the difficulties the First Officer encountered trying to reprogram the FMS for landing at East Midlands Airport. It was inferred from the cockpit voice recorder that the pilot possibly attempted to enter East Midlands Airport as the next en route point without first selecting the route page and entering it as an arrival airfield. The problems encountered in trying to reprogram the FMS are reflected in some of the scales developed within this study. For example, improved **Structure & Sequence** cues may

have assisted the pilot in programming the FMS. Similarly, improved **Feedback** may have assisted the pilots in the identification of the errors they had made in shutting down the incorrect engine.

Evidence from the Cali accident report (Aeronautica Civil, 1996), suggested that a number of the usability scales developed in Studies One and Two may also have been applicable to this accident in which 159 passengers and crew were killed. Inadequate **Feedback** from the FMS, for example, meant that the pilots were unaware of their current location relative to their destination location in the minutes preceding the accident. **Information Presentation** was a contributory factor since the waypoints were not in a logical order, which increased the likelihood of the pilots selecting the incorrect waypoint as occurred on that day. **Error Prevention & Correction** was also an important factor since the FMS either did not have the facility to perform, or did not correctly perform, validity checks which could have informed the pilots that the waypoint entered was taking the aircraft away from the required destination.

The scales produced in this thesis could assist in identifying system weaknesses and therefore could to some extent be useful in the prevention of accidents and incidents. Each incident and accident, however, occurs in a unique set of circumstances, and both of the accidents described in this section occurred in a complex set of circumstances. It is therefore and that makes it difficult to comment on how useful the scales would be in reality.

Whilst psychometrically robust scales could be useful in identifying usability problems, they would not provide FMS manufacturers with design solutions. Qualitative methods such as interviewing typical end users, conducting focus group sessions with end users and conducting iterative prototyping with alternative design configurations are more useful mechanisms for assisting with improved design solutions.

9.7 FMS CDU Satisfaction

The final analysis conducted during this study was a Multiple Regression analysis which examined if ratings on the scales developed could predict pilots' ratings of the overall usability item "The FMS is pleasant to use; I like it". The results indicated that pilots' ratings of the **Structure & Sequence**, **Appropriate Functionality**, **Predictability**, **Workload** and **Flexibility for Navigation** scales were all significant predictors of ratings on the questionnaire item "The FMS is pleasant to use I like it".

The satisfaction element of usability, which has been discussed in many papers including (ISO 9241 Part 11 and Thomas, 1996), was defined by Preece et al (1994) as the positive attitude engendered in users by using a system, was not elicited as a usability scale in this study. It was, however, noted from the interview transcripts that although many usability problems exist in interacting with FMS CDUs, pilot attitudes towards FMSs were favourable. There are a number of possible reasons for something that might seem contradictory. Many of the functions performed by the FMS are extremely useful and accurate and take a significant workload off the pilots particularly during the cruise phase of flight. Life without the FMS would be more difficult for

those operating modern two-crew flight decks. A further reason may be related to the fact that use of the FMS is an essential part of the pilot's job. As discussed by Noyes, Starr and Frankish (1996) line pilots can exhibit a loyalty towards their current aircraft due to the inherent risk associated with flying their (or any) aircraft. They therefore may be reluctant to question features of the flight deck in which they operate, and may justify design features regardless of their suitability for the task being undertaken.

9.8 Conclusions

The overall aim of this study was to confirm the factor structure derived from Study One and to develop a series of reliable and valid FMS CDU usability scales. It was demonstrated that the usability factors derived were consistent and stable since the underlying factor structure was valid for both of the independent data sets derived from ratings of the Cruise and Descent and Approach flight phases. Usability scales were then produced and subjected to analyses to demonstrate their reliability and validity.

The usability scales produced were internally consistent as shown from the Cronbach's Alpha analyses, which demonstrated that the scales had an acceptable level of internal consistency to be used as scales. Evidence of face validity, content validity, and construct validity was also demonstrated.

Evidence of discriminant validity was demonstrated since the usability scale ratings differed significantly by FMS and aircraft Type, but did not differ significantly by personal factors such as participant rank. The rank ordering of FMS and aircraft manufacturers based on the usability scales ratings were largely consistent with those identified by the factors produced in Study One. This demonstrated that the scales could discriminate between manufacturers in a consistent pattern, but also showed that little discriminating capability was lost in the conversion of factors into the more simplified scales used in this second study. Although it was possible to predict participants' type-ratings at a rate significantly better than chance ($p.<0.01$), no practical evidence of criterion validity was demonstrated in this study. The analyses indicated that Press's Q statistic was very sensitive, and was shown to be significant even when a large proportion of cases were incorrectly classified.

Analyses demonstrated that ratings of usability were not significantly different based on the flight phase the scales were applied to. This demonstrated that FMS CDU usability was a product of the FMS and its operating platform (the aircraft), and was not significantly affected by the workload pilots were under while using the equipment.

Finally, it was shown that that ratings on the usability scales **Structure & Sequence**, **Appropriate Functionality**, **Predictability**, **Workload** and **Flexibility for Navigation** were significant predictors of ratings on the overall satisfaction in using the FMS as assessed by ratings on the questionnaire item "The FMS is pleasant to use; I like it".

10 STUDY THREE INTRODUCTION

10.1 Purpose

The purpose of this third study was to collate the qualitative comments elicited during Studies One and Two, with the goals of (i) identifying problems associated with current FMSs; (ii) making recommendations for improving FMSs; (iii) providing evidence of validity for the usability scales developed during Studies One and Two and (iv) aiding interpretation of those scales.

10.2 Qualitative Research Approach

Qualitative research usually refers to the collection and analysis of descriptive data – written or spoken words and observable behaviour rather than numbers (Cassell & Symon, 1994). The process may be viewed as a method of imposing structure on chaos. Some of the sources from which qualitative data can be elicited are qualitative interviews, questionnaires and observation studies.

Cassell & Symon (1994) argued that qualitative research had a number of defining characteristics which included "*a focus on interpretation rather than quantification; an emphasis on subjectivity rather than objectivity; flexibility in the process of conducting research; an orientation towards process rather than outcome; a concern with context – regarding behaviour and situation as inextricably linked in forming experience; and finally, an explicit recognition of the impact of the research process on the research situation.*" (p.7).

The analysis of qualitative data is very dependent on the quality of the researcher conducting both the data collection and analysis. Qualitative analysis usually occurs through a process of data reduction, data display, conclusion drawing and verification. The main advantage to using a qualitative research approach is that an opportunity exists to access "*well grounded, rich descriptions and explanations of process in identifiable local contexts*" (Miles & Huberman, 1994, p.1). The main disadvantage to the use of qualitative research methods is the traditional doubt surrounding the reliability and validity of the findings from qualitative research (Cassell & Symon, 1994). A second disadvantage is that the analysis of qualitative data can be very time consuming and thus expensive (Robson, 1993).

In the qualitative research to be described in the subsequent section, a large volume of comments were elicited from the questionnaires returned from respondents in Studies One and Two. In order to fulfil this study's goals as outlined in Section 10.1, it was necessary to impose structure on the data, and this procedure is described in Section 11.2. The process of imposing structure on the data is outlined according to the procedures outlined by many authors (e.g. Patton, 1990), and includes the techniques of data reduction, data display, conclusion drawing and verification.

Data reduction refers to the qualitative process of "*selecting, focusing, simplifying, abstracting and transforming*" data from transcriptions or field notes (Miles & Huberman, 1994, p.10). Data reduction is part of qualitative analysis and does not necessarily imply quantification. Prior to qualitative data collection, data reduction occurs as the researcher decides which conceptual framework and which data collection approach to choose. During and after data collection data reduction occurs through summarising, coding, and deriving themes and categories (Patton, 1987).

The data display process permits data to be organised and presented in a format that permits conclusions to be drawn. Qualitative analysis typically requires a researcher to examine a large, unwieldy set of text. Since humans are not very powerful at processing large quantities of information, displaying the data in a more manageable format permits trends in the data to be examined without oversimplification. Data display is again part of, rather than a precursor to, qualitative analyses.

The final phase of qualitative analysis is conclusion drawing and verification, and refers to the process of drawing meaning from the data. Throughout the process of qualitative analysis, researchers note trends and explanations but final conclusions may not appear until data collection is completed and coding has occurred. The goal of verification is to establish the validity of the meaning derived from the data set. The verification phase may require a review of field notes or a consensus being established between researchers (Miles & Huberman, 1994).

10.2.1 Methodological Considerations

The main methodological consideration that must be made in conducting qualitative research is the dependence on the quality of the researcher(s) conducting the analysis. The biggest threats, therefore, to the reliability and validity of the analyses conducted are the reliability of the raters conducting the analysis, and their associated consistency of interpretation.

Patton (1990) argued that the first problem that must be overcome in the analysis of qualitative data is "*convergence*" (p.402), which refers to the problem of determining which elements of data fit together to create a classification system. The process of converging data into systematic categories is performed by looking for recurring regularities in the data. The regularities can then be converted into categories. The categories should then be judged by two criteria, which he referred to as "*internal*" homogeneity and "*external*" heterogeneity (p.403). Internal homogeneity concerns how well the data in a category fit together meaningfully. External heterogeneity refers to the extent that the categories are distinct from each other. Guba (1978) argued that "*The existence of a large number of unassignable or overlapping data items is good evidence of some basic fault in the category system*" (p.53). Two researchers performed the qualitative analysis described in this study to ensure that the data fitted together meaningfully and the categories were distinct from each other.

In order to elicit trends in qualitative data it is necessary to code the data into a framework. Harris (1995), however, argued that such *post hoc* coding of qualitative

information does raise methodological issues. Harris argued that for categorisation to be reliable and to reduce subjective bias, there should be agreement between at least two independent assessors about the factors involved. As well as requiring inter-rater reliability, intra-rater reliability in which raters are consistent in their allocation of data to categories, is also essential to the coding of qualitative data.

Qualitative research typically involves the interrogation of in-depth information from a relatively small sample due to the costs incurred and time taken to analyse the data elicited. It therefore provides a narrow but deep look at the subject of interest. The study to be described here is unusual in that it involved a larger sample of participants than is usually associated with qualitative research. The use of a "broad but shallow" approach means that the study findings have an increased generalizability.

One potential approach to the analysis of the data within this study was to use the scales derived from Studies One and Two and to structure the qualitative data within that framework. An alternative methodology was to adopt a grounded theory approach (Glaser & Strauss, 1968) in which the researcher remained expectation free in the initial stages of research to allow the key issues to emerge via systematic grounded analyses. This latter approach was adopted for this study as a further check on the construct validity of the usability scales produced. The analysis undertaken examined all the qualitative data elicited from the questionnaires completed by participants in Studies One and Two, without reference to the scales developed within those studies.

11 STUDY THREE METHOD

11.1 Study Aim

In order to satisfy the goals of the study as set out in Section 10.1, a content analysis was undertaken. The procedure adopted is described in this section. Content analysis refers to the process of identifying, coding and categorising data and was defined by Krippendorff (1980) as "*a research technique for making replicable and valid inferences from data to their context*" (p.21). Content analysis begins with the identification of a research question. In this study the researcher set out to identify the problems pilots encountered using current FMSs, and the improvements sought by pilots to future FMSs. The remaining steps in content analysis involve identifying a sampling strategy, defining the recordings units, constructing categories for analysis, testing the coding on samples of text and assessing readability, and finally carrying out the analysis Krippendorff (1980).

A grounded theory approach was adopted in that the researcher did not seek to fit the data to a pre-developed conceptual framework such as that elicited in the scale development process described in the previous two studies. As discussed in Section 10.2.1, this latter approach was adopted for this study as a further check on the construct validity of the usability scales produced. The researcher on the whole abandoned preconceptions through the process of analysis and aimed to build up an explanatory framework through conceptualisation of the data as described by Chell (1998). The stages of content analysis conducted within this study are described in the following sections.

11.2 Procedure

In light of the first two distinctive study aims, the initial stage of this qualitative study involved dividing the data elicited from the questionnaires into two general categories based on the content of each comment. The two categories were:

(i) Comments reported by participants that concerned problems encountered with the FMS on which they were currently type-rated.

(ii) Comments that concerned improvements participants would wish to see to FMSs on future flight decks.

In total 496 discrete comments were derived from the 224 participants who chose to make qualitative comments. After reading each comment a number of times the researcher decided which category the comment was most pertinent to and classified it as such. The allocation of a comment to either the category of "suggested improvement" or "problem encountered" was largely based on how the comment was phrased.

Based on this classification a total of 289 of the comments were categorised as referring to issues participants raised concerning their current FMS and each comment was numbered. These comments have all been included in the "Problems with Current FMS" category although some of the comments could better be regarded as observations concerning the participants' current FMS and are thus not necessarily negative.

A total of 207 comments were categorised as referring to improvements the participants would like to see to future FMSs, and each comment was numbered. The raw data illustrating each comment has been included at Appendix M. Having collated all the data referring to either "Problems with Current FMS" or "Suggested Improvements for Future FMSs", each discrete comment was placed on a piece of card. The "Problems" and "Improvements" categories could be considered two sides of the same coin, since problems encountered with current FMSs could potentially lead to suggested improvements for future FMSs. All comments referring to "problems encountered" with current FMSs were reviewed one by one, until the researcher felt sufficiently familiar with the data to be able to group the comments into categories. Initially, nineteen categories were used in the analysis as shown in Table 11-1. To ensure intra-rater reliability, the researcher allocated each comment to a category repeatedly until the allocation of comments to categories could be replicated with 95% consistency when the procedure was conducted on sessions a week apart. The process was logged by allocating each comment to the desired category. Once the researcher was satisfied that the process was complete, the comment numbers were recorded in tables against the category to which they had been allocated. The process was conducted on three occasions before the desired 95% level of consistency was achieved on sessions a week apart.

During this procedure the comments that had initially been allocated to the categories labelled "Two-step entries", "Language", "Flexibility & Control" and "Too heads down", were re-distributed amongst other categories, as it became evident that the comments within these categories were more appropriately placed in other categories. The number of categories was thus reduced to 15. Table 11-1 shows the categories elicited concerning problems encountered with current FMSs. The italicised categories were those excluded from the final analysis.

Table 11-1 - Qualitative Categories of Problems with current FMSs

Coding Number	*Category*	*Issues*
1	Small Database / Memory size	
2	Slow Processor speed	
3	Inadequate feedback, error prevention & correction	Including error messages and corrective actions.
4	Poor Display quality & Information Presentation	Including Layout, size, colour, angle, location, static/dust/scratchpad, interface, visual clarity.
5	Poor Consistency & Compatibility	Including standardization of inputs, display layouts.
6	Observations	
7	Functional Limitations	Including system limitations.
8	*Flexibility and Control*	
9	High / Low Workload	
10	Inadequate Control Devices	Including data entry, control devices, QWERTY keyboard, undo button.
11	*Two-step entries*	*Including requirement for "execute" button.*
12	*Language*	*English instead of International Civil Aviation Organization (ICAO) codes.*
13	Poor CDU Architecture & Logic	Including content.
14	Accuracy of parameters	
15	Insufficient User Guidance & Support	Including manuals.
16	Lack of Intuitiveness	
17	Inadequate Training	
18	*Too heads down*	
19	Crew Co-ordination	

The procedure outlined for classifying problems encountered with current FMSs was replicated to classify all comments that made reference to the improvements respondents reported wishing to see in future FMSs.

The comments referring to "improvements" were reviewed one by one, until the researcher felt sufficiently familiar with the data to be able to group the comments into categories. Seventeen categories of improvements were derived, as shown in Table 11-2. The raw data illustrating each comment has been included at Appendix M. During this procedure the comments that had initially been allocated to the "Two-step entries" category was re-distributed amongst other categories, as it became evident that the comments within these categories were more appropriately placed in other categories. The number of improvement categories that remained was thus sixteen.

Table 11-2 - Qualitative Categories for Suggested Improvements to FMSs

Coding Number	*Category*	*Issues*
1	Bigger database or memory	
2	Faster Processor	
3	Clear feedback, error prevention & correction	Including error messages and corrective actions.
4	Display quality & Information Presentation	Including layout, size, colour, angle, location, static/dust/scratchpad, interface, visual clarity.
5	Consistency & Compatibility	Standardization in inputs, display layouts.
6	Observations	
7	Additional Functionality	
8	Improved "What if?" capability	
9	Improved barriers to erroneous operation	
10	Control Devices	Including data entry, control devices, QWERTY keyboard, undo button.
11	*Two-step entries*	*Including requirement for "execute" button.*
12	Language	English instead of ICAO codes.
13	CDU Architecture & Logic	Including content.
14	Accuracy of parameters	
15	User Guidance & Support	Including manuals.
16	Intuitiveness	
17	Training	

In order to validate the structure imposed on the data, and to determine an inter-rater reliability coefficient for the ratings made by the researcher, as second human factors expert also conducted the process of allocating comments to categories. The second human factors expert had more than fifteen years experience of conducting research in the aviation automation domain.

The second researcher was asked to repeat the process conducted by the first researcher until an intra-rater reliability rate of 95% could be established. The second researcher also took three attempts to achieve the 95% consistency level. The process was logged and recorded as per the first researcher's methodology. Once the second researcher had completed the process for both sets of comments, an analysis was performed to establish the percentage of inter-rater agreement. The comments included in the subsequent results section include *only* those comments on which *both* of the raters agreed on the comment's classification.

11.3 Participants

It is important to consider the number of participants who chose to make comments when interpreting the data contained within the results section for this study. A lack of due consideration of this issue could lead the reader to grossly misinterpret the data. At first glance, a large number of problems associated with one particular fleet could be misinterpreted as being an issue of usability for that aircraft or FMS. The true underlying reason might only reflect the fact that there were more participants who either could have or chose to make comments concerning that fleet or FMS type.

The percentages as shown on Table 11-3 were calculated as a proportion of the total number of pilots for each fleet who chose to take part in the study. The percentages thus reflect the proportion of, for example, B-757 pilots who chose to make comments out of those B-757 pilots who participated and thus could have made comments.

Table 11-3 - Sample Description

Category	*Total N who could have made comments (% of sample)*	*N who made Problem Comments (% of N in Column 2)*	*N who made improvement Comments*	*N who made both Problem and Improvement Comments*	*Total N who made Comments*
B-757/767	258 (42.4%)	53 (20.5%)	53 (20.5%)	13 (5.0%)	93 (36.0%)
B-747	41 (6.7%)	7 (17.0%)	9 (22.0%)	3 (7.3%)	13 (31.7%)
B-737	179 (29.4%)	42 (23.5%)	30 (16.8%)	10 (5.6%)	62 (34.6%)
B-777	20 (3.3%)	4 (20.0%)	6 (30.0%)	1 (5.0%)	9 (45.0%)
A320/A321/A330	88 (14.5%)	29 (33.0%)	15 (17.0%)	7 (8.0%)	37 (42.0%)
F-100	22 (3.6%)	9 (40.9%)	1 (4.5%)	0 (0.0%)	10 (45.5%)
Total	608	144 (23.7%)	114 (18.8%)	34 (5.6%)	224 (36.8%)

Examining the participants across fleets a mean of 39.1% (sd=5.8%) of respondents per fleet chose to make comments. A mean of 25.8% (sd=9.2%) of respondents across individual fleets chose to make "problem" comments relating to their current FMS CDU. While a mean of 18.5% (sd=8.4%) of respondents across individual fleets chose to make "improvement" comments.

Table 11-4 shows the total number of participants who chose to make comments, and the number of comments made for both the problem and improvement categories. The percentages shown indicate the proportion of respondents that made comments by each fleet.

Table 11-4 - Number of Comments by Participants' Current Type-rating

Fleet	*FMS Manufacturer*	*Total number of participants who made comments (% of sample)*	*Number of problem comments made*	*Number of improvements comments made*
B-757/767	Honeywell	93 (41.5%)	114 (39.4%)	106 (51.3%)
B-747	Honeywell	13 (5.8%)	29 (10.0%)	15 (7.2%)
B-737	Smiths	62 (27.7%)	71 (24.6%)	52 (25.1%)
B-777	Honeywell	9 (4.0%)	8 (2.8%)	9 (4.3%)
A320/A321/A330	Honeywell	37 (16.5%)	52 (18.0%)	24 (11.6%)
F-100	Honeywell	10 (4.5%)	15 (5.2%)	1 (0.5%)
Total		224 (100.0%)	289 (100.0%)	207 (100.0%)

12 STUDY THREE RESULTS

12.1 Overview

This section presents the results of the content analyses conducted on the qualitative data elicited during Studies One and Two. The categories of comments derived from both problems associated with current FMSs and improvements sought to future FMSs are presented. Table 12-1 shows the number of comments concerned with problems with current FMSs as allocated to each category by the first and second raters. Table 12-2 shows the number of comments concerned with improvements to future FMSs as allocated to each category by the first and second raters. The number of individual comments which both assessors gave the same ratings to, and the associated percentage agreement between the two assessors are also shown.

The formula used to calculate the percentage agreement between assessors was:

$$\frac{A \cap B}{A+B-(A \cap B)}$$

Where: A = Number of Comments from Assessor A

B = Number of Comments from Assessor B

$A \cap B$ = Comments that both Assessors agreed on

Since the assessors agreed on 237 of the 289 ratings, the overall agreement between the assessors was 82.0%. Some of the percentages indicating agreement between assessors within individual categories appear quite low but this is due to the relatively small number of comments allocated to some categories. The category Crew Co-ordination was eliminated at this stage, since the assessors did not agree on any of the comments allocated to this category. Comments allocated to the "Observations" categories referred to specific properties of respondents' aircraft or FMS and were not directly related to usability. Since such comments were not contributing to the overall objectives of this thesis, they were excluded from the analysis described in this section.

Table 12-1 - Qualitative Categories of Problems with current FMSs

Coding Number	*Category*	*No of Comments from Rater One*	*No of Comments from Rater Two*	*Number of Comments both Assessors Agreed on*	*Percentage Agreement between Assessors*
1	Small Database / Memory size	8	8	8	100.0%
2	Slow Processor speed	43	42	42	97.7%
3	Inadequate Feedback, Error Prevention & Correction	19	16	14	66.7%
4	Poor Display Quality & Information Presentation	14	16	13	76.4%
5	Poor Consistency & Compatibility	8	5	5	62.5%
6	*Observations*	*67*	*97*	*66*	*67.0%*
7	Functional Limitations	32	17	17	53.1%
8	High/Low Workload	9	2	2	22.2%
9	Inadequate Control Devices	17	18	16	84.2%
10	Poor CDU Architecture & Logic	29	34	25	65.8%
11	Accuracy of parameters	19	15	15	78.9%
12	Insufficient User Guidance & Support	7	7	7	100.0%
13	Lack of Intuitiveness	3	3	2	50.0%
14	Inadequate Training	4	7	4	57.1%
15	*Crew Co-ordination*	*4*	*0*	*0*	*0.0%*
	Total	289	289	237	

Table 12-2 - Qualitative Categories of Suggested Improvements to FMSs

Coding Number	*Category*	*No of Comments from Rater 1*	*No of Comments from Rater 2*	*Number of Comments both Assessors Agreed on*	*Percentage Agreement between Assessors*
1	Bigger database or memory	7	7	7	100.0%
2	Faster Processor	20	20	20	100.0%
3	Improved feedback, error prevention & correction	14	13	13	92.9%
4	Display quality & Information presentation	44	44	43	95.6%
5	Consistency & Compatibility	7	8	7	87.5%
6	*Observations*	*5*	*5*	*5*	*100.0%*
7	Additional Functionality	49	49	48	96.0%
8	Improved "What if?" capability	5	5	5	100.0%
9	Improved barriers to erroneous operation	3	3	3	100.0%
10	Control Devices	29	29	28	93.3%
11	Language	7	6	6	85.7%
12	CDU Architecture & Logic	4	4	4	100.0%
13	Accuracy of parameters	7	7	7	100.0%
14	User Guidance & Support	2	3	2	66.7%
15	Intuitiveness	2	2	2	100.0%
16	Training	2	2	2	100.0%
	Total across categories	207	207	201	

Since the assessors agreed on 202 of the 207 suggested improvement ratings, the overall agreement between assessors was 97.6%.

12.2 Categories of Reported FMS Problems & Suggested Improvements

This purpose of this section is to present and interpret the results of the content analysis conducted. Thirteen categories of problems associated with current FMSs and fifteen categories of suggested improvements were derived. The categories derived during the study are put in the context of existing research. Where possible the categories of FMS problems and suggested improvements raised in this study are related to the scales developed from Studies One and Two in the process of conclusion drawing and verification described by Miles & Huberman (1994). The complete set of comments is presented at Appendix M and extracts will be used in this section for illustrative purposes. The titles used for the categories are relatively generic and a number of issues were identified within some categories. The comments have been included in the respondents' own words. Where the respondents' current type-rating is not explicit from the comment, it has been included in parentheses. The categories of suggested improvements to FMSs are strongly related to those problems associated with current FMSs. This is to be expected since problems in using the FMS have led respondents to suggest related improvements.

Based on the 170 "problem" and 197 "improvement" comments that both the assessors agreed upon, the individual comments were divided by fleet. The purpose of this was to examine if any of the differences in usability ratings by aircraft and FMS as reported in Studies one and two were reflected in the comments made by the assessors in this study. It is important to reiterate that only those comments on which both assessors agreed were included in this stage of the qualitative analysis. The distribution of problem comments across fleet types has been included at Table 12-3, while the distribution of improvement comments across fleets has been included at Table 12-4. The number of comments per category is displayed by fleet type, while the percentages of comments within a particular category and fleet are displayed in parentheses.

The "total" shown the end of each column in Table 12-3 and Table 12-4 represents the total number of comments made by participants type-rated on a particular fleet. The "total" shown at the end of each row represents the total number of comments made within a particular category of FMS "Problems" or "Improvements sought". Comments related to problems with current FMSs have been listed in order by the total number of comments within a category (Table 12-3). To aid the reader, the improvements categories have been listed in the same order as the problems categories where possible. General observation comments have been included at the end of each table.

Caution must be exercised in any interpretation of fleet or FMS effects in this phase of the research since (1) small numbers of comments were allocated to some categories; and (2) there were large variations in both number of comments made by participants from each fleet and in the number of participants from each fleet.

Table 12-3 - Problems with Current FMSs

Category	*B-757/767 Honeywell FMS* *No of Comments*	*B-737 Smiths FMS* *No of Comments*	*B-777 Honeywell FMS* *No of Comments*	*A320/A321/A330 Honeywell FMS* *No of Comments*	*F-100 Honeywell FMS* *No of Comments*	*B-747 Honeywell FMS* *No of Comments*	*Total number of comments*
Slow Processor speed	20 (29.4%)	12 (28.6%)	4 (50.4%)	3 (10.0%)	0 (0.0%)	3 (20.0%)	42 (24.8%)
Poor CDU Architecture & Logic	11 (16.2%)	5 (11.9%)	0 (0.0%)	5 (16.7%)	3 (42.8%)	1 (6.7%)	25 (14.7%)
Functional Limitations	6 (8.8%)	4 (9.5%)	0 (0.0%)	4 (13.3%)	0 (0.0%)	3 (20.0%)	17 (10.0%)
Inadequate Control Devices	5 (7.4%)	1 (2.4%)	1 (12.5%)	5 (16.7%)	1 (14.3%)	3 (20.0%)	16 (9.4%)
Accuracy of parameters	9 (13.2%)	3 (7.1%)	1 (12.5%)	0 (0.0%)	2 (28.6%)	0 (0.0%)	15 (8.8%)
Inadequate feedback, Error Prevention & Correction	2 (2.9%)	1 (2.4%)	2 (25.0%)	4 (13.3%)	1 (14.3%)	4 (26.6%)	14 (8.2%)
Poor Display quality & Information Presentation	3 (4.4%)	5 (11.9%)	0 (0.0%)	4 (13.3%)	0 (0.0%)	1 (6.7%)	13 (7.6%)
Small Database/Memory size	7 (10.3%)	1 (2.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	8 (4.7%)
Insufficient User Guidance & Support	3 (4.4%)	3 (7.1%)	0 (0.0%)	1 (3.3%)	0 (0.0%)	0 (0.0%)	7 (4.1%)
Poor Consistency & Compatibility	0 (0.0%)	2 (4.8%)	0 (0.0%)	3 (10.0%)	0 (0.0%)	0 (0.0%)	5 (2.9%)
Inadequate Training	0 (0.0%)	3 (7.1%)	0 (0.0%)	1 (3.3%)	0 (0.0%)	0 (0.0%)	4 (2.4%)
Lack of Intuitiveness	1 (1.5%)	1 (2.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.2%)
High/Low Workload	1 (1.5%)	1 (2.4%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.2%)
Total number of comments	68 (100%)	42 (100%)	8 (100%)	30 (100%)	7 (100%)	15 (100%)	170 (100%)

Table 12-4 - Improvements Sought to Future FMSs

Category	B-757/767 Honeywell FMS	B-737 Smiths FMS	B-777 Honeywell FMS	A320/A321/A330 Honeywell FMS	F-100 Honeywell FMS	B-747 Honeywell FMS	Total number of comments
	No of Comments	No of Comments	No of Comments	No of Comments	No of Comments	No of Comments	
Faster Processor	11 (10.4%)	7 (14.4%)	0 (0.0%)	1 (4.8%)	0 (0.0%)	1 (7.1%)	20 (10.2%)
CDU Architecture & Logic	2 (1.9%)	1 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.1%)	4 (2.0%)
Additional Functionality	25 (23.6%)	9 (18.5%)	2 (33.3%)	8 (38.0%)	0 (0.0%)	4 (28.7%)	48 (24.4%)
Control Devices	14 (13.2%)	8 (16.4%)	1 (16.7%)	2 (9.5%)	0 (0.0%)	3 (21.4%)	28 (14.2%)
Accuracy of parameters	7 (6.6%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	7 (3.6%)
Feedback, Error Prevention & Correction	6 (5.7%)	3 (6.1%)	0 (0.0%)	4 (19.1%)	0 (0.0%)	0 (0.0%)	13 (6.6%)
Display Quality & Information Presentation	21 (19.8%)	10 (20.5%)	3 (50%)	5 (23.8%)	0 (0.0%)	4 (28.6%)	43 (21.8%)
Bigger Database / memory	6 (5.7%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (7.1%)	7 (3.6%)
User Guidance & Support	1 (0.9%)	1 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.0%)
Consistency & Compatibility	3 (2.8%)	2 (4.0%)	0 (0.0%)	1 (4.8%)	1 (100%)	0 (0.0%)	7 (3.6%)
Training	0 (0.0%)	2 (4.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.0%)
Intuitiveness	1 (0.9%)	1 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	2 (1.0%)
"What if?" capability	4 (3.8%)	1 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	5 (2.5%)
Improved barriers to erroneous operation	2 (1.9%)	1 (2.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	3 (1.5%)
Language	3 (2.8%)	3 (6.1%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (3.0%)
Total number of comments	106 (100%)	49 (100%)	6 (100%)	21 (100%)	1 (100%)	14 (100%)	197 (100%)

12.3 Processor Speed

The results of analysis on the Processor Speed categories of FMS problems and suggested improvements are presented and discussed in the following three sections.

12.3.1 Processor Speed Problems

Almost twenty-five percent of all the problem comments made concerned issues to do with FMS processor speed. It is worth noting that 50.4% of the comments made by B-777 pilots concerned processor speed. Attention is drawn to this fleet because it is the newest fleet in the sample showing that this issue is not becoming less of a problem on newer fleets. On a cautionary note, the B-777 fleet had the smallest number of participants in the study making any conclusions drawn less generalizable than those of other fleets on which larger numbers of participants were type-rated. However, a high proportion of B-777 type-rated participants chose to make comments (45.0%).

A sample of the comments concerning processor speed are listed here for illustrative purposes only – the complete set of problems encountered with current FMSs as reported by participants has been included at Appendix M.

"Processor agonisingly slow" (B-737, Smiths FMS).

"Can take pilot away from immediate task ("head in cockpit") due to too slow speed of it at times" (B-737, Smiths FMS).

"Processor speed on FMS is predictably slow" (B-757/767, Honeywell FMS).

"Limited and slow memory affect FMS performance with occasional re-synch "lock-ups" at inopportune times e.g. on approach" (B-747, Honeywell FMS).

"The main disadvantage of this FMS is its speed – it processes information very slowly compared to computers elsewhere these days" (A320/A321/A330, Honeywell FMS).

Comparing Airbus and Boeing aircraft, pilots type-rated on Airbus aircraft made notably fewer comments concerning processor speed than those type-rated on Boeing aircraft. Only 10% of Airbus type-rated pilots made comments concerning processor speed compared with a mean of 32.1% (sd=12.9%) of Boeing type-rated pilots. No Fokker pilots chose to make comments concerning processor speed. No notable difference was identified between Honeywell and Smiths FMSs, as one respondent commented "All FMSs are far too slow".

12.3.2 Processor Speed Suggested Improvements

Twenty suggested improvement comments were made concerning processor speed. Pilots type-rated on the F-100 or B-777 aircraft made no processor speed improvement comments. Eleven of the comments were made by B-757/767 type-rated pilots, while

seven comments were made by B-737 type-rated pilots. One comment was made by a B-747 pilot and one comment was made by an Airbus A320/A321/A330 type-rated pilot. The improvement comments raised reflected the processor speed problems identified in Section 12.3.1. All comments have been included at Appendix M, a representative sample of the comments made have been included here:

"The FMS is extremely slow in adapting to new inputs in the vertical profile. This causes lag in response, therefore you almost always have to override it with another vertical mode. Faster chip please." (B-737, Smiths FMS).

"Faster processing" (B-757/767, Honeywell).

"Retrieval and processing should be much faster" (B-757/767, Honeywell).

As reported in the previous section, processor speed is clearly an area in which FMS improvements can be made.

12.3.3 Processor Speed Discussion

One third of all comments relating to problems with current FMSs were related to the FMS processor being too slow. The speed of operation for FMSs depends on the task being performed. Most current FMSs have a 16-bit processor, although 32 bit processors are also available now. The B-757/767 software certificated in 1992, the Product Improvement Package (PIP), had a 16-bit processor. The B-747-400 incorporating two 16-bit processors was certificated in 1997. Similar to the issues associated with database size, the processor fitted to the FMS on an aircraft is largely down to the airline that purchases that aircraft, and the specification of their requirements. So while 32-bit processors are found on the latest version (U8) of Smiths B737 FMSs, most airlines currently flying B-737 will not be using FMSs fitted with that processor.

Processor speed was one of the usability scales developed during Studies One and Two and included the items "The FMS processes information for other activities in the flight deck" and "It is irritating to have to input information at a rate slow enough for the FMS processor". In both Studies One and Two, the **Processor Speed** scale discriminated between usability ratings on the aircraft/FMS manufacturer combined and on the basis of aircraft type when examined on its own. In Studies One and Two, Airbus aircraft with Honeywell FMSs, were rated more favourably than Boeing aircraft with Smiths FMSs, which in turn were rated more favourably than Boeing aircraft with Honeywell FMSs. In Study One, where a number of participants were type-rated on Fokker aircraft, they received the least favourable ratings of processor speed. The mean ratings for **Processor Speed** items were well below the ratings scale midpoints indicating poor usability – see Tables 4-7 and 8-7. The number of comments made by aircraft types supported the rank order findings in Studies One and Two in that Airbus type-rated pilots made the least number of problem comments related to processor speed (10%). Over 28% percent of pilots type-rated on Boeing aircraft with Smiths FMSs made problem comments related to processor speed, while a mean of 33.3% (sd=15.6%) of

pilots type-rated on Boeing aircraft with Honeywell FMSs made such comments. Evidence that the usability scale developed was measuring that which it purported to measure was therefore derived (construct validity). A similar trend was noted in the suggested improvements as shown on Table 12-4, with Airbus type-rated pilots proportionally making a smaller number of comments than either those pilots type-rated on Boeing aircraft with Honeywell or Smiths FMSs. A caveat of caution must be made in interpreting the qualitative data due to the variations in sample size across the fleet groups.

The fact that the processors within FMSs are very slow is compounded by inadequate feedback such that pilots are unaware as to whether the system is processing information or whether the inputs they have made are incorrect. This was evident from the comment made by a B-737 pilot - "Quite often after some value has been entered and executed on the FMS CDU, the FMS goes into a calculation mode. The wait period is often perplexing because the pilots want to verify the input but can't tell if the FMS is 'cooking' something or for some reason the input was incorrect. It would be very helpful to have a symbol something like the hourglass used with computers to indicate that the FMS is busy".

The findings of this research reflect those of Curry (1985) who even sixteen years ago found that a large number of pilots felt that the response times for Flight Management Systems were *"excessive"* (p.22).

12.4 CDU Architecture & Logic

In the following three sections, the results of analysis of the CDU Architecture & Logic categories of problems and suggested improvements are presented. A discussion of CDU Architecture & Logic is put in the context of usability literature in Section 12.4.3.

12.4.1 CDU Architecture & Logic Problems

The researchers agreed that 25 of the comments made were concerned with CDU architecture & logic issues. Comments were made by participants type-rated on every fleet included in the sample, except the B-777. It must be reiterated that caution must be applied when interpreting any fleet or FMS effects from this study. It was, however, noted that although Fokker type-rated pilots only made up 3.6% of the sample who were given the opportunity to make comments, 42.8% of the comments they made were concerned with CDU architecture issues.

A sample of the comments made have been included here for illustrative purposes:

"Even my mobile phone has a more logical menu/submenu set-up" (B-757/767, Honeywell FMS).

"This should run on a 'Next Page' logic but does not. Time constraints in the pre-departure phase caused by lengthy FMS loading procedures lead to poor crew situational awareness immediately after take-off " (B-747, Honeywell FMS).

"Engine out information awkward to obtain" (B-757/767, Honeywell FMS).

"The flight plan format has unnecessary information, while neglecting distance" (F-100, Honeywell FMS).

"Key label entry sequences of FMS units should follow that of ATC clearances. (B-737, Smiths FMS)".

The comments made by participants indicated that the current structure and logic of FMS CDUs caused usability problems for participants.

12.4.2 CDU Architecture & Logic Suggested Improvements

Four of the improvement comments made referred to CDU architecture issues. Two of the comments were made by B-757/767 type-rated pilots, while the remaining two comments were made by B-737 and B-747 type-rated pilots. The areas identified for improvement reflect the problems associated with current FMSs as discussed in Section 12.4.1. All comments referring to CDU architecture can be viewed at Appendix M, three of the four comments have been included here:

"Simplified Logic" (B-757/767, Honeywell FMS).

"To change the arrival data, say runway change, when part way into a STAR is far too cumbersome at times. It makes the workload peak unnecessarily as sometimes all of the elements, Approach type & runway/STAR/Transition, need changing. Surely this can be made a much more intuitive function by combining these elements" (B-747, Honeywell FMS).

"Ease of access to range/bearing direct to a waypoint without having to go through 'Fix' page" (B-737, Smiths FMS).

The comments made by participants indicated that design improvements are required to the structure and logic of current generation FMS CDUs in order to improve usability.

12.4.3 CDU Architecture & Logic Discussion

Almost fifteen percent of respondents that identified problems with their current FMS asserted that CDU architecture & logic was a concern. The comments made were largely concerned with the way in which the data within the CDU were structured and the logic that the designers had applied to sequencing pages and menus. Difficulties in accessing information were also noted. Similarly, the suggested improvement comments referred to the requirement for a simplified logic, and ease of access to particular information in a particular set of circumstances. On reviewing the comments

made, it became evident that the issues raised within this category largely reflected those of the Structure & Sequence scale developed during Studies One and Two. Items loading onto this scale included "The structure of the FMS is well organised from the user's point of view" and "There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts". The distribution of comments across fleet groups meant that no definitive conclusion could be drawn on fleet effects. It was evident, however, from the scales derived in Studies One and Two that Boeing aircraft achieved more favourable ratings on the Structure and Sequence usability scale than either Airbus or Fokker aircraft.

The display area of an FMS CDU is a five-inch CRT that can typically display up to 14, 24 character lines. In order to fit all the required data, it has been necessary to provide something in the order to fifty pages of information. Individual functions may have more than five pages, and it may be necessary to navigate through ten or fifteen pages in order to retrieve the required data on some CDU designs (Billings, 1997).

Hidden functions also exist within the CDU architecture. Singer (1999), for example, argued that the function of transferring a change from one CDU to the other on a Dual FMS system, was a hidden function that was required for safe operation of the aircraft. He argued that on one particular FMS design it was necessary to page through the entire flight plan in order to perform this function. This, he argued, could require up to six inputs. Sarter (1991) noted that "*The opaqueness of the system also appears in the page architecture of the CDU which makes it difficult to find all the relevant data*" (p.1309). When Curry (1985) asked pilots whether they felt that aircraft designers' logic differed from pilot users' logic, 88% of the 104 participants agreed that the logic differed.

Where space is scarce as on the flight deck, it is undoubtedly necessary to present information in a hierarchical manner. Wherever data have to be stored hierarchically, there is inevitably a trade-off between the breath of top level information and the depth of the information stored beneath it. The key to success however is to ensure that these pages are organised in a logical, task oriented manner and that the pilot is kept aware of their current position within the database at all time. Hidden options should be eliminated. The number of steps required to perform tasks should be minimised and perhaps a map of the database with associated short cuts should also be considered. Without radical redesign it is unlikely that many of the CDU architecture issues may be difficult to resolve, but improvements are certainly possible (Billings, 1997).

12.5 Appropriate Functionality

Functionality related problems have in this section been described as Functional Limitations, while additional functionality proposed by participants in the improvements section has been described as Additional Functionality. Appropriate Functionality is discussed in Section 12.5.3.

12.5.1 Functional Limitations Problems

Seventeen of the comments made concerned functional limitations of FMSs. Since Accuracy of Parameters is not a usability dimension but rather reflected functional limitations, it was not possible to draw parallels between the scales derived and the qualitative data derived concerning accuracy of parameters. No functional limitation issues were raised for the B-777 and Fokker 100 aircraft, but these were also the two fleet types on which there were the least number of participants and comments made. Comments concerning functional limitations included:

"A newly created waypoint has to be inserted fully i.e. HON.144/12 = Place/Bearing/Distance, which is given the designator PBD01 on the flight plan. To review this new waypoint at a later stage the position must be identified from its position relative to the adjacent waypoints, as the PBD identified provides no useful information" (A320/A321/A330, Honeywell FMS).

"Descent speeds etc. not deleted when VNAV parameters changed" (B-757/767, Honeywell FMS).

"Apart from the waypoint directly behind, you cannot check for times etc. for any other waypoint" (A320/A321/A330, Honeywell FMS).

"It doesn't have missed approach information or the capability to store multiple flight plans at one time" (B-737, Smiths FMSs).

"As ex-Smiths users we all miss the FIX page" (A320/A321/A330, Honeywell FMS).

The comments made by participants indicated that participants viewed their FMS to have many functional limitations. The associated improvements suggested by participants are described in the subsequent section.

12.5.2 Additional Functionality Suggested Improvements

The Additional Functionality category was used to group all comments made by respondents that referred to functionality they would like to see added to future FMSs. A large proportion of the improvement comments made fell into this category (24.4%). The distribution of comments by fleet types can be seen from Table 12-4. Only F-100 type-rated pilots made no comments concerning the requirement for additional functionality. Some of the issues raised were very generic while many others were very aircraft and FMS specific. A sample of the Additional Functionality sought by respondents is presented here, while the complete set of comments can be reviewed at Appendix M.

"Provision of calculator facility within FMS" (B-757/767, Honeywell FMS).

"Greater functionality with secondary fight plan" (A320/A321/A330, Honeywell FMS).

"More than two fixes would be useful" (B-737, Smiths FMS).

"I would like to see time constraints available for waypoints" (A320/A321/A330, Honeywell FMS).

"Take-off and landing times would be helpful" (B-737, Smiths FMS).

"Can we have more fix pages?" (B-777, Honeywell FMS).

"Frequently given direct routing by one ATC controller to have them cancelled by the next controller. Would be a useful facility to be able to recover old route" (B-757/767, Honeywell FMS).

"The well known problem of holding entries which I think is one of the most 'what is it doing now' things, should be resolved for once and for all with a totally different holding page. The same applies for hold exits" (B-757/767, Honeywell FMS).

The large number of Additional Functionality comments made by participants demonstrated the diversity of functionality participants would wish to see in future FMSs and the difficulty of the task facing the manufacturers of future generation FMSs.

12.5.3 Appropriate Functionality Discussion

The Functional Limitations category of comments represents the mismatch between the needs and requirements of FMS users and the actual functionality available from the FMS CDU. Examining this category within the context of existing research, functional limitations maps onto the appropriate functionality dimension of usability discussed by Ravden & Johnson (1989) and the task match dimension of usability proposed by Shackel (1986). Shackel (1986) defined task match in terms of their being an acceptable match between the functions provided by a product and the user's requirement. Ravden & Johnson (1989) argued that if the user interface of a system failed to provide the information or options which a user required to carry out a task, it would lead to frustration in users and could impair their efficiency in carrying out the task. It was also noted in the taxonomy of flight deck automation problems / concerns produced by Funk et al (1996) that FMSs may lack reasonable functionality. Other studies have noted that pilots use only 20% of current functionality (Sherman, Helmreich & Merritt, 1997).

Consistent with the findings of this study (see Table 12-3), analysis conducted in Study One and Study Two found that pilots' type-rated on aircraft with Smiths FMSs, made more favourable Appropriate Functionality ratings than those type-rated aircraft with Honeywell FMSs (see Section 8.7.2). Further evidence for the validity of the usability scales produced was thus derived.

The functional limitations raised in this study were largely aircraft and FMS specific with comments made such as "It doesn't have missed approach information or the capability to store multiple flight plans at once". A pilot type-rated on a Boeing 737 aircraft with a Smiths FMS made this comment. Other aircraft and FMSs do have this capability. This does not however reduce the frustration for current B-737 users. Every current FMS and aircraft has functional limitations and it is likely that this will always

be the case, since there are always functions that can be added. New functionality should not be added on the basis that it can be added, attention should be focused on whether it should be added. Attention should also be focused on ensuring that the functionality that has been implemented is usable.

While there are many cases of specific functionality that could usefully be added to particular FMSs and aircraft, excessive functionality will not in itself resolve usability problems. In fact it may only compound them. The page architecture of the FMS CDU already raises a considerable number of operational issues, as a huge quantity of data is visible through only a small display requiring pilots to page through numerous pages in order to access the information required (Billings, 1997). It is unlikely that the "wish list" of additional functionality that respondents identified will or should be incorporated into the current FMS generation without considerable redesign of the CDU architecture.

12.6 Control Devices

In the following three sections, the results of analysis on the Control Devices categories of problems and suggested improvements are presented. A discussion of input devices is put in the context of usability literature in Section 12.6.3.

12.6.1 Control Device Problems

The researchers allocated a total of sixteen comments, made by participants from every included fleet, to the Control Devices category. The distribution of the comments indicated that Control Devices issues were pertinent to every fleet, although no conclusions based on differences between fleet and FMS types could be drawn. Comments included in the Control Devices category included:

"Poor keyboard" (B-757/767, Honeywell FMS).

"Input keys are in alphabetical order. Most people are now familiar with QWERTY" (B-747, Honeywell FMS).

"Buttons seem to be randomly positioned" (A320/A321/A330, Honeywell FMS).

"CDU Line select keys are too close together" (F-100, Honeywell FMS).

"Flight plan scroll keys appear to operate in the 'wrong sense' i.e. 'up' arrow key scrolls down the flight plan (unlike the B-737 or home computer mouse)" (A320/A321/A330, Honeywell FMS).

"Location of CDUs should not involve stretching to modify inputs – disrupts scan and strains back" (A320/A321/A330, Honeywell FMS).

The comments made by participants indicated that almost every aspect of the FMS CDU input devices caused usability problems.

12.6.2 Control Devices Suggested Improvements

A total of 28 of the improvement comments made referred to Control Devices issues. This constituted 14.2% of all improvement comments. Participants from every fleet except the F-100 made control device related comments as can be seen from Table 12-4. Eleven of the comments referred to the desire to have a QWERTY keyboard rather than an alphabetical one. Other issues raised included the desire to have alternative input devices. The improvements sought to FMS control devices corresponded to the problems associated with control devices in current FMSs as discussed in Section 12.6.1. Some of the issues raised are included here while the entire set of comments can be viewed at Appendix M.

"I would like to see a more comfortable location of the keyboard" (B-757/767, Honeywell FMS).

"More ergonomic keyboard" (B-757/767, Honeywell FMS).

"Keyboard repeat rates need standardizing" (B-747, Honeywell FMS).

"In this day and age I think a QWERTY keyboard would be better" (B-777, Honeywell FMS).

"More logical keyboard layout – better grouping of keys" (B-757/767, Honeywell FMS).

"Mouse/trackball" (B-747, Honeywell FMS).

The comments made by participants indicated that almost every aspect of the FMS CDU input devices required design improvements.

12.6.3 Control Devices Discussion

The Control Devices category in this study was used to group the comments made by participants related to the physical concerns about CDU control (input) devices. The specific issues raised related to keyboard layout, intuitiveness of arrow keys used to scroll through menus, requirement for other input devices such as touchscreen and rollerball controls, and physical discomfort associated with the location from which the FMS is used. Many of the comments referred to the desire to have QWERTY keyboard rather than an alphabetical one. As one respondent Airbus type-rated respondent phrased it "Buttons seem to be randomly positioned".

Respondents from every participant fleet raised control device problem issues, but the distribution of comments across fleets/FMSs meant that no definitive conclusions concerning control devices could be drawn. This may, again, be due to the fact that control device issues are pertinent to every FMS type. Suggested improvement comments relating to control devices were made by participants from every fleet except the F-100, but this was also a fleet category with a low number of participants.

Current CDU designs are made up of a screen, line select keys, alphanumeric keys, specific function keys and warning lights. As previously discussed, although the design is typical to all manufacturers, there is no standard position for letter keys (Singer, 1999). The fact that the FMS now performs functions beyond that for which it was initially designed underlies many of the problems identified within this study (Billings, 1997). Since FMSs were introduced, desktop PCs have become common place in homes. The result has been a generation of people familiar with operating QWERTY keyboards and many of the comments made about control devices referred to users' preference for such a keyboard in the FMS context. Examining the issue of the physical location, CDUs may be installed in different locations and angles relative to the pilot's reference eye point. One of the deficiencies, however, is that the line select keys may not be aligned to the display. Many other specific issues, for example, related to the size of text on the keyboard, were raised as can be seen from the entire set of comments at Appendix M.

12.7 Accuracy of Parameters

In the following three sections, the results of analysis on the Accuracy of Parameters (e.g. Wind, fuel or ETA) categories of problems and suggested improvements are presented and discussed.

12.7.1 Accuracy of Parameters Problems

A total of fifteen comments were made concerning problems associated with the accuracy associated with FMS parameters. Those pilots type-rated on Airbus or B-747 aircraft made no comments concerning accuracy of parameters.

"FMS places too much emphasis on current wind when making predictions" (B-757/767, Honeywell FMS).

"Fuel predictions are poor" (B-757/767, Honeywell FMS).

"One weakness is VNAV descent which usually gets the aircraft high resulting in a demand for 'more drag'. I can work out a better descent profile in my head than the profile that the FMS comes up with" (B-737, Smiths FMS).

"VNAV descent profile (path) generated by FMS is far too vague to be relied upon" (B-777, Honeywell FMS).

Comments indicated that a number of usability problems are created for users by poor parameter accuracy within the FMS calculations.

12.7.2 Accuracy of Parameters Suggested Improvements

Seven of the improvement comments made were concerned with the accuracy of parameters. It is interesting to note that B-757/767 type-rated pilots made all the

comments in this category. The issue of Accuracy of Parameters was also raised as one of the problems associated with current FMSs as discussed at section 12.7.1. The entire set of comments has been included at Appendix M. Some of the comments made are included here for illustrative purposes:

"More accurate fuel and ETA estimates using inputted winds would improve system. It is difficult to know what information it is using to make these calculations" (B-757/767, Honeywell FMS).

"More accurate prediction: fuel/time/climb/descent profile" (B-757/767, Honeywell FMS).

Comments made by participants indicated that parameter estimates made by FMSs should be more accurate.

12.7.3 Accuracy of Parameters Discussion

As discussed in Section 1.4.2 the data stored in an FMS database includes accurate aircraft drag and engine model data, maximum altitudes, and maximum and minimum speeds. Computations related to lateral navigation are based on a stored or manually entered flight plan which includes items such as courses to be flown, Estimated Times of Arrival (ETA) and distances to go. For vertical navigation, the FMS can compute items such as fuel burn data, optimum speeds and altitudes based on an aircraft's gross weight, cost index, predicted winds at cruise altitudes, and specific ATC constraints.

As with any technology the FMS has some performance limitations, and the limitations discussed within this section are concerned with inaccuracies associated with some FMS parameters. The comments made by participants concerning the accuracy of parameters were quite specific to particular FMSs and aircraft. Almost nine percent of the problem comments made concerned accuracy of parameters and it was interesting to note that none of these comments were made by pilots type-rated on Airbus or B-747 aircraft. However, it is not possible, within the constraints of this study to draw definitive conclusions concerning the Accuracy of Parameters associated with the FMS on board these aircraft.

Since Accuracy of Parameters is not a usability dimension but rather reflected functional limitations, it was not possible to draw parallels between the scales derived and the qualitative data derived concerning accuracy of parameters. However, it is suggested from the findings of this study that inaccuracies do occur within the FMS when predicting parameters such as wind, fuel, and ETA. The associated implication is that the predicted flight profiles can be adversely affected and can thus be misleading to pilots. The implication from the 'suggested improvements' comments made by participants for future FMSs is that increased accuracy in terms of parameter predictions, and improved pilot awareness of the information used in the calculation of these parameters would reduce pilot workload and would increase trust in the automation.

12.8 Feedback, Error Prevention & Correction

In the following three sections Feedback, Error Prevention & Correction problems made by participants are described, along with associated suggested improvements. In examining the comments that were categorised as Feedback, Error Prevention & Correction, the relationship between Feedback, Error Prevention & Correction becomes evident. The comments made are two sides of the same coin, the feedback dimension relating to current FMS status with the error prevention & correction dimension referring to future FMS status. The relationship between these two dimensions was discussed in Section 1.7.8.

12.8.1 Feedback, Error Prevention & Correction Problems

A total of 8.2% of the problem comments made referred to Feedback and Error Prevention & Correction. Participants from every fleet involved in the study made comments related to this category. Due to the relatively small number of comments made and the distribution of comments across participants from every fleet, it was not possible to draw any conclusions concerning effects of FMS or fleet.

Comments relating to the current status of the system and thus feedback included:

"It appears that waypoints have hidden attributes (i.e. definite glide paths to the waypoint) which are not clear to the pilot" (B-777, Honeywell FMS).

"Error messages need clarifying" (B-747, Honeywell FMS).

"Generally speaking a lot of valuable time is wasted in format errors when the FMS returns with "Not in Database" instead of "Format Error" in lateral inputs, and "Invalid Entry", with no clue as to what is wrong in other areas" (B-747, Honeywell FMS).

Four of the comments made referred to the problems associated with FMSs which did not permit pilots to confirm FMS inputs prior to those inputs being activated." This is similar to "Confirm File Delete" function in word-processor applications which asks the user questions such as "Are you sure you want to permanently delete this file?".

"The 'Go direct to' a particular waypoint, you get no second chance or execute key, but to change to alternates takes at least six steps" (A320/A321/A330, Honeywell FMS).

"There is not a strong logic line between those functions which require confirmation before execution and those that do not. If the processing unit is slow the individual will have moved onto another page before the EXEC light illuminates leaving the crew with no way of instantly knowing what is waiting to be executed (B-747, Honeywell FMS).

Those messages relating to error prevention and correction included:

"Frequently there is no advance warning that critical information has been omitted" (A320/A321/A330, Honeywell FMS).

"Error messages are not accompanied by reason and recovery strategy" (B-757/767, Honeywell FMS).

The comments demonstrated that both a lack of informative feedback and poor error recovery strategies within the FMS caused usability problems for participants.

12.8.2 Feedback, Error Prevention & Correction Suggested Improvements

Thirteen comments were made concerning Feedback, Error Prevention & Correction improvements. Participants type-rated on B-777, B-747 or F-100 aircraft made no comments pertinent to this category, but these were also the fleets which had the least number of participants. Table 12-4 shows that six comments were made by B-757/767 pilots, three comments were made by B-737 pilots and A320/A321/A330 pilots made four comments. The improvement comments categorised in this section correspond to the Feedback, Error Prevention & Correction problems with current FMSs identified in Section 12.8.1. Four comments referred to the lack of opportunity to review entries prior to their execution. It is interesting to note that three of these comments were made by Airbus A320/A321/A330 pilots typed-rated on Honeywell FMSs. Unlike the Boeing-737 Smiths FMS, the Airbus A320/A321/A330 Honeywell FMSs do not permit pilots to review entry errors prior to their execution. An opportunity for feedback is thus not created, nor is it possible for changes to be cross-checked by the other pilot on the flight deck.

The first five example comments refer to Feedback while the latter comment refers to Error Prevention & Correction improvements.

"Use of colour on the display would highlight important information." (B-757/767, Honeywell FMS).

"More definitive fault messages" (A320/A321/A330, Honeywell FMS).

"Most entries are single button push. It would be helpful to be offered more two-step entries i.e. enter, then confirm or erase" (A320/A321/A330, Honeywell FMS).

"I think there should be an enter key to confirm the important functions and executions to allow a second chance to evaluate the implications of the function carried out as is the case with the B-737 'exec key'" (A320/A321/A330, Honeywell FMS).

"A more obvious system when a change has been made but not executed as the execute button is sometimes pushed but does not execute and the light in it is often very dim or does not work" (B-737, Smiths FMS).

"Error messages should describe the problem, and give information on correct entry format" (B-757/767, Honeywell FMS).

The comments indicated that improved Feedback and Error Prevention & Correction strategies would benefit FMS usability.

12.8.3 Feedback, Error Prevention & Correction Discussion

Feedback and Error Prevention & Correction have been considered within a single category in this study as they are interrelated. Good feedback aids error prevention and correction, but they can at the same time be considered as distinct elements in that feedback refers to current status where Error Prevention & Correction is related to future system status. The problems associated with inadequate feedback and error prevention and correction in human-computer interfaces have been well documented (e.g. Shneiderman, 1992). In discussing the problems associated with flight deck automation Norman (1991) noted that it was feedback rather than automation which was the culprit.

The respondents to this study reported both Feedback and Error Prevention and Correction to be problems associated with current FMSs, and made recommendations that both dimensions should be improved. The problems associated with inadequate Feedback and Error Prevention and Correction are evident from feedback messages such as "Invalid Entry" which neither state why the entry was invalid, nor provides any information on how it might be amended. Sarter and Woods (1992) found similar problems related to inadequate Feedback and Error Prevention & Correction and argued that such data entry problems frequently occurred when the pilots were working under time pressure. Sarter and Woods reported that many examples of inadequate feedback existed in use of the FMS including those associated with integrating data on FMS status which was distributed on different CDU pages. They argued that weak feedback could increase mental workload by increasing the demands on pilots to remember information and by additionally requiring them to rely on mental models of the FMS structure and to predict future FMS behaviour.

Two distinct factors, **Feedback**, and **Error Prevention & Correction** were elicited during Study One, both factors were confirmed during Study Two and scales were produced. Analyses conducted in Study One demonstrated that **Feedback** was one of the factors which could be used to discriminate between FMS manufactures, with Honeywell FMSs receiving more favourable ratings than Smiths FMSs. Similarly in Study Two, the **Feedback** scale discriminated between FMS manufacturers. Honeywell FMSs were again rated more favourably than Smiths FMSs.

In Study Two, it was demonstrated that **Error Prevention & Correction** was one of the scales that could discriminate between the combined impact of FMS and aircraft manufacturers. Airbus aircraft with Honeywell FMSs were rated most favourably, followed by Boeing aircraft with Smiths FMSs then Boeing aircraft with Honeywell FMSs. **Error Prevention & Correction** also discriminated between aircraft manufacturers when they were examined independent of FMS manufacturer. Airbus type-rated pilots made more favourable usability ratings than Boeing type-rated pilots.

In Study Three, comments concerning **Feedback, Error Prevention & Correction** problems associated with current FMSs were made by participants from every fleet.

Relating the qualitative study to the findings of Studies One and Two in which ratings of Honeywell FMSs were more favourable than Smiths FMSs on the Feedback dimension of usability. It is interesting to note that three of the nine respondents that made comments concerning improvements sought to Feedback were type-rated on Smiths FMSs. Examining Error prevention and Correction, four comments referred to the lack of opportunity to review entries prior to their execution. It is interesting to note that three of these comments were made by Airbus A320/A321/A330 pilots typed-rated on FMS. Unlike the Boeing 737 FMS, the Airbus A320/A321/A330 Honeywell FMS does not permit pilots to review entry errors prior to their execution. An opportunity for feedback is thus not created, nor is it possible for changes to be cross-checked by the other pilot on the flight deck. This defies the monitor cross-check principle, which should be associated with all flight deck automation (Dekker & Orasanu, 1999), and raises a pertinent issue for consideration in the design of flight deck philosophies.

Dekker & Woods (1999) argued that new feedback design targets were needed for all flight deck design, while Singer (1999) noted that current FMSs were particularly deficient in terms of the feedback given to the pilot. Some FMSs, for example, permit the pilot to review entries made to the CDU prior to that change being executed. Others implement the change as soon has it has been entered into the CDU. Since data entry errors are quite common (Singer, 1999), pilots should be able to review and accept their entries prior to those entries being implemented. Another deficiency in many CDUs highlighted by this study and noted by Singer (2001) is the absence of a facility that allows the user to undo the last function executed by the system. This merits particular attention in light of the fact that some FMSs do not incorporate the function where by the user is asked confirmatory questions prior to the execution of risky commands. The 1996 report of the FAA Human factors team argued that there were many reasons why an "undo" function was not available on modern FMSs including those related to cost, but asserted that "*a contributing factor is insufficient knowledge about how to design systems to be human-centred*" (p.102).

The issues raised here defy the basic Feedback HCI principle outlined by Nielsen (1993) which argued that at all times feedback should be provided *"to the user about the system's state and actions"* (p.92). Dismukes & Tullo (2000) made a more radical statement considering Error Prevention & Correction by arguing that "*Designers need to do much more. The entire aviation system should be designed to help humans detect and correct error*" (p.63).

12.9 Display Quality & Information Presentation

The results of the analysis of the FMS Display Quality and Information Presentation categories of problems and suggested improvements are presented and discussed in the following three sections.

12.9.1 Display Quality & Information Presentation Problems

Thirteen comments were made which related to problems with FMS Display Quality and Information Presentation. The comments in this category were concerned with (i) pilots' inability to physically see the display in certain conditions and (ii) information presentation issues. It is interesting to note that pilots type-rated on the Fokker 100 or the Boeing 777 raised no display issues. The largest number of comments made by pilots type-rated on any one fleet was five comments made by pilots type-rated on the Boeing 737 (Smiths FMS). Comments relating to pilots' physical inability to see the display included:

"The display is often difficult to read in bright daylight" (B-737, Smiths FMS).

"Screen often obscured by sunlight (B-737, Smiths FMS)".

"Cannot see screen because of sunlight reflection (A320/A321/A330, Honeywell FMS)".

Comments relating to information presentation issues included:

"Presentation of data inconsistent" (B-737, Smiths FMS).

"Main criticism is that "FLT Plan" pages are cluttered with small characters – not good for fast assimilation of data" (A320/A321/A330, Honeywell FMS).

"The layout of information on the Boeing FMS is not easy to read" (B737, Smiths FMS).

The comments made by participants indicated that both the display quality and the presentation of information were problematic.

12.9.2 Display Quality & Information Presentation Suggested Improvements

Improved Display Quality and Information Presentation issues were raised for every fleet except the F-100. For the remaining fleets of aircraft, over a fifth of comments made were concerned with Display and Information Presentation issues. Comments allocated to this category were concerned with two main issues. The first was the physical issues associated with the quality of display. The pilots in this study reported that a larger screen was required, and that both fingerprints left on the display and the

sun shining on it meant that improvements to its physical location and quality were required. The pilots also reported wanting improvements to information presentation issues such as use of colour and larger font sizes for highlighting or emphasising information. It is interesting to note that even though the B-777 FMS has had some significant changes to improve the quality of the CDU and its location, improvements similar to those of other fleets were still being sought by participants. The improvements identified reflected the display quality and information presentation problems associated with current FMSs (See Section 12.9.1). Comments which are representative of the display and information presentation issues raised are included here:

"I am a training pilot and teach FMS skills. The biggest improvement I'd like to see is use of colour/type size to highlight essential information" (B-737, Smiths FMS).

"I would prefer the display to be larger so that one page could provide more information (e.g. more waypoints on LEGS page). This would reduce the time spent 'paging up/down'" (B-757/767, Honeywell FMS).

"Screens need angling to be at ninety degrees to line of sight, even if keyboard isn't" (B-757/767, Honeywell FMS).

"My eyesight is okay at 42, but the buttons on the CDU are sometimes not easy to read in certain light" (B-777, Honeywell FMS).

"Windows style GUI using large screen and colour co-ordinated with other functions on the B-777" (B-777, Honeywell FMS).

"On progress page make 'last position' in same size font" (B-777, Honeywell FMS).

The comments made by participants indicated that both the display quality and the presentation of information could be improved.

12.9.3 Display Quality & Information Presentation Discussion

Comments related to Display Quality & Information Presentation problems were largely concerned with two issues. The first was the pilots' inability to physically see the display. This was reported to be due to its physical location and environmental conditions such as sunlight and the display having finger print marks on it. The second set of issues raised was related to information presentation concerns which focused on important information not being adequately highlighted, the font sizes being too small, pages being cluttered etc. The findings of this study were consistent with the findings of Funk et al (1996) who noted that in aviation flight deck automation *"Displays may not be designed for adequate visibility, legibility, and readability, possibly resulting in misinterpretation of display information"* (p.11).

Five of the thirteen comments concerned with problems related to current FMSs were made by B-737 type-rated pilots who work with Smiths FMSs. This was a large number proportionally since B-737 type-rated pilots made up 27.7% of the participant sample that made comments. It was interesting that no B-777 or F-100 type-rated pilots made problem comments, since both of these aircraft have displays that have had many changes to improve both their display quality and the style of Information Presentation. B-757 pilots being type-rated on the one of the older FMS-equipped aircraft, made the largest number of display improvement comments. It should, however, be noted that they were also the largest participant group.

It is significant that Information Presentation was one of the scales developed to assess the usability of FMS CDUs during Studies One and Two. The items that loaded onto this scale largely reflected the issues raised in this qualitative study, for example, "Important information is adequately highlighted on the screen". Some evidence of construct validity has thus been provided.

Looking back to Study One, the Information Presentation factor successfully discriminated between the combined effects of aircraft and FMS manufacturer. Consistent with the findings of this study, Boeing aircraft with Smiths FMSs achieved the least favourable ratings, while Fokker Honeywell achieved the most favourable ratings. Although in Study Two, Boeing aircraft with Honeywell FMSs achieved less favourable ratings than those achieved by Boeing aircraft with Smiths FMSs.

In Study Two ratings of the information presentation scale also permitted successful discrimination between aircraft manufacturers with Airbus type-rated pilots making more favourable Information Presentation ratings that Boeing type-rated pilots. Evidence in support of this finding has been found in this qualitative study with one pilot, for example, making the comment "Comparing Airbus and Boeing – the use of colour (in Airbus) is good and assists in speed of data entry, especially in emergency/abnormal situations" (see Appendix M).

Consistent with the findings of this study Billings (1997) noted that "*considerable attention*" (p.117) in terms of research and development was warranted to CDU displays.

12.10 Database / Memory Size

The results of analysis on the FMS database and memory size categories of FMS problems and suggested improvements are presented and discussed in the following three sections.

12.10.1 Database / Memory Size Problems

The Database / Memory Size category was used to group comments concerned with issues such as the FMS memory being too small and thus the navigation database being too small. FMS databases are typically used to store company routes, navigation waypoints and destination airfields. One UK respondent, for example, commented that the FMS on board his B-757 was "too small to put a return route to Rhodes (into)". Pilots type-rated on Boeing 757/767 aircraft with Honeywell FMSs made seven of the eight comments concerning database issues. It is interesting to note, however, that the B-737 pilot who did make a comment pertinent to this section noted that the database was the "worst feature" of the Smiths FMS that his airline was currently using.

12.10.2 Database / Memory Size Suggested Improvements

Seven suggested improvement comments were made concerning database size. Six of the comments made referred to the B-757/767, while the other comment referred to the B-747. Both of these aircraft have Honeywell FMSs. The improvement comments made by participants reflected the problems outlined in Section 12.10.1, which identified problems associated with the FMS memory, and thus the database, being too small.

The following comments represent the improvement issues raised:

"Increase size of database. My Psion has one hundred times the capacity of a 200k FMS" (B-757/767, Honeywell).

"Bigger database – standard route, safe altitudes" (B-757/767, Honeywell).

"Needs far bigger database" (B-747, Honeywell).

The issue of memory size and its associated impact on the FMS database was noted as a problem associated with current FMSs and comments made in the improvements category indicated that participants wished to see FMS memory and database sizes increased.

12.10.3 Database / Memory Size Discussion

FMS memory sizes have been increased over the years. The initial B767 hardware, which went into service in 1982, had 100k of memory. The B-757/767 software certificated in 1992, the Product Improvement Package (PIP) had 1 Mb of memory. There are advantages to increasing database sizes, in that for example, the number of keystrokes a pilot has to make would be reduced if navigation databases were bigger and permitted a greater amount of information to be stored. Foreman (1996) argued that "*Almost without exception, navigation databases do not have enough memory for airline operations. Database sizes are increasing, but many FMSs have only 200 kilobytes of memory. Ten times that amount are really required to provide for ease of operation over even a modest route network*" (p.19).

The memory and database size in any particular FMS is largely down to the operating airline. An airline will choose how much memory and what size of database they wish to fit to the aircraft they operate. Some airlines, for example, will choose to have many more waypoints and approaches, included in their database than do other airlines. The routes to be flown by a particular aircraft will also impact upon the size of database required. Where a route is not pre-programmed into the FMS database, the pilot will be required to input the route manually.

The database issue is not as black and white as it may first appear. Today, databases of waypoints, navigation aids and procedures can be enormous. The fact that the respondents to this study outlined the problems associated with the FMS databases in their airline, reflects the technology that airlines choose to purchase rather than the technology available. Having a large FMS database does not, however, solve all problems as evidenced from Flight 965 in which a Boeing 757 crashed near Cali Columbia on December 20th, 1995 (Aeronautica Civil, 1996). Having been cleared to fly the "Rozo 1 arrival", the non-handling pilot enter the letter "R", the charted identifier code for Rozo, into the FMS CDU. The FMS database however interpreted "R" as "Romeo" which was a beacon 132 miles away. Once the command had been input, the aircraft turned towards the Romeo beacon, which was located in high terrain. Minutes later, the aircraft crashed near the summit of El Deluvio approximately ten miles east of course. Of the 163 persons on board, four passengers survived the accident.

One database issue highlighted by the accident at Cali, was the lack of consistency in naming conventions between FMS databases and paper based charts. The database producers Jeppersen Sanderson had followed one set of conventions for creating the FMS databases while a different set of standards had been adopted for naming points on the paper charts (Endsley & Strauch, 1997). The FMS naming convention thus allocated "R" to a beacon close to the larger airport at Bogota, the Rozo beacon was named Rozo in the FMS database. The lack of consistency between chart identifiers and FMS identifiers can create a significant problem for pilots as was evident at Cali (Aeronautica Civil, 1996). It is however at this point worth examining the Cali accident in terms of the broader usability issues. Evidence from the accident report (Aeronautica Civil, 1996), suggests that a number of the other usability scales developed in Studies One and Two may also have been applicable to the accident in which 159 passengers

and crew were killed. Inadequate **Feedback** from the FMS, for example, meant that the pilots were unaware of their current location relative to their destination location in the minutes preceding the accident. **Information Presentation** was a contributory factor since the waypoints were not in a logical order, which increased the likelihood of the pilots selecting the incorrect waypoint as occurred on that day. **Error Prevention & Correction** was also an important factor since the FMS either did not have the facility to perform, or did not correctly perform, validity checks which could have informed the pilots that the waypoint entered was taking the aircraft away from the required destination.

A further issue that needs to be considered is the structure of the database. Navigation databases may have a number of waypoints with the same name of identifier. To resolve this ambiguity the pilot selects the waypoint from an FMS list of waypoints having the same identifier. The flaw with this mechanism was highlighted by Foreman (1996) who argued that "*unfortunately, the list is not in a logical order, increasingly the possibility is that the pilot may pick the wrong waypoint*" (p.19). Clearly some logic should be applied to the presentation of alternatives (Singer, 2001). For example, when a pilot inserts a waypoint identifier that has many possible locations, the waypoints on the database should be displayed in order of proximity to the aircraft's current location or the closest choice to the destination airport could be displayed first. Alternatively, an updated map indication could be provided on the navigation display.

Furthermore, the CDU should make it evident in the user's language which identifier has been selected, prior to the user executing its selection (Singer, 1999). The participants in this study repeatedly reinforced the problems associated with small FMS databases, which resulted in waypoints having to be input manually.

The small sample sizes included in this category of the content analysis make it difficult to draw definitive conclusions based on fleet or FMS types, although consistent with Studies One and Two it was noted that thirteen of the fourteen comments made in this category referred to problems or improvements required to Honeywell FMSs.

Examining the scales produced during Studies One and Two, the database concerns were reflected in the Appropriate Functionality scale which focused on the FMS allowing users to perform tasks as efficiently as possible, and having access to all information required. From a user's perspective, a small FMS database may mean that routes have to be constructed manually thereby increasing workload and the potential for data entry errors. Both the **Appropriate Functionality** factor produced in Study One and the **Appropriate Functionality** scale produced in Study Two discriminated between FMS manufacturers. Smiths FMSs were rated more favourably than Honeywell FMSs in both studies. The absolute mean scale score for Smiths FMSs was 4.63 (sd=1.12), while being 4.29 (sd=1.32) for Honeywell FMSs (see Table 8-68). Smiths FMSs were also rated more favourably than Honeywell FMSs on the Appropriate Functionality dimension in Study One. Since the qualitative comments supported the applicability of the theoretical construct being measured, evidence of the construct validity associated with the usability scales was demonstrated.

The results of this study indicated that database and memory size were important issues to participants. Some evidence for the validity of this finding has been derived from the aviation automation literature (e.g. Foreman 1996). This section has shown that significant human factors issues are also associated with selecting waypoint identifiers from a list of options. Research to examine methods of reducing this type of human error is warranted.

12.11 User Guidance & Support

In the following three sections, the results of the analysis on the User Guidance & Support categories of problems and suggested improvements are presented and discussed in the context of the usability and aviation automation literature.

12.11.1 User Guidance & Support Problems

Seven of the comments made were categorised by the researchers as referring to User Guidance & Support problems. Pilots type-rated on the B-777, B-747 or Fokker 100 aircraft made no comments concerning User Guidance & Support. It should, however, be noted that these were also the fleet groups that had the least number of participants. Comments relating to the User Guidance & Support facility are given here to demonstrate the types of issues raised. The entire set of comments has been included at Appendix M.

"No online help facility and limited published on methodology of software is unhelpful" (B-757/767, Honeywell FMS).

"Our FMSs do not have an online help feature at all. As far as printed help, we only have our Pilot Handbooks and they are woefully inept as they are not written in a user friendly manner" (B-737, Smiths FMS).

"Manuals tend to be poorly presented by both Boeing and Airbus – can lead to a lack of understanding. Not easy for reference to 'how to do a particular task' or 'how can we do that?'" (B-757/767, Honeywell FMS).

Comments indicated that usability problems are created for pilots by inadequate help facilities.

12.11.2 User Guidance & Support Improvements

Two comments were made concerning improvement sought to User Guidance & Support facilities. A B-757/767 type-rated pilot made one comment, while a B-737 type-rated pilot made the other comment. The comments made were:

"Handbook is generic – would prefer to see a company SOP specific handbook" (B-737, Smiths FMS).

"Access to on-line help" (B-757/767, Honeywell FMS).

The improvements sought in this category clearly reflect the problems associated with accessing help in either on-line or manual forms for current FMSs as discussed in Section 12.11.1.

12.11.3 User Guidance and Support Discussion

User guidance and support, in terms of documentation (users manuals) and on-line help is part of the user interface, and users can form impressions of how usable a system is by flicking through a manual or accessing its on-line help (Mayhew, 1992). It is true to say that when manuals for user interfaces are considered as a whole, they are typically poorly designed, full of difficult jargon, difficult to navigate and difficult to read (Mayhew, 1992). The findings of this study provided no evidence to suggest that the FMS CDU interface is any different. This issue reflects the Gulf of Execution cited by Norman (1986), which reflected the gap between a user knowing what they wanted to have happen and knowing what is required to bring that. The importance of user guidance and support in interactions with any Human-Machine Interface, was discussed at Section 1.8.12.

Over four percent of the problem comments made by participants concerned User Guidance & Support. The comments made by participants suggested that FMSs offer little or no on-line help. Manuals were reported to be "woefully" inept and difficult to navigate through. It was also asserted that little was published on the working of the FMS software, and that the FMS manuals were written for engineers rather than the pilots who have to use them. The improvement comments made reflected the same issues.

Since User Guidance and Support was not one of the usability scales developed during Studies One and Two, it is not possible to draw direct parallels. However, based on the qualitative evidence derived from this study it is suggested that many improvements could be made to the user guidance and support documentation available to pilots. Particular attention is merited to this in light of the difficulties associated with navigating through databases which have many layers, and much complex functionality which is used infrequently and when it is used it may be in the context of abnormal or emergency situations e.g. single engine operations.

Further, in light of the resistance to change by the aircraft manufacturers and the airlines themselves, this is one dimension that could be easily remedied. The findings of this study are consistent with the findings of research conducted by Funk, Lyall and Riley (1996) who also reported that the support documentation for FMSs was inadequate.

12.12 Consistency and Compatibility

Consistency refers to problems encountered within a single FMS type. Two issues were raised in considering compatibility between FMSs and these were issues also raised during the interview stage of Study One (see Appendix C). Issues raised in this study concerned (1) differences between FMS CDUs within a single airline and single fleet, and (2) differences between FMS CDUs when moving between fleets and thus potentially FMS manufacturer. The entire set of comments may be viewed at Appendix M.

12.12.1 Consistency and Compatibility Problems

Five comments were made concerning consistency within and compatibility between FMSs. A comment made concerning compatibility issues between FMSs within a single fleet indicated that "Some FMSs perform slightly differently in obscure ways and some perform at slightly different speeds. Even though the FMS model and operating program are identical, there are some differences in different aircraft" (Boeing 737, Smiths FMS). Two comments were made concerning consistency within the FMS and both were concerned with the Airbus A320/A321/A330 Honeywell FMS. The comments made were:

"Some pages take four letter codes, others take three letter codes for airports" (A320/A321/A330, Honeywell FMS).

"Some pages use "next page", others use prompts at LSK (Line Select Keys)" (A320/A321/A330, Honeywell FMS).

A comment made considering the lack of compatibility between fleets stated "Recently converted to 737-800 (Smiths FMS) from 757/767. There are many subtle differences in FMS operation between the types for no apparent reason. This makes type conversion harder and can lead to incorrect FMS usage during the re-learning process" (Boeing 737, Smiths FMS).

The comments indicated that both the lack of Consistency within FMSs and Compatibility across FMSs was problematical for participants.

12.12.2 Consistency and Compatibility Suggested Improvements

Seven comments were allocated to this category by the researchers. The comments concerned both the lack of consistency within FMSs and the lack of compatibility across FMSs when pilots undertook new type-ratings as discussed in Section 12.12.1. The distribution of comments across fleets can be seen from Table 12-4. The comments made relating to improvements sought to consistency within FMSs and Compatibility across FMSs included:

"General standardization of procedures i.e. abeam fixes incorporated into legs page yet non-standard procedure when abeam fix is to be next / active waypoint" (B-757/767, Honeywell FMS).

"Due to fleets having three types of FMS – standard, PIP and Pegasus, the additional functions and methods of use between the three types (which can occur in one duty with aircraft changes) can cause additional workload. A more common thread between the three would aid this problem as it only adds to having to remember the differences with the aircraft types as well" (B-757/767, Honeywell FMS).

"Standardization throughout all FMSs would make transitions easier!" (F-100, Honeywell FMS).

"Would be nice if B-737 FMS was similar to B-767 with Route 2 etc." (B-737, Smiths FMS).

"Standardization of pages, input methods and display of information is vital. An international protocol needs to be agreed" (B-757/767, Honeywell FMS).

The comments indicated that participants sought improvements to both Consistency within FMSs and Compatibility across FMSs.

12.12.3 Consistency and Compatibility Discussion

Consistency and Compatibility are amongst the most basic usability principles (Nielsen, 1993). The Consistency and Compatibility issues raised in this study were focused on the lack of consistency within single FMSs and fleets of aircraft and the lack of compatibility between FMSs when pilots undergo a new type-rating onto a new aircraft and FMS. Funk et al (1996) raised the lack of standardization within and between fleets as a concern. Even though the FMSs within a single fleet may be of the same model, differences in, for example, response speeds have been noted by participants. Many of the differences that occur within a single fleet are a product of aircraft being bought from different sources. This is almost inevitable unless the whole fleet is bought direct from the manufacturer when new. The technology fitted within a single fleet may have initially been specified by many different airlines when they were initially purchased.

As aircraft are sold on to other airlines, a single airline may end up with a fleet from numerous different sources.

Comments analysed in this study and noted in Appendix M highlighted the difficulties pilots have in moving between fleets e.g. "Having been on Boeings for ten years prior to the F-100, I am probably biased towards to Boeing philosophy and therefore am not really comfortable with the F-100. Standardization throughout all FMSs would make transitions easier" (F-100, Honeywell FMS).

The Consistency scale developed in Studies One and Two did not discriminate between FMS usability ratings by either FMS or aircraft manufacturer. However, the Consistency scale scores from Study Two showed that that Airbus aircraft with Honeywell FMSs achieved higher usability ratings than Boeing aircraft with Honeywell FMSs or Boeing aircraft with Smiths FMSs (see Table 8-62). Consistent with the findings of Study One, Honeywell FMSs were rated as being more usable than Smiths FMSs on the Consistency scale in Study Two (see Table 8-68). Similarly Airbus aircraft achieved higher Consistency scores than Boeing aircraft in both studies One and Two (see Tables 4-28 and 8-74). Although, all of the Consistency scale scores were above the mid-rating scale point, the findings of the content analysis suggested that consistency and compatibility is a problem across manufacturers and fleet.

From a usability perspective, all FMS CDUs should as far as possible contain the same information, presented in the same location and formatted in the same style to facilitate recognition (Nielsen, 1993). The problem from the FMS CDU perspective is that little is standard. In the PC environment, a certain level of usability standard exists in the form of ISO 9241 Part 11 guidelines and use of the Windows operating systems. Looking across FMS CDU manufacturers, little is standard. While the CDU design is typical across manufacturers, there is no standard position for the letter keys, and each function key has a different purpose in each design. As reported by Singer (1999), the navigation key is used for navigation on one FMS design, while only being used for planning on another design. Similarly the Flight Plan key can only be used for planning on one design while it can be used for navigating on another design. The lack of consistency within FMSs and compatibility across FMS raises the potential for pilots to make errors. Further, a significant opportunity for negative transfer is also created when pilots move between aircraft and FMS manufacturers as they undertake new type-ratings.

The findings of this study were reinforced by Gilson & Abbott (2000) who noted that the chief disadvantage of flight deck interfaces, including those of the FMS, was the loss of standardization of symbology, format, sequencing priorities, etc. used. The authors argued that varying designs and customisations had been implemented by systems and aircraft manufacturers and by airlines. The outcome, Gilson & Abbott (2000), argued was that such variations had come to exist among flight decks that pilots type-rated on an aircraft "*may find flight safety compromised by their unfamiliarity with the particular interface system available to them*" (p.131).

12.13 Training

In the following three sections, problems and suggested improvement comments related to FMS training are presented and discussed. Although training is not strictly a usability issue, the requirement for increased user training can be a consequence of poor interface design and usability (Nielsen, 1993).

12.13.1 Training Problems

Four respondents made comments concerning problems associated with FMS training. It is interesting to note that pilots type-rated on B-737 aircraft which has a Smith's FMS made three of these comments. This is both a large proportion of the small sample who made comments concerning training, and is also a notable proportion of B-737 respondents, when it is considered that B-737 type-rated pilots constituted only 34.6% of the respondents who made comments. The comments made were:

"Airlines tend to spoil the FMS by poor initial training and confusing information contained in flight manuals" (B-737, Smiths FMS).

"There is a culture of computer illiterate trainers, passing it off as 'you'll get used to it on the line'" (B-737, Smiths FMS).

"The system requires a great deal of training/familiarity during the early stage of line flying. There is completely inadequate access to training systems on computer courses. We are perfectly able to cope with these systems often learning by experiment rather than consultation with manuals" (A320/A321/A330, Honeywell FMS).

"The greatest problem in use of the FMS is the poor standard of training in its use. Apart from one or two slow computers there is no means of getting hands on experience with an FMS other than during line training, which is far too busy in general for in-depth discussion about 'what is it doing now?'" (B-737, Smiths FMS).

Some participants clearly felt the FMS training they received to be inadequate.

12.13.2 Suggested Training Improvements

Two comments were made concerning improvements sought to training in FMS usage. B-737 type-rated pilots made both comments. Training issues were also raised in the problems with current FMSs as discussed at Section 12.13.1. The improvement comments made were:

"Operators should have access to a 'real' CDU linked to a suitable computer which emulates all the subtleties and idiosyncrasies of the aircraft concerned, so that the student can punch the keys and make all the customary mistakes long before he/she gets into the aircraft for the first time." (B-737, Smiths FMS).

"Considering the complexity of recent game consoles and the latest generation Pentiums, it should not be beyond the wit of man to produce a system with say a miniature MCP and joystick linked with the CDU and computer to actually TEACH the students that which he NEEDS to know." (B-737, Smiths FMS).

The comments indicated that a number of participants felt that some updating of the FMS training process was required.

12.13.3 Training Discussion

A total of six comments were made concerning the inadequacy of current FMS CDU training and suggested improvements to future FMS training. The comments were concerned with the lack of hands-on FMS training which is currently received by pilots in many airlines. It was interesting to note that pilots type-rated on B-737 aircraft with Smiths FMSs made five of the six comments. This is a large proportion considering B-737 pilots constituted 29.4% of the participants who could have made comments. The number of comments made was too small to draw any definitive conclusions, but was none the less worth noting. The large proportion of comments from B-737 type-rated pilots may be related to the fact that the B-737 is one of the few large passenger aircraft fitted with Smiths FMSs as standard. Therefore, a pilot who has been type-rated on any other type of FMS-equipped aircraft will come face-to-face with the problems caused by the lack of compatibility and consistency across FMS manufacturers.

FMS related training issues have been discussed by many authors (e.g. Rudisill, 1995). Research conducted by Curry (1985) asked airline pilot participants directly which material they would like more or less training in. Similar to the findings of this study, the greatest responses were for more FMS / CDU training generally, more 'hands on' experience and training for the CDU, and a greater quantity of line-oriented CDU experiences. Similar, to the findings of this study, the research conducted by Curry also found that training should not teach just CDU functions, but rather a "*philosophy of computer applications and programming*" (p.29) as applicable to a particular aircraft.

It is important to note that some participants in this research programme will have been type-rated on their current aircraft a few years ago when Computer Based Training (CBT) was not generally available as part of type-rating programmes. This situation has now changed to some extent with some type-rating courses now allowing participants to undertake self-paced CBT for much of the ground school training including training in FMS CDU interactions.

12.14 Intuitiveness

In the following three sections, the results of analysis on the Intuitiveness categories of FMS problems and suggested improvements are presented and discussed in the context of the usability and aviation automation literature.

12.14.1 Intuitiveness Problems

Although only two of the comments made concerned the intuitiveness, or rather lack of intuitiveness, associated with operation of the FMS it has been considered as an independent category. The comments made were:

"I would not say that the FMS is intuitive" (B-737, Smiths FMS).

"In particular it is hard to understand the concept of progress via legs pages, and how to go through the index pages" (B-757/767, Honeywell FMS).

It was important to include this category as it was an also an issue that was also raised at the interview stage described in Study One (see Appendix C). The small number of comments included in this section has meant that no conclusions based on fleet or FMS type may be drawn.

12.14.2 Intuitiveness Suggested Improvements

Two comments were explicitly related to improvements sought the intuitiveness of the FMS. A B-757/767 type-rated pilot made one of the comments, while a B-737 type-rated pilot made the other. The comments made were:

"More intuitive operating system" (B-757/767, Honeywell FMS).

"With the technology available today, the FMS CDU could be made much more intuitive" (B-737, Smiths FMS).

Other comments have referred to the lack of intuitiveness associated with the FMS but since the content of the comment was explicitly referring to the intuitiveness of a dimension of the FMS, the comment was allocated to that specific category. For example the comment, "To change the arrival data, say runway change, when part way into a STAR is far too cumbersome at times. It makes the workload peak unnecessarily as sometimes all of the elements, Approach type & runway/ STAR/Transition, need changing. Surely this can be made a much more intuitive function by combining these elements", made by a B-747 pilot referred to intuitiveness but was specifically related to CDU architecture. It was, however, determined by the researchers that this separate category should be derived for Intuitiveness specific comments since the issue had also been raised at the interview stage.

12.14.3 Intuitiveness Discussion

Intuitiveness in the HCI domain loosely refers to user interfaces that are easy to understand without training. Although the term 'Intuitiveness' is a not a component of usability discussed in the research literature, it is in itself a product of usable interfaces and is related to learnability. Components of usability such as feedback make it explicit to the user how interaction with the interface should be performed. The findings of this study reflect Billings (1997) comment which stated that "*many pilots now flying FMS-equipped aircraft would welcome a simpler, more intuitive flight management system with which they could interact more easily than is possible with today's CDUs*" (p.143).

In this study, a small number of comments were made concerning the lack of intuitiveness associated with operation of the FMS. Only B-737 and B-757/767 type-rated pilots made comments related to intuitiveness, but the small number of comments made meant that it was not possible to draw any fleet or FMS related conclusions. It must be noted that the B-737 and B-757/767 were also the aircraft that had the largest number of participants involved in this study.

Highly intuitive systems have a steep incline at the first part of the learning curve and permit users to be reasonably proficient within a short time. Nielsen (1993) argued that almost all user interfaces have learning curves associated with them and in the case of many human-machine interfaces, users have insufficient experience to perform tasks immediately. The exception, Nielsen argued, was for walk-up-and-use systems such as those used to provide tourist information which must essentially have a zero learning time permitting users to successfully interact with the system in one attempt. Nielsen (1993) went on to assert that there are some systems for which "*one can afford to train users extensively to overcome a hard-to-learn interface, but in most cases, systems need to be easy to learn* (p.28). It is argued that FMSs have traditionally fallen into this category, with training being used to compensate for an interface that is not intuitive. The important point to note from the comments made in this study is that training is a poor substitute for good design. This conclusion is in agreement with Sarter, Woods & Billings (1997) who noted that it was important to realise that "*training cannot solve all observed problems – training cannot and should not be a fix for bad design*" (p.1933).

12.15 Workload

In the following two sections, the results of analysis on the Workload problems category is presented and discussed in the context of the usability and aviation automation literature.

12.15.1 Workload Problems

It was surprising that only two of the comments made were directly concerned with workload although many of the other categories discussed within this section impact on both mental and physical workload (e.g. Feedback, Error Prevention & Correction and Display Quality & Information Presentation). A Boeing 737 type-rated pilot made one of the comments, while a Boeing 757/767 type-rated pilot made the other comment. The comments made were:

"Workload is busy in the terminal areas" (B-757/767, Honeywell FMS).

"Can be time consuming just at the wrong time" (B-737, Smiths FMSs).

Although other comments were made which mentioned workload, the researchers did not agree on the category the comment should be allocated to. The following comments are examples of the latter:

"Very high workload with one pilot flying and the other pilot monitoring" (B-747, Honeywell FMS).

"FMS operation tends to increase workload in the TMA, especially in arrival phase" (B-737, Smiths FMSs).

Workload problems in use of the FMS have been well documented as discussed in the subsequent section.

12.15.2 Workload Discussion

It has been well documented that FMSs have redistributed workload, increasing it where it was already high (Foreman, 1996). Wiener (1989) was amongst the first to discuss workload in the context of modern aviation technology. He concluded that many highly automated systems supported pilots in the traditionally low-workload phases of flight but were of little use or even hindered the pilot when help was most needed, in time critical, highly dynamic circumstances. The workload issues associated with use of the FMS also have been widely documented (e.g. FAA, 1996; Sarter & Woods, 1994). As one comment on the Batelle Aviation Safety Reporting System (ASRS) noted *"ON A FLT such as this, the only thing computers do for PLTS* (pilots) *is double the workload"* (ASRS report number 132299).

Although only a small number of the comments made by respondents directly mentioned workload issues, many of the categories of comments discussed within Section 12 impact on both mental and physical workload for pilots. For example, Feedback provided in terms of "Invalid Entry" without provision of Error Prevention and Correction information can leave a pilot scrolling through pages of navigation information looking for a value that requires amendment. This would not necessarily be a problem in a non-busy phase of flight, but can detract from other activities during a time critical phase of flight such as the descent and approach.

Workload is distinct in that it is not a product of usability frequently discussed in traditional usability literature, although it has been referred to in ISO 9241 Part 11. Since usability has traditionally focused on concerns with PCs and consumer goods, mental and physical workload have achieved less attention than is merited when safety critical products are concerned (Bevan, 1995). On the other hand, workload has achieved much attention in the human factors and aviation automation literature. Sarter & Woods (1992) for example argued that the weak feedback aspect of usability associated with FMS CDUs, could increase "*cognitive workload by increasing the demands on pilots to remember information and by increasing the need to rely on mental models of FMS structure and function to assess or project FMS behaviour*" (p.317). Increased workload is a consequence of poor HCI design and usability rather than a dimension of usability or human-machine interfaces (Williges, Williges & Fainter, 1988).

FAR/JAR 25.1523(a) asserts that in determining the minimum number of flight crew required to operate an aircraft safely, the workload on individual crewmembers should be considered. However, there are no criteria set out for determining the acceptable levels of workload associated with the communications and navigation domain involved in using the FMS (Singer, 1999).

The workload scales developed in Studies One and Two did not discriminate between either aircraft or FMS manufacturer. It was further not possible to draw specific conclusions on the workload associated with particular FMS or aircraft variants from the qualitative study. This may, however, be due to the fact that the workload issues associated with FMS CDUs are applicable across all aircraft and FMS variants (FAA, 1996).

12.16 Other Categories of Suggested Improvements reported by respondents

A further five categories of suggested improvements to FMSs, which did not have corresponding problem categories, were raised by participants. The results of analysis on these categories are presented and discussed in Sections 12.16.1 to 12.16.6.

12.16.1 Improved "What if?" Capability

Five comments were made referring to the "What if?" category which grouped comments that referred to the usefulness of a hypothetical capability function. B-757/767 type-rated pilots made four of these comments, while a B-737 type-rated pilot made the final comment. The comments included here reflect the improvements sought.

"More 'hypothetical' capability would be useful e.g. 'if cleared direct to xxx what ETA?' without dumping current route and wind data" (B-757/767, Honeywell FMS).

"I would like to see better wind/ altitude/ temperature related 'what if' scenarios when considering step climb" (B-757/767, Honeywell FMS).

"Ability to make changes and see results without executing" (B-737, Smiths FMS).

A discussion of improved hypothetical capability is discussed in the subsequent section.

12.16.2 Improved "What If?" Capability Discussion

One of the procedures most emphasised in multicrew flight decks is that of monitoring, cross-monitoring, challenge and response. While such a procedure, makes for good theory its implementation on the flight deck is far from straightforward. In interacting with the FMS CDU, for example, the two pilots have effectively their own private workspace, and can make flight critical changes without involving the other pilot. One pilot can, for example, be typing an Aviation Communication And Reporting System (ACARS) message, while the other makes flight plan changes. Dekker & Orasanu (1999) described the practice of fight deck cross monitoring of CDU inputs as a "*procedural band-aid that doesn't stick in practice*" (p.70). Compounding this problem, is the fact that many FMSs do not have an execute function, thus commands are executed as soon an inputs are made leaving no opportunity for cross-monitoring or challenging the other pilots' inputs. The obstacles in the way of cross-checking are even greater when they are most critical, in dynamic phases of flight such as descent and approach where pilots have to perform multiple tasks in a very short time frame.

While an overlap exists between improved "What If?" capability and improved Feedback, Error Prevention & Correction, the existence of a number of comments referring to this specific dimension meant that the researchers felt it merited a focus and was thus treated as an independent category. Five distinct improvement comments

made in this study referred to the requirement for FMS CDUs to permit hypothetical data entries to be made without those inputs being executed. Although B-757/767 type-rated pilots made four of the comments, the small sample size meant that it was not possible to draw any definitive conclusions based on fleet or FMS. It is true, however, that the B-737 Smiths FMS does have an execute button allowing inputs to be made without those inputs being processed. This may be reflected in the fact that the majority of comments concerned the B-757/767 Honeywell FMS which does not have an execute function. Further to the cross-monitoring issue, the comments made also referred to the desire to be able to see changes without those inputs being implemented such that scenarios could be built without being implemented. The hypothetical capability referred to would permit pilots to determine the effects of a parameter such as wind or temperature changes on performance. In light of the lack of opportunity for cross-monitoring, it is understandable that pilots would wish to see the impact of changes they make prior to executing those changes.

12.16.3 Improved Barriers to Erroneous Operation

Three comments were made which directly referred to improved barriers to erroneous operation. B-757/767 type-rated pilots made two of the comments, while a B-737 type-rated pilot made the final comment. The comments that directly made reference to barriers to erroneous operation were:

"During descent the FMS should question the selection of a waypoint which is not part of the approach procedure" (B-757/767, Honeywell FMS).

"Better alerting of change from VNAV path to VNAV speed descent so that pilot knows in advance" (B-737, Smiths FMS).

"Better 'fail-safes' if silly errors are put in" (B-757/767, Honeywell FMS).

Improvements in Barriers to Erroneous Operation are discussed further in the subsequent section.

12.16.4 Improved Barriers to Erroneous Operation Discussion

The improved barriers to erroneous operation category of comments had significant overlap with other categories discussed in this section in that improved usability in some of the other categories (e.g. Feedback / Error Prevention & Correction), would also lead to improved barriers to erroneous operation. The importance of its subject matter meant that the researchers decided that "Improved Barriers to Erroneous Operation" merited being a category in itself. Since many inputs made to the FMS CDU by one pilot are not cross-checked by the other pilot, it is argued that barriers to erroneous operation should be built into the CDU design to minimise the number of errors pilots make. Improved feedback, for example, should be provided when uncommanded mode changes occur. Dekker & Orasanu (1999) argued that *"Time pressure, multiple interleaving activities,*

the backdrop of unambiguous normal practice, and even Flight Management System design" (p.72) can all conspire against the pilot.

Three of the comments made referred directly to the requirement for improved barriers to erroneous operation. The types of barriers to erroneous operation raised by the participants included the questioning of waypoints that were not part of the approach. The small number of comments allocated to this category meant that it was not possible to draw any definitive conclusion based on fleet or FMS type.

Since the FMS is a highly complex system it is highly unlikely that designers could ever conceive a method of trapping every potential error that pilots might make. It is thus necessary to provide an alternative method of detecting such errors in the form barriers to erroneous operation.

Evidence from the accident at Cali, illustrated how although the aircraft was on the approach to Cali, Colombia, the entry of an incorrect beacon 132 miles away was not detected by the FMS (Aeronautica Civil, 1996). Even a basic change, such as the listing of possible beacons listed in order of distance from current location, could have been sufficient to draw the pilots' attention to the erroneous data entered into the CDU.

12.16.5 Language Improvements

Six comments were made concerning the use of language for the FMS. These comments were made by B-757 and B-737 pilots and concerned the conflict that exists between the language used to program the FMS and the language used by Air Traffic Control Officers. A sample of the comments made have been included here:

"Plain language option – if you are not familiar with a route, three letter codes and four letter airport IDs do not mean much – ATC give you clearances in plain language so it is difficult to correlate it with the waypoint on the legs page – often have to ask ATC for three letter code." (B-737, Smiths FMS).

"List of nearest airfields in plain language" (B-757/767, Honeywell FMS).

The language concerns raised in this study were focused on the mismatch that exists between the language or codes used by Air Traffic Control Officers (ATCOs) and those used by pilots as discussed in the subsequent section.

12.16.6 Language Discussion

English is the international standard language of aviation. The language related concerns raised in the aviation automation literature are usually focused on the use of systems and documents, and on pilots who do not have English as their native language and who have varying levels of competency in English usage. The concerns raised in this study were, however, of a different nature and focused on the mismatch that exists between the language or codes used by Air Traffic Control Officers (ATCOs) and those used by pilots to enter information into the FMS CDU. Standard International Civil

Aviation Organization (ICAO) phraseology does exist and should help the potential for miscommunications (FAA, 1996). It is true, however, that existing ICAO standards are outdated for *"many new types of operation now occurring in the aviation environment including FMS approaches"* (FAA, 1996, p.105).

The issue of concern was made explicit by a B-737 pilot's comment "ATC give you clearances in plain language so it is difficult to correlate it with the waypoint on the legs page - often have to ask ATC for three letter code". The incompatibility between flight deck automation and the ATC environment has been discussed by a small number of researchers (e.g. Courteney, 1999), and typically refers to programming of the FMS being too time consuming to cope with ATC instructions in busy phases of flight such as during the approach. The issue of language raised in this study extends that incompatibility. Incompatibilities also exist between paper based charts and FMS databases such that fix and bearing names do not correspond between FMS and chart identifiers (see Section 12.10.3), and clearly some standard between information sources could ease pilots' workload in the busier phases of flight.

12.17 Specific Recommendations for FMS Designers & Certification Authorities

Kelly et al (1992) indicated that *"some improvements to Flight Management Systems have been difficult to justify because they did not appear to provide new capabilities which would result in return on investment"* (p.2). The view that improving Flight Management Systems is not viable economically has meant that many of the valuable HCI lessons learned since FMSs have come into operation have not been promulgated into re-design. This section presents a number of specific recommendations that may be applied by FMS designers and certification authorities. The recommendations are based on current aviation automation and usability literature and the conclusions of this qualitative study.

- FMS response speeds need to be improved. This study suggested that system response speeds need to be faster to prevent pilots becoming frustrated. It is also important, however, that response speeds are not so quick that pilots cannot follow the data that has been processed. The evidence suggests that FMS feedback to the pilot is poor, and even that poor feedback is slow in being displayed. The slow system response may also hinder pilots monitoring and cross-checking each other.
- The number of pages that must be navigated through in order to perform critical activities should be minimised.
- The number of CDU layers should be minimised, and shortcuts provided to frequently used functions.

- Control device issues need to be addressed to ensure data entry methods are improved. This could include the re-location of the CDUs, and the introduction of more intuitive QWERTY style keyboards. It is likely that flight deck space will be the limiting factor for this recommendation unless radical flight deck changes are implemented.
- It was noted in this research that CDUs frequently refuse to accept incorrectly formatted entries, but do not provide feedback as to why an entry was rejected. Telling the pilot that an "Invalid Entry" has been made without providing information on how to correct the erroneous entry, wastes time and increases the pilot's workload and frustration levels. Timely feedback should also be provided to the pilots when commands are entered. It should be explicit to the user whether the FMS is processing the data entered, or whether the data entered was incorrect and was thus not processed. Prompts should be provided to assist CDU data entry, this would assist pilots in making more accurate and speedy data inputs and would aid the predictability of the FMS CDU inputs required. This would also assist users when it is necessary to perform rarely required data inputs which are more likely to be prone to slips and lapses (Billings, 1997).
- Automation levels and automation philosophies need to be considered by designers and certification authorities. The B767, for example, is less automated than the A340 or B777. Further analysis of both the automation level and flight deck philosophy is warranted when retrofitting FMS equipment onto older aircraft.
- FMS designers should assume that pilots will make routine data entry errors and should build in barriers to erroneous operation to accommodate this user characteristic (Courteney, 1999a). Error tolerance and resistance should be maximised and the FMS CDU should, where possible, monitor the pilot's actions and warn of mistakes, slips and lapses.
- The absence of an "undo" function on many FMS variants violates a very basic usability principle and should be implemented on all FMSs. It should be possible for most interactions for pilots to 'undo' the last function they executed, although careful consideration should be taken when applying this principle to interactions that involve the flight path of the aircraft.
- Related to the previous recommendation, prompts should be made available to assist pilots to perform their tasks as rapidly as possible, but also to provide barriers to erroneous data entry. It is rare for current CDUs to provide such assistance thereby increasing pilot workload for even simple inputs.

- Although present on some FMS types the provision of a 'Two-step' entry, or 'Execute' function would permit pilots to monitor and cross validate each other's inputs prior to those inputs being executed. Such a function would also give pilots an opportunity to re-check their own inputs in situations where no other cross-checking will be made. In usability terms, the ability to review and cross-check data entries is vital from both a feedback and error prevention and correction perspective. This must however be considered within the HCI principle that it is desirable to minimise number of pages that must be navigated through in order to perform critical activities.
- The operation of responses from the FMS should be predictable and intuitive. As noted by Billings (1997) "If a system is predictable and reasonably intuitive, it may not need to be particularly simple, for operators will understand and remember it. Complexity is an enemy of comprehensibility" (Billings, 1997, p.246).
- Pilots using current FMS CDUs have reported that the use of ambiguous waypoint identifiers is a problem, with one identifier being used to define more than one waypoint. This problem needs to be addressed by FMS designers and certification authorities. A convention needs to be established for the ordering of database information, to guide pilots in waypoint selection. The closest waypoint could, for example, be listed first and the waypoint name should be listed prior to its execution.
- The quality of FMS CDU displays could be improved so that they are not so difficult to see in, for example, bright sunlight. Similarly, information presentation formats could be improved to assist the assimilation of information e.g. through highlighting of important information. CDUs could also make better use of colour coding such that the pilot is assisted in integrating information from different sources.
- The HCI principle of consistency should be applied to CDU page design such that the same types of information are presented at the same location of every page.
- FAR/JAR guidelines need updating to include human factors guidelines applicable to modern flight decks. For example, some guidance could be supplied to avionics manufacturers on how acceptable levels of workload and usability should be defined.
- Some standardization across avionics manufacturers needs to be put in place to maximise compatibility when pilots undertake new type-ratings, and to reduce any negative transfer of training effects. Such standardization would also eliminate the difficulties encountered by pilots in having to use numerous FMS variants within one duty period.
- The introduction of ISO 9241 Part 11 type usability guidelines would be beneficial such that a certain level of CDU architecture and layout could be standardized across manufacturers.
- Designers should use simple and natural language that is consistent with ATC operations and navigation charts.

- Pilot, not engineering, language should be used in the design of CDUs and their supporting aids.
- The pilot's memory load should be minimised, particularly during the time critical and dynamic phases of flight, e.g. in the terminal area, and during the take-off and Descent and Approach phases of flight.
- The qualitative comments elicited in this study showed the depth and breadth of information that can be elicited from end users. The process is cost-effective and could be used in conjunction with detailed scenarios.
- FMS designers should consider ATC requirements when determining which functions need to be rapidly performed in busy flight phases.

12.18 Summary

The purpose of this third study was to perform a content analysis on all the qualitative comments generated during Studies One and Two. The study goals, as set out in Section 10.1, were to (i) analyse the problems associated with current FMSs (ii) make recommendations for improving FMSs based on participants' comments (iii) use the comments made to interpret the scales developed during Studies One and Two and (iv) provide evidence of construct validity.

Thirteen categories of problems associated with current FMSs, and 15 categories of suggested improvements were identified in this content analysis. The discussion of results put the findings in the context of current literature and illustrated how recent research supported the results of this study. Evidence of construct validity was therefore demonstrated. Specific recommendations for FMS designers and certification authorities were made based on the findings of this and the previous studies.

13 CONCLUSIONS & RECOMMENDATIONS

13.1 Conclusions

As described in Section 1.8, usability is a multidimensional concept, which has been associated with many components over the past twenty years (Lewis, 1995). Most of the early work on usability was concerned with the design of computers and software within the context of Human-Computer Interaction (HCI). Usability, however, has now been adopted in many other areas and has come to prominence over the past few years. This is largely due to the increasing number of sophisticated products, from complex Information Technology (IT) systems to everyday household items, being used by a growing user population. Booth (1989) drew attention to the fact that manufacturers had focused on continually increasing the number of functions that their products could perform, while users had become increasingly frustrated, annoyed and confused by not being able to operate products they had purchased. Poorly designed interfaces can seriously hamper a user's ability to use a system and can lead to "*confusion and frustration, difficulty in learning how to use the system, misunderstanding of what the system is doing and of what the user should do, errors, and difficulty in using the system to carry out tasks successfully*" (Ravden & Johnson, 1989, p.1). The essential point is that every user interface should meet the requirements of those who will carry out tasks with that system.

Not only has usability become crucial issue within product marketing and sales with more and more products being advertised as being "user friendly", it is also increasingly being seen as a component of industry's responsibility to its customers (Jordan, 1998). Usability has become a critical component in the design of technical equipment and designers are being forced to consider the needs and limitations of those who will use their products.

A review of the literature suggested that a gap may exist between what academia is providing and what industry finds helpful. Jordan, McClelland & Thomas (1996) suggested that a lack of understanding of industrial practices and pressures could be a contributory factor to this. The fundamental notion of usability is that a product should be easy to use. The task is to ensure that problems of usability do not arise.

In 1996 the FAA Human Factors Team noted that "*Existing methods, data, and tools are inadequate to evaluate and resolve many of the important human performance issues*" (p.3), and subsequently made the recommendation that "*new tools and methods need to be developed and existing ones improved to accompany the process improvements*" (p.4). This thesis sought to develop such a tool to assess the usability of FMS CDUs.

A review of existing usability checklists, guidelines and subjective measures revealed that most usability assessment tools had been derived for use with desktop computers. Even those usability scales that claimed to be generic were not found to be applicable to

the flight deck domain for a number of reasons. The Software Usability Measurement Inventory (SUMI), developed by Kirakowski & Corbett (1993) within the Metrics for Usability Standards in Computing programme, for example, demonstrated good psychometric properties but was developed using office-based applications such as word-processing and spreadsheet packages. Of the fifty-items in the questionnaire numerous were not directly applicable to the FMS or flight deck (see Section 1.11). It was evident that in order to assess the usability of FMS CDUs, a new series of scales was required.

Both the European Joint Airworthiness Authorities (JAA) and the US Federal Aviation Administration (FAA) have proposed the introduction of Human Factors Certification Criteria for civil aircraft flight decks. While the regulatory authorities have supported the philosophy behind the human factors certification process, no tools or methods have been proposed as techniques for evaluating the flight deck from a human factors perspective. This series of studies produced a series of reliable and valid scales for assessing the usability of FMS CDUs. The specific conclusions from each study are outlined in the following subsections.

13.1.1 Study One Conclusions

The first study set out to define which usability components or factors underpinned FMS CDU usability. A series of in-depth semi-structured interviews with airline pilots demonstrated to the researcher that a number of the usability components discussed in the literature were cited repeatedly as being critical to FMS usability. Following, comprehensive pre-piloting and piloting, a 49-item self-completion questionnaire was developed which could be distributed to an international sample of airline pilots. A total of 252 completed questionnaires were received from respondents working for five airlines across the UK, USA and Canada. The respondents were type-rated on aircraft produced by three of the largest aircraft manufacturers - Boeing, Airbus and Fokker.

A factor is said to have content validity when the items it contains are a randomly chosen subset of the universe of possible appropriate items. A rigorous methodology was used throughout the scale development process to ensure that the factor structure demonstrated content validity. The questionnaire items were developed through an iterative process of reviewing relevant academic literature, talking to Subject Matter Experts (SMEs), observing the pilots interacting with the FMS, and pre-testing and piloting the questionnaire items extensively in both North America & the UK. This ensured that the items elicited were appropriate and looked plausible to SMEs. Since the factors developed should have content validity, they almost certainly should have face validity (Cook, 1998).

Using a Principal Components Analysis, it was shown that FMS CDU usability comprised ten individual factors. The factors derived were concerned with Structure & Sequence, Appropriate Functionality, Feedback, Predictability, Information Presentation, Consistency, Flexibility & Control, Error Prevention & Correction and Processor Speed. In order to examine the internal consistency of the factors, a series of Cronbach's Alphas were conducted. It was concluded from these analyses that the

internal consistency within the factors was satisfactory for the scale development process to continue.

With the goal of demonstrating discriminant validity, a series of Discriminant Function Analyses (DFA) were conducted. DFA analyses permitted the researcher to examine if participants' ratings of the items, from which the usability factors were derived, would discriminate between aircraft and FMS types. Since the usability of the FMS CDU is intrinsically linked to the aircraft in which it resides, the analyses were designed to examine if the factors elicited could discriminate between aircraft and FMS manufacturers when these were examined together and independently. It was hypothesised that if the factors could discriminate between these different aircraft and FMS CDUs, then evidence of discriminant validity would be demonstrated.

It was further hypothesised that respondents' personal attributes should not significantly affect the factors derived since usability should be a product of the FMS CDU and its operating context not culture or personal experience. The ratings of the usability factors should therefore not discriminate significantly between the (i) pilots' culture (characterised by the country in which the pilot was based) or (ii) pilots' rank (First Officer or Captain).

The series of Discriminant Function Analyses demonstrated that ratings of the usability factors derived discriminated between aircraft manufacturers and FMS manufacturers, but did not discriminate between personal factors such as the pilot's rank or culture as hypothesised. It was therefore concluded that evidence of discriminant validity existed since the usability factors discriminated between that which they should discriminate, but did not discriminate between the personal attributes such as rank which should not influence FMS usability. Statistical evidence of criterion validity was demonstrated by assessing the predictive power of the usability factor structure elicited and showing that respondents could be classified by their type-ratings at a rate significantly better than chance. From a practical perspective, however, no evidence of criterion validity was demonstrated. It was not possible to predict participants' current aircraft or FMS type accurately from their ratings of questionnaire items. The evidence indicated that Press's Q statistic was very sensitive and that no practical evidence of criterion validity was demonstrated in this study.

Construct validity is the extent to which an assessment is measuring something meaningful. Guion (1977) argued that since test creators always have a theory or construct that they are trying to assess, all test validation is necessarily construct validation. Murphy and Davidshofer (1991) defined the goal of construct validation as being "*to determine whether test scores provide a good measure of a specific construct*" (p.116). In this study, therefore, it referred to the extent to which the theoretical construct of usability was being assessed. Ideally, evidence of construct validity would be achieved by comparing the usability factors developed in this study with existing measures of FMS CDU usability. Since no objective measure of FMS CDU usability existed, construct validity was difficult to establish. Evidence of construct validity was, however, derived from a content analysis conducted on participants' comments. The findings of the content analysis demonstrated the theoretical meaningfulness of the usability scales produced since the comments made by participants reflected the same

issues as the scales were measuring. Further evidence of construct validity was derived from the fact that the factors did not discriminate between pilots' usability ratings on the basis of their rank or culture. Both rank and culture are known to influence pilot attitudes to automation (see Section 2.2.4). Evidence in this study therefore indicated that the factors elicited were assessing usability and not attitudes to FMSs. Evidence toward construct validity was therefore demonstrated.

By examining the components derived in the context of current aviation automation and usability literature, it was possible to conclude that the factors elicited were generalizable and could be validated within the context of existing literature.

The critical conclusions from the first study were that:

- A factor structure of FMS CDU usability was elicited which looked plausible to Subject Matter Experts (SMEs), evidence of face validity was therefore derived (Cook, 1998).
- A rigorous process of questionnaire development through interviewing, pre-piloting, piloting with SMEs in both the UK and North America should have ensured content validity (Anastasi, 1988).
- The factor structure was appropriate for usability ratings made by pilots from both the UK and North America and was therefore independent of culture.
- The factor structure was appropriate for usability ratings made by both First Officers and Captains and was thus independent of pilot rank.
- The factor structure demonstrated an acceptable level of internal consistency for scale development (Hair et al, 1998).
- Although ratings of the questionnaire allowed statistically significant predictions to be made of the aircraft and FMS on which respondents were currently type-rated, no practical evidence of criterion validity was demonstrated. It was not possible to predict respondents' current aircraft and FMS type accurately from their ratings of the questionnaire items.
- Evidence toward construct validity was derived from the aviation automation and usability literature and through comparing the factors elicited with the categories of data derived from the content analysis conducted in Study Three. Ratings of the questionnaires were not dependent on either pilot culture or rank. Evidence therefore also suggested that it was usability, and not for example attitudes to automation, that was being measured. Further evidence of construct validity was therefore derived.
- The factor structure demonstrated good discriminant validity in that it discriminated between that which it should discriminate between, but did not discriminate between variables that it should not discriminate.

Table 13-1 shows a summary of the psychometric properties of the usability factors derived.

Table 13-1 - Psychometric Properties of Usability Factors

Property	
Usability Factor structure was internally consistent.	✓
Usability Factor structure demonstrated evidence of face validity.	✓
Usability Factor structure demonstrated evidence of content validity.	✓
Usability Factor structure demonstrated evidence of criterion validity.	X
Usability Factor structure demonstrated evidence of discriminant validity.	✓
Usability Factor structure demonstrated evidence toward construct validity.	✓

13.1.2 Study Two Conclusions

The purpose of Study Two was to verify the factor structure produced during Study One and to ensure it was stable and reliable when applied to new independent data. A reliable and valid factor structure would permit the researcher to produce a series of reliable and valid scales from that factor structure fulfilling the ultimate goal of this thesis.

It was demonstrated that not only did the usability factor structure underlie the new data set when it was examined in its entirety, but it also underpinned the data from both the Cruise and Descent & Approach flight phases when the data were split into two based on those flight phases. A series of Confirmatory Factor Analyses produced a more parsimonious model of FMS CDU usability. The more parsimonious solution of FMS CDU usability was constituted of the following components: Structure & Sequence, Appropriate Functionality, Feedback, Information Presentation, Error Prevention & Correction, Processor Speed, Workload, Consistency and Flexibility for Navigation. *Since the factor structure was shown to be replicable and stable, usability scales were produced with confidence.*

In order to determine if the scales produced still had discriminant capabilities, and to demonstrate evidence of discriminant validity, a further series of DFAs were conducted using the newly produced scales. It was demonstrated that the scales developed in this study still had discriminant capabilities and the scales showed the same general pattern of results in discriminating between FMS manufacturers and aircraft manufacturers as the factors had done in the previous study. It was therefore concluded that minimal discriminant capability was lost in the conversion of usability *factors* to usability *scales*. A series of Cronbach's Alphas analyses demonstrated that the scales were internally consistent.

This research did not intend to demonstrate that any one FMS or fleet type was more usable than another, however it was noted in the course of the research that different FMSs and aircraft achieved more favourable ratings on particular usability scales. The pattern of scale ratings was different according to FMS and aircraft types but the pattern

was consistent across both studies. Further evidence of the reliability and stability of the findings was therefore provided.

In an attempt to demonstrate criterion validity the predictive power of the usability scales was assessed (Ghiselli et al, 1981). The classification matrices (based on pilots' type-ratings) were compared with a chance model using Press's Q statistic and were found to be statistically significant ($p<0.01$). However, no practical evidence of criterion validity was demonstrated since the scales were better at predicting some types of aircraft and FMS than other types. It was not possible to predict participants' current aircraft or FMS type accurately from their ratings of the usability scales. Some of the classification problems associated with the DFAs could have been due to the weighting in the analyses being biased towards certain aircraft and FMS types. The majority of participants in this research were type-rated on Boeing aircraft and Honeywell FMSs. It is therefore possible that the weighted linear combinations from the DFAs and the associated classification matrices could have reflected this bias. In the DFA classification matrices, Airbus and Fokker aircraft were, for example, frequently incorrectly classified as being Boeing aircraft but this may have been at least partially due to the fact that the sample was biased towards Boeing aircraft. Of the pilots who participated in Study Two, 88.5% were type-rated on Boeing aircraft. The classification problems could also have been due to the sensitivity of the Press's Q statistic (Hair et al, 1998).

The analyses demonstrated that it was not possible to discriminate between flight phases on the basis of the pilots' ratings of the usability scales. Ratings of the usability scales were not a function of the flight phase that was being rated by participants, but rather were solely a rating of usability in terms of the equipment being used and the operational environment in which that equipment resided. The scales were not, for example, measuring cognitive workload alone which is known to be higher in the Descent & Approach phase of flight than in the Cruise phase of flight (Sarter, Woods, Billings, 1997). Although the scales developed in this research demonstrated some good psychometric properties, there are some limitations associated with a scale-based approach to usability. These limitations were outlined in Section 9.6.

The final analysis conducted during Study Two was a multiple regression, from which it was concluded that participants' ratings of the Structure & Sequence, Appropriate Functionality, Predictability, Workload and Flexibility for Navigation usability scales were all significant predictors of satisfaction in using the FMS.

The critical conclusions from the second study were that:

- The factor structure elicited in Study One was stable and replicable. It was not a statistical artefact, but was real.
- The factor structure elicited in Study One was not dependent on flight phase, the factor structure held true for both the Cruise phase of flight and the Descent & Approach phase of flight.
- On the basis of the evidence derived, it was feasible to develop usability scales.

- Extensive consultation with SMEs, and pre-piloting and piloting the questionnaires with line pilots during the scale development process ensured that the scales produced had face validity. The scales appeared superficially to measure that which they claimed to measure (Anastasi, 1988).
- Using Cronbach's Alpha (1951), it was demonstrated that the scales developed were internally consistent.
- Using appropriate scale development procedures (e.g. extensive pre-piloting and piloting of questionnaires), such as those outlined by DeVellis (1991) ensured that the scale items were a randomly chosen subset of the universe of possible appropriate items (Anastasi, 1988). Evidence of content validity was therefore demonstrated.
- The scales demonstrated discriminant validity since they discriminated between that which they should theoretically discriminate (e.g. FMS types), but did not discriminate between that which they should not theoretically discriminate (e.g. pilots' culture).
- Although ratings of the usability scales allowed statistically significant predictions to be made of the aircraft and FMS on which respondents were currently type-rated, no practical evidence of criterion validity was demonstrated. It was not possible to predict respondents' current aircraft and FMS type accurately from their ratings of the usability scales.
- Evidence of discriminant validity consistent with that elicited during Study One was demonstrated in Study Two using new and independent data (Hair et al, 1998). Since the analyses showed the same general pattern of results in discriminating between FMS and aircraft manufacturers in both Studies One and Two, it was demonstrated that little discriminant capability had been lost in the conversion of factors into scales. Evidence for the generalizability and stability of the scales produced was therefore demonstrated.
- The usability ratings reflected the equipment (FMS) and its operational environment (the flight deck), and not user characteristics. It was therefore FMS usability and NOT workload, attitudes to automation or phase of flight that was being measured. Evidence of construct validity was therefore demonstrated (Anastasi, 1988).
- Finally, ratings of the Structure & Sequence, Appropriate Functionality, Predictability, Workload and Flexibility for Navigation usability scales were all significant predictors of self-reported satisfaction in using the FMS.

Table 13-2 shows a summary of the psychometric properties associated with the usability scales derived.

Table 13-2 - Psychometric Properties of Usability Scales

Property	
Usability Factor structure was stable and replicable.	✓
Usability Factor Structure could be generalised across flight phase (Cruise and Descent & Approach).	✓
Usability Scales were internally consistent.	✓
Usability Scales demonstrated evidence of face validity.	✓
Usability Scales demonstrated evidence of content validity.	✓
Usability Scales demonstrated evidence towards construct validity.	✓
Usability Scales demonstrated some evidence of discriminant validity.	✓
Usability Scales demonstrated evidence of criterion validity.	X
Usability Scale ratings reflected FMS usability and not user characteristics.	✓

13.1.3 Study Three Conclusions

The content analysis conducted within Study Three elicited categories of problems associated with current Flight Management Systems, as well as categories of improvements respondents reported wishing to see to future FMSs. The issues raised by participants in this study further aided the researcher in interpreting the scales developed during Studies One and Two. In Study Two it was demonstrated that it was possible to discriminate between FMS and aircraft manufactures on the basis of ratings on the usability scales. In Study Three, comments made by participants allowed ratings of the scales to be interpreted and provided evidence of construct validity for the scales produced. The content analysis further suggested where the usability problems were and related them to specific aspects of FMSs. It was concluded that problems associated with current FMSs and suggested improvements for future FMSs were associated with a number of areas as described in the following subsections. The critical conclusions from the third study are summarised at the end of this section.

- Database / Memory Size

Database / Memory Size problems and suggested improvements were reflected in the Appropriate Functionality scale which focused on the FMS allowing users to perform tasks as efficiently as possible, and giving users access to all the information they required. From a user's perspective, a small FMS database may mean that routes have to be constructed manually thereby increasing workload and the potential for data entry errors. Both the **Appropriate Functionality** factor produced in Study One and the **Appropriate Functionality** scale produced in Study Two discriminated between FMS manufacturers. Smiths FMSs were rated more favourably than Honeywell FMSs in both studies. Consistent with the pattern of scale ratings, Honeywell type-rated pilots made almost all of the comments related to database problems and suggested improvements. The studies therefore suggested that it was possible to discriminate between FMSs on the basis of ratings on the Appropriate Functionality scale, and evidence for the discriminant validity shown in scale ratings was demonstrated by the comments made. Evidence of construct validity was therefore also demonstrated.

- Processor Speed

Comments related to FMS processor speeds being too slow or needing to be faster were made by participants from every fleet involved in the study except the F-100. **Processor speed** was also one of the usability scales developed during Studies One and Two and included the items "The FMS processes information for other activities in the flight deck" and "It is irritating to have to input information at a rate slow enough for the FMS processor".

In both Studies One and Two, ratings of the **Processor Speed** scale discriminated between participants' on the basis of the aircraft / FMS on which they were currently type-rated. Ratings of the Processor Speed scale also discriminated between aircraft types, when the impact of participants' current aircraft type was examined on its own. Airbus aircraft with Honeywell FMSs, were rated more favourably than Boeing aircraft with Smiths FMSs, which in turn were rated more favourably than Boeing aircraft with Honeywell FMSs. These findings were consistent across Studies One and Two. Similarly, the number of processor speed related comments made by aircraft type supported the scale ratings in that Airbus type-rated pilots made the least number of problem comments related to processor speed. Construct validity evidence for the **Processor Speed** scale was therefore demonstrated and it was suggested that the pattern of comments made reflected the differences in usability scale ratings by aircraft/FMS manufacturers described in Studies One and Two.

- Feedback / Error Prevention & Correction

Ratings of the **Feedback** scale developed during Studies One and Two demonstrated that Honeywell FMSs were rated more favourably than Smiths FMSs. It was also interesting to note that three of the nine respondents that made comments concerning improvements sought to Feedback were type-rated on Smiths FMSs. This was a large proportion of participants considering the sample distribution by fleet. The content analysis demonstrated that comments concerning FMS Feedback / Error Prevention & Correction problems were made by participants from every fleet included in this study. Due to the distribution of comments across aircraft / FMS types, it was not possible to draw definitive conclusions based on fleet or FMS types. Examining the absolute scale scores, it was noted that ratings of **Feedback** were below the rating midpoint indicating poor usability across all aircraft and FMS types.

In Study Two, it was demonstrated that ratings of the **Error Prevention & Correction** scale could be used to discriminate between the combined impact of FMS and aircraft manufacturers. Airbus aircraft with Honeywell FMSs were rated most favourably, followed by Boeing aircraft with Smiths FMSs then Boeing aircraft with Honeywell FMSs. Ratings on the **Error Prevention & Correction** scale also discriminated between aircraft manufacturers when they were examined independent of FMS manufacturer. Airbus type-rated pilots again made more favourable usability ratings than Boeing type-rated pilots did. Although, Airbus aircraft with Honeywell FMSs were rated most favourably on the **Error Prevention & Correction** scale, a relatively large proportion of Airbus type-rated pilots made related suggested improvement comments.

Although, it was not possible to draw any definitive conclusions based on FMS / aircraft types, it is suggested that **Feedback / Error Prevention & Correction** is a problem across all FMS / fleet types, even for the aircraft / FMS combination which achieved the most positive rating on the applicable usability scales.

- Display Quality & Information Presentation

Boeing-737 type-rated pilots who use Smiths FMSs made a large number of comments related to display quality and information presentation when the B-737 participant sample was considered as a proportion of the total pilot sample. Information Presentation was also one of the scales developed to assess the usability of FMS CDUs during Studies One and Two. The items that loaded onto this scale largely reflected the issues raised in this qualitative study, for example, "Important information is adequately highlighted on the screen". Looking back to Study One, the Information Presentation factor successfully discriminated between the combined effects of aircraft and FMS manufacturer. Consistent with the findings of the content analysis, Boeing aircraft with Smiths FMSs achieved the least favourable ratings, while Fokker aircraft with Honeywell FMSs achieved the most favourable ratings. In Study Two, however, Boeing aircraft with Honeywell FMSs achieved less favourable ratings than those achieved by Boeing aircraft with Smiths FMSs. In Study Two, ratings of the **Information Presentation** scale also permitted successful discrimination between aircraft manufacturers with Airbus type-rated pilots making more favourable **Information Presentation** scale ratings than Boeing type-rated pilots. Evidence in support of this finding has been found in this qualitative study with one pilot, for example, making the comment "Comparing Airbus and Boeing – the use of colour (in Airbus) is good and assists in speed of data entry, especially in emergency/abnormal situations" (see Appendix M). The content analysis provided construct validity evidence for the Information Presentation scale produced in Study Two, and further provided guidance to aid the researcher in interpreting scale ratings.

- Consistency within & Compatibility between FMSs

Consistency and Compatibility are amongst the most basic usability principles (Nielsen, 1993). The consistency and compatibility issues raised in the content analysis were focused on the lack of consistency within single FMSs, and the lack of compatibility across FMSs in a single fleet, and across FMS types when pilots undergo a new type-rating onto a new aircraft and FMS.

The **Consistency** scale developed in Studies One and Two did not discriminate between FMS usability ratings by either FMS or aircraft manufacturer. However, the **Consistency** scale scores from Study Two showed that Airbus aircraft with Honeywell FMSs achieved higher usability ratings than Boeing aircraft with Honeywell FMSs or Boeing aircraft with Smiths FMSs (see Table 8-62). Consistent with the findings of Study One, Honeywell FMSs were rated as being more usable than Smiths FMSs on the **Consistency** scale in Study Two (see Table 8-68). These results must be interpreted with caution, however, since there were more Honeywell FMSs than Smiths FMSs in the sample. Similarly Airbus aircraft achieved more favourable Consistency scale scores than Boeing aircraft in both studies One and Two (see Tables 4-28 and 8-74). It

may be that it was not possible to discriminate between aircraft and FMS types on the basis of ratings on the **Consistency** scale because consistency is a problem across FMS manufacturers and fleet types.

Funk et al (1996) raised the lack of standardization within and between fleets as a concern. Even though the FMSs within a single fleet may be of the same model, differences in, for example, response speeds have been noted by participants. Many of the differences that occur within a single fleet are a product of aircraft being bought from different sources. This is almost inevitable unless the whole fleet is bought direct from the manufacturer when new. The technology fitted within a single fleet may have initially been specified by many different airlines when they were initially purchased. As aircraft are sold on to other airlines, a single airline may end up with a fleet containing aircraft from numerous different sources.

- Functional Limitations / Additional Functionality

The Functional Limitations / Additional Functionality was not something that could be related back to any of the scales produced during Studies One and Two but was rather a description list of the functional limitations associated with current FMSs as well as a wish-list of functionality for inclusion in future FMSs. It was concluded that future FMS designers seeking ideas from "operational" airline pilots could usefully consider the functional limitations, and additional functionality comments derived in this study.

- Workload

The types of workload comments made by participants referred to workload being busy in the terminal area as well as being busy just at the wrong time. Increased workload is a consequence of poor HCI design and usability, rather than a dimension of usability or human-machine interfaces (Williges, Williges & Fainter, 1988). It has been well documented that FMSs have redistributed workload, increasing it where it was already high (Wiener, 1989; Foreman, 1996). The workload scales developed in Studies One and Two did not discriminate between either aircraft or FMS manufacturer. It was further not possible to draw specific conclusions on the workload associated with particular FMS or aircraft variants from the qualitative study. This may, however, be due to the fact that the workload issues associated with FMS CDUs are applicable across all aircraft and FMS variants (FAA, 1996). Evidence from the literature (e.g. FAA, 1996), and from the poor mean workload scale scores made by participants, indicate that the workload issues associated with FMS CDUs are applicable across all aircraft and FMS variants.

- Control Devices

The Control Devices category in this study was used to group the comments made by participants related to the physical concerns related to the CDU control (input) devices. The specific issues raised related to keyboard layout, keys used to scroll through menus, the requirement for other input devices such as touchscreen and rollerball controls, and the physical discomfort associated with the location from which the FMS is used. Many of the comments referred to the desire to have a QWERTY keyboard rather than an alphabetical one. As one Airbus type-rated respondent phrased it "Buttons seem to be randomly positioned". Since Control Devices was not one of the scales developed in

Studies One and Two, it was not possible to make direct comparisons with those studies.

Respondents from every participant fleet raised control device problem issues, but the distribution of comments across fleets/FMSs meant that no definitive conclusions based on fleet could be drawn. This could, again, be due to the fact that control device issues are pertinent to every FMS/aircraft type.

- CDU Architecture & Logic

The comments made by participants were largely concerned with the way in which the data within the CDU were structured and the logic that the designers had applied to sequencing pages and menus. Difficulties in accessing information were also noted. Similarly, the suggested improvement comments referred to the requirement for a simplified logic, and ease of access to particular information in a particular set of circumstances. On reviewing the comments made, it became evident that the issues raised within this category largely reflected those of the Structure & Sequence scale developed during Studies One and Two. Items loading onto this scale included "The structure of the FMS is well organised from the user's point of view" and "There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts". The distribution of comments across fleet groups meant that no definitive conclusion could be drawn on fleet effects. It was evident, however, from the scales derived in Studies One and Two that Boeing aircraft achieved more favourable ratings on the Structure and Sequence usability scale than either Airbus or Fokker aircraft.

- Accuracy of Parameters

As with any technology the FMS has some performance limitations, and the limitations discussed within this section are concerned with inaccuracies associated with some FMS parameters. The comments made by participants concerning the accuracy of parameters were quite specific to particular FMS and aircraft types. Almost nine percent of the problem comments concerned accuracy of parameters and it was interesting to note that none of these comments were made by pilots type-rated on Airbus or B-747 aircraft. However, it is not possible, within the constraints of this study to draw definitive conclusions concerning the Accuracy of Parameters associated with the FMS on board these aircraft.

Since Accuracy of Parameters is not a usability dimension but rather reflected functional limitations of the FMS, it was not possible to draw parallels between the scales derived and the qualitative data derived concerning accuracy of parameters. However, it is suggested from the findings of this study that inaccuracies do occur within the FMS when predicting parameters such as wind, fuel, and ETA and the associated implication is that the predicted flight profiles can be adversely affected and can be misleading to pilots. The implication from the 'suggested improvements' comments made by participants for future FMSs is that increased accuracy in terms of parameter predictions, and improved pilot awareness of the information used in the calculation of these parameters would reduce pilot workload and therefore potentially improve usability.

- User Guidance and Support

The User Guidance & Support comments made by participants suggested that FMSs offer little or no on-line help. Manuals were reported to be "woefully" inept and difficult to navigate through. It was also asserted that little was published on the working of the FMS software, and that the FMS manuals were written for engineers rather than the pilots who have to use them. The improvement comments reflected the same issues.

Since User Guidance and Support was not one of the usability scales developed during Studies One and Two, it is not possible to draw direct parallels between the comments made in Study Three and the scales developed in Studies One and Two. However, based on the qualitative evidence derived from this study it is suggested that many improvements could be made to the user guidance and support documentation available to pilots. Particular attention is merited to this in light of the difficulties associated with navigating through databases which have many layers, much complex functionality which is used infrequently and when it is used it may be in the context of abnormal or emergency situations (e.g. single engine operations). Further, in light of the resistance to change by the aircraft manufacturers and the airlines themselves, this is one dimension that could be easily remedied. The findings of this study are consistent with the findings of research conducted by Funk, Lyall and Riley (1996) who also reported that the support documentation for FMSs was inadequate.

- Intuitiveness

It was concluded that a lack of intuitiveness was a concern for participants using current generation FMSs. Intuitiveness was also an area proposed for suggested improvements. Although the term 'Intuitiveness' is a not a component of usability frequently discussed in the research literature, it is in itself a product of usable interfaces and is related to learnability (see Figure 1-8). Components of usability such as feedback make it explicit to the user how interaction with the interface should be performed. The findings of this study reflect Billings (1997) comment which stated that "*many pilots now flying FMS equipped aircraft would welcome a simpler, more intuitive Flight Management System with which they could interact more easily than is possible with today's CDUs*" (p.143).

- Training

Training was not one of the scales developed during Studies One and Two, since training is traditionally beyond the scope of the usability concept. Training is, however, related to usability in that training is often used as an inadequate fix for systems that demonstrate poor usability. The training related comments made by participants were concerned with the lack of hands-on FMS training which is currently received by pilots in many airlines. It was interesting to note that pilots type-rated on B-737 aircraft with Smiths FMSs made five of the six comments. This may well be related to the fact that the B-737 is the only aircraft to be fitted with Smiths FMSs as standard. Therefore, a pilot who has been type-rated on any other type of FMS-equipped aircraft will come face-to-face with the problems caused by a lack of compatibility and consistency across FMS manufacturers.

- Other suggested improvements

Finally, it was concluded that a number of other categories of suggested improvements were elicited which did not map onto comments associated with current FMSs nor to the scales developed during Studies One and Two. The conclusions related to these other suggested improvements are described in the following sub-sections.

Improved "What-If?" or hypothetical capability

The content analysis suggested that a number of participants wished to make hypothetical inputs to their FMS without those inputs being executed. Although B-757/767 type-rated pilots made four of the comments, the small sample size meant that it was not possible to draw any definitive conclusions based on fleet or FMS. It is true, however, that the B-737 Smiths FMS does have an execute button allowing inputs to be made without those inputs being processed. The absence of an execute button on the B-757/767 Honeywell FMS may be reflected in the fact that the majority of Improved "What-if?" or hypothetical capability comments were made by B-757/767 type-rated pilots. It was concluded that the potential to have extended hypothetical capability should be considered by future FMS manufacturers.

Improved Barriers to Erroneous Operation

A number of participant comments referred directly to the requirement for improved barriers to erroneous operation. The types of barriers to erroneous operation raised by the participants included the questioning of waypoints that were not part of the approach. The small number of comments allocated to this category meant that it was not possible to draw any definitive conclusions based on fleet or FMS type. Since the FMS is a highly complex system it is highly unlikely that designers could ever conceive of a method of trapping every potential error that pilots might make. It is therefore necessary to provide an alternative method of detecting such errors in the form of barriers to erroneous operation. Since many inputs made to the FMS CDU by one pilot are not cross-checked by the other pilot, it is argued that barriers to erroneous operation should be built into the CDU design to minimise the number of errors pilots make.

On the basis of participants' comments and a review of the aviation automation literature (e.g. Dekker & Orasanu, 1999), it was concluded that FMS designers should consider improving FMS Barriers to Erroneous Operation. Even a basic change, such as the listing of possible beacons listed in order of distance from current location, could be sufficient to draw the pilot's attention to the erroneous data entered into the CDU.

Standardization of Language used to program the FMS & to co-ordinate with ATC commands.

The concerns raised in this study were focused on the mismatch that exists between the language and codes used by Air Traffic Control Officers (ATCOs) and that used by pilots to enter information into the FMS CDU. Standard ICAO phraseology does exist and should help the potential for miscommunications (FAA, 1996). It is true, however, that existing ICAO standards are outdated for *"many new types of operation now occurring in the aviation environment including FMS approaches"* (FAA, 1996, p.105). Inconsistencies also exist between paper based charts and FMS databases such that fix

and bearing names do not correspond between the FMS and chart identifiers, and clearly some standard between the information sources could ease the pilot's workload in the busier phases of flight. The findings of this research indicated that now might be a good time to update such ICAO standards particularly in light of anticipated ATC changes in the near future.

The critical conclusions from the third study were that:

- The comments made by participants and the content analysis conducted on those comments aided the researcher in interpreting the scales developed, and provided evidence of construct validity by reflecting the content of the scales developed in Studies One and Two (Anastasi, 1988).
- On the basis of the comments made, and the content analysis conducted on those comments, it was possible suggest explanations for differences in usability scale scores achieved during Study Two.
- It was noted that many participants found the processor speed of their FMS to be too slow. HCI theory would indicate that this is likely to be a source of frustration for users (Nielsen, 1993). The comments made by participants reflected this frustration.
- Participants reported that CDU architectures were overly complex, and also noted that the CDU logic was not obvious to the user.
- Many participants reported that they wished to see extra functionality added to their FMS CDU.
- Participants from every included fleet reported the existence of usability problems with the control devices used to interact with the FMS, and made associated suggested improvements.
- Analysis of comments concerned with Feedback, Error Prevention & Correction found that these dimensions of usability were inadequate. Feedback should be more informative, while improved error prevention & correction mechanisms should be provided.
- The content analysis indicated that pilots had difficulty physically seeing the FMS CDU display due to both its location and screen quality. It was also reported that the FMS CDU screens were too cluttered, the text sizes were too small, and information was not adequately highlighted. Participants also made associated suggested improvements.
- Some participants noted that the FMS database / memory size was too small, and therefore created extra workload for pilots since data such as waypoint information had to be inserted manually.
- Some participants also noted that User Guidance and Support was deficient, both in terms of the absence of on-line help and the poor quality of reference manuals.
- The lack of consistency in FMSs within a single fleet, and the lack of compatibility in FMSs across fleets and FMS manufacturers was cited as a usability problem by some participants. Participants also made the associated suggested improvement that standardization across FMS variants would be desirable.

- Some participants noted that FMS training is inadequate and training programmes should be improved to include more CBT-based training. Although training is not strictly a usability issue, the requirement for increased user training can be a consequence of poor interface design and usability (Nielsen, 1993).
- It was also noted by some participants that FMS CDUs are not as intuitive as they might be. Participants consequently made "suggested improvement" comments that indicated that FMS CDUs should be made more intuitive.
- Consistent with recent literature (e.g. Singer, 1999), the results of the content analysis indicated that interactions with the FMS can increase workload during phases of flight which are already high workload.
- It was reported that in some FMSs, it was not possible to generate "What if?" scenarios, without executing them. This is problematical from a usability perspective in that one pilot cannot verify the other pilot's inputs to the CDU. Some pilots also suggested that the inclusion of this hypothetical capability would be an improvement since the effects of parameter changes such as wind could be examined without being executed.
- Participants recommended that FMS CDUs should include improved barriers to erroneous operation, since at least some operational Flight Management Systems do not even question the inclusion of a waypoint that is not part of the descent & approach procedure.
- The final category of the content analysis noted that a mismatch existed between the language or codes used by Air Traffic Control Officers and those used by pilots to enter data into the FMS CDU. It was suggested that improved standardization of phraseology would aid pilots making FMS data inputs, particularly in busy phases of flight.

13.2 Overall Conclusions

The first study in this thesis resulted in the production of a series of factors, from which scales to assess the usability of FMS CDUs were developed. The factors were shown to be internally consistent and evidence of face validity, construct validity and content validity was demonstrated. As hypothesised, the factors derived from participants' questionnaire ratings, discriminated between aircraft manufacturers and FMS manufacturers but did not discriminate between participants' based on their rank or culture. Evidence of discriminant validity was therefore also demonstrated. While participants' type-ratings (in terms of aircraft and FMS type) could be predicted at a rate statistically better than chance ($p<0.01$), no practical evidence of criterion validity was demonstrated since the factors were better at predicting some FMS and aircraft types than others. It was not possible to predict respondents' current aircraft and FMS type accurately, from their ratings of the usability questionnaire.

In the second study it was demonstrated that the usability factors derived were consistent and stable. It was demonstrated that the underlying factor structure was valid for both the Cruise and Descent and Approach flight phases. Internally consistent usability scales were produced and evidence of face and content validity was demonstrated. While participants' type-ratings (in terms of aircraft and FMS type) could be predicted from scale ratings at a rate statistically better than chance ($p<0.01$), no practical evidence of criterion validity was demonstrated. It was again not possible to predict respondents' current aircraft and FMS type accurately, from their ratings of the usability scales.

Evidence of discriminant validity was also demonstrated since the usability scale ratings differed significantly by FMS and aircraft Type, but did not differ significantly by personal factors such as participant rank. The rank ordering of FMS and aircraft manufacturers based on the usability scale ratings were largely consistent with the results obtained using the factors in Study One. This demonstrated that the scales could discriminate between FMS and aircraft manufacturers in a consistent pattern, but also showed that little discriminating capability was lost in the conversion of factors into the simplified scales used in the second study. Analyses also demonstrated that ratings of usability were not significantly different based on the flight phase the scales were applied to. This demonstrated that FMS CDU usability was a product of the FMS and its operating platform (the aircraft).

In the final study, a content analysis conducted on the qualitative comments generated during Studies One and Two elicited categories of problems associated with current FMSs and suggested improvements for future FMSs. The categories elicited provided construct validity evidence for the scales derived, and allowed differences in scale ratings in Studies One and Two to be explained. It was also possible to make usability recommendations for FMS manufacturers and certification authorities on the basis of the study's findings.

This thesis therefore largely achieved its aim of producing a series of reliable and valid scales for assessing FMS usability, although the scales were not shown to have criterion validity in any practical sense.

13.3 Research Contributions

The FAA reported in 1996 that in the previous twenty years the number of world-wide commercial aircraft departures had risen from eight million departures per year, to fifteen million departures per year. Accident rates have remained constant, with two or three accidents occurring per million departures. However, the predicted continued growth in air travel will mean that more accidents will occur each year unless accident rates are reduced. It is currently estimated that between 65% and 80% of civil air transport accidents are attributable in whole or in part to human error. In order to reduce human error, this research sought to contribute to current knowledge by examining the usability issues associated with the flight deck's core computer, the FMS.

The FAA document, *The Interfaces Between Flightcrews and Modern Flight Deck Systems*, published in 1996 noted that significant deficiencies existed in the aviation system including insufficient criteria, methods, and tools for design, training and evaluation to aid the goal of achieving more human-centred aviation automation. FMSs have been criticised for not incorporating good human factors considerations and for being "*difficult to use, vulnerable to crew error, creating high workload peaks, and presenting inadequate feedback*" (Courteney, 1999, p.112). Many authors including Billings (1997) and Singer (1999) highlighted the requirement for improvements to the usability of current generation FMS CDUs. The outlined requirements provided the justification for the research studies presented within this thesis. This research has contributed to filling a void in the aviation automation and usability research literature by developing a series of reliable and valid scales to assess the usability of FMS CDUs. The product of this research programme has been a series of reliable and valid scales for assessing the usability of FMS CDUs.

Within the usability domain, this thesis concluded that existing usability measures such as the Software Usability Measurement Inventory (SUMI) (Kirakowski & Corbett, 1993) were not applicable to the flight deck. This was due to the fact that such tools were developed and validated using desktop applications such as word-processing and spreadsheet packages which have interaction mechanisms, logic, and displays, which are very distinct from those pertinent to flight deck automation. This research, is thus novel in providing a tool which is specifically designed to assess the usability of a flight deck component.

Many designers and human factors experts would welcome an FMS interface re-design. Certainly both the avionics and human factors knowledge exists to create an improved interface. The limiting factor remains that today's FMSs perform a function and to replace them would be extremely expensive. In the commercial airline market, requirements such as the desire for common type-ratings encourage manufacturers to adopt a philosophy of "*change as little as possible*" (Rogers & Schutte, 1997, p.860).

Moderate changes to current aircraft offer a number of advantages over radically altered new designs in terms of reducing design risk, reducing design costs, minimising changes to meet requirements, maintenance of common type-ratings and reducing both training costs and certification risks. Any significant re-design will require a very significant quantity of training for pilots, potentially sufficient to deter many airlines from purchasing the new technology. Running along side the return-on-investment issue is the safety issue that is inherent in the re-training of pilots familiar with old technology and introducing them to something radically different. Billings (1997) argued that the costs associated with a radical re-design of the FMS, both in terms of the financial risk and the associated implications for safety, made it likely that FMSs and their interfaces would look as they currently do for the foreseeable future. While a radically redesigned FMS is unlikely to come into service in the next few years, changes to the CDU formats, map displays and mode control panels may well be viable. This research has provided a tool for assessing the usability of FMS CDU interfaces, which would allow alternative CDU configurations to be compared and iterative assessments to be conducted during the FMS development lifecycle. This study has also contributed to current research by highlighting potential changes to improve the usability of current generation FMSs, without imposing the radical re-designs resisted by airlines, aircraft manufacturers and avionics manufacturers.

The process of certification, from a human factors perspective, was discussed in Section 1.5 and it was argued that current civil flight deck certification requirements are very equipment focused. Although this is due to change to some extent under the JAA's interim policy document "Human Factors Aspects of Flight Deck Design", even the new policy does not consider usability (JAA, 2001). It was also asserted that flight deck interfaces had not been subjected to "*a rigorous and critical examination based on existing human factors and human-computer interaction methods, standards, or guidelines*" (McClumpha & Rudisill, 2000, p.338). Singer (1999) argued that from a human factors point of view FAR/JAR 25 regulations were "*general, vague, and do not aid the manufacturer*" (p.93). While FAR/JAR 25.671 mandates that controls, for example, must operate with ease, smoothness and positiveness appropriate to their functions, the guidelines are less than useful in the design of FMS CDUs. The problems associated with current FMSs and the improvements for future FMSs suggested by participants in this study could be used by certification authorities in the derivation of guidelines that can be applied by FMS designers.

As discussed in Section 1.5, the regulatory authorities such as the JAA and the FAA have proposed the introduction of Human Factors Certification Criteria for civil flight decks. However, no tools or methods have been proposed as techniques for evaluating the flight deck or its components from a Human Factors perspective. Singer (1999), for example, discussed the requirement for a measure of "*ease and simplicity*" of use (p.93). The tool developed within this study demonstrates how carefully developed, reliable and valid assessment tools could be used to assess the usability of flight deck components and could further be used to discriminate between flight deck components such as FMS CDUs.

The results of this research have highlighted how quantitative and qualitative methods can be used together in the same study, while simultaneously demonstrating the relative strengths and weaknesses of each approach. The quantitative analysis conducted during Studies One and Two permitted the elicitation of a series of usability scales that demonstrated appropriate reliability and validity psychometric properties. What was also shown in these studies was how some quantitative statistical procedures can be misleading. Using Press's Q in this research it was possible to demonstrate that participants' type ratings could be predicted at a rate significantly better than chance, based on their ratings of the usability scales. The Press's Q statistic, however, produced a significant ($p<0.01$) result even when over a third of participants were incorrectly classified using this analysis. While the capability to predict participant group membership from scale ratings was a positive finding in terms of criterion validity, the Press's Q statistic is very sensitive and therefore must be interpreted with appropriate consideration of that sensitivity (Hair et al, 1998). No practical evidence of criterion validity was demonstrated in this research.

It could perhaps be considered a weakness of the usability scale approach adopted, that usability problems could not be prioritised. Whilst the scales produced by this research would not necessarily aid the prioritisation of usability problems, it would be possible to identify the attributes of a product that were rated exceptionally and would therefore be useful both in terms of system development and certification. Through identifying which attributes of a product are rated exceptionally it could be possible to focus more cost-intensive methodologies such as interviews and observation on the usability problems identified. The methodology would therefore be of value to both certification authorities and manufacturers in achieving certification whilst simultaneously improving the usability of the FMS produced.

It must again be reiterated that the usability scales developed were not intended for use as a sole method of certificating FMSs. The qualitative analysis of comments enhanced the research by providing a richness of data that could not have been elicited through use of usability scales alone. The depth and richness of the comments provided guidance as to "why?" participants rated particular usability scales favourably or unfavourably. It was also possible to trace some of the issues raised in this research to aviation accidents such as Cali (Aeronautica Civil, 1996) & Kegworth (AAIB, 1991).

This research demonstrated the applicability of well-accepted HCI guidelines to FMS CDUs. The importance of providing users with informative feedback, and providing users with cues for error prevention & correction, is well documented within the usability literature (e.g. Preece et al, 1994; Nielsen, 1993). The comments made by participants indicated that usability dimensions such as feedback and error prevention & correction are as important in the flight deck environment as they are for desktop packages. The importance of the workload and processor speed components of usability, however, are perhaps even more important in the dynamic flight deck environment than in the office-based environment, since safety as well as user frustration levels are affected.

13.4 Future Developments

Based on the findings of this research, a number of recommendations for future research can be made.

- It would be useful, for example, if a researcher working within an airline could examine the sensitivity of the scales developed within this thesis and determine if the scales could discriminate between software variants within FMS types and aircraft fleets.
- Similarly, it would be advantageous to elicit information on the test-retest reliability of the scales. It is likely that this would have to be conducted within an airline, as it would be difficult for non-airline personnel to gain access to the same participants twice. Further research is also required to examine the practical evidence of criterion validity associated with the scales.
- Future research is required to examine ratings of the scales when combined with more detailed scenario information. Usability scale ratings could, for example, be examined across different routes, approach patterns etc.
- The scales developed within this study could be adapted for other aviation automation components to further aid avionics manufacturers in achieving more usable flight deck components, but also to aid certification authorities in specifying usability requirements for the flight deck.
- Further research is required to determine how tools such as the usability scales developed within this research programme will be incorporated into the Human Factors Certification of future flight decks and avionics components.
- In order to be useful in the certification process a "cut-score" in terms of acceptability would need to be identified for each scale produced in this research. Larger scale validation of the usability scales developed needs to be conducted in order to identify what those cut-scores should be. This will not, however, negate the requirement for scale data to be enhanced and complemented with the performance, mental workload and effectiveness data frequently elicited during system development and certification.
- Future research is required to determine methods of making the FMS more error tolerant and providing better methods of aiding the pilot in error recovery. Acknowledging that human error is inevitable; designers of future FMSs, charts and procedures should examine methods of improving analysis of potential error paths. This would serve to reduce vulnerability and to facilitate error recovery. As asserted by Dismukes & Tullo (2000), "the entire aviation system should be designed to help humans detect and correct error" (p.63).
- Acknowledging that FMSs are unlikely to change dramatically in the near future due to the associated costs and airline resistance to change, further research is required to examine methods of improving current FMSs and their associated displays including the CDU format, the map display and the mode annunciation panel.

- As outlined by Woods (1996), today's FMSs are very mode rich and their functions can be difficult for pilots to keep track of. Research needs to be conducted to improve the simplicity and transparency of FMSs, and to aid pilots in programming the FMS.
- Research also needs to be conducted on how best to standardize FMSs such that for pilots undertaking new type-ratings, workload is minimised and the potential to make errors based on experience of a different aircraft/FMS will be reduced.
- Associated with the previous recommendation, research should be conducted to determine the feasibility of standardizing the language and codes used within the aviation system to minimise the amount of information re-coding pilots are required to perform in order to program FMS CDUs.
- The future of global Communications, Navigation, Surveillance / Air Traffic Management (CNS/ATM) systems developed by the ICAO Future Navigation Services (FANS) Special Committee, will have a significant influence on all new systems developed for civil aviation (Maurino & Galotti, 2000). The impact of these changes on pilot interactions with the FMS and flight deck automation requires through investigation to determine how usability and pilot workload will be influenced.

14 REFERENCES

Aeronautica Civil of the Republic of Colombia (1996). *Controlled Flight Into Terrain, American Airlines Flight 965, Boeing 757-223, N651AA, Near Cali, Colombia, December 20, 1995.* Santafe de Bogota, DC, Colombia: Author.

Air Accidents Investigation Branch (1990). *Report on the accident to Boeing 737-400 G-OBME near Kegworth, Leicestershire on 8 January 1989.* London: HMSO Publications Centre.

Amalberti R. & Wibaux F. (2000). Advanced Automation Glass Cockpit Certification. In, J. Wise & V.D. Hopkin (Eds.), *Human Factors in Certification* (pp. 361-374). Mahwah, NJ: Lawrence Erlbaum Associates.

Anastasi A. (1988). *Psychological Testing.* New York: Macmillan.

Bailey R.W. (1982). *Human Performance Engineering: A guide for System Designers.* New Jersey: Prentice-Hall.

Batelle Aviation Safety Reporting System (ASRS). ASRS office, Ohio: Author.

Bevan N., Kirakowski J. & Maissel J. (1991). What is Usability? In, H.J. Bullinger (Ed.), *Human Aspects in Computing: Design and Use of Interactive Systems and Work with Terminals* (pp. 651-655). The Netherlands: Elsevier.

Bevan N. & Macleod M. (1994). Usability measurement in context. *Behaviour & Information Technology, 13,4,* 132-145.

Bevan N. (1995). Measuring usability as quality of use. *Software Quality Journal, 4,* 115-130.

Billings C. (1991). *Human-centered aircraft automation: A concept and guidelines (NASA Technical Memorandum 103885).* Moffett Field, CA: NASA-Ames Research Center.

Billings C. (1997). *Aviation Automation - The Search for a Human-Centred Approach.* Mahwah, NJ: Lawrence Erlbaum Associates.

Boeing Commercial Airplane Group (1997). *Statistical Summary of Commercial Jet Aircraft Accidents 1959-1997.* Seattle, WA: Author.

Boeing Commercial Airplane Group (2002). B747-400 Flight Deck Picture. *http://www.boeing.com/companyoffices/gallery/images/commercial/747400-04.html*

Bollen K.A. (1989). *Structural Equations with Latent Variables.* London: Wiley.

Booth P. (1989). *An Introduction to Human-Computer Interaction.* London: Lawrence Erlbaum Associates.

Boy G.A. & Cacciabue P.C. (1997). The optimisation of automation on civil flight decks. In, D. Harris (Ed.), *Human Factors for Flight Deck Certification* (pp. 57-73). Cranfield: Cranfield University Press.

Braune R. (1989). *A survey of Boeing / Deutsche Lufthansa pilots flying a mixed Fleet of 737-200/300.* Seattle, WA: Boeing Inc.

Brooke J. (1996). SUS: a 'quick and dirty' usability scale. In, P. Jordan, B. Thomas, B. Weerdmeester & I. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 179-188). London: Taylor & Francis.

Bulfer B. & Gifford S. (1997). *FMC User's Guide B 757, B 767, B 747.* Texas: Leading Edge Libraries.

Bulfer B. (1996). *FMC User's Guide B 737.* Texas: Leading Edge Libraries.

Bureau of Air Safety Investigation (1998). *Advanced Technology Aircraft Safety Survey Report.* Australia: Department of Transport and Regional Development.

Campbell D.T. & Fiske D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Cassell C. & Symon G. (1994). Qualitative research in work contexts. In, G. Symon & C. Cassell (Eds.), *Qualitative Methods and Analysis in Organizational Research* (pp. 1-13). London: Sage.

Chell E. (1998). The Critical Incident Technique. In, G. Symon & C. Cassell (Eds.), *Qualitative Methods and Analysis in Organizational Research* (pp. 51-72). London: Sage.

Chidester T. (1999). Introducing FMS aircraft into airline operations. In, S. Dekker & E. Hollnagel (Eds.), *Coping with Computers in the Cockpit* (pp. 153-194). Aldershot: Ashgate.

Cook M. (1998). *Personnel Selection: Adding Value through People.* UK: Wiley.

Courteney H. (1997). Practising what we preach. In, D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Volume One* (pp. 111-120). Aldershot: Ashgate.

Courteney H. (1999). Human Factors of Automation: The Regulator's Challenge. In, S. Dekker & E. Hollnagel (Eds.), *Coping with Computers in the Cockpit* (pp. 109-130). Aldershot: Ashgate.

Courteney H. (1999a). Assessing Error Tolerance in Flight Management Systems. In, D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Volume Three* (pp.43-50). Aldershot: Ashgate.

Cronbach L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Curry R.E. (1985). *The Introduction of New Technology: A Human Factors Study.* (NASA TM 86659). Moffett Field: NASA Ames Research Center.

Dekker S. & Orasanu J. (1999). Automation and Situation Awareness – Pushing the Research Frontier. In, S. Dekker & E. Hollnagel (Eds.), *Coping with Computers in the Cockpit* (pp. 69-85). Aldershot: Ashgate.

Dekker S. & Woods D. (1999). Automation and its impact on human cognition. In, S. Dekker & E. Hollnagel (Eds.), *Coping with Computers in the Cockpit* (pp. 7-27). Aldershot: Ashgate.

Desurvire H.W. (1994). Faster Cheaper!! Are Usability Inspection Methods as Effective as Empirical Testing. In, J. Nielsen & R.L. Mack (Eds.), *Usability Inspection Methods* (pp. 173-202). New York: John Wiley & Sons.

DeVellis R. (1991). *Scale Development Theory and Applications.* California: Sage.

Dismukes K. & Tullo F. (2000, July 17). Aerospace Forum: Rethinking Crew Error. *Aviation Week and Space Technology, 153,* 63.

Dodd R., Eldredge D. & Mangold S. (1992). *A review and discussion of Flight Management System Incidents Reported to the Aviation Safety Reporting System. Report Number DOT/FAA/RD-92/2.* Washington DC: US Department of Transport.

Douglas Aircraft Co. (1990). *Functional Decomposition of the Commercial Flight Domain for Function Allocation.* Final report to NASA-Langley Research Center, Contract No, NAS1-18028. Briefing paper, December 10. Hampton VA: NASA-Langley Research Center.

Dourish P. & Bellotti V. (1992). Awareness and Co-ordination in Shared Workspaces in J. Turner & R. Kraut (Eds.), *Proceedings ACM conference on Computer Supported Co-operative Work CSCW '92* (pp. 107-114). New York: Association for Computing Machinery.

Dumas J. & Redish J. (1993*). A Practical Guide to Usability Testing.* New Jersey: Ablex Publishing.

Dutt A., Johnson, H. & Johnson P. (1994). Evaluating evaluation methods. In, G. Cockton, S.W. Draper & G.R.S. Weir (Eds.), *People and Computer IX. Proceedings of the British Computer Society HCI'94* (pp. 109-121). Cambridge: Cambridge University Press.

Dyck J., Abbott D. & Wise J. (1993). HCI in Multi-Crew Aircraft. In, G. Salvendy & M. Smith (Eds.), *Human-Computer Interfaces – Software and Hardware Interfaces. Advances in Human Factors/Ergonomics 19B* (pp. 151-156). Amsterdam: Elsevier.

Eason K.D. (1984). Towards the experimental study of usability. *Behaviour & Information Technology, 3, 2,* 133-143.

Eason K.D., Damodaran L. & Stewart T. (1975). Interface problems in man-computer interaction. In, E. Mumford & H. Sackman (Eds.), *Human Choice and Computers* (pp. 91-105). Amsterdam: North-Holland.

Eldredge D., Mangold S. & Dodd R. (1992). *Review and Discussion of Flight Management System Incidents Reported to the Aviation Safety Reporting System. Report No. DOT/FAA/RD-92-2.* Washington, DC: U.S. Department of Transportation.

Endsley M. (1993). Aviation System Certification: Challenges and Opportunities. In, J.A. Wise, V.D. Hopkin & D. Garland (Eds.), *Human Factors Certification of Advanced Aviation Technologies* (pp. 9-12). Daytona Beach, Florida: Embry Riddle Aeronautical University Press.

Endsley M. (1996). Automation and Situation Awareness. In, R. Parasuraman & M. Mouloula (Eds.), *Automation Technology and Human Performance* (pp. 163-182). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley M. & Kiris E.O. (1995). The Out-of-the-Loop Performance Problems and Level of Control in Automation. *Human Factors, 37, 2,* 381-394.

Endsley M. & Strauch B. (1997). Automation and Situation Awareness: The Accident At Cali, Columbia. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 877-881). Ohio: Ohio State University.

Federal Aviation Administration (1987). *Transport Category Airplane Electronic Display Systems. Advisory Circular 25-11.* Washington, DC: Federal Aviation Administration. Washington, DC: Author.

Federal Aviation Administration (1989). *Approval of Flight Management Systems in Transport Category Airplanes. Advisory Circular 25-15.* Washington, DC: Author.

Federal Aviation Administration (1996). Human Factor Team Report on *The Interfaces Between Flightcrews and Modern Flight Deck Systems.* Washington D.C.: Author.

Federal Aviation Administration (1996). *Federal Aviation Regulation Part 121-Operating Requirements: Domestic, Flag, And Supplemental Operations.* Washington, DC: Author.

Federal Aviation Administration (1998). *Code of Federal Regulations, Aeronautics and Space, Part 25 (Large Airplanes).* Washington, DC: Federal Aviation Administration.

Flight International (1998, 24-30 March), World Airline Directories.

Foreman P.M. (1996). A potential boon for the pilot, many current automated systems produce unwanted complexity. *International Civil Aviation Organization Journal, 51, 5,* 15-19.

Fuld R.B. (1997). The Fiction of Function Allocation Revisited. In, E.F Fallon, L. Bannon & J. McCarthy (Eds.), *ALLFN'97 – Revisiting the Allocation of Functions Issues. Proceedings of the First International Conference on Allocation of Functions. Volume 1* (pp. 17-30). KY: IEA.

Funk K., Lyall B. & Riley V. (1996). *Taxonomy of Flightdeck Automation Problems and Concerns. A comparative Analysis of Flightdecks with Varying Levels of Automation.* Federal Aviation Administration Grant G 93-G-039 Phase 1 Final Report. Washington DC: US Department of Transportation.

Galitz W.O. (1980). *Human Factors in Office Automation.* Atlanta, GA: Life Office Management Association.

Ghiselli E., Campbell J. & Zedeck S. (1981). *Measurement Theory for the Behavioural Sciences.* San Francisco, CA: Freeman.

Gilson R. & Abbott D. (2000). Certification of Flight Crews for Advanced Flight Control Systems. In, J. Wise & V.D. Hopkin (Eds.), *Human Factors in Certification* (pp. 131-138). Mahwah, NJ: Lawrence Erlbaum Associates.

Glaser B.G. & Strauss A.L. (1968). *The Discovery of Grounded Theory.* Chicago: Aldine.

Goteman O. (1999). Automation Policy or Philosophy? Management of Automation in the Operational Reality. In, S. Dekker & E. Hollnagel, *Coping with Computers in the Cockpit* (pp. 215-224). Aldershot: Ashgate.

Graeber R.C. & Mumaw R.J. (1999). Realising the benefits of cognitive engineering in commercial aviation. In, D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Volume Three* (pp. 3-26). Aldershot: Ashgate.

Guba E.G. (1978). Toward a Methodology of Naturalistic Inquiry in Educational Evaluation. *CSE Monograph Series in Evaluation no.8.* Los Angeles: University of California, Center for the Study of Evaluation.

Guion R.M. (1977). Content validity – Three years of talk – what's the action? *Public Personnel Management, 6,* 407-414.

Hair J.E, Anderson R., Tatham R. & Black W. (1998). *Multivariate Data Analysis.* New Jersey: Prentice Hall.

Harris D. (1995). The quantitative analysis of qualitative data: methodological issues in the derivation of trends from incident reports. In, N. McDonald, N. Johnson & R. Fuller (Eds.), *Applications of Psychology to the Aviation System. Volume One* (pp. 117-122). Aldershot: Ashgate.

Helmreich R. & Merritt A. (1998). *Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences.* Aldershot: Ashgate.

Helmreich R., Merritt A., Sherman P., Gregorich S. & Wiener E. (1993). *The Flight Management Attitude Questionnaire.* NASA/UT/FAA Technical Report 93-4. Austin, TX: University of Texas.

Herberlein T. & Baumgartner R. (1978). Factors affecting response rates to mailed questionnaires: a quantitative analysis of the published literature. *American Sociology Review, 43, 4,* 447-462.

Hines W.E. (1997). Flight Crew Performance in Standard and Advanced Technology Aircraft. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 1066-1072). Ohio: Ohio State University.

Hughes D. (1995, February 6). Studies highlight automation surprises. *Aviation Week and Space Technology, 142, 6,* 48-49.

International Standards Organisation (ISO, 1998). *ISO 9241 Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) Part 11: Guidance on Usability.* Switzerland: Author.

James M., McClumpha A., Green R., Wilson P. & Belyavin A. (1991). Pilot Attitudes to Cockpit Automation. In, R. Jensen (Ed.), *Proceedings of the Sixth International Symposium on Aviation Psychology* (pp. 192-197). Ohio: Ohio State University.

Javaux D., Masson M. & De Keyser V. (2000). Beware of Agents when Flying Aircraft: Basic Principles Behind a Generic Methodology for the Evaluation and Certification of Advanced Aviation Systems. In, J. Wise & V.D. Hopkin (Eds.), *Human Factors Certification* (pp. 375-405). Mahwah, NJ: Lawrence Erlbaum Associates.

Johnson G.I. (1996). The usability checklist approach revisited. In, P. Jordan, B. Thomas, B. Weerdmeester & I. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 179-188). London: Taylor & Francis.

Joint Airworthiness Authorities (1994). *JAR 25 Joint Aviation Requirements (Large Aeroplanes),* JAR-OPS 1. The Netherlands: Author.

Joint Aviation Authorities (1996). *Joint Aviation Requirements (All Weather Requirements), JAA-AWO.* The Netherlands: Author.

Joint Aviation Authorities (2001). Human Factors Aspects of Flight Deck Design. *JAA Administrative & Guidance Material Section Three: Certification Part 3: Interim Policies & Temporary Guidance Material.* The Netherlands: Author.

Jordan P. (1998). *An Introduction to Usability.* London: Taylor & Francis.

Jordan P. (1998a). Human Factors for pleasure in product use. *Applied Ergonomics, 29,1,* 25-33.

Jordan P., McClelland I. & Thomas B. (1996). Introduction. In, P. Jordan, B. Thomas, B. Weerdmeester & I. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 1-3). London: Taylor & Francis.

Jordan P., Thomas B. & McClelland I. (1996). Issues for Usability Evaluation in Industry: Seminar Discussions. In, P. Jordan, B. Thomas, B. Weerdmeester & I. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 237-243). London: Taylor & Francis.

Jöreskog K.G. (1970). *A General Method for Analysis of Covariance Structures. Biometrika, 57*, 239-251.

Kaiser H.F. (1974). Little Jiffy, Mark IV. *Educational and Psychology Measurement 34,* 111-117.

Kantowitz B.H. & Campbell J.L. (1996). Pilot Workload and Flightdeck Automation. In, R. Parasuraman & M. Mouloula (Eds.), *Automation Technology and Human Performance* (pp. 117-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Karat C., Campbell R. & Fiegel T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In P. Bauersfeld, J. Bennett & G. Lynch (Eds.), *Proceedings of ACM CHI '92 Conference on Human Factors in Computing Systems* (pp. 397-404). New York: ACM.

Kelly B., Graeber C. & Fadden D. (1992). Applying Crew-centered Concepts to Flight deck Technology: The Boeing 777. Paper presented at *The Flight Safety Foundation 45th International Air Safety Seminar.* CA: Flight Safety Foundation.

Kirakowski J. (1996). The Software Usability Measurement Inventory: background and usage. In, P. Jordan, B. Thomas, B. Weerdmeester & I. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 169-188). London: Taylor & Francis.

Kirakowski J. (2001). *Questionnaires in Usability Engineering – A List of Frequently Asked Questions* (3rd Edition). University College Cork: Human Factors Research Group. http://www.ucc.ie/hfrg/resources/qfaq1.html/.

Kirakowski J. & Cierlik B. (1998). Measuring the usability of web sites. *Proceedings of the Human Factors & Ergonomics Society 42nd International Meeting* (pp. 424-429). CA: Human Factors & Ergonomics Society.

Kirakowski J. & Corbett M. (1993). SUMI: The Software Usability Measurement Inventory. *British Journal of Educational Technology, 24,3,* 210-212.

Krippendorff K. (1980). *Content Analysis: An Introduction to its Methodology*. London: Sage.

Last S. (1994). Hidden origins to crew-caused accidents. In, Proceedings of IFALPA Conference, *Interpilot,* June, 5-15.

Lavery D., Cockton G. & Atkinson M. (1997). Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology, 16, 4/5,* 246-266.

Lewis J. (1995). IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. *International Journal of Human Computer Interaction, 7,1,* 57-78.

Lin H., Choong Y. & Salvendy G. (1997). A proposed index of usability: a method for comparing the relative usability of different software systems. *Behaviour & Information Technology, 16, 4/5,* 267-278.

Macleod M. (1992). *An Introduction to Usability Evaluation.* Teddington: National Physical Laboratory.

Maguire M. (1982). An evaluation of published recommendations for the design of man-computer dialogues. *International Journal of Man-Machine Studies, 16,* 237-261.

Marshall C., Nelson C. & Gardiner M. (1987). Design Guidelines. In, M. Gardiner & B. Christie (Eds.), *Applying Cognitive Psychology to User-Interface Design* (pp. 221-278). Chichester: John Wiley.

Marshall H. (1994). Discourse Analysis in an Occupational Context. In, C. Cassell & G. Symon (Eds.), *Qualitative Methods in Organizational Research – A Practical Guide* (pp. 91-106). London: Sage.

Maurino D. & Galotti V. (2000). Integrating Human Factors Knowledge into Certification: The Point of View of the International Civil Aviation Organisation. In, J. Wise & V.D. Hopkin (Eds.), *Human Factors in Certification* (pp. 295-313). Mahwah, NJ: Lawrence Erlbaum Associates.

Mayhew D. (1992). *Principles and Guidelines in Software User Interface Design.* New Jersey: Prentice Hall.

McClumpha A. & Rudisill M. (2000). Certification for civil flight decks and the human-computer interface. In, J. Wise & V.D. Hopkin (Eds.), *Human Factors in Certification* (pp. 335-347). Mahwah, NJ: Lawrence Erlbaum Associates.

Merritt A. & Helmreich R. (1995). Culture in the cockpit: A multi-airline study of pilot attitudes. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 676-687.). Ohio: Ohio State University.

Microsoft Corporation (1995). *The Windows Interface Guidelines for Software Design.* Washington: Microsoft Press.

Miles M. & Huberman A. (1994). *Qualitative Data Analysis.* California: Sage.

Miller R.B. (1968). Response time in man-computer conversational transactions. *Proceedings of the American Federation of Information Processing Societies (AFIPS) Spring Joint Computer Conference* (pp. 267-277). Monvale, NJ: AFIPS Press.

Miller D.C. (1991). *Handbook of Research Design and Social Measurement.* California: Sage.

Molich R. & Nielsen J. (1990). Improving a human-computer dialogue. *Communications of the Association for Computing Machinery, 33,3,* 338-348.

Mosier K.L. & Skitka L.J. (1996). Human Decision Makers and Automated Decision Aids: Made for Each Other? In, R. Parasuraman & M. Mouloula (Eds.), *Automation Technology and Human Performance* (pp. 201-220). Mahwah, NJ: Lawrence Erlbaum Associates.

Murphy K.R. & Davidshofer C.O. (1991). *Psychological Testing: Principles & Applications.* New Jersey: Prentice-Hall.

MUSiC (1992). *Metrics for Usability Standards in Computing.* ESPRIT II Project 5429), Product information. Teddington: National Physical Laboratory.

Nielsen J. (1993). *Usability Engineering.* London: Academic press.

Nielsen J. (1994). Usability Laboratories. *Behaviour & Information Technology, 13, 1 & 2,* 3-8.

Nielsen J. (1994a). Usability Inspection Methods. In, J. Nielsen & R. Mack (Eds.), *Usability Inspection Methods* (pp. 25-62). New York: Wiley & Sons.

Nielsen J. (1997). Usability Testing. In, G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1543-1568). New York: Wiley.

Norman D. (1986). Cognitive Engineering. In, D.A. Norman & S. Draper (Eds.), *User-centred design: New perspectives on human-computer interaction* (pp. 31-61). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Norman D. (1988). *The Psychology of Everyday Things.* New York: Basic Books.

Norman D. (1989). *The problem of automation: Inappropriate feedback and interaction not over automation* (ICS 8904). San Diego, CA: Institute for Cognitive Science, University of California.

Norman K. & Shneiderman B. (1989). *Questionnaire for User Interaction Satisfaction (QUIS 5.0).* Maryland: University of Maryland HCI Lab.

Norman D. A. (1991). Cognitive Science in the Cockpit. *Crew System Ergonomics Information Analysis Center (CSERIAC) Gateway, 11,2,* 1-6.

Norman D. A. (1992). Where Human Factors Fails: Ergonomics Versus the World of Design and Manufacture. *Arnold M. Small Safety Lecture. Human Factors Society Conference.* Atlanta, GA: Human Factors Society.

Noyes J., Leggatt A. & Irwin M. (1997). Altimeter Design in Avionics: Analogue Versus Digital? In, D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Volume One* (pp. 39-45). Aldershot: Ashgate.

Noyes J., Starr A. & Frankish C. (1996). Aircraft Warning System Development, *Behaviour & Information Technology, 15,2,* 67-75.

Oppenheim A. (1966). *Questionnaire Design and Attitude Measurement.* London: Heinemann.

Oppermann R. & Reiterer H. (1997). Software Evaluation Using the 9241 Evaluator. *Behaviour & Information Technology, 16, 4/5,* 232-245.

Parasuraman R., Mouloua M., Molloy R. & Hilburn B. (1996). Monitoring of Automated Systems. In, R. Parasuraman & M. Mouloula (Eds.), *Automation Technology and Human Performance* (pp. 91-115). Mahwah, NJ: Lawrence Erlbaum Associates.

Parasuraman R. & Riley V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors, 39, 2,* 230-253.

Paries J. (2000). Some inadequacies of current human factors certification process of advanced aircraft technologies. In, J. Wise & V.D. Hopkin (Eds.), *Human Factors in Certification* (pp. 349-361). Mahwah, NJ: Lawrence Erlbaum Associates.

Patton M. (1987). *How to Use Qualitative Methods in Evaluation.* California: Sage.

Patton M. (1990). *Qualitative Evaluation and Research Methods.* California: Sage.

Pedrali M., Cojazzi G. & Cacciabue P.C. (1995). A methodology for retrospective analyses of accidents involving human factors. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 1278-1283). Ohio: Ohio State University.

Porteous M., Kirakowski J. & Corbett M. (1993). *SUMI User Handbook.* Cork, Ireland: Human Factors Research Group, University College Cork.

Preece J., Rogers Y., Sharp H., Benyon D., Holland S. & Carey T. (1994). *Human-Computer Interaction.* Wokingham: Addison-Wesley.

Schiratti E. (1999). Project Magenta – Boeing Smiths Type FMS CDU. *http://www.schiratti.com/*

Ravden S. & Johnson G. (1989). *Evaluating Usability of Human-Computer Interfaces – A Practical Method.* Chichester: Ellis Horwood.

Remington R.J. (1998, 12 June). Usability Evaluation Trends: The Usability Approach. Lockheed Martin Missiles & Space Company. http://targetsoft-ware.com/aiaa/htm.

Reuzeau F. (1997). The Human Factors Program at Aerospatiale. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 853-858.). Ohio: Ohio State University.

Robson C. (1993). *Real World Research.* Oxford: Blackwell.

Rogers W. & Schutte P. (1997). Clean Slate Flight Deck Design: Theory Based Solutions Provide Useful Design Goals. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Ninth International Symposium on Aviation Psychology* (pp. 859-864). Ohio: Ohio State University.

Romahn S. & Schafer D. (1997). Error analysis as a means for user interface evaluation: a comparison of graphically interactive and traditional FMS user interfaces. In, D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Volume One* (pp. 101-109). Aldershot: Ashgate.

Rubin J. (1994). *Handbook of Usability Testing.* New York: John Wiley.

Rudisill M. (1995). Line pilots' attitudes about and experience with flight deck automation: results of an international survey and proposed guidelines. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 288-293.). Ohio: Ohio State University.

Sarter N.B. (1991). The Flight Management System - Pilots' interaction with cockpit automation. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1307-1310). New York: Institute of Electrical and Electronics Engineers.

Sarter N., & Woods D. (1992). Pilot Interaction with cockpit automation: operational experiences with the Flight Management System. *The International Journal of Aviation Psychology, 2,4,* 303-321.

Sarter N., Woods D. & Billings C. (1997). Automation Surprises. In, G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (pp. 1926-1943). New York: Wiley.

Sarter N. & Woods D. (1994). Pilot interaction with cockpit automation II: an experimental study of Pilots' model and awareness of the Flight Management System. *The International Journal of Aviation Psychology, 4,1,* 1-28.

Sarter N. & Woods D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37,1,* 5-19.

Scapin D. & Bastien L. (1997). Ergonomic criteria for evaluating the ergonomic quality of interactive systems. *Behaviour & Information Technology, 16,4/5*, 220-231.

Scerbo M.W. (1996). Theoretical Perspectives on Adaptive Automation. In, R. Parasuraman & M. Mouloula (Eds.), *Automation Technology and Human Performance* (pp. 37-63). Mahwah, NJ: Lawrence Erlbaum Associates.

Shackel B. (1986). Ergonomics in design for usability. In, M.D. Harrison & A.F. Monk (Eds.), *People & computers: Designing for Usability. Proceedings of the Second Conference of the BCS HCI Specialist Group* (pp. 44-64). Cambridge: Cambridge University Press.

Sherman P. & Helmreich R. (1995). Attitudes toward automation: the effect of national culture. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 682-687). Ohio: Ohio State University.

Sherman P., Helmreich R. & Merritt A.C. (1997). National Culture and Flight deck Automation: Results of a Multi-nation Survey. *International Journal of Aviation Psychology, 7,4,* 311-329.

Shneiderman B. (1992). *Designing the User Interface: Strategies for Effective Human Computer Interaction.* Reading, MA: Addison-Wesley.

Singer G. (1999). Filling the Gaps in the Human Factors Certification Net. In, S. Dekker & E. Hollnagel (Eds.), *Coping with Computers in the Cockpit* (pp. 87-107). Aldershot: Ashgate.

Singer G. (2001). Minimising crew error by design: are test pilots doing a good enough job. *Human Factors and Aerospace Safety 1*, (4), 301-321.

Smith S. & Mosier J. (1986). *Guidelines for designing user interface software.* Report No. MTR – 100090; ESD-TR-86-278. Bedford, MA: The Mitre Corporation.

Sparaco P. (1994, January 3). Human Factors Cited in French A320 Crash. *Aviation Week & Space Technology, 140,* 30-31.

Stager P. (2000). Validation as Certification: Methodological Issues for Human Factors. In, J. Wise & V.D. Hopkin (Eds.), *Human Factors in* Certification (pp. 91-103). Mahwah, NJ: Lawrence Erlbaum Associates.

Stanton N. & Baber C. (1996). Factors Affecting the Selection of Methods and Techniques Prior to Conducting a Usability Evaluation. In, P. Jordan, B. Thomas, B. Weerdmeester & I. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 39-57). London: Taylor & Francis.

Stanton N. & Young M. (1999). Utility Analysis in Cognitive Ergonomics. In, D. Harris (Ed.), *Engineering Psychology and Cognitive Ergonomics. Volume Four* (pp. 411-417). Aldershot: Ashgate.

Stewart S. (2000). *Flying the Big Jets.* UK: Airlife.

Sutcliffe A. (1988). *Human-Computer Interface Design.* London: Macmillan.

Sweeney M., Maguire M. & Shackel B. (1993). Evaluating user-computer interaction: a framework. *International Journal of Man-Machine Studies, 38*, 689-711.

Sweet W. (1995). The Glass Cockpit. *IEEE Spectrum, September.*

Tabachnick B. & Fidell L. (1983). *Using Multivariate Statistics.* New York: Harper & Row.

Tenney Y.J., Rogers W.H. & Pew R.W. (1998). Pilot Opinions on Cockpit Automation. *The International Journal of Aviation Psychology, 8, 2*, 103-120.

Thomas B. (1996). Quick and dirty usability tests. In, P. Jordan, B. Thomas, B. Weerdmeester & I. McClelland (Eds.), *Usability Evaluation in Industry* (pp. 107-114). London: Taylor & Francis.

Wharton C., Rieman J., Lewis C. & Polson P. (1994). The Cognitive Walkthrough Method: A Practitioner's Guide. In, J. Nielsen & R. Mack (Eds.), *Usability Inspection Methods* (pp. 105-140). New York: Wiley & Sons.

Wickens C. (1992). *Engineering Psychology & Human Performance.* New York: Harper Collins.

Wiener E.L. (1985). *Human Factors of Cockpit Automation: A field study of flight crew transition (NASA contractor report No 177333).* Moffett Field, CA: NASA Ames Research Center.

Wiener E.L. (1988). Cockpit Automation. In, E.L. Wiener & D.C. Nagel (Eds.), *Human Factors in Aviation* (pp. 433-461). San Diego, CA: Academic Press.

Wiener E.L. (1989). *Human factors of advanced technology (glass cockpit) transport aircraft (NASA contractor report No 177528).* Moffett Field, CA: NASA Ames Research Center.

Williams C.L. (1995). Potential conflicts of time sharing the Flight Management System Control Display Unit with Data Link Communications. In, R. Jensen & L. Rakovan (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 341-346.). Ohio: Ohio State University.

Williges R.C., Williges B. & Elkerton J. (1987). Human Factors in the Design & Use of Computing Systems. In, G. Salvendy (Ed.), *Handbook of Human Factors* (pp. 1416-1449). New York: Wiley.

Williges R.C., Williges B. & Fainter R.G. (1988). Software Interfaces for Aviation, In, E.L. Wiener & D.C. Nagel (Eds.), *Human Factors in Aviation* (pp. 463-493). San Diego CA: Academic Press.

Woods D. (1995). Towards a theoretical base for representation design in the computer medium: Ecological perception and aiding in human cognition. In, J. Flach, P. Hancock, J. Caird & K. Vicente (Eds.), *Global Perspectives on the Ecology of Human-Machine Systems* (pp. 157-188). Hillsdale, NJ: Lawrence Erlbaum Associates.

Woods D. (1996). Decomposing automation: Apparent simplicity, real complexity. In, R. Parasuraman & M. Mouloula (Eds.), *Automation Technology and Human Performance* (pp. 3-17). Mahwah, NJ: Lawrence Erlbaum Associates.

Woods D., Johannesen L., Cook R.I. & Sarter N. (1994). *Behind Human Error. Cognitive Systems, Computers & Hindsight.* State of the Art Report. Dayton, OH: Crew Systems Ergonomic Information & Analysis Center.

Woods D. & Sarter N. (1998). *Learning from Automation Surprises and "Going Sour" Accidents: Progress on Human-Centered Automation.* Report ERGO-CSEL-98-02. Columbus, OH: Institute for ergonomics, The Ohio State University.

Woods T. & Courteney H. (2001). Current Regulatory Approach to Human Factors, Interim Measures and Introduction to Human Factors Working Group. *Presentation at the Civil Avionics Support Group Workshops for the Avionics Industry – Human Factors for Flight Deck Certification (19th November 2001).* Leatherhead, Surrey: ERA Technology.

Wyatt J. & Emerson P. (1990). A pragmatic approach to knowledge engineering with examples of use in a difficult domain. In, D. Berry & A. Hart (Eds.), *Expert Systems: Human Issues* (pp. 65-78.). London: Chapman and Hall.

Cranfield UNIVERSITY

COLLEGE OF AERONAUTICS

HUMAN FACTORS GROUP

PHD THESIS

ACADEMIC YEAR 2001-02

KAREN P. LANE

THE DEVELOPMENT OF USABILITY SCALES FOR FLIGHT MANAGEMENT SYSTEMS (VOLUME TWO)

SUPERVISOR: D. HARRIS

2001-02

THIS THESIS IS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

© Cranfield University 2002. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

TABLE OF CONTENTS

APPENDIX A – SUMMARY OF OTHER USABILITY APPROACHES CONSIDERED 313

APPENDIX B - INTERVIEW SCHEDULE DERIVED FROM USABILITY LITERATURE 318

APPENDIX C - SUMMARY OF INTERVIEW THEMES 322

APPENDIX D - STUDY ONE PRE-PILOT QUESTIONNAIRE 327

APPENDIX E - STUDY ONE PILOT QUESTIONNAIRE 331

APPENDIX F - STUDY ONE FINAL QUESTIONNAIRE 335

APPENDIX G - PARTICIPATION INVITATION TO AIRLINES 339

APPENDIX H - STUDY TWO FINAL QUESTIONNAIRES 340

APPENDIX I - LETTER TO FLIGHT CREW 348

APPENDIX J - ANALYSIS OF USABILITY RATINGS BY FLIGHT PHASE 349

APPENDIX K - CURVILINEAR REGRESSION ANALYSES FOR PILOTS' EXPERIENCE ... 353

APPENDIX L - MODIFICATION INDICES 366

APPENDIX M - STUDY THREE QUALITATIVE COMMENTS 367

M1 IMPROVEMENTS SOUGHT TO FUTURE FMSS 368
M2 PROBLEMS WITH CURRENT FMSS 381

Appendix A – Summary of other Usability Approaches Considered

Usability Method	Description	Critical Analysis for Applicability
Inquiry		
Field Observation	Field observation is constituted of both watching users interacting with the product of interest in its real world context, and talking to the users about how they use the product on a day to day basis (Nielsen, 1993).	Some field observation was undertaken in this study to permit the researcher to understand the complexities of the operational environment, and to observe the FMS context of use. The researcher was permitted to discuss operational and contextual issues with the pilots during low workload periods of the operational flights. While field observation is a useful technique for evaluating the usability of current products, it is likely to be of little value in evaluating the usability of future FMSs that will not be installed in their real world context.
Interviews & Focus Groups	Interview & Focus Group usability methodologies involve an experienced interviewer querying a user or group of users about their experiences of using a product.	Interview & Focus Group usability methodologies can be used at any point during a product's development lifecycle. Interviews are useful in that real end users, rather than HCI expert opinion is gained during the assessment. A further advantage is that interviews may be flexible and can therefore be shaped to the assessor's responses and be highly informative (Oppermann & Reiterer, 1997). In-depth individual interviews were conducted in this research as a foundation for the usability questionnaire development process. The major drawback to using this method is the time taken to conduct the interview and analyse its content. Conducting an interview takes a significant amount of planning and reasonable expertise on the part on the interviewer. The time and cost taken to conduct and analyse interviews therefore limits the number of participants that may be involved in the assessment. In-depth interviews were conducted in this research programme in order to gain a detailed understanding of general FMS usability issues and to derive the questionnaire items from which the usability scales would ultimately be created. The key distinguishing factor of focus group research from interviews is that it employs more

Usability Method	Description	Critical Analysis for Applicability
		than one user in a single assessment. The Focus group methodology was not used in this research for a number of reasons. The prime advantage of group discussions is that they help to determine the "*universality of comments*" (Shneiderman, 1992, p.484). However, the main disadvantage to conducting focus group research is the time and cost taken to conduct the interviews and analyse the data. Although costs are likely to be lower for focus group research than for individual interviews, the analysis can still be very time consuming. A further disadvantage to group interviews is that it is sometimes difficult to pursue the views of any one participant (Oppermann & Reiterer, 1997). Since it is not possible to give a product a "score", or evaluate the product against a set of criteria, this methodology would not be adequate for certification.
Inspection		
Heuristic Evaluation	The Heuristic Evaluation involves having a small set of usability experts examine the user interface under investigation and judge its compliance with recognised usability principles such as: *"Even better than good error messages is a careful design which prevents a problem from occurring in the first place."* (Nielsen, 1994a, p.30] The goal of the methodology is to identify usability problems so that they can be attended to as part of an iterative design process.	The Heuristic Evaluation methodology was not adopted for this research since it does not involve real end users but rather is dependent on a small number of expert usability specialists. Since use of the FMS requires significant training and is operated in a complex dynamic environment, methodologies that did not involve real-end users were not considered to be appropriate. Even if the use of usability specialists rather than line pilots was acceptable, access to FMSs would have been very difficult to achieve and the operational context would not be incorporated. The current certification process reflects this methodology in its use of expert "test" pilots, rather than typical end users.

Testing		
Thinking Aloud	The thinking aloud methodology involves asking participants to vocalise their thoughts and opinions while interacting with a product (Nielsen, 1993).	This technique can be used at any stage of product development and is a very cost-effective method of eliciting qualitative feedback during usability assessment. The limitation of the methodology for the purposes of this study was that its application would have been too intrusive for use during operational flights on the flight deck or in the simulator. Similarly, for safety reasons airlines would be very reluctant to allow this methodology to be utilised on operational flights. Gaining access to the large number of participants required to elicit the usability framework sought would in addition have been very difficult logistically.
Co-Discovery Method	Co-discovery is a method of usability testing in which two participants attempt to interact and perform tasks with a system while being observed.	This technique can be used at any point during system development and evaluation and has advantages over the thinking aloud protocol in that in many work environments a user can seek assistance can be sought from another individual (e.g. a work colleague). The interactions between the participants, and the vocalisation of thoughts can also provide insights into usability problems. The limitation of the co-discovery methodology for the purposes of this study was that its application would have been too intrusive for use during operational flights or in the simulator. Similarly, for safety reasons airlines would be very reluctant to allow this methodology to be utilised on operational flights. Gaining access to the large number of participants required to elicit the usability framework sought would in addition have been very difficult logistically.

Question Asking Protocol	The Question Asking Protocol is an extension of the Thinking Aloud methodology in that rather than waiting for users to vocalise their thoughts, the researcher prompts the participants with questions. The rationale behind this extension of the Thinking Aloud methodology is that is allows the researcher to probe the user about areas of the interface under assessment that are opaque or cause usability problems.	This technique can be used during any phase of development but can be intrusive on the participants when they are trying to conduct multiple tasks or during periods of high workload. It would be difficult to conduct this methodology on the flight deck since safety could be compromised if the pilots were trying to respond to questions, while simultaneously flying the aircraft, talking to Air Traffic Control etc.
Performance Measurement	Performance Measurement usability methodologies are oriented at the elicitation of quantitative performance data such as how long it takes a user to perform a particular task or sub-task, or how many errors a user makes in performing a task (Nielsen, 1993).	Performance testing can be used to create usability benchmarks in the early stages of the product development, and can be used to compare for example, error rates in different products or different configurations of the same product. Performance testing requires rigorous experimental designs and the performance criteria have to be measurable. Such performance criteria were not thought to be an effective method of establishing stable usability properties of FMSs.

Appendix B - Interview Schedule Derived From Usability Literature

Study Aim: To develop a measure for assessing the usability of Flight Management System CDUs.

Rationale: FAA/JAA requirements; Human Factors Flight Deck Certification; Pilot input to design and development of future FMSs.

Visual Clarity & Information Presentation

Is information displayed on the screen clear?

Is the FMS in colour or monochrome?

Is highlighting and colour (if applicable) effectively used?

Does the CDU contain ambiguities?

Is it CDU easy to read?

Is the information well organised?

In terms of inputting information does the FMS require data in the order in which you would expect?

Consistency

Does the system look consistent at all times?

Are all the screens structured in a consistent format?

Do you enter information in a consistent manner?

Does the system respond in a consistent manner?

Compatibility

If you have undertaken other type-ratings, did you find the FMS on board this aircraft to be compatible with that on your previous aircraft?

Was there much commonality between the FMS CDUs?

Are the data entry mechanisms on this FMS CDU similar to your previous FMS CDU?

In your experience, do the FMS CDUs within the fleet of aircraft on which you are currently type-rated respond in a similar manner?

Feedback

Is the Feedback provided by the FMS CDU informative?

Does the feedback provide information about where you are in the FMS CDU page structure?

Is it clear what actions they have taken, whether these actions have been successful and what actions should be taken next?

If you input information incorrectly does the FMS give you an informative error message?

Explicitness

Is the way the FMS CDU works clear to you?

Are the ways in which the FMS CDU is structured, and the reasons why it is structured as it is, evident to you?

Appropriate Functionality

Does the FMS do everything that you need it to do, does it meet your requirements?

Are there any tasks that you think the FMS CDU on board your aircraft should or could do that it doesn't do at the moment?

Effectiveness

How effective is the CDU at performing the tasks for which it was designed?

How effective is the CDU in assisting you achieve the tasks you wish to / or need to perform?

Flexibility

Are there parts of the FMS that you use frequently that are difficult to access?

Is it easy to change a waypoint in flight?

Are there parts of the FMS that you never use?

Do you feel in control of the FMS CDU at all times?

How difficult is it to recover from errors in programming the FMS?

Does the FMS ever surprise you?

Do you feel you have any difficulty tracking FMS status in any flight contexts?

If something goes wrong would your reaction be to attempt to make a corrective action on the FMS or to fly manually?

Error Prevention & Correction

Does the FMS minimise the possibility of you making an error?

Does the FMS have facilities for detecting and handling those that do occur?

Can pilots check their inputs and correct errors, or potential error situations before the inputs are processed?

Does it tell you what corrective action is required, when an error has been made?

Help Facilities

Are the FMS manuals you have on board the aircraft easy to use?

Are they over-complicated, over-simplified or pitched at the right level?

Do you have any "on-line" help available with the FMS on board your aircraft? If so, if it easy to use.

Workload

Does the FMS ever add to your mental workload at an inappropriate time?

Does the FMS ever add to your physical workload at an inappropriate time?

Can you identify any factors that make the FMS on board your aircraft difficult to use?

Can you think of any mechanisms of making the FMS on board your aircraft easier to use?

How applicable are the factors discussed to the FMS on board your aircraft?

Can you think of any factors pertinent to FMS usability that we have not discussed?

Learnability

Was use of the CDU easy to learn so that you could rapidly start getting some work done with the system?

Did you feel competent in using the FMS at the end of line training?

Would you say its structure is intuitive?

Is it easy to remember how to use the FMS after being away from it for some time?

How easy is it to remember where information is contained within the menu structure?

Efficiency

Does the FMS CDU allow you to do the required tasks as efficiently as possible?

Does the FMS CDU allow a high level of productivity?

How efficient do you find the input device?

Does the FMS CDU allow changes at an appropriate rate for other activities going on in the flight deck?

How efficient do you find the FMS menu structure through which you navigate?

Are short cuts available to navigating through the page structure?

Attitude

Do you like using the FMS?

Do you find the FMS difficult to use?

Do you ever find it frustrating to use?

Would you say the FMS is pleasant to use, is it satisfying to use?

Are you aware of some pilots finding the FMS easier to use than others?

Appendix C - Summary of Interview Themes

Usability Theme	*Comments Extracted from Interview Transcripts*
Visual Clarity	"Quality reduced by dust". "Quality reduced by static". "Sun can reduce visibility". "Screen is very small which is tedious". "Display could do with being angled better".
Feedback	"Errors are not defined". "Causes of errors have to be inferred". "Feedback is very subtle". "Companies have to put procedures in place for checking, to compensate for poor feedback". "Because pulling out a circuit breaker can frequently fix a problem – danger of becoming complacent when things break down". "Can be harsh e.g. "invalid entry"". "Important to keep everything interactive e.g. problem with autocall "radio altimeter"". "We've had to come up with a whole procedure for the operation of the aircraft to overcome the fact that the FMS is quite capable of giving you a wrong navigation waypoint." "You can have the same annunciations for different modes". "When new items appear on the FMS they are boxed for fifteen seconds, you can miss them if you are not looking at that precise moment". "You have to learn by making mistakes". "Can be misleading e.g. displaying less fuel than exists because when route is updated previous waypoints are not deleted automatically". "You could have waypoints backtracking on the route and it does not identify them".
Structure	"No logic or relationship identified between frequently used elements and ease of access to those elements." "Can't put you finger on button/page that you need".
Compatibility / Consistency	"Depends on fleet background". "The same pilot does not always use the FMS consistently e.g. they do not always do the set-up in the same pattern". "Things are not always in the same place on every screen, it's not consistent" "Should be able to use waypoints in "English" rather than searching for the three letter code". "If you come from a Boeing background it is quite a leap to go onto an Airbus aircraft."

Usability Theme	*Comments Extracted from Interview Transcripts*
CDU Structure	"Could be improved". "The sequence of inputting data does not follow as your would expect". "Navigation tree to illustrate page architecture would be useful". "Doesn't seem to have any logic to it". "There is some information you cannot access depending on the flight phase (A320) e.g. altering weight once doors are closed has to be done in a roundabout manner". "A lot of information is duplicated and doesn't need to be there". "Frequently many steps required to get information desired". "Have to scroll through too many pages". "There is some information you cannot access depending on the flight phase (A320) e.g. altering weight once doors are closed has to be done in a roundabout manner" "It can take ten or twelve button presses on the 757/767 to find IRS positions". "Can't always think where things are". "Not always practical e.g. A300 can't program IRSs without zero fuel weight, information which is received very late".
Workload	"It can increase your workload at inappropriate times especially with holding patterns" "It's not just mental workload, it is physical workload too"
Appropriate Functionality	"Good en-route". "Poor in descent". "Not used below 15,000ft." "There are parts of FMS rarely used because they are poorly implemented". "Sometimes you can't do things because it will have "page update in progress"". "Could do with inclusion of ATC frequencies" "If you want to add a waypoint, it is sometimes difficult to work out which waypoints it goes between".
Flexibility	"Good laterally". "Not good vertically - may be a good thing". "Not good enough to meet ATC requirements". "Sometimes height restriction means you have to do things like put in a fix a busy point". "You have to adapt to its limitations". "You can over write any error message on the 737-400, you cannot on the 757/767".

Usability Theme	*Comments Extracted from Interview Transcripts*
Error Prevention & Correction	"Poor". "It only does basic range checks". "Complete operating procedures have been designed to overcome weaknesses in FMS design". "The error prevention is in the procedures we use for operating it" "Some FMSs can identify some inconsistencies e.g. you cannot put in a westerly departure on an easterly runway on a 737-400". "The FMS does always tell you where it was shut down, so if you put in something at a distance it won't accept it first time."
Help Facilities	"Systems too basic for on-line help". "Manuals written for software experts not pilots". "Manuals only referred to in training". "Manuals are a waste of space"."
Memorable	"Don't want it to be too memorable because of fleet transfers". "Can't always remember where things are within the structure". "People forget where stuff is".
Speed	"Slow". "Not quick enough for ATC".
Attitude	"Job would be a lot harder without FMS" "No. Very irritating." "It can make you edgy when a pilot is flying it in a way that you wouldn't" "Tedious to scroll through". "Frustrating." "It can be frustrating e.g. when you can't get rid of a height restriction".
Learnability	"Not easy to learn to begin with".
Intuitiveness	"Switch/button facilities could be located more intuitively". "The way it works is not very intuitive".
Control Devices	"Could certainly be better". "QWERTY Vs Alphabetical Keyboard". "Box around NESW keys make them difficult to find". "Mouse as an optional input device". "Input device not always satisfactory". "Keyboard is a disaster". "Keypad is upside-down".

Suggested Improvements	"Letter search facility for waypoints". "Inclusion of FIR boundary as a waypoint should be consistent." "It would be good if ATIS information could be received automatically as ACARS is. Inclusion of Frequencies would be very helpful". "It could be structured better e.g. when you enter a route, it could ask you which runway and give you a list of options". "Index would be good. Inclusion of an overview of the FMS structure". "Reduce number of alternatives provides for waypoints or at least list them such that the closest is at the top of the list" "Feedback from the FMA is critical so maybe it should be more prominent - arguably a panel by itself". "Graphical User Interface". "Improvements to keyboard/input devices". "Including the Minimum Safe Altitude (MSA) would be good. It can be dangerous working off a legs page while the MSA is on a piece paper." "Having ATIS information automatically available would be good. If the aircraft has an engine failure it automatically notifies maintenance, but it doesn't automatically receive weather updates!" "Routes aren't always coded well". "MSA in red would be good." "Adding colour for emphasis would be good". "For the FMS to work well you really need two variants - one for short haul and one for long haul". "Bigger databases e.g. inclusion of taxi charts".

System Usability and Limitation	"Doesn't update winds". "On 757/767 you can put in a fixed speed & alt., or alt. only. You cannot however put in speed only which would be useful." "Errors in software still exist e.g. In A320 if you choose to go down at 3,000 ft a minute it will only go to 2,800ft. a minute". "Even for long haul flights the cruise page isn't big enough to contain all the information you need, so you have to go back to the route pages" "Have to input information at a rate slow enough for the processor or it freezes" "The 757/767 does not visually show you what it is doing in the hold. So you can't feel in control of it." "The 757/767 has an anomaly that if you manually tune in a beacon, the FMS will always dump the entire route". "In heading select mode the hold is not updated (757/767)". "On performance page (757/767) you can't get gross weight and reference speeds until 200 miles into, or half way through flight. Very late for a short sector." "Does not allow for turbulence in its calculation of max./optimum flight level." "No temperature deviation is available on A320". "On 757/767 if you press "direct to" a waypoint you have passed the aircraft will turn 180° and go in wrong direction." "Sometimes company routes have errors or miss out waypoints". "User expectations come to match known limitations of FMS". "On some aircraft there are figures in the database that have to be deleted every time."
Training	"Focus is on failure modes rather than standard operation".

Appendix D - Study One Pre-Pilot Questionnaire

Flight Management System Usability Questionnaire

This questionnaire has been devised by the College of Aeronautics, Cranfield University, UK to assess the usability of Flight Management Systems across a range of aircraft fleets. The information elicited from this study will be used to compare the usability of current FMSs & will assist manufacturers in developing and evaluating future FMSs such that they are more usable than the existing generation. Participation in this study is voluntary, and anonymity is guaranteed. Please answer all questions as honestly as possible.

When you have completed this questionnaire please return it using the FREEPOST envelope which is enclosed.

If you have any questions regarding the purpose or findings of this study, please feel free to contact the Human Factors Technology Group, College of Aeronautics, Cranfield University, Cranfield, Bedford MK43 OAL, UK. (Tel: 00 44 1234 750111 x 5196; e-mail: k.p.lane@cranfield.ac.uk).

Instructions: Please complete the remainder of this questionnaire by rating all of the statements below such that 1 = Strongly Disagree & 7 = Strongly Agree. Tick the number which you feel is most appropriate for each statement.

1. The information in the FMS display (CDU) is not easy to see and read.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

2. The FMS display (CDU) is sometimes cluttered.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

3. Important information is highlighted on the screen.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

4. Information is organised logically on the screen.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

5. The way the FMS looks & works is consistent at all times.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

6. Abbreviations, acronyms & codes are not used consistently throughout the FMS.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

7. The format & method of entering information is not consistent throughout the FMS.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

8. The way the FMS responds to a particular user action is consistent at all times.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

9. The FMS sometimes makes unexpected actions.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

10. I prefer to stick to the FMS facilities I know best.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

11. Established conventions are not followed for the format in which information is displayed.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

12. The sequence of activities required in use of the FMS as I would expect.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

13. The FMS works the way I think it should work.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

14. FMS buttons/switches could be located more intuitively.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

15. The feedback messages displayed by the FMS are frequently inadequate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

16. The FMS makes it clear what changes will occur as a result of an input being made.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

17. When using the FMS I sometimes lose track of where I am in the system; of what I am doing or have done; or of what the other pilot is doing or has done.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

18. The FMS clearly indicates which modes the system is currently in. It is informative & accurate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

19. FMS error messages do not explain clearly where errors are, what the errors are & why they have occurred.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

20. It is clear to the user what should be done to correct any error that occurs.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

21. The structure of the FMS is not obvious to the user.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

22. It is clear why the system is organised & structured as it is
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

23. It is clear why the screens are sequenced as they are.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

24. The structure of the FMS is well organised from the user's point of view.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

25. The FMS does not perform all the functions that I need it to.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

26. FMS users can access all the information they require in all flight phases.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

27. The FMS screens provide all the information which is relevant to the task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

28. The system allows users to do what they feel is necessary in order to carry out a task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

29. There is no logic or relationship between frequently used parts of the FMS & ease of access to those parts.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

30. The FMS processes information too slowly or other activities going on in the flight deck.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

31. FMS users have control over the order in which they request information, or carry out a series of activities.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

32. The FMS allows you to do the required tasks as efficiently as possible.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

33. The FMS is not adequately flexible to meet Air Traffic Control requirements.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

34. It is easy to return to the main menu from any part of the FMS architecture.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

35. I do not feel in control of the FMS at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

36. The FMS is adequately flexible for all navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

37. It is not always possible to check inputs to the FMS & to correct errors (or potential errors) before inputs are processed.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

38. The FMS has facilities for detecting and handling errors that occur.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

39. The FMS is not protected against common trivial errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

40. The FMS is protected against possible knock-on effects of changes in one part of the system.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

41. It was easy to learn how to use the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

42. It is sometimes difficult to remember where information is located within the FMS architecture.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

43. The FMS keyboard is difficult or awkward to use.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

44. The help facility (either on-line or manuals) available is adequate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

45. It is irritating to have to input information at a rate slow enough for the FMS processor.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

46. The FMS can increase mental workload at inappropriate times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

47. The FMS is pleasant to use; I like it.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

Personal Details

Sex: Male/Female Age:_____
Aircraft Fleet: _____ Hours on Type:_____
Please delete as appropriate: Captain / First Officer / Second Officer
Other Type Ratings (include Hours): _____
Total Hours:_____
Type of FMS (e.g. Honeywell 757-200 PIP) on your current flight deck:_____
Do you use a personal computer at home: Yes / No

Please outline any usability problems you have encountered with the FMS on board your aircraft in the space below or attached to the completed questionnaire.

Please list any improvements to the FMS which you would like to see in future flight decks.

Thank you for taking the time to complete this questionnaire. Please return the completed questionnaire in the enclosed FREEPOST envelope as soon as possible.

Appendix E -
Study One Pilot Questionnaire

Flight Management System Usability Questionnaire

This questionnaire has been devised by the College of Aeronautics, Cranfield University, UK, to assess the usability of Flight Management Systems (FMSs) across a range of aircraft fleets. The information elicited from this study will be used to compare the usability of current FMSs and will assist manufacturers in developing and evaluating future FMSs such that they are more usable than the existing generation. Participation in this study is voluntary, and anonymity is guaranteed. The airline you work for will have no access to your responses. Please answer all questions as honestly as possible. When you have completed this questionnaire please return it using the FREEPOST envelope which is enclosed.

If you have any questions regarding the purpose or findings of this study, please feel free to contact the Human Factors Technology Group, College of Aeronautics, Cranfield University, Cranfield, Bedford MK43 OAL, UK. (Tel: 00 44 1234 750111 x 5196; e-mail: k.p.lane@cranfield.ac.uk).

Instructions: Please complete this questionnaire by rating all of the statements below such that 1 = Strongly Disagree & 7 = Strongly Agree. Tick the number which you feel is most appropriate for each statement (e.g.✓). Please remember at all times that you are rating the FMS Control and Display Unit (CDU) only, NOT the navigation display.

1. The information in the FMS display (CDU) is not easy to see and read.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

2. The FMS display (CDU) is sometimes cluttered.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

3. Important information is adequately highlighted on the screen.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

4. Information is organised logically on the CDU screen (familiarity aside).
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

5. The way the FMS (CDU) looks and works is consistent at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

6. Abbreviations, acronyms and codes are not used consistently throughout the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

7. The format and method of entering information is not consistent throughout the operation of the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

9. The FMS sometimes makes unexpected actions.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

10. I prefer to stick to the FMS functions and facilities I know best.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

11. Established conventions are not followed for the format in which information is displayed (e.g. layout of dates).
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

12. The sequence of activities required in the use of the FMS is as I would expect.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

13. The FMS works the way I think it should work.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

14. The FMS buttons/switches could be located more intuitively.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

15. The feedback messages displayed by the FMS are frequently inadequate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

16. The FMS makes it clear what changes will occur as a result of an input being made.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

17. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

18. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

19. The FMS clearly indicates which modes the system is currently in. It is informative and accurate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

20. FMS error messages do not explain clearly where errors are, what the errors are and why they have occurred.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

21. It is clear to the user what should be done to correct any error that occurs.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

22. The structure of the FMS is not obvious to the user.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

23. It is clear why the system is organised and structured as it is.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

24. It is clear why the FMS screens are sequenced as they are.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

25. The structure of the FMS is well organised from the user's point of view.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

26. The FMS does not perform all the functions that I need it to.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

27. FMS users can access all the information they require in all flight phases.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

28. The FMS screens provide all the information which is relevant to the task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

29. The system allows users to do what they feel is necessary in order to carry out a task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

30. There is no logic or relationship between frequently used parts of the FMS and ease of access to those parts.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

31. The FMS processes information too slowly for other activities going on in the flight deck.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

32. FMS users have control over the order in which they request information, or carry out a series of activities.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

33. The FMS allows the user to do the required task as efficiently as possible.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

34. It is easy to return to the main menu from any screen of the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

35. I do not feel in control of the FMS at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

36. The FMS is not adequately flexible for all lateral navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

37. The FMS is adequately flexible for all vertical navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

38. It is not always possible to check or correct inputs to the FMS before they are processed.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

39. The FMS has facilities for detecting and handling input errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

40. The FMS is not protected against common trivial entry errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

41. Changes in one part of the FMS may produce subsequent adverse effects on other parts of the system.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

42. It was easy to learn how to use the FMS.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

43. It is sometimes difficult to remember where information is located within the FMS architecture.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

44. The FMS keyboard is difficult or awkward to use.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

45. The FMS help facility (either on-line or manuals) is adequate.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

46. It is irritating to have to input information at a rate slow enough for the FMS processor.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

47. The FMS can increase mental workload at inappropriate times.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

48. The FMS can increase physical workload at inappropriate times.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

49. The FMS is pleasant to use; I like it.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

Personal Details (for analysis purposes only)

Gender: Male/Female Age:_____ Airline:_____

Aircraft Fleet: _____ Hours on Type:_____

Please delete as appropriate: Captain / First Officer / Second Officer

Other Type Ratings (include Hours): _____

Total Hours:_____

Type of FMS (e.g. Honeywell 757-200 PIP) on your current flight deck:_____

Do you use a personal computer at home: Yes / No

If you wish you may use the space below or attach extra paper to outline any problems you have encountered with the FMS on board your aircraft. You may also wish to list any improvements to the FMS you would like to see in future flight decks.

Thank you for taking the time to complete this questionnaire. Please return the completed questionnaire in the enclosed FREEPOST envelope as soon as possible.

Appendix F -
Study One Final Questionnaire

Flight Management System Usability Questionnaire

This questionnaire is the product of a number of development studies undertaken by the College of Aeronautics, Cranfield University, UK, to assess the usability of Flight Management Systems (FMSs) across a range of aircraft fleets. The information elicited from this study will be used to compare the usability of current FMSs and will assist manufacturers in developing and evaluating future FMSs such that they are more usable than the existing generation. Participation in this study is voluntary, and anonymity is guaranteed. The airline you work for will have no access to your responses. Please answer all questions as honestly as possible. When you have completed this questionnaire please return it using the FREEPOST envelope which is enclosed.

If you have any questions regarding the purpose or findings of this study, please feel free to contact the Human Factors Technology Group, College of Aeronautics, Cranfield University, Cranfield, Bedford MK43 OAL, UK. (Tel: 00 44 1234 750111 x 5196; e-mail: k.p.lane@cranfield.ac.uk).

Instructions: Please complete this questionnaire by rating all of the statements below such that 1 = Strongly Disagree & 7 = Strongly Agree. Tick the number which you feel is most appropriate for each statement (e.g.☑). Please remember at all times that you are rating the FMS Control and Display Unit (CDU) only, NOT the navigation display.

1. The information in the FMS display (CDU) is easy to see and read.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

2. The FMS display (CDU) is sometimes cluttered.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

3. Important information is adequately highlighted on the screen.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

4. Information is organised logically on the CDU screen (familiarity aside).
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

5. The way the FMS (CDU) looks and works is consistent at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

6. Abbreviations, acronyms and codes are used consistently throughout the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

7. The format and method of entering information is consistent throughout the operation of the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

9. The FMS sometimes makes unexpected actions.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

10. I prefer to stick to the FMS functions and facilities I know best.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

11. Established conventions are followed for the format in which information is displayed (e.g. layout of dates).
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

12. The sequence of activities required in the use of the FMS is as I would expect.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

13. The FMS works the way I think it should work.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

14. The FMS buttons/switches could be located more intuitively.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

15. The feedback messages displayed by the FMS are frequently inadequate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

16. The FMS makes it clear what changes will occur as a result of an input being made.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

17. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

18. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

19. The FMS clearly indicates which mode the system is currently in. It is informative and accurate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

20. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

21. It is clear to the user what should be done to correct any error that occurs.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

22. The structure of the FMS is obvious to the user.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

23. It is clear why the system is organised and structured as it is.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

24. It is clear why the FMS screens are sequenced as they are.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

25. The structure of the FMS is well organised from the user's point of view.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

26. The FMS performs all the functions that I need it to.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

27. FMS users can access all the information they require in all flight phases.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

28. The FMS screens provide all the information which is relevant to the task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

29. The system allows users to do what they feel is necessary in order to carry out a task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

30. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

31. The FMS processes information too slowly for other activities in the flight deck.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

32. FMS users have control over the order in which they request information, or carry out a series of activities.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

33. The FMS allows the user to do the required task as efficiently as possible.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

34. It is easy to return to the main menu from any screen of the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

35. I feel in control of the FMS at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

36. The FMS is adequately flexible for all lateral navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

37. The FMS is adequately flexible for all vertical navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

38. It is always possible to check or correct inputs to the FMS before they are processed.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

39. The FMS has facilities for detecting and handling input errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

40. The FMS is protected against common trivial entry errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

41. Changes in one part of the FMS may produce subsequent adverse effects in other parts of the system.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

42. It was easy to learn how to use the FMS.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

43. It is sometimes difficult to remember where information is located within the FMS architecture.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

44. The FMS keyboard is difficult or awkward to use.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

45. The FMS help facility (either on-line or manuals) is adequate.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

46. It is irritating to have to input information at a rate slow enough for the FMS processor.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

47. The FMS can increase mental workload at inappropriate times.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

48. The FMS can increase physical workload at inappropriate times.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

49. The FMS is pleasant to use; I like it.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

Personal Details (for analysis purposes only)

Gender: Male/Female Age:_____ Airline:_____

Aircraft Fleet: _____Series_____ Hours on Type:_____

Please delete as appropriate: Captain / First Officer / Second Officer

Total Hours:_____

Type of FMS (e.g. Honeywell 757-200 PIP) on your current flight deck:_____

If you wish you may use the space below or attach extra paper to outline any problems you have encountered with the FMS on board your aircraft. You may also wish to list any improvements to the FMS you would like to see in future flight decks.

Thank you for taking the time to complete this questionnaire. Please return the completed questionnaire in the enclosed FREEPOST envelope as soon as possible.

Appendix G - Participation Invitation to Airlines

HUMAN FACTORS GROUP

Human Factors Group,
College of Aeronautics,
Cranfield University,
Bedfordshire MK43 OAL

Tel.: +44 1234 750111 x 5196
Fax.: +44 1234 750192
e-mail: k.p.lane@cranfield.ac.uk

Capt. (Name),
Chief Pilot,
Airline Co. Ltd.,
Airport,
Post Code.

15 March 1999

Dear Capt. Name,

The Human Factors Group, College of Aeronautics, is currently involved in a research project aimed at developing a measure for assessing Flight Management System Control and Display Unit (CDU) usability.

In order to validate the questionnaire that has been developed, access is required to a sample of line pilots. A number of airlines in the US and Canada have already participated in this study, but to ensure completeness it is essential that some participation from airlines in the UK is also achieved.

The usability tool being developed could potentially be used by manufacturers to examine the usability of early FMS CDU prototypes and to compare alternative configurations in a very cost-effective manner. The more important benefit, from the pilot perspective, is that this research is feeding into a process to create a more user friendly and safe flight deck.

Should Airline1 choose to participate in this study, a questionnaire and FREEPOST envelope will be supplied to each pilot. All questionnaire responses will be treated in a completely confidential manner. I have enclosed a copy of the questionnaire, and hope you will appreciate the value of this research.

Thanking you in advance for your assistance,

Yours sincerely,

Karen P. Lane BA MSc
Don Harris, PhD MSc

Appendix H -
Study Two Final Questionnaires

Flight Management System Usability Questionnaire

This questionnaire is the product of a number of development studies undertaken by the College of Aeronautics, Cranfield University, UK, to assess the usability of Flight Management System Control and Display Units across various flight phases and aircraft type. The information elicited from this study will be used to compare the usability of current FMSs and will assist manufacturers in developing and evaluating future FMSs such that they are more usable than the existing generation. *This questionnaire is concerned with the CRUISE phase of flight only.* Some of your colleagues will complete a similar questionnaire for other phases of flight, but it is important that you complete the questionnaire making all ratings applicable to the cruise phase of flight only. Participation in this study is voluntary, and anonymity is guaranteed. The airline you work for will have no access to your responses. Please answer all questions as honestly as possible. When you have completed this questionnaire please return it using the FREEPOST envelope which is enclosed.

If you have any questions regarding the purpose or findings of this study, please feel free to contact the Human Factors Group, College of Aeronautics, Cranfield University, Cranfield, Bedford MK43 OAL, UK. (Tel: 00 44 1234 750111 x 5196; e-mail: k.p.lane@cranfield.ac.uk).

Instructions: This questionnaire is concerned with the CRUISE phase of flight only. Please complete this questionnaire by rating all of the statements below such that 1 = Strongly Disagree & 7 = Strongly Agree. Tick the number which you feel is most appropriate for each statement (e.g. ⊘). Please remember at all times that you are rating the FMS Control and Display Unit (CDU) only, NOT the navigation display.

1. The information in the FMS display (CDU) is easy to see and read.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

2. The FMS display (CDU) is sometimes cluttered.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

3. Important information is adequately highlighted on the screen.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

4. Information is organised logically on the CDU screen (familiarity aside).
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

5. The way the FMS (CDU) looks and works is consistent at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

6. Abbreviations, acronyms and codes are used consistently throughout the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

7. The format and method of entering information is consistent throughout the operation of the FMS.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

9. The FMS sometimes makes unexpected actions.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

10. The sequence of activities required in the use of the FMS is as I would expect.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

11. The FMS works the way I think it should work.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

12. The FMS buttons/switches could be located more intuitively.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

13. The feedback messages displayed by the FMS are frequently inadequate.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

14. The FMS makes it clear what changes will occur as a result of an input being made.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

17. The FMS clearly indicates which mode the system is currently in. It is informative and accurate.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

19. It is clear to the user what should be done to correct any error that occurs.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

20. The structure of the FMS is obvious to the user.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

21. It is clear why the system is organised and structured as it is.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

22. It is clear why the FMS screens are sequenced as they are.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

23. The structure of the FMS is well organised from the user's point of view.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

24. The FMS performs all the functions that I need it to.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

25. FMS users can access all the information required during this flight phase (cruise).
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

26. The FMS screens provide all the information which is relevant to the task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

27. The system allows users to do what they feel is necessary in order to carry out a task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

29. The FMS processes information too slowly for other activities in the flight deck.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

30. The FMS allows the user to do the required task as efficiently as possible.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

31. It is easy to return to the main menu from any screen of the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

32. I feel in control of the FMS at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

33. The FMS is adequately flexible for all lateral navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

34. The FMS is adequately flexible for all vertical navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

35. It is always possible to check or correct inputs to the FMS before they are processed.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

36. The FMS has facilities for detecting and handling input errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

37. The FMS is protected against common trivial entry errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

38. The FMS keyboard is difficult or awkward to use.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

39. The FMS help facility (either on-line or manuals) is adequate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

40. It is irritating to have to input information at a rate slow enough for the FMS processor.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

41. The FMS can increase mental workload at inappropriate times.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

42. The FMS can increase physical workload at inappropriate times.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

43. The FMS is pleasant to use; I like it.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

Personal Details (for analysis purposes only)

Gender: Male/Female Age:_____ Airline:_____

Aircraft Fleet: _____Series_____ Hours on Type:_____

Please delete as appropriate: Captain / First Officer / Second Officer

Total Hours:_____

Type of FMS (e.g. Honeywell 757-200 PIP) on your current flight deck:_____

If you wish you may use the space below or attach extra paper to outline any problems you have encountered with the FMS on board your aircraft. You may also wish to list any improvements to the FMS you would like to see in future flight decks.

Thank you for taking the time to complete this questionnaire. Please return the completed questionnaire in the enclosed FREEPOST envelope as soon as possible.

Flight Management System Usability Questionnaire

This questionnaire is the product of a number of development studies undertaken by the College of Aeronautics, Cranfield University, UK, to assess the usability of Flight Management System Control and Display Units across various flight phases and aircraft type. The information elicited from this study will be used to compare the usability of current FMSs and will assist manufacturers in developing and evaluating future FMSs such that they are more usable than the existing generation. *This questionnaire is concerned with the DESCENT & APPROACH phase of flight only.* Some of your colleagues will complete a similar questionnaire for other phases of flight, but it is important that you complete the questionnaire making all ratings applicable to the cruise phase of flight only. Participation in this study is voluntary, and anonymity is guaranteed. The airline you work for will have no access to your responses. Please answer all questions as honestly as possible. When you have completed this questionnaire please return it using the FREEPOST envelope which is enclosed.

If you have any questions regarding the purpose or findings of this study, please feel free to contact the Human Factors Group, College of Aeronautics, Cranfield University, Cranfield, Bedford MK43 OAL, UK. (Tel: 00 44 1234 750111 x 5196; e-mail: k.p.lane@cranfield.ac.uk).

Instructions: This questionnaire is concerned with the DESCENT & APPROACH phase of flight only. Please complete this questionnaire by rating all of the statements below such that 1 = Strongly Disagree & 7 = Strongly Agree. Tick the number which you feel is most appropriate for each statement (e.g. ⊘). Please remember at all times that you are rating the FMS Control and Display Unit (CDU) only, NOT the navigation display.

1. The information in the FMS display (CDU) is easy to see and read.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

2. The FMS display (CDU) is sometimes cluttered.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

3. Important information is adequately highlighted on the screen.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

4. Information is organised logically on the CDU screen (familiarity aside).
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

5. The way the FMS (CDU) looks and works is consistent at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

6. Abbreviations, acronyms and codes are used consistently throughout the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

7. The format and method of entering information is consistent throughout the operation of the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

9. The FMS sometimes makes unexpected actions.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

10. The sequence of activities required in the use of the FMS is as I would expect.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

11. The FMS works the way I think it should work.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

12. The FMS buttons/switches could be located more intuitively.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

13. The feedback messages displayed by the FMS are frequently inadequate.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

14. The FMS makes it clear what changes will occur as a result of an input being made.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

17. The FMS clearly indicates which mode the system is currently in. It is informative and accurate.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

19. It is clear to the user what should be done to correct any error that occurs.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

20. The structure of the FMS is obvious to the user.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

21. It is clear why the system is organised and structured as it is.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

22. It is clear why the FMS screens are sequenced as they are.

Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

23. The structure of the FMS is well organised from the user's point of view.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

24. The FMS performs all the functions that I need it to.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

25. FMS users can access all the information required during this flight phase (descent & approach).
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

26. The FMS screens provide all the information which is relevant to the task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

27. The system allows users to do what they feel is necessary in order to carry out a task.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

29. The FMS processes information too slowly for other activities in the flight deck.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

30. The FMS allows the user to do the required task as efficiently as possible.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

31. It is easy to return to the main menu from any screen of the FMS.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

32. I feel in control of the FMS at all times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

33. The FMS is adequately flexible for all lateral navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

34. The FMS is adequately flexible for all vertical navigation.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

35. It is always possible to check or correct inputs to the FMS before they are processed.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

36. The FMS has facilities for detecting and handling input errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

37. The FMS is protected against common trivial entry errors.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

38. The FMS keyboard is difficult or awkward to use.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

39. The FMS help facility (either on-line or manuals) is adequate.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

40. It is irritating to have to input information at a rate slow enough for the FMS processor.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

41. The FMS can increase mental workload at inappropriate times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

42. The FMS can increase physical workload at inappropriate times.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

43. The FMS is pleasant to use; I like it.
Strongly Disagree ① ② ③ ④ ⑤ ⑥ ⑦ Strongly Agree

Personal Details (for analysis purposes only)

Gender: Male/Female Age:_____ Airline:_____
Aircraft Fleet: _____Series_____ Hours on Type:_____
Please delete as appropriate: Captain / First Officer / Second Officer
Total Hours:_____
Type of FMS (e.g. Honeywell 757-200 PIP) on your current flight deck:_____

If you wish you may use the space below or attach extra paper to outline any problems you have encountered with the FMS on board your aircraft. You may also wish to list any improvements to the FMS you would like to see in future flight decks.

Thank you for taking the time to complete this questionnaire. Please return the completed questionnaire in the enclosed FREEPOST envelope as soon as possible.

Appendix I - Letter to Flight Crew

HUMAN FACTORS GROUP

Human Factors Group,
College of Aeronautics,
Cranfield University,
Bedfordshire MK43 OAL

Tel.: +44 1234 750111 x 5196
Fax.: +44 1234 750192
e-mail: k.p.lane@cranfield.ac.uk

14 December 1999

Dear Flightcrew Member,

The Human Factors Group, College of Aeronautics, is currently involved in an international research project aimed at developing a questionnaire for assessing the usability of Flight Management Systems (FMS) Control and Display Units (CDUs). The phase of research currently being conducted is concerned with the usability of FMS CDUs across flight phases and aircraft type.

The usability questionnaire being developed could potentially be used by FMS manufacturers to examine the usability of early FMS CDU prototypes, and to compare alternative configurations in a very cost-effective manner. The more important benefit, from the pilot perspective, is that this research is feeding into a process to create a more user friendly and safer flight deck.

Airline 1 has chosen to participate in this study. All that is required of you is that you complete the questionnaire and return it to Cranfield University as soon as possible. A FREEPOST envelope is enclosed for your convenience. Your participation in this research is voluntary and your responses will be treated in complete confidence. If you require any further information on this research, please feel free to contact us at the above address.

Thanking you in advance for your assistance,

Yours sincerely,

Karen P. Lane BA MSc

Dr. Don Harris

Appendix J - Analysis of Usability Ratings by Flight Phase

Independent t-test Analyses Questionnaire Item Ratings Based on Flight Phase (Coding shown at End of Table)

Questionnaire Item	*Mean*	*sd*	*Mean*	*sd*	*t*	*df*	*Sig. (2-tailed)*
	Cruise		Descent & Approach				
1. The information in the FMS display (CDU) is easy to see and read.	5.24	1.14	5.11	1.19	1.05	352	0.30
2. The FMS display (CDU) is sometimes cluttered.	4.38	1.55	4.15	1.56	1.38	352	0.17
3. Important information is adequately highlighted on the screen.	**3.85**	**1.56**	**3.76**	**1.57**	**0.55**	**353**	**0.58**
4. Information is organised logically on the CDU screen (familiarity aside).	**4.45**	**1.43**	**4.26**	**1.36**	**1.24**	**351**	**0.22**
5. The way the FMS (CDU) look and works is consistent at all times.	**4.30**	**1.67**	**4.12**	**1.35**	**1.11***	**339***	**0.27***
6. Abbreviations, acronyms & codes are used consistently throughout the FMS.	**5.18**	**1.27**	**5.28**	**1.17**	**-0.83**	**352**	**0.41**
7. The format and method of entering information is consistent throughout the operation of the FMS.	4.66	1.56	4.63	1.48	0.15	353	0.88
8. The way the FMS responds to a particular user action is consistent at all times irrespective of mode.	**4.03**	**1.63**	**3.98**	**1.49**	**0.34**	**353**	**0.74**
9. The FMS sometimes makes unexpected actions.	3.93	1.84	3.73	1.74	1.06	351	0.29
10. The sequence of activities required in the use of the FMS is as I would expect.	**4.57**	**1.39**	**4.57**	**1.37**	**-0.01**	**350**	**0.99**
11. The FMS works the way I think it should work.	4.37	1.44	4.35	1.46	0.18	353	0.86
12. The FMS buttons could be located more intuitively.	3.84	1.68	3.66	1.77	0.96	350	0.34

Questionnaire Item	*Mean*	*sd*	*Mean*	*sd*	*t*	*df*	*Sig. (2-tailed)*
	Cruise		Descent & Approach				
13. The feedback messages displayed by the FMS are frequently inadequate.	4.02	1.60	3.73	1.66	1.64	351	0.10
14. The FMS makes it clear what changes will occur as a result of an input being made.	3.71	1.73	3.26	1.61	2.50*	350*	0.01*
15. When using the FMS I sometimes lose track of where I am in the system, of what I am doing, or have done.	**4.92**	**1.59**	**4.86**	**1.72**	**0.29**	**351**	**0.77**
16. When using the FMS I sometimes lose track of what the other pilot is doing, or has done.	**3.72**	**1.81**	**3.69**	**1.82**	**0.14**	**353**	**0.89**
17. The FMS clearly indicated which mode the system is currently in. It is informative and accurate.	4.50	1.49	4.21	1.53	1.79	351	0.07
18. FMS error messages explain clearly where errors are, what the errors are and why they have occurred.	**2.94**	**1.46**	**2.93**	**1.52**	**0.11**	**353**	**0.91**
19. It is clear to the user what should be done to correct any error that occurs.	**2.80**	**1.51**	**2.65**	**1.29**	**1.03***	**345***	**0.31***
20. The structure of the FMS is obvious to the user.	3.68	1.45	3.47	1.51	1.32	352	0.19
21. It is clear why the system is organised and structured as it is.	**3.72**	**1.50**	**3.58**	**1.50**	**0.83**	**352**	**0.41**
22. It is clear why the FMS screens are sequenced as they are.	4.39	1.38	4.19	1.46	1.31	353	0.19
23. The structure of the FMS is well organised from the user's point of view.	**4.35**	**1.37**	**4.07**	**1.41**	**1.84**	**352**	**0.07**
24. The FMS performs all the functions that I need it to.	**4.00**	**1.76**	**3.94**	**1.67**	**0.31***	**351***	**0.76***

Questionnaire Item	*Mean*	*sd*	*Mean*	*sd*	*t*	*df*	*Sig. (2-tailed)*
	Cruise		Descent & Approach				
25. FMS users can access all the information they require in this flight phase (Cruise or Descent & Approach).	**4.84**	**1.65**	**4.59**	**1.47**	**1.49**	**350**	**0.14**
26. The FMS screens provide all the information which is relevant to the task.	**4.56**	**1.53**	**4.14**	**1.50**	**2.64**	**354**	**0.01**
27. The system allows users to do what they feel is necessary in order to carry out a task.	4.63	1.32	4.39	1.40	1.63	351	0.11
28. There is a clear logic or relationship between frequently used parts of the FMS and ease of access to those parts.	**4.53**	**1.40**	**4.28**	**1.37**	**1.66**	**352**	**0.10**
29. The FMS processes information too slowly for other activities in the flight deck.	**3.05**	**1.89**	**3.02**	**1.85**	**0.17**	**351**	**0.87**
30. The FMS allows the user to do the required task as efficiently as possible.	**3.92**	**1.52**	**3.80**	**1.44**	**0.80**	**354**	**0.43**
31. It is easy to return to the main menu from any screen of the FMS.	5.22	1.62	5.03	1.61	1.11	353	0.27
32. I feel in control of the FMS at all times.	5.07	1.52	4.82	1.58	1.47	354	0.14
33. The FMS is adequately flexible for all lateral navigation.	**4.55**	**1.64**	**4.46**	**1.57**	**0.53**	**354**	**0.60**
34. The FMS is adequately flexible for all vertical navigation.	**3.81**	**1.64**	**3.53**	**1.55**	**1.68**	**354**	**0.09**
35. It is always possible to check or correct inputs to the FMS before they are processed.	4.45	1.74	4.15	1.66	1.63	354	0.10
36. The FMS has facilities for detecting and handling input errors.	**3.84**	**1.56**	**3.80**	**1.62**	**0.28**	**354**	**0.78**
37. The FMS is protected against common trivial entry errors.	**3.87**	**1.67**	**3.71**	**1.77**	**0.91**	**354**	**0.37**
38. The FMS keyboard is difficult or awkward to use.	4.59	1.70	4.27	1.76	1.73	350	0.08

Questionnaire Item	*Mean*	*sd*	*Mean*	*sd*	*t*	*df*	*Sig. (2-tailed)*
	Cruise		Descent & Approach				
39. The FMS help facility (either on-line or manuals) is adequate.	**2.97**	**1.60**	**2.79**	**1.56**	**1.03**	**348**	**0.30**
40. It is irritating to have to input information at a rate slow enough for the FMS processor.	**2.85**	**1.78**	**2.95**	**1.82**	**-0.50**	**350**	**0.62**
41. The FMS can increase mental workload at inappropriate times.	**3.15**	**1.75**	**2.79**	**1.60**	**2.04**	**349**	**0.04**
42. The FMS can increase physical workload at inappropriate times.	**3.79**	**1.79**	**3.56**	**1.86**	**1.14**	**349**	**0.26**
43. The FMS is pleasant to use; I like it.	5.14	1.36	4.86	1.42	0.92	348	0.06

* = Equal variances not assumed.
Bold Typeface = Included in Final Usability Scales
Highlighted rows = Ratings not significantly different based on flight phase rated ($p.<0.05$).

Appendix K - Curvilinear Regression Analyses for Pilots' Experience

Analysis of Pilots' experience (Hours on Current Aircraft Type) against Usability Scale Ratings

Structure & Sequence Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Structure & Sequence	Linear	.001	342	.17	.676	$4.2592 + (-5x10^{-6} \times hours)$
Structure & Sequence	Log.	.003	342	.87	.352	$4.9164 + (-0.0788 \times \log_{10} hours)$
Structure & Sequence	Quad.	.011	341	1.85	.159	$4.5747 + (-9x10^{-5} \times hours) + (4.6x10^{-9} \times hours^2)$
Structure & Sequence	Cubic	.011	340	1.27	.284	$4.6656 + (-1x10^{-4} \times hours) + (9.7x10^{-9} \times hours^2) + (-2x10^{-13} \times hours^3)$

Appropriate Functionality Scale & Pilots' Total Flying Hours

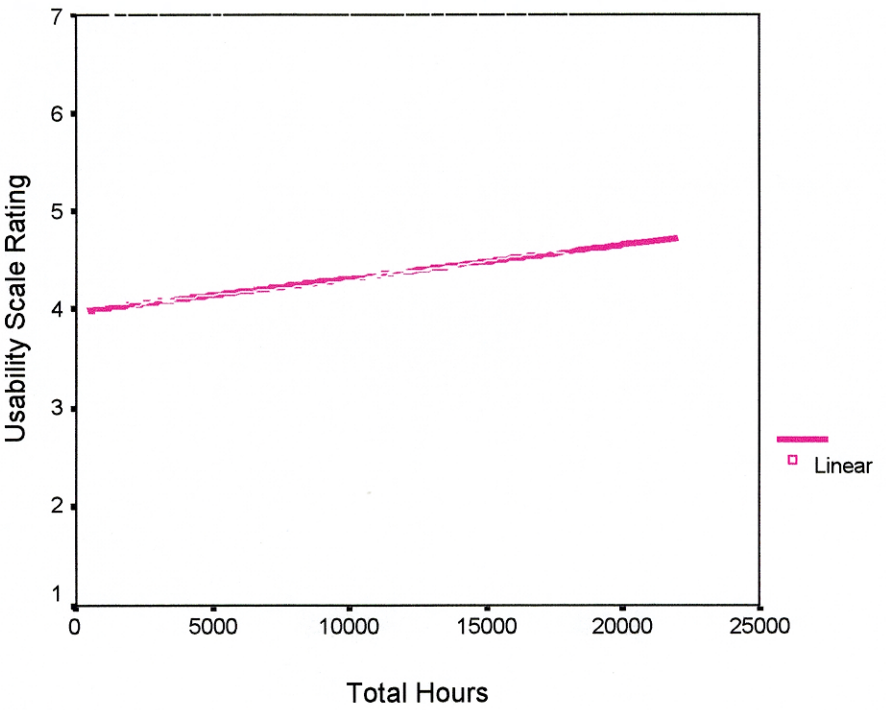
Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Appropriate Functionality	Linear	.000	341	.04	.839	$4.3636 + (2.8x10^{-6} \times hours)$
Appropriate Functionality	Log.	.000	341	5.9×10^{-4}	.981	$4.3694 + (0.0023 \times \log_{10} hours)$
Appropriate Functionality	Quad.	.001	340	.25	.779	$4.4924 + (-3x10^{-5} \times hours) + (1.9x10^{-9} \times hours^2)$
Appropriate Functionality	Cubic	.001	339	.17	.918	$4.4725 + (-2x10^{-5} \times hours) + (7.5x10^{-10} \times hours^2) + (3.6x10^{-14} \times hours^3)$

Feedback Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Feedback	Linear	.000	341	.15	.702	$2.8883 + (-5x10^{-6} \times hours)$
Feedback	Log.	.001	341	.36	.548	$3.3187 + (-0.0532 \times \log_{10} hours)$
Feedback	Quad.	.005	340	.85	.427	$3.1088 + (-7x10^{-5} \times hours) + (3.2x10^{-9} \times hours^2)$
Feedback	Cubic	.005	339	.59	.620	$3.0365 + (-3x10^{-5} \times hours) + (-9x10^{-10} \times hours^2) + (1.3x10^{-13} \times hours^3)$

Predictability Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Predictability	Linear	.012	346	4.21	.041	$3.9778 + (3.4 \times 10^{-5} \times \text{hours})$
Predictability	Log.	.010	346	3.43	.065	$2.4106 + (0.2111 \times \log_{10}\text{hours})$
Predictability	Quad.	.012	345	2.13	.121	$4.0303 + (1.9 \times 10^{-5} \times \text{hours}) + (7.7 \times 10^{-10} \times \text{hours}^2)$
Predictability	Cubic	.016	344	1.87	.135	$4.4178 + (-2 \times 10^{-4} \times \text{hours}) + (2.3 \times 10^{-8} \times \text{hours}^2) + (-7 \times 10^{-13} \times \text{hours}^3)$

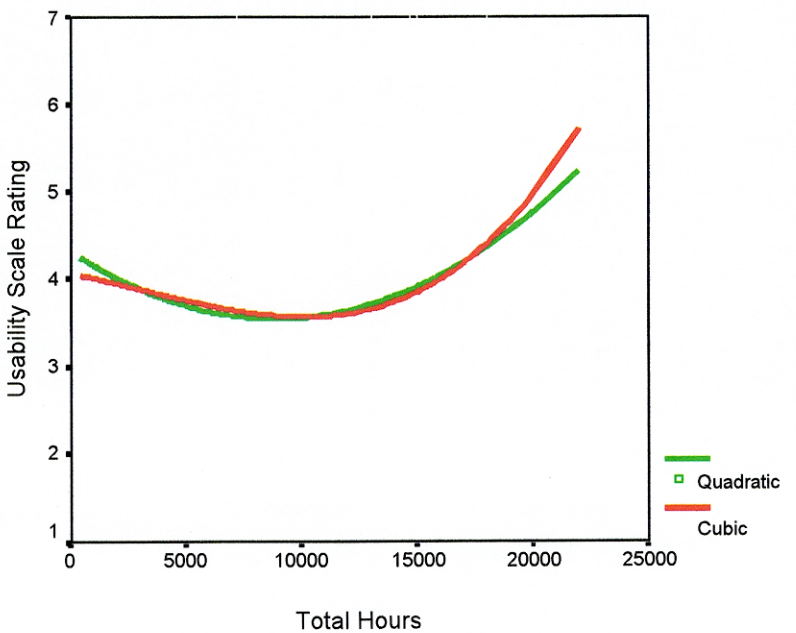


Information Presentation Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Information Presentation	Linear	.001	345	.37	.542	$4.4005 + (6.5 \times 10^{-6} \times \text{hours})$
Information Presentation	Log.	.000	345	1.2×10^{-5}	.997	$4.4643 + (-3 \times 10^{-4} \times \log_{10}\text{hours})$
Information Presentation	Quad.	.005	344	.80	.452	$4.5613 + (-4 \times 10^{-5} \times \text{hours}) + (2.4 \times 10^{-9} \times \text{hours}^2)$
Information Presentation	Cubic	.017	343	2.00	.114	$5.0058 + (-3 \times 10^{-4} \times \text{hours}) + (2.8 \times 10^{-8} \times \text{hours}^2) + (-8 \times 10^{-13} \times \text{hours}^3)$

Error Prevention & Correction Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Error Prevention & Correction	Linear	.003	349	1.05	.307	$3.6479 + (1.6x10^{-5} \times hours)$
Error Prevention & Correction	Log.	.000	349	.04	.834	$3.5984 + (0.0223 \times \log_{10} hours)$
Error Prevention & Correction	Quad.	.031	348	5.65	.004	$4.3174 + (-2x10^{-4} \times hours) + (9.8x10^{-9} \times hours^2)$
Error Prevention & Correction	Cubic	.033	347	3.98	.008	$4.0665 + (-5x10^{-5} \times hours) + (-4x10^{-9} \times hours^2) + (4.6x10^{-13} \times hours^3)$



Processor Speed Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Processor Speed	Linear	.000	342	.02	.884	$3.0088 + (-3x10^{-6} \times hours)$
Processor Speed	Log.	.000	342	.02	.880	$3.1510 + (-0.0186 \times \log_{10} hours)$
Processor Speed	Quad.	.000	341	.01	.989	$3.0162 + (-5x10^{-6} \times hours) + (1.1x10^{-10} \times hours^2)$
Processor Speed	Cubic	.000	340	.04	.990	$3.1279 + (-6x10^{-5} \times hours) + (6.4x10^{-9} \times hours^2) + (-2x10^{-13} \times hours^3)$

Workload Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Workload	Linear	.004	343	1.44	.232	$3.5135 + (-2x10^{-5} \times hours)$
Workload	Log.	.006	343	2.16	.142	$4.8610 + (-0.1719 \times \log_{10} hours)$
Workload	Quad.	.009	342	1.54	.215	$3.8131 + (-1x10^{-4} \times hours) + (4.3x10^{-9} \times hours^2)$
Workload	Cubic	.009	341	1.06	366	$3.7014 + (-5x10^{-5} \times hours) + (-2x10^{-9} \times hours^2) + (2.0x10^{-13} \times hours^3)$

Consistency Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Consistency	Linear	.004	347	1.42	.234	$4.2640 + (-2x10^{-5} \times hours)$
Consistency	Log.	.005	347	1.71	.192	$5.2401 + (-0.1269 \times \log_{10} hours)$
Consistency	Quad.	.006	346	1.03	.357	$4.4196 + (-6x10^{-5} \times hours) + (2.3x10^{-9} \times hours^2)$
Consistency	Cubic	.006	345	.70	.554	$4.3681 + (-4x10^{-5} \times hours) + (-6x10^{-10} \times hours^2) + (9.5x10^{-14} \times hours^3)$

Flexibility for Navigation Scale & Pilots' Total Flying Hours

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Flexibility for Navigation	Linear	.002	349	.67	.415	$3.9944 + (1.2x10^{-5} \times hours)$
Flexibility for Navigation	Log.	.001	349	.36	.551	$3.5737 + (0.0596 \times \log_{10} hours)$
Flexibility for Navigation	Quad.	.004	348	.70	.496	$4.1660 + (-4x10^{-5} \times hours) + (2.5x10^{-9} \times hours^2)$
Flexibility for Navigation	Cubic	.004	347	.52	.669	$4.2824 + (-9x10^{-5} \times hours) + (9.1x10^{-9} \times hours^2) + (-2x10^{-13} \times hours^3)$

Analysis of Pilots' experience (Hours on Current Aircraft Type) against Usability Scale Ratings

Structure & Sequence Scale & Pilots' Hours on Current Aircraft Type

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Structure & Sequence	Linear	.007	342	2.46	.117	$4.1035 + (4.1x10^{-5} \times hours)$
Structure & Sequence	Log.	.005	342	1.63	.202	$3.6870 + (0.0712 \times \log_{10} hours)$
Structure & Sequence	Quad.	.014	341	2.43	.090	$4.2249 + (-7x10^{-5} \times hours) + (1.5x10^{-8} \times hours^2)$
Structure & Sequence	Cubic	.021	340	2.46	.062	$4.0635 + (2x10^{-4} \times hours) + (-6x10^{-8} \times hours^2) + (5.0x10^{-12} \times hours^3)$

Appropriate Functionality Scale & Pilots' Hours on Current Aircraft Type

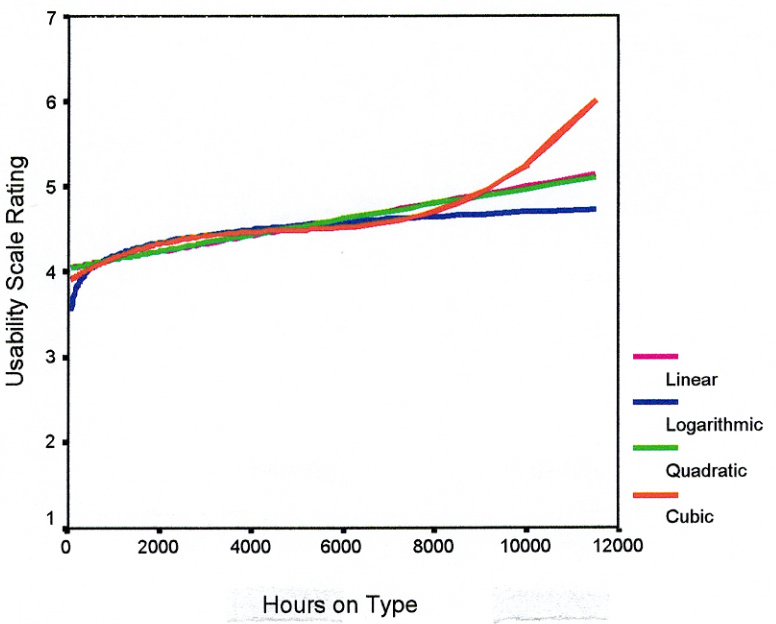
Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Appropriate Functionality	Linear	.004	341	1.30	.254	$4.2969 + (3.3x10^{-5} \times hours)$
Appropriate Functionality	Log.	.001	341	.21	.645	$4.1742 + (0.0289 \times \log_{10} hours)$
Appropriate Functionality	Quad.	.011	340	1.83	.162	$4.4328 + (-1x10^{-4} \times hours) + (1.7x10^{-8} \times hours^2)$
Appropriate Functionality	Cubic	.011	339	1.23	.298	$4.4085 + (-6x10^{-5} \times hours) + (6.3x10^{-9} \times hours^2) + (7.4x10^{-13} \times hours^3)$

Feedback Scale & Pilots' Hours on Current Aircraft Type

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Feedback	Linear	.001	341	.37	.542	$2.7953 + (1.7x10^{-5} \times hours)$
Feedback	Log.	.003	341	1.02	.313	$2.4063 + (0.0584 \times \log_{10} hours)$
Feedback	Quad.	.001	340	.20	.816	$2.7797 + (3.2x10^{-5} \times hours) + (-2x10^{-9} \times hours^2)$
Feedback	Cubic	.013	339	1.44	.231	$2.5709 + (4x10^{-4} \times hours) + (-9x10^{-8} \times hours^2) + (6.5x10^{-12} \times hours^3)$

Predictability Scale & Pilots' Hours on Current Aircraft Type

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Predictability	Linear	.020	346	7.13	.008	$4.0489 + (9.5x10^{-5} \times hours)$
Predictability	Log.	.024	346	8.44	.004	$2.6814 + (0.2187 \times \log_{10} hours)$
Predictability	Quad.	.020	345	3.55	.030	$4.0421 + (1x10^{-4} \times hours) + (-8x10^{-10} \times hours^2)$
Predictability	Cubic	.023	344	2.74	.043	$3.8958 + (3x10^{-4} \times hours) + (-6x10^{-8} \times hours^2) + (4.5x10^{-12} \times hours^3)$



Information Presentation Scale & Pilots' Hours on Current Aircraft Type

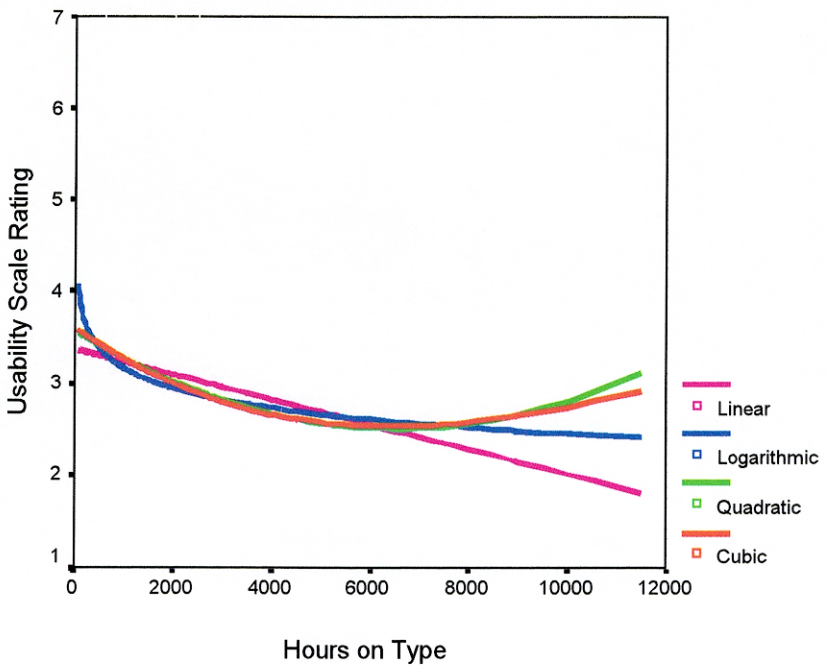
Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Information Presentation	Linear	.004	345	1.49	.223	$4.3860 + (2.8x10^{-5} \times hours)$
Information Presentation	Log.	.001	345	.46	.496	$4.2175 + (0.0331 \times \log_{10} hours)$
Information Presentation	Quad.	.009	344	1.52	.221	$4.4716 + (-5x10^{-5} \times hours) + (1.1x10^{-8} \times hours^2)$
Information Presentation	Cubic	.009	343	1.07	.364	$4.4353 + (2.3x10^{-6} \times hours) + (-5x10^{-9} \times hours^2) + (1.2x10^{-12} \times hours^3)$

Error Prevention & Correction Scale & Pilots' Hours on Current Aircraft Type

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Error Prevention & Correction	Linear	.002	349	.53	.469	$3.8560 + (-2x10^{-5} \times hours)$
Error Prevention & Correction	Log.	.003	349	.89	.346	$4.2852 + (-0.0665 \times \log_{10} hours)$
Error Prevention & Correction	Quad.	.007	348	1.26	.286	$3.9964 + (-2x10^{-4} \times hours) + (1.8x10^{-8} \times hours^2)$
Error Prevention & Correction	Cubic	.007	347	.87	.455	$3.9528 + (-9x10^{-5} \times hours) + (-1x10^{-9} \times hours^2) + (1.3x10^{-12} \times hours^3)$

Processor Speed Scale & Pilots' Hours on Current Aircraft Type

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Processor Speed	Linear	.036	342	12.94	.000	$3.3679 + (-1x10^{-4} \times hours)$
Processor Speed	Log.	.042	342	14.85	.000	$5.2785 + (-0.3067 \times \log_{10} hours)$
Processor Speed	Quad.	.045	341	8.06	.000	$3.5654 + (-3x10^{-4} \times hours) + (2.5x10^{-8} \times hours^2)$
Processor Speed	Cubic	.045	340	5.37	.001	$3.5961 + (-4x10^{-4} \times hours) + (3.8x10^{-8} \times hours^2) + (-1x10^{-12} \times hours^3)$



Workload Scale & Pilots' Hours on Current Aircraft Type

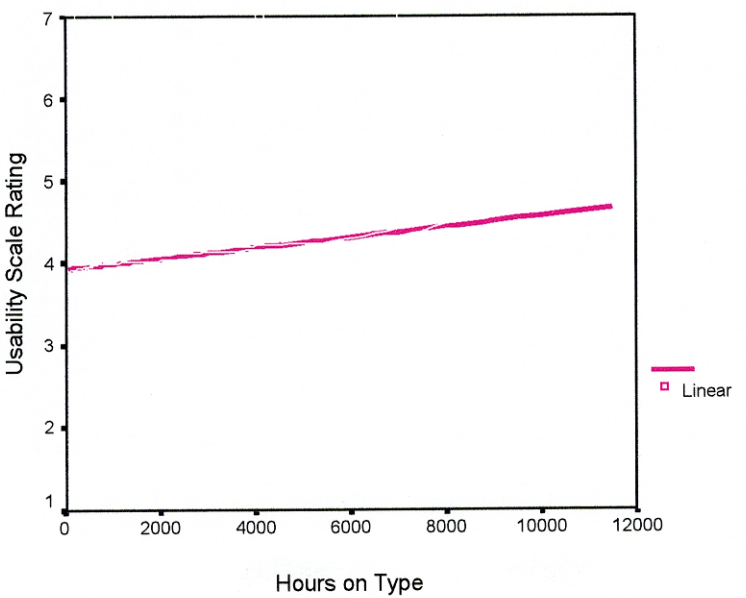
Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Workload	Linear	.003	343	1.07	.301	$3.4390 + (-4x10^{-5} \text{ x hours})$
Workload	Log.	.005	343	1.66	.199	$4.0732 + (-0.0990 \text{ x } \log_{10}\text{hours})$
Workload	Quad.	.014	342	2.49	.084	$3.6534 + (-2x10^{-4} \text{ x hours}) + (2.7x10^{-8} \text{ x hours}^2)$
Workload	Cubic	.015	341	1.72	.163	$3.5924 + (-1x10^{-4} \text{ x hours}) + (2.9x10^{-11} \text{ x hours}^2) + (1.9x10^{-12} \text{ x hours}^3)$

Consistency Scale & Pilots' Hours on Current Aircraft Type

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Consistency	Linear	.002	347	.60	.438	$4.1925 + (-2x10^{-5} \text{ x hours})$
Consistency	Log.	.003	347	1.18	.279	$4.6529 + (-0.0707 \text{ x } \log_{10}\text{hours})$
Consistency	Quad.	.007	346	1.25	.289	$4.3197 + (-1 \text{ x}10^{-4} \text{ x hours}) + (1.6x10^{-8} \text{ x hours}^2)$
Consistency	Cubic	.007	345	.85	.467	$4.3496 + (-2x10^{-4} \text{ x hours}) + (3.0x10^{-8} \text{ x hours}^2) + (-9x10^{-13} \text{ x hours}^3)$

Flexibility for Navigation Scale & Pilots' Hours on Current Aircraft Type

Scale	*Method*	*Rsq*	*df*	*F*	*Sig F*	*Regression Equations for Usability Rating*
Flexibility for Navigation	Linear	.012	349	4.15	.042	$3.9315 + (6.4x10^{-5} \times hours)$
Flexibility for Navigation	Log.	.010	349	3.42	.065	$3.1915 + (0.1229 \times \log_{10} hours)$
Flexibility for Navigation	Quad.	.013	348	2.36	.096	$4.0023 + (-4x10^{-6} \times hours) + (8.9x10^{-9} \times hours^2)$
Flexibility for Navigation	Cubic	.016	347	1.82	.142	$3.8954 + (2x10^{-4} \times hours) + (-4.0x10^{-8} \times hours^2) + (3.3x10^{-12} \times hours^3)$



Appendix L - Modification Indices

	Appropriate Functionality	*Feedback*	*Predictability*	*Information Presentation*	*Flexibility & Control*	*Error Prevention & Correction*	*Processor Speed*	*Workload*	*Consistency*
Structure & Sequence	0.619	0.605	0.590	0.880	0.820	0.428	0.374	0.487	0.747
Appropriate Functionality		0.506	0.319	0.589	0.770	0.356	0.386	0.308	0.555
Feedback			0.501	0.560	0.578	0.646	0.266	0.265	0.468
Predictability				0.563	0.769	0.394	0.255	0.498	0.484
Information Presentation					0.802	0.393	0.413	0.583	0.920
Flexibility & Control						0.501	0.413	0.684	0.748
Error Prevention & Correction							0.199	0.213	0.312
Processor Speed								0.439	0.525
Workload									0.405

Appendix M -
Study Three Qualitative Comments

M1 Improvements Sought to Future FMSs

Fleet	*Comments*	*Category*
B757/767-200/300	More accurate predictions: fuel/time/climb/descent profile	Accuracy of Parameters
B757/767-200/300	More accurate fuel and ETA estimates using inputted winds would improve system. It is difficult to know what information it is using to make these calculations.	Accuracy of Parameters
B757/767-200/300	More wind direction and speed information/prediction on flight levels above or below current flight level.	Accuracy of Parameters
B757/767-200/300	More accurate descent information. Greater standardisation of procedures i.e. abeam fixes incorporated into legs page yet non-standard procedure when abeam fix is to be next/active waypoint.	Accuracy of Parameters
B757/767-200/300	More accurate descent information.	Accuracy of Parameters
B757/767-200/300	More accurate fuel diagnostics.	Accuracy of Parameters
B757/767-200/300	More accurate optimum alt and max alt info.	Accuracy of Parameters
B757/767-200/300	Provision of calculator facility within FMS.	Additional Functionality
A320/A321/A330	Greater functionality with secondary flight plan.	Additional Functionality
A320/A321/A330	Increased ability to make FM Predictions.	Additional Functionality
B737-300/800	Progress report – how you are going versus the plan "are you up on or losing time/fuel?	Additional Functionality
B737-300/800	Lateral offset navigation – (e.g. 3nm left of track)	Additional Functionality
B737-300/800	Retention of en route altitude/speed restrictions if given a direct route, by-passing the restricting waypoint.	Additional Functionality
A320/A321/A330	I would like to see time constraints available for waypoints.	Additional Functionality
B737-300/800	More than two "Fixes would be useful".	Additional Functionality
B757/767-200/300	Automatic generation of abeam points and time and fuel predictions.	Additional Functionality
A320/A321/A330	Offset and Abeam facility would be useful, as would an execute button – so you can check inputs before going ahead with them.	Additional Functionality
B737-300/800	Take off and landing times would be helpful.	Additional Functionality

Fleet	*Comments*	*Category*
B757/767-200/300	More abilities to pre-select different approaches to Runways for the same airport.	Additional Functionality
B747-400	Why can't published holds be in memory?	Additional Functionality
B777-200	Can we have more Fix pages?	Additional Functionality
B757/767-200/300	Frequently given direct routings by one ATC controller to have them cancelled by the next controller. Would be a useful facility to be able to recover old route.	Additional Functionality
A320/A321/A330	Would like to have a "Fix" page facility with display on.	Additional Functionality
B757/767-200/300	I would also suggest a final approach page giving distance to touchdown, wind, go around procedure, landing speed, ILS/VOR selected and approach minimums.	Additional Functionality
B757/767-200/300	On descent page, it would be an advantage if we could see the track distance planned without having to revert to progress page.	Additional Functionality
B757/767-200/300	The well known problem of holding entries which I think is one of the most "what is it doing now" things, should be resolved once and for all with a totally different holding page. The same applies for hold exits.	Additional Functionality
B757/767-200/300	On descent forecast, the engine anti-ice prompt should not have an "on" function only but an "off" function as well.	Additional Functionality
B757/767-200/300	The map facility is not yet sufficiently exploited. A big improvement is the fix pages on later FMSs (i.e. range circles). VNAV, which could help on visual approaches, would be useful – by providing a glide path on RoD.	Additional Functionality
B757/767-200/300	On descent and approach it would be useful to have direct distance to the touchdown point, distance along the anticipated routing and vertical displacement from the VNAV profile all displayed on the same page i.e. in 757/767 a combined progress page 2 and fix page.	Additional Functionality
B757/767-200/300	List of nearest waypoints	Additional Functionality
B747-400	Needs many more stored routes and facility to Join New Routes.	Additional Functionality
A320/A321/A330	A repeat key to re-enter last entry would be useful.	Additional Functionality
B737-300/800	A constant readout of distance from P. Position to Centre fix would be more useful at such times.	Additional Functionality

Fleet	*Comments*	*Category*
B757/767-200/300	Ability to transfer VOR stations Left to Right when operating aircraft using raw data.	Additional Functionality
B757/767-200/300	Allow for setting bands of TAI 2on" in desc forecast rather than just when on waypoints	Additional Functionality
B757/767-200/300	Three routes available would allow two approach options and the next sector.	Additional Functionality
B757/767-200/300	LH/RH downwind positions readily available for best airfield.	Additional Functionality
B737-300/800	Could do with 2 "Route" Features.	Additional Functionality
B757/767-200/300	Pick list of info/parameters to be displayed on custom screen.	Additional Functionality
B737-300/800	Take off and landing times would be helpful.	Additional Functionality
A320/A321/A330	A suggested improvement on the A321-200 would be to retain abeam waypoints when cleared direct to a further downpath waypoint. In this example of a typical routing from Dublin [Drawing included here], via UBI to Heathrow, we create the waypoint at 45d before BNN (FL200 or below) and enter an altitude constraint with the waypoint. There is a stored constraint at BNN25 (FL150 or below]. If cleared by ATC from Hon direct to BNN, if we enter than into the CDU we loose the intermediate points with their associated height constraints. We then calculate and monitor our descent profile mentally which negates the value of entering the waypoints in the first place.	Additional Functionality
A320/A321/A330	Also databases need to reflect what tends to happen in practice e.g. the constraint at 45nm before Bovingdon happens every time we go to Heathrow, but it is not stored in the database.	Additional Functionality
B747-200	An easier method of obtaining the extended centreline of the runway would be helpful.	Additional Functionality
B747-400	Give full information for "gear down" diversions (Alt, Fuel, Spds, ETAs and ETC).	Additional Functionality
B757/767-200/300	Addition of lat/long information on progress page would help for internal met reports and gross error checks.	Additional Functionality
B757/767-200/300	New precision approaches added for longer runways, so you don't have to build the approach when time is short and workload is high.	Additional Functionality
B737-300/800	A change would be nice in the vertical VNAV crossing restrictions. Do not have a program to allow certain crossing restrictions to be put in.	Additional Functionality

Fleet	*Comments*	*Category*
B757/767-200/300	A separate page for off track bearing and distance between two or more points would be useful.	Additional Functionality
B757/767-200/300	"Nearest airport" facility.	Additional Functionality
B757/767-200/300	On progress page 1 (used in cruise), it would be good to have track information displayed.	Additional Functionality
A320/A321/A330	It would be nice to have vertical nav. Capability in HDG mode. Only available in NAV mode or a required RoD to reach on vertical Nav. Constraint.	Additional Functionality
B757/767-200/300	FMS could easily be programmed to recognise waypoints appropriate to airway, rather than frequently presenting choices.	Additional Functionality
A320/A321/A330	The ability to see the vertical bearing from a user defined waypoint.	Additional Functionality
A320/A321/A330	More vertical information particularly when in HDG mode and not following the FMS route (i.e. when being radar vectored by ATC).	Additional Functionality
B737-300/800	More fix points (>2).	Additional Functionality
B757/767-200/300	Better "Fail safes" if silly errors are put in.	Barriers to Erroneous Operation
B757/767-200/300	During descent the FMS should question the selection of a waypoint which is not part of the approach procedure.	Barriers to Erroneous Operation
B737-300/800	Better alerting of change from VNAV path to VNAV speed descent so that pilot knows in advance.	Barriers to Erroneous Operation
B757/767-200/300	Increase size of database. My Psion has one hundred times the capacity of a 200k FMS.	Bigger Database / Memory
B757/767-200/300	Bigger database – standard route, safe altitudes.	Bigger Database / Memory
B757/767-200/300	More capacity for analysing operations and consequences in LNAV/VNAV.	Bigger Database / Memory
B757/767-200/300	Big memory.	Bigger Database / Memory
B757/767-200/300	Increased database would be good.	Bigger Database / Memory
B757/767-200/300	Larger databases would be useful.	Bigger Database / Memory
B747-400	Needs far bigger database.	Bigger Database / Memory
B757/767-200/300	Simplified Logic.	CDU Architecture & Logic

Fleet	*Comments*	*Category*
B757/767-200/300	Menu Structure.	CDU Architecture & Logic
B747-400	I flew the TriStar for many years, its FMS had some very good features which do not appear in "modern" FMSs. Arrival Data: to change the arrival data, say runway change, when part way into a STAR is far too cumbersome at times. It makes the workload peak unnecessarily as sometimes all of the elements, Approach type & runway/Star/Transition, need changing. Surely this can be made a much more intuitive function by combining these elements.	CDU Architecture & Logic
B737-300/800	Ease of access to range/bearing direct to a waypoint without having to go through "Fix" page.	CDU Architecture & Logic
B757/767-200/300	Due to fleets having three types of FMS – standard, PIP and Pegasus, the additional functions and methods of use between the three types (which can occur in one duty with aircraft changes) can cause additional workload. A more common thread between the these would aid this problem as it only adds to having to remember the differences with the aircraft types as well.	Consistency & Compatibility
B757/767-200/300	Standardisation of pages, input methods and display of information is vital. An international protocol needs to be agreed.	Consistency & Compatibility
B737-300/800	Symbology - This may be a personal preference but I'll include it. On some of our FMS descent pages displays, the variation from programmed altitude during descent is indicated by either the word "HI" or "LO". On others this variation is indicated by a "+" or a "-". I personally find the plus or minus symbols easier to interpret than the words, especially when tired.	Consistency & Compatibility
F-100	Having been on Boeings for ten years prior to the F-100 I am probably biased towards the Boeing philosophy and therefore am not really comfortable with the F-100. Standardisation throughout all FMSs would make transitions easier!	Consistency & Compatibility
A320/A321/A330	A common standard of abbreviations etc. agreed between manufacturers.	Consistency & Compatibility
B737-300/800	Would be nice if B737 FMS was similar to B767 with Route 2 etc.	Consistency & Compatibility
B757/767-200/300	Greater standardisation of procedures i.e. abeam fixes incorporated into legs page yet non-standard procedure when abeam fix is to be next/active waypoint.	Consistency & Compatibility
B757/767-200/300	I would like to see QWERTY keyboard.	Control Devices
B737-300/800	An "undo" button – undo the last executed change.	Control Devices

Fleet	*Comments*	*Category*
A320/A321/A330	I would like to see a more comfortable location of the keyboard.	Control Devices
B747-400	QWERTY Keyboard.	Control Devices
B747-400	Mouse/Trackball.	Control Devices
B747-400	Keyboard repeat rates need standardising.	Control Devices
B757/767-200/300	More ergonomic keyboard.	Control Devices
B757/767-200/300	QWERTY keyboard.	Control Devices
B777-200	In this day and age I think a QWERTY keyboard would be better.	Control Devices
B757/767-200/300	A QWERTY keyboard would assist in entering data.	Control Devices
B757/767-200/300	Touchscreen Nav display would be very helpful for entries – instead of entering xxx341/9.6 one could touchscreen at appropriate point.	Control Devices
B737-300/800	Have keyboard in QWERTY format rather than A-Z.	Control Devices
B737-300/800	I'd like to see a fast touchscreen CDU.	Control Devices
B757/767-200/300	Colour coded/differentiated keyboard.	Control Devices
B757/767-200/300	More logical keyboard layout – better grouping of keys.	Control Devices
B757/767-200/300	We all use QWERTY keyboards with various computers. Why do aircraft FMS keyboards have to be different?	Control Devices
B757/767-200/300	Layout of buttons can be improved to be more self explanatory, by variation in shape/size/position etc.	Control Devices
B757/767-200/300	QWERTY keyboard and separated numeric keypad.	Control Devices
B757/767-200/300	Pointing device and pointable screen – No touchscreen either IBM thinkpad device or touchpad or trackball.	Control Devices
B737-300/800	QWERTY keyboard.	Control Devices
B737-300/800	Pointer Device.	Control Devices
B757/767-200/300	Would prefer QWERTY keyboard.	Control Devices
B757/767-200/300	Location of FMS on centre pedestal could be better.	Control Devices

Fleet	*Comments*	*Category*
A320/A321/A330	Why can't we have a QWERTY keyboard like other computers. I'm right handed so having the keyboard on the left slows me down.	Control Devices
B757/767-200/300	Easier methods of information input.	Control Devices
B737-300/800	QWERTY Keyboard.	Control Devices
B737-300/800	Larger keys.	Control Devices
B737-300/800	Line Select Keys alignment with line on screen.	Control Devices
B737-300/800	I am a training pilot and teach FMS skills. The biggest improvement I'd like to see is use of colour/type size to highlight essential information.	Display Quality & Information Presentation
B737-300/800	Coloured screens would help. Green screens are difficult to focus on after long night flights.	Display Quality & Information Presentation
B757/767-200/300	Screens need angling to be at ninety degrees to line of sight, even if keyboard isn't.	Display Quality & Information Presentation
B757/767-200/300	Clearer colour coded CDU.	Display Quality & Information Presentation
B757/767-200/300	Colour CDU would be nice.	Display Quality & Information Presentation
A320/A321/A330	Larger screens.	Display Quality & Information Presentation
A320/A321/A330	Larger scratchpad.	Display Quality & Information Presentation
B737-300/800	Colour Screen.	Display Quality & Information Presentation
A320/A321/A330	I would like to see increased use of colour in displays.	Display Quality & Information Presentation
B747-400	Use of Colour.	Display Quality & Information Presentation
B757/767-200/300	Better display required when FMS is being used for LNAV.	Display Quality & Information Presentation

Fleet	*Comments*	*Category*
B757/767-200/300	MSAs should be displayed at all times.	Display Quality & Information Presentation
B757/767-200/300	Multicoloured displays would be useful.	Display Quality & Information Presentation
B757/767-200/300	More use of colour (as in A320 series).	Display Quality & Information Presentation
B737-300/800	Use of colour would greatly improve FMS and draw attention to key facts.	Display Quality & Information Presentation
B747-200	The use of different colours on the FMS CDU would be useful.	Display Quality & Information Presentation
B777-200	Slightly larger screens and larger print size. If you wear glasses it can be difficult to optimise general cockpit and instrument lighting with FMS display at night and when tired.	Display Quality & Information Presentation
B777-200	"Windows" style GUI using large screen and Colour Co-ordinated with other functions on the B777.	Display Quality & Information Presentation
B777-200	My eyesight is okay at 42 but the buttons on the CDU are sometimes not easy to read in certain light.	Display Quality & Information Presentation
B757/767-200/300	Colour screen.	Display Quality & Information Presentation
B757/767-200/300	Screens difficult to read in direct sun; all aircraft TVs like this.	Display Quality & Information Presentation
B757/767-200/300	Dust on screen especially opposite screen in F/Os with light sun impossible to read.	Display Quality & Information Presentation
B757/767-200/300	The use of colour in FMS presentation would be welcome.	Display Quality & Information Presentation
B757/767-200/300	Colour display.	Display Quality & Information Presentation
B757/767-200/300	Colour display.	Display Quality & Information Presentation
B757/767-200/300	Bold or colour code distance to destination (it is a key piece of information).	Display Quality & Information Presentation

Fleet	*Comments*	*Category*
B757/767-200/300	Colour code final fix and Runway/Map.	Display Quality & Information Presentation
B757/767-200/300	Different colours i.e. active waypoint in red would be an improvement.	Display Quality & Information Presentation
B757/767-200/300	More user friendly interface.	Display Quality & Information Presentation
B757/767-200/300	Use colour in the Boeing FMS systems.	Display Quality & Information Presentation
B757/767-200/300	Colours as highlighters of certain parameters.	Display Quality & Information Presentation
B737-300/800	Use of colour would greatly enhance the FMS drawing your attention to important information.	Display Quality & Information Presentation
B747-400	On progress page make "last position" in same size font.	Display Quality & Information Presentation
B747-400	Show compulsory way points with a symbol beside them on "legs page" and "Prog page".	Display Quality & Information Presentation
B757/767-200/300	Colour code information red (immediate attention), yellow (non-immediate actions) and green (general information).	Display Quality & Information Presentation
A320/A321/A330	Colour coding of database airfields open 24 hours.	Display Quality & Information Presentation
B757/767-200/300	I would prefer the display to be larger so that one page could provide more information (e.g. more waypoints on LEGS page). This would reduce the time spent "paging up/down".	Display Quality & Information Presentation
A320/A321/A330	Greater range of colours to more easily identify information.	Display Quality & Information Presentation
B737-300/800	Larger Colour Display.	Display Quality & Information Presentation
B737-300/800	Colour Display.	Display Quality & Information Presentation
B737-300/800	Larger screen.	Display Quality & Information Presentation

Fleet	*Comments*	*Category*
B737-300/800	Fingerprints are left on CDU screen and vision is often obscured especially in sunlight.	Display Quality & Information Presentation
B737-300/800	Biggest single improvement would be multi-coloured display, so finding information would be quicker.	Display Quality & Information Presentation
B757/767-200/300	Error messages should describe the problem, and give information on correct entry format.	Feedback, Error Prevention & Correction
B737-300/800	Error code to be more accurate and indicative of the actual problem.	Feedback, Error Prevention & Correction
B737-300/800	A more obvious system when a change has been made but not executed as the execute button is sometimes pushed but does not execute and the light in it is often very dim or does not work.	Feedback, Error Prevention & Correction
B757/767-200/300	Use of colour on the display would highlight important information.	Feedback, Error Prevention & Correction
B757/767-200/300	Relay of more information to other display would be useful.	Feedback, Error Prevention & Correction
B737-300/800	Calculation Indicator - Quite often after some value has been entered and executed on the FMS CDU the FMS goes into a calculation mode. The wait period is often perplexing because the pilot wants to verify the input but can't tell if the FMS is "cooking" something or for some reason the input was incorrect. It would be very helpful to have a symbol something like the hourglass used with computers to indicate that the FMS is busy.	Feedback, Error Prevention & Correction
B757/767-200/300	Error messages need to be clearer and more detailed.	Feedback, Error Prevention & Correction
A320/A321/A330	More definitive fault messages.	Feedback, Error Prevention & Correction
B757/767-200/300	The FMS should show more details of calculation criteria.	Feedback, Error Prevention & Correction
B757/767-200/300	Waypoint wind page on Pegasus. When initially inputting wind speed and direction on 4R – this could inadvertently be entered on 5R ALT/OAT Line. It would be better to separate these two slightly more.	Feedback, Error Prevention & Correction
A320/A321/A330	Most entries are single button push. It would be helpful to be offered more two step entries i.e. enter, then confirm or erase.	Feedback, Error Prevention & Correction

Fleet	*Comments*	*Category*
A320/A321/A330	I think there should be an enter key to confirm the important functions and executions to allow a second chance to evaluate the implications of the function carried out as is the case with the B737 "exec" key.	Feedback, Error Prevention & Correction
A320/A321/A330	An "execute" button is invaluable for avoiding silly mistakes.	Feedback, Error Prevention & Correction
B757/767-200/300	More intuitive operating system.	Intuitiveness
B737-300/800	With the technology available today, the FMS CDU could be made much more intuitive.	Intuitiveness
B757/767-200/300	Ability to use three letter codes rather than ICAO four letter codes or even ability to input airports by name.	Language
B737-300/800	Plain language option would be of great use – often ATC give you clearances in full name but FMS only contains three letter codes (GMH) – if you are not familiar with the route it would be helpful to have full name available plus airport IDs in plain language.	Language
B757/767-200/300	List of nearest airfields in plain language.	Language
B757/767-200/300	Natural waypoint naming rather than HON01,02 etc. which are meaningless, would rather be able to name points and perhaps italicise or colours these named points	Language
B737-300/800	Plain language option – if you are not familiar with a route, three letter codes and four letter airport Ids do not mean much – ATC give you clearances in plain language so it is difficult to correlate it with the waypoint on the legs page – often have to ask ATC for the three letter code.	Language
B737-300/800	Our version of the ops manual is written for engineers - the flow charting is hard to follow. The technical writing should be written more in lay man terms.	Language
B777-200	Boeing needs input from Airline pilots not test pilots.	Observations
B777-200	Shouldn't crash during taxi or on power change over on the ground.	Observations
B777-200	Exiting holds is average.	Observations
B757/767-200/300	There should be limitations in the FMS. We need to exercise out brains and use basic skills like mental arithmetic. This latter skill is become worryingly eroded.	Observations
B737-300/800	FMS are designed by electrical engineers not pilots. The next generation FMS systems need to have line pilot input from the start, not as an after thought.	Observations

Fleet	*Comments*	*Category*
B757/767-200/300	Retrieval and processing should be much faster.	Processor Speed
B757/767-200/300	Quicker process times.	Processor Speed
A320/A321/A330	Increased processing speed.	Processor Speed
B737-300/800	Faster processors – current processing power is seriously outdated.	Processor Speed
B737-300/800	FMS needs faster processor and more modern computer hardware.	Processor Speed
B757/767-200/300	Faster Processor.	Processor Speed
B747-400	Processor must be speeded up.	Processor Speed
B757/767-200/300	Faster processing.	Processor Speed
B757/767-200/300	Fast processor!!	Processor Speed
B737-300/800	Fast processor – it is so slow.	Processor Speed
B757/767-200/300	Faster Processor.	Processor Speed
B757/767-200/300	Faster processor.	Processor Speed
B757/767-200/300	Faster processor.	Processor Speed
B737-300/800	The FMS is extremely slow in adapting to new inputs in the vertical profile. This causes lag in response, therefore you almost always have to override it with another vertical mode. Faster chip please.	Processor Speed
B757/767-200/300	Faster operation.	Processor Speed
B757/767-200/300	Faster Processor.	Processor Speed
B757/767-200/300	Faster computers are necessary in the FMS.	Processor Speed
B737-300/800	Increased processor speed.	Processor Speed
B757/767-200/300	Faster processing	Processor Speed
B737-300/800	Speed of processing information.	Processor Speed
B737-300/800	Operators should have access to a "Real" CDU linked to a suitable computer which emulates all the subtleties and idiosyncrasies of the aircraft concerned, so that the student can punch the keys and make all the customary mistakes long before he/she gets into the aircraft for the first time.	Training

Fleet	*Comments*	*Category*
B737-300/800	Considering the complexity of recent game consoles and the latest generation Pentiums, it should not be beyond the wit of man to produce a system with say a miniature MCP and joystick linked with the CDU and computer to actually TEACH the student that which he NEEDS to know.	Training
B757/767-200/300	Access to online "help".	User Guidance & Support
B737-300/800	Handbook is generic - would prefer to see a company SOP specific handbook.	User Guidance & Support
B757/767-200/300	More "hypothetical" capability would be useful e.g. "if cleared direct to xxx what ETA? Without dumping current route and wind data.	'What-if?' Capability
B757/767-200/300	I would like to see Easier entry of non-active waypoints to highlight things such as ETOPS Equal Time Points.	'What-if?' Capability
B757/767-200/300	I would like to see better wind/altitude/temperature related "what if" scenarios when considering step climb.	'What-if?' Capability
B757/767-200/300	Ability to work "offline" could be improved i.e. using it to check different routes and wind solutions etc.	'What-if?' Capability
B737-300/800	Ability to make changes and see results without executing.	'What-if?' Capability

M2 Problems with Current FMSs

Fleet	*Comments*	*Category*
B757/767-200/300	Fuel predictions are poor.	Accuracy of Parameters
B757/767-200/300	FMS places too much emphasis on current wind when making predictions.	Accuracy of Parameters
B757/767-200/300	FMS program on B757 package is poor at accurately computing vertical track error and frequently would allow the user to become high on profile unless pilot carries out mental cross check and intervenes.	Accuracy of Parameters
B737-300/800	FMS is the best thing since sliced bread, however, one weakness is VNAV descent which usually gets the aircraft high resulting in a demand for "more drag". I can work out a better descent profile in my head than the profile that the FMS comes up with.	Accuracy of Parameters
B737-300/800	FMS can give "silly" incorrect vertical profiles on descent (even with wind and QNH entered on descent forecast page).	Accuracy of Parameters
B757/767-200/300	In descent, rarely takes account of programmed forecast winds.	Accuracy of Parameters
B757/767-200/300	This FMS does not contain an adequate model of airplane performance during descent.	Accuracy of Parameters
B757/767-200/300	By design, on final approach in a crosswind, the FMS will calculate zero crosswind.	Accuracy of Parameters
B757/767-200/300	VNAV calculations do not match –E4 idle parameters.	Accuracy of Parameters
B737-300/800	VNAV profiles not good.	Accuracy of Parameters
B777-200	Vnav descent profile (path) generated by FMS is far too vague to be relied upon.	Accuracy of Parameters
F-100	Sometimes vertical Nav misses the mark and a crossing restriction will be missed w/o an input.	Accuracy of Parameters
F-100	In VNAV descent with a tail wind and minimum time speed, altitude crossing restrictions are busted.	Accuracy of Parameters
B757/767-200/300	Sometimes VNAV errors are gross when changing track.	Accuracy of Parameters
B757/767-200/300	VNAV profile sometimes inaccurate in descent when route/ALTS are modified due to re-clearances.	Accuracy of Parameters
B757/767-200/300	They are not as logical and require the use of several pages to find what is wanted.	CDU Architecture & Logic

Fleet	*Comments*	*Category*
B757/767-200/300	Ability to change default setting –especially transition default of 18000' and descent speed 240 below 10000. It's not just used in the USA.	CDU Architecture & Logic
B757/767-200/300	Latest versions do not show fuel at dest. for amended flight level unless you execute the new flight level.	CDU Architecture & Logic
B757/767-200/300	Constant interruption with "insufficient fuel" during route and wind profile input.	CDU Architecture & Logic
B737-300/800	Unable to retain "Abeam" waypoints with "Direct to...." on our FMS.	CDU Architecture & Logic
A320/A321/A330	This FMS has very little logical sequencing for programming etc. No next step prompt or hint.	CDU Architecture & Logic
B757/767-200/300	Engine out information awkward to obtain.	CDU Architecture & Logic
B757/767-200/300	Even my mobile phone has a more "logical" menu/submenu etc setup.	CDU Architecture & Logic
B737-300/800	It is necessary to delete the FA hold before it will accept another en-route hold.	CDU Architecture & Logic
B757/767-200/300	Program Fault: Waypoints more than one with same ident. FMS asks you to select which lat/long, if you then go dir int and line select the waypoint it asks you to verify which one again giving you the choices if you then select a different waypoint then FMS accepts that and is difficult to erase. It therefore has no benefit as a safety check.	CDU Architecture & Logic
B757/767-200/300	Legs page – unable to enter a selected speed without a hard altitude during descent.	CDU Architecture & Logic
B757/767-200/300	Descent page – doesn't allow enough speed restrictions.	CDU Architecture & Logic
B757/767-200/300	One must select a special page to see which VOR/DMEs are auto tuned rather than having this displayed on PROG, which is usually selected for descent/approach to give track miles to touchdown.	CDU Architecture & Logic
B737-300/800	Unable to hold and display specific Nav aid on RMI e.g. to have RMI needle and DME required for an arc approach allowing ILS frequency and inbound course ready for intercept.	CDU Architecture & Logic
B747-400	The Point of No Alternate (PNA) can only be calculated by changing the destination on the progress page. Equal time points cannot be inserted for automatic updating.	CDU Architecture & Logic
A320/A321/A330	The FMS system is organised illogically – excellent in some facets; irritatingly slow in others.	CDU Architecture & Logic

Fleet	*Comments*	*Category*
B757/767-200/300	I'm a new user – so it is getting easier with time but the logic is not always straightforward.	CDU Architecture & Logic
B737-300/800	The FMS on the -400 will not allow certain crossing restriction entries. For the most part it is a great system.	CDU Architecture & Logic
F-100	The flight plan format has unnecessary information, while neglecting distance.	CDU Architecture & Logic
F-100	Climb, cruise and descent modes require separate inputs, for example, a cruise mach number will not transfer to the descent mode. One descent is initiated, Mach number must be re-entered.	CDU Architecture & Logic
F-100	Tactical mode parameters are lost during mode changes i.e. assigned airspeed in climb does not carry over to cruise mode.	CDU Architecture & Logic
A320/A321/A330	Changing runway drops all descent crossing restrictions.	CDU Architecture & Logic
A320/A321/A330	Idle TOD point only computed for first crossing restriction.	CDU Architecture & Logic
A320/A321/A330	The FMS CDU on my previous type of aircraft (B737-400) was far more logical to use. It contained very useful functions on the 'FIX" page i.e. the "Abeam" prompt and DME range arc.	CDU Architecture & Logic
B737-300/800	Key label entry sequences for FMS units should follow that of ATC clearances.	CDU Architecture & Logic
B737-300/800	Recently converted to 737-800 from 757/767. There are many subtle differences in FMS operation between the types for no apparent reason. This makes type conversion harder and can lead to incorrect FMS usage during the re-learning process.	Consistency & Compatibility
B737-300/800	Unit Variation - Some FMSs perform slightly different in obscure ways and some perform at slightly different speeds. Even though the FMS model and operating program, are identical, there are some subtle differences in different aircraft.	Consistency & Compatibility
A320/A321/A330	In an industry where so many flight instruments and general ergonomics have become standardised between type manufacturers, it is remarkable how dissimilar the Airbus type FMS is from the Boeing. It is almost as if they did this deliberately to be difficult. I believe the FMS concept to be superb - but the continuing development and improvement rates are a disgrace.	Consistency & Compatibility
A320/A321/A330	Q5,6,7 Some pages take four letter codes others take three letter codes for airports.	Consistency & Compatibility
A320/A321/A330	Some pages use "next page", others use prompts at LSK.	Consistency & Compatibility

Fleet	*Comments*	*Category*
B757/767-200/300	Location of CDUs should not involve stretching to modify inputs – disrupts scan and strains backs.	Control Devices
A320/A321/A330	Flight plan scroll keys appear to operate in the "wrong sense" i.e. "up" arrow key scrolls down the flight plan (unlike the B737 or home computer mouse.)	Control Devices
A320/A321/A330	The small diameter of heading and latitude setting knobs on the glareshield FCU, makes accurate setting unnecessarily difficult, with the pilots arm outstretched, particularly if the aircraft is experiencing turbulence. Conversely setting radio frequencies is easier as the knobs are larger in diameter and are adjacent to the pilot.	Control Devices
B757/767-200/300	Poor keyboard.	Control Devices
B757/767-200/300	Retrofit palm rests onto CDUs (as fitted to 747-400).	Control Devices
A320/A321/A330	It can sometimes be very difficult to undo an entry mode in error.	Control Devices
A320/A321/A330	A constant irritant is the physical location and orientation of the MCDU and screen which is not within easy reach or aligned normally with the crewmembers' eyes.	Control Devices
A320/A321/A330	Buttons seem to be randomly positioned (Q12).	Control Devices
B757/767-200/300	In B757 need several button presses: INIT – NAV DATA – enter name of wypt – check lat/long.	Control Devices
B747-400	Tuning of Radio aids via FMS is a retrograde step from radio select panels. RTE 2 Diversion information is limited until active.	Control Devices
B747-400	Speed brake lever directly over keypad affects input.	Control Devices
B777-200	My eyesight is okay at 42 but the buttons on the CDU are sometimes not easy to read in certain light.	Control Devices
B737-300/800	Keyboards on some aircraft are better than others.	Control Devices
B747-400	Input keys are in alphabetical order. Most people are now familiar with QWERTY.	Control Devices
F-100	CDU line select keys are too close together.	Control Devices
B757/767-200/300	The keyboard is not tactical enough.	Control Devices
B757/767-200/300	Insufficient memory.	Database/Memory Size

Fleet	*Comments*	*Category*
B757/767-200/300	Often have a database too small to put a return route into Rhodes.	Database/Memory Size
B757/767-200/300	Poor memory space.	Database/Memory Size
B757/767-200/300	Not enough memory.	Database/Memory Size
B757/767-200/300	Databases are not big enough.	Database/Memory Size
B737-300/800	The worst feature is the database that we are currently using.	Database/Memory Size
B757/767-200/300	Main problems are memory size e.g. company routes, airfields etc.	Database/Memory Size
B757/767-200/300	It lacks capacity.	Database/Memory Size
B757/767-200/300	Poor display.	Display Quality & Information Presentation
B737-300/800	The screen is often difficult to read in bright daylight.	Display Quality & Information Presentation
A320/A321/A330	Display brightness is a problem in sunny conditions.	Display Quality & Information Presentation
B757/767-200/300	Lack of colour on 757/767 CDUs is unhelpful.	Display Quality & Information Presentation
B747-400	With ageing eyes the screen is further away than the instruments. Very difficult to obtain a focus, unless using varifocals – only now endorsed by the CAP.	Display Quality & Information Presentation
B737-300/800	Some of the screens are very hard to read in sunlight.	Display Quality & Information Presentation
B737-300/800	Lighting/presentation of data inconsistent and hard to see clearly.	Display Quality & Information Presentation
A320/A321/A330	Cannot see screen because of sunlight reflection.	Display Quality & Information Presentation
B737-300/800	It is inconvenient we do not have EFIS displays. It is like buying a computer without a monitor.	Display Quality & Information Presentation
A320/A321/A330	Main criticism is that "FLT Plan" pages are cluttered with small characters - not good for fast assimilation of data.	Display Quality & Information Presentation
A320/A321/A330	Just one example of this is that the screen cannot be read in direct and bright sunlight on the A320.	Display Quality & Information Presentation

Fleet	*Comments*	*Category*
B757/767-200/300	Comparing Airbus and Boeing: The use of colour is good and assists in speed of data entry, especially in emergency/abnormal situations.	Display Quality & Information Presentation
B737-300/800	In general, the layout of information on the Boeing FMS is not easy to read.	Display Quality & Information Presentation
A320/A321/A330	FMS descent program intercepts descent into an increasing wind as going above profile.	Functional Limitations
B757/767-200/300	Sometimes on approach to Reus, FMS changes arc from just a segment of VOR-DME Arc to the remaining part of the arc at TOD leaving you about 10,000' too high unless you are expecting it.	Functional Limitations
A320/A321/A330	A newly created waypoint has to be inserted fully i.e. HON/144/12 =Place/Bearing/Distance, which is given the designator PBD01 on the flight plan. To review this new waypoint at a later stage, the position must be identified from its position relative to the adjacent waypoints, as the PBD identifier provides no useful information. (The B737 EFIS "scratch pad" system is much more user friendly.)	Functional Limitations
B757/767-200/300	Descent Speeds etc not deleted when V Nav parameters changed.	Functional Limitations
B757/767-200/300	Our FMSs do not permit moving waypoint to top of list to become active waypoint.	Functional Limitations
A320/A321/A330	Apart from waypoint directly behind, you cannot check for times etc for any other waypoints behind.	Functional Limitations
B737-300/800	"Direct to..." turns very slow – ATC must wonder if we have forgotten.	Functional Limitations
B757/767-200/300	FMS calculates "unable to achieve cruise altitude" when it clearly can.	Functional Limitations
B737-300/800	However, one area where there is a problem is when an unexpected hold is required an on hold is programmed into the GA procedure.	Functional Limitations
A320/A321/A330	Q23,24,25,26,27,33 and 34 - as ex-Smith's users we all miss the Fix page.	Functional Limitations
B757/767-200/300	Altitude/Speed constraint x miles before the active to waypoint does not use the direct to button. Limited offset range need greater than or equal to thirty nm. Lateral LSK only selects the waypoint in A320 series gives lat/long – very useful for long range / North Atlantic waypoint checking.	Functional Limitations

Fleet	*Comments*	*Category*
B737-300/800	On this particular type of FMS it is not possible to fly an "offset" airway/track as sometimes required by ATC. When descending in vertical Nav (VNAV) mode and after capturing an initial height restriction, the FMS will reduce the vertical below the 1000FPM required by ATC and level off until it becomes necessary to descend again and capture a subsequent altitude restriction.	Functional Limitations
B747-400	I flew the TriStar for many years, its FMS had some very good features which do not appear in "modern" FMSs. Standard set up button: this deleted/amended all the previous user's setting to "known" base. I am surprised that the Honeywell system does not do more of this on initialisation. I would really expect it, in addition to setting the existing default parameters to: Delete Route 2 information, freeing memory for faster functioning. Delete existing fixes. Return any manually tuned Radio Aids to Automatic.	Functional Limitations
B747-400	I flew the TriStar for many years, its FMS had some very good features which do not appear in "modern" FMSs. Legs Page Format: the legs page presented cumulative distance i.e. you could see at a glance the cumulative distance to any waypoint along the route from present position. This is not available anywhere on the Honeywell system. It made descent monitoring much more intuitive. Also Any waypoint could be selected with a line select key [LSK] which turned the presentation in "Fix" page information. Much slicker than routing to either (choice of 2) Fix pages.	Functional Limitations
B747-400	I flew the TriStar for many years, its FMS had some very good features which do not appear in "modern" FMSs. Naming of Created waypoints: you could create a waypoint then name it whatever you wanted. This again makes it much clearer than accepting "ABC04" etc.	Functional Limitations
B757/767-200/300	Current FMS does not have "Abeam positions".	Functional Limitations
B737-300/800	It also doesn't have missed approach information or the capability to store multiple flight plans at one time.	Functional Limitations

Fleet	*Comments*	*Category*
B757/767-200/300	Windows 95/98, error messages accompanied by reason and recovery strategy. Error messages are not accompanied by reason and recovery strategy.	Inadequate Feedback, Error Prevention & Correction
B757/767-200/300	Frequently there is no advance warning that critical information/data has been omitted.	Inadequate Feedback, Error Prevention & Correction
B747-400	Input error messages need clarifying.	Inadequate Feedback, Error Prevention & Correction
B747-400	I flew the TriStar for many years, its FMS had some very good features which do not appear in "modern" FMSs. Input Errors: generally speaking a lot of valuable time is wasted in format errors when the FMS returns with "Not in Database" instead of "Format Error" in lateral inputs but "Invalid Entry", with no clue as to what is wrong in other areas.	Inadequate Feedback, Error Prevention & Correction
B747-400	The "Execute" facility can be confusing i.e. it can remain illuminated after the user has selected another programme – here confusion from both pilots as to what the execute light refers to.	Inadequate Feedback, Error Prevention & Correction
B777-200	It appears that waypoints have hidden attributes (i.e. definite glide paths to the waypoint) which are not clear to the pilot.	Inadequate Feedback, Error Prevention & Correction
B777-200	It is impossible to know directly what factors affect the path. Altitude constraints entered can affect the path in indeterminate ways such that removal of those constraints can throw out the path completely.	Inadequate Feedback, Error Prevention & Correction
B737-300/800	Autopilot often inadvertently left in heading mode after direct clearance given – particularly when new track matches current track.	Inadequate Feedback, Error Prevention & Correction
B757/767-200/300	System will go direct to new fix without confirmation	Inadequate Feedback, Error Prevention & Correction
F-100	When making an input to the FMS the screen display will shift as line positions as the FMS re-calculates or the FMS will display a message "input ignored", frequent independent operation causes unnecessary and disruptive work sequence alterations.	Inadequate Feedback, Error Prevention & Correction
B757/767-200/300	The "Go Direct to" a particular waypoint, you get no second chance or execute key, but to change to alternates takes at least six steps.	Inadequate Feedback, Error Prevention & Correction

Fleet	*Comments*	*Category*
B747-400	I flew the TriStar for many years, its FMS had some very good features which do not appear in "modern" FMSs. Execute of Not?: there is not a strong logic line between those functions which require confirmation before execution and those that do not. If the FPU is slow the individual will have moved on to another page before the EXEC light illuminates leaving the crew with no way of instantly knowing what is waiting to be executed! This requires real discipline particularly with dual FMS installations.	Inadequate Feedback, Error Prevention & Correction
B757/767-200/300	The "Dir To" function is too easy to engage in nav.	Inadequate Feedback, Error Prevention & Correction
B757/767-200/300	The "DIR TO" function on Boeing system always gives the option to erase or execute which is good because it prevents errors and allows predictions to be assess before going ahead with it. NB Boeing requires all information to be executed first after entry.	Inadequate Feedback, Error Prevention & Correction
B737-300/800	I would not say that the FMS is intuitive.	Intuitiveness
B757/767-200/300	In particular it is hard to understand the concept of progress via legs pages, and how to go through the index pages.	Intuitiveness
B757/767-200/300	The later 757 aircraft FMS are not as "easy" to use as earlier ones.	Observations
A320/A321/A330	Functions which are rarely used such as flying an offset on an airway or diverting to an alternate, do not cater for the lack of familiarity of the pilot.	Observations
A320/A321/A330	I would like to emphasise that the FMS is of great help in operating the aircraft particularly in a busy air traffic environment, however the points raised would not have been incorporated into the system, if Apple or Microsoft software engineers had designed it in a more user friendly way.	Observations
B757/767-200/300	There has got to be a better tool than this one.	Observations
B757/767-200/300	One gets the impression that it was designed by a committee of non-pilots.	Observations
B757/767-200/300	I would fill this in differently for the descent phase.	Observations
B757/767-200/300	In the cruise it is user friendly and provides all the information clearly.	Observations
B757/767-200/300	Frequent re-synchs between #1 and #2 FMS	Observations
B757/767-200/300	The FMS is far more limited than I had previously thought, an automatic washing machine for the flight deck, but nothing more radical.	Observations

Fleet	*Comments*	*Category*
B737-300/800	I find CDU generally very user friendly.	Observations
B757/767-200/300	Tendency for too much reliance on FMSs in terminal areas can be distracting.	Observations
A320/A321/A330	As a trainer, I have noticed pilots trying to type too fast for input of data and annoyed when their expected input is not accurate. My own experience is that hands blurred over keyboards is as alarming as hands moving fast anywhere in the flightdeck. Any actions should be deliberate, accurate and with some clear thought. Speed is seldom required, accuracy is! I have gone from DC3 to A340 and just love the new technology.	Observations
B757/767-200/300	Occasionally FMS will stop processing information while it resynchronises the other pilot's FMS. This can be followed by one of the two shutting down.	Observations
A320/A321/A330	Boeing 737 FMS is easier to use that A320.	Observations
B757/767-200/300	I never trust "VNAV" in decent phase below 10,000FT. I always revert to "FL Ch".	Observations
B747-400	Software problems require special input restrictions otherwise failures may occur.	Observations
B747-400	Inputting cruise winds etc is a real pain!	Observations
B777-200	I have had it crash during taxi and another time on power change over on the ground.	Observations
B777-200	Exiting holds is average.	Observations
A320/A321/A330	Boeing FMS far more user friendly Airbus too fancy and badly thought out. Designed by Boffins not by flightcrew.	Observations
B757/767-200/300	Inputs to the FMS during descent and approach phases should be actively discouraged.	Observations
B757/767-200/300	Better to programme the FMS in the cruise and if changes are required following top of descent then use alternative modes and use the FMS for reference only i.e. the "Big Picture". Information should be readily available on the FMS CDU.	Observations
B757/767-200/300	The FMS with GPS has been an excellent upgrade. When feeding in the route – the ability to use just the airways and not the RP's is a great improvement.	Observations
B757/767-200/300	Overall, much easier than previous mental calculations.	Observations

Fleet	*Comments*	*Category*
B757/767-200/300	The FMS that we use has many little glitches. I feel that this makes you more aware that the FMS should be used only as a useful aid to vertical and lateral navigation, and it shouldn't be running the flight deck. This would detract from the most important part of cross-checking and monitoring the flight.	Observations
B757/767-200/300	A fool proof FMS may lead to a deterioration of these basic points of airmanship.	Observations
B737-300/800	The descent page is a total waste of time.	Observations
B737-300/800	Boeing FMS "Interface" much more user friendly than Airbus FMS which is overly labour intensive and complicated for no visible or apparent advantage.	Observations
B747-400	Winds have to be inserted manually and individually.	Observations
B737-300/800	It took me two months on line to feel comfortable with FMS.	Observations
B737-300/800	In an ideal world with no ATC constraints the FMS might be useful for descent and early approach, but especially in VNAV it cannot cope with many ATC imposed requirements in any useful way. It is often far simpler and less demanding to revert to basics.	Observations
A320/A321/A330	I believe the current state of development of FMS is rudimentary – but a valuable aid in operation.	Observations
A320/A321/A330	While the system on my current aircraft (A321) is faster at processing information than my previous experience on B737-300/400/500 aircraft, it offers much less flexible usage in lateral and vertical navigation (absence of pilot selected and modified fixes; use of DME arc for ATC compliance).	Observations
B757/767-200/300	FMS "Hold" page is very unfriendly.	Observations
B757/767-200/300	FMS requires lots of "notice" before entering a FMS hold.	Observations
B737-300/800	Descent profiles during the later stage of descent i.e. vectoring for final approach, can become unnecessary task or re-programming of expected lateral path. This is a dangerous distraction especially for less experienced crewmembers.	Observations
B757/767-200/300	For descent and approach, I always use mental arithmetic first, and back it up with VNAV, I advise others to do the same.	Observations
B757/767-200/300	VNAV often gives power changes in order to maintain the profile precisely – uncomfortable ride.	Observations

Fleet	*Comments*	*Category*
B757/767-200/300	I like the limitations of the FMS. We need to exercise out brains and use basic skills like mental arithmetic. This latter skill is become worryingly eroded.	Observations
B757/767-200/300	Pilots should be encouraged not to rely too heavily on FMS and to always be aware of raw data.	Observations
B737-300/800	VNAV path has to be monitored carefully to prevent overspeeds when changing waypoints.	Observations
B737-300/800	FMS on B737 is very easy to use; gives relevant information to phase of flight to back up pilot's own judgement.	Observations
B737-300/800	It is up to pilot to fly aircraft as he/she wants and not how FMS necessarily wants to.	Observations
B737-300/800	The INIT Ref/Approach page is compulsory on Finals for [Airline Name] and is well displayed.	Observations
B737-300/800	V/P, FPA and V/B info is never referred to.	Observations
A320/A321/A330	The "Fix" page on the B737-400 is very useful.	Observations
B757/767-200/300	The FMS is far from clever, when given a steep descent or lower cruise level as it completely loses the plot.	Observations
B777-200	Use of Vnav for approach has been abandoned due to programming problems.	Observations
B777-200	Vnav leads to unexpected performance problems on approach.	Observations
B737-300/800	"FMS error messages explain clearly where errors are, what the errors are and why they have occurred", FMS position depends on combination of radio position, left IRS, right IRS – automatically (on certain occasions) biasing position/excluding an erroneous source.	Observations
F-100	Our FMS has a software problem where it often "crashes" when entering take-off weight. System goes into "Independent Operation" and often must be reset by a "Self Test" This is frustrating when you are having to wait to take off because of this.	Observations
B737-300/800	I use home PCs extensively. I have a degree in computer information systems and can write computer programs (code) I had great difficult learning the FMS as it did not "work like a computer".	Observations
A320/A321/A330	The B767 FMS was much more logical.	Observations

Fleet	*Comments*	*Category*
A320/A321/A330	System is okay after initial learning curve, however, it is much different and more difficult than the Smith's unit.	Observations
A320/A321/A330	Uptrack, down track and abeam waypoint difficult to construct.	Observations
F-100	In general, I feel that the FMS used on Boeing aircraft is much more user friendly and is laid out in a much more logical manner than the FMS used by Fokker.	Observations
B757/767-200/300	B757 superb, can't fault it. However, previous experience "Saab 340" suggests that where an FMS is an "Add-on" i.e. not original equipment that confusion or difficulties arise, not least ergonomically. If an aircraft was not designed to have an FMS then a bolt on one that won't talk to the other aircraft systems usually means doing everything twice.	Observations
B757/767-200/300	If a climb is selected on the B757 it knows as soon as the ALT selector is moved on the autopilot panel.	Observations
B757/767-200/300	The "Hold Page" is a nightmare.	Observations
A320/A321/A330	No co-ordination between common ATC constraints and "standard database constraints".	Observations
A320/A321/A330	On previous types that I have flown, I have found the B737-400 (CDU) to be considerably more user friendly and easier to use than this Airbus CDU. The Boeing system was also easier to learn.	Observations
B757/767-200/300	Once chosen as a fix, constant presentation of choice for a given route (e.g. ABM) is irritating.	Observations
B757/767-200/300	Some systems e.g. Pegasus works best with data uplink. Manual data insertion e.g. winds - very time consuming.	Observations
B757/767-200/300	Some pages are useless e.g. Airbus prog page - what is its function.	Observations
B737-300/800	Reduces basic lookout/airmanship and does not encourage thinking.	Observations
B737-300/800	FMS is very good at distracting pilot's attention at inoperative moments. Too much reliance made on FMS.	Observations
B757/767-200/300	Increasingly slow between press of buttons (letters missed out).	Processor Speed
B757/767-200/300	All the FMSs are far too slow.	Processor Speed
B757/767-200/300	Poor speed of operation.	Processor Speed

Fleet	*Comments*	*Category*
B757/767-200/300	Some systems are so slow.	Processor Speed
B757/767-200/300	All too slow.	Processor Speed
B757/767-200/300	Computer is very slow although not so much of a problem in cruise!!	Processor Speed
B757/767-200/300	Later models can be quite slow.	Processor Speed
B757/767-200/300	Slow processing of entered data.	Processor Speed
B737-300/800	Processor agonisingly slow.	Processor Speed
B757/767-200/300	Slow to respond.	Processor Speed
B757/767-200/300	FMS CDU too slow to calculate intermediate entries where a sequence of keys are pressed to achieve a desired modification to a lateral profile.	Processor Speed
B757/767-200/300	Main problems are speed of operation.	Processor Speed
B757/767-200/300	Processor speed on FMS is predictably slow.	Processor Speed
B757/767-200/300	By modern standards of computing power, it lacks speed.	Processor Speed
B747-400	Limited and slow memory affect FMS performance with occasional re-sync "lock-ups" at inopportune times e.g. on approach.	Processor Speed
A320/A321/A330	They are very slow at processing information.	Processor Speed
B737-300/800	Time from input of data, to successful completion of calculation is too slow.	Processor Speed
A320/A321/A330	The main disadvantage of this FMS is its "speed" – it processes information very slowly (compared to computers elsewhere these days.)	Processor Speed
B737-300/800	Too slow.	Processor Speed
B737-300/800	The FMS is very quick and produces almost instant response to any calculation required of it. However, the FMS on the 737–400 is tediously slow.	Processor Speed
B757/767-200/300	FMS often slow.	Processor Speed
B757/767-200/300	FMSs often slow to talk to each other, and occasionally lose information in between.	Processor Speed
B747-400	FMS incredibly slow and only one person can make inputs at a time (setting up on the ground thus is a nuisance).	Processor Speed
B737-300/800	Very slow in operation. How about upgrading the processor?	Processor Speed

Fleet	*Comments*	*Category*
B757/767-200/300	B757-767 Too slow, Too slow, Too Slow! 737-300/400 and L1011 all faster.	Processor Speed
A320/A321/A330	Processor too slow for re-calculations.	Processor Speed
B757/767-200/300	It is just damned slow!	Processor Speed
B737-300/800	Very Slow.	Processor Speed
B757/767-200/300	B757 CDU is robust, simple but antiquated, small memory and slow.	Processor Speed
B737-300/800	Can take pilot away from immediate task ("head in cockpit") due to slow speed of it at times.	Processor Speed
F-100	Slow processing, will fault often when inputting weights.	Processor Speed
F-100	Process speed very slow.	Processor Speed
F-100	Main problem is the slow processor unit.	Processor Speed
B737-300/800	The computer is much too slow when calculation top of descent after data has been entered.	Processor Speed
F-100	FMS data input is often slow - often takes a long time to process data.	Processor Speed
B737-300/800	It is extremely slow in the calculation at descent points.	Processor Speed
F-100	Slow.	Processor Speed
B737-300/800	Processing is very slow.	Processor Speed
B757/767-200/300	The processor is too slow.	Processor Speed
B757/767-200/300	Processor speed very slow in computing VNAV Vs Track updates.	Processor Speed
B737-300/800	Having operated the FMS on the B747-400 I find the B737 model to be very slow.	Processor Speed
B737-300/800	FMS is slow to process information.	Processor Speed
B737-300/800	Airlines tend to spoil the FMS by poor initial training and confusing information contained in flight manuals. There is a culture of computer illiterate trainers, passing it off as "you'll get used to it on the line"!!	Training

Fleet	*Comments*	*Category*
A320/A321/A330	The system requires a great deal of training/familiarity during early stage of line flying. There is completely inadequate access to training system on computer courses. We are perfectly able to cope with these systems often learning by experiment rather than consultation with manuals.	Training
B737-300/800	The greatest problem in the use of the FMS is the poor standard of training in its use. Apart from one or two slow computers there is no means of getting hands on experience with an FMS other than during line training, which is far too busy in general for in depth discussion about "what is it doing now?"	Training
B737-300/800	The main problem with the use of FMSs is poor CRM or bad SOPs.	Training
A320/A321/A330	There is no on-line help as on home PC.	User Guidance & Support
B757/767-200/300	No online help facility and limited published on methodology of software is unhelpful.	User Guidance & Support
B737-300/800	No clear explanation of FMS position in any of our manuals – so can be confusing when IRS problems arise i.e. unclear how position is being calculated under these conditions.	User Guidance & Support
B737-300/800	Help - Our FMSs do not have an online help feature at all. As far as printed help, we only have our Pilot Handbooks and they are woefully inept as they are not written in a user-friendly manner.	User Guidance & Support
A320/A321/A330	Also the Standard of manuals suck!!	User Guidance & Support
B757/767-200/300	Manuals tend to be poorly presented by both Boeing and Airbus - can lead to a lack of understanding. Not easy for reference to "how to do a particular task" or "can we do that?"	User Guidance & Support
B737-300/800	Not much "help" info. available on FMS.	User Guidance & Support
B757/767-200/300	Workload is busy in terminal areas.	Workload
B737-300/800	Can be time consuming just at the wrong time.	Workload