



**VALUE FOCUSED THINKING APPROACH USING MULTIVARIATE
VALIDATION FOR JUNIOR ENLISTED PERFORMANCE REPORTING IN
THE UNITED STATES AIR FORCE**

THESIS

Christopher M. Jones, Captain, USAF

AFIT-ENS-14-M-13

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT-ENS-14-M-13

**VALUE FOCUSED THINKING APPROACH USING MULTIVARIATE
VALIDATION FOR JUNIOR ENLISTED PERFORMANCE REPORTING IN
THE UNITED STATES AIR FORCE**

THESIS

Presented to the Faculty

Department of Operations Research

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Operations Research

Christopher M. Jones, MBA

Captain, USAF

March 2014

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

**VALUE FOCUSED THINKING APPROACH USING MULTIVARIATE
VALIDATION FOR JUNIOR ENLISTED PERFORMANCE REPORTING IN
THE UNITED STATES AIR FORCE**

Christopher M. Jones, MBA

Captain, USAF

Approved:

//signed//
Jennifer L. Geffre, Major, USAF (Co-Chairman)

3 Mar 2014
Date

//signed//
Dr. Sharon G. Heilmann (Co-Chairman)

3 Mar 2014
Date

Abstract

From United States Air Force (USAF) doctrine, Air Force Instruction 1-1 lists three purposes for the USAF Enlisted Evaluation System. The first purpose is to provide feedback to individuals on how well they are meeting expectations. The second purpose is to provide a cumulative record of performance and potential based on observations. The third purpose is to identify the best qualified personnel. However, current Air Force leadership has expressed a need to revamp the enlisted appraisal process, requesting consistency in identifying the best performers, reduction in ratings inflation, and better delineation between “near peer” performers.

This research proposes utilizing Value-Focused Thinking to perform junior enlisted performance reports, to better align with Air Force doctrine and values. Moreover, the multivariate Management Science techniques of Exploratory and Confirmatory Factor Analysis are applied to statistically validate the accuracy and defensibility of the design. Finally, Artificial Neural Networks are employed to showcase the classification accuracy of the proposed system. In addition to providing consistency, inflation reduction, and delineation during appraisals, this research advocates the use of a web-based design to reduce administrative demands and to provide query capability of appraisal data to the Air Force Personnel Center for trend and force management decisions.

Acknowledgments

I would like to express my sincere appreciation to my faculty advisor, Major Jennifer L. Geffre, and my co-advisor Dr. Sharon Heilmann for their guidance and support throughout the course of this thesis effort. Their insight and experience was certainly appreciated. I would also like to thank Captain Andrew Cooper for his input in refining the requirements of the proposal and for allowing his organization to participate in data gathering and testing during development of the model. Finally, I would like to thank Chief Master Sergeant Jason Price who was the Senior Non-Commissioned Officer liaison for this project. The insight provided by Chief Price into both the enlisted corps value structure and the current EPR system proved vital in the development and refinement of this project.

Christopher M. Jones

Table of Contents

	Page
Abstract	iv
Table of Contents	vi
I. Introduction	1
General Issue	1
Problem Statement.....	12
Research Approach.....	12
Research Goals	13
Preview	15
II. Literature Review	18
Chapter Overview.....	18
Desired Traits of an Appraisal System	19
Military Research in Appraisal System Design.....	25
Mitigation Techniques to Address Appraisal System Concerns	27
Why use a Value-Focused Thinking Approach?.....	32
Validating a VFT Framework Using Multivariate Management Science Methods...37	
III. Value Model Construction	40
Chapter Overview.....	40
Purpose	42
VFT Values and Objectives.....	44
VFT Evaluation Hierarchy	49
VFT Weight Solicitation	56
VFT Attribute Function Development	61

VFT Attribute Function Revisions	67
VFT Data Collection	70
VFT Deterministic Analysis (Notional Dataset)	72
IV. Model Validation	81
Chapter Overview.....	81
Sensitivity Analysis (Notional Dataset)	83
Data Solicitation Process (Training Dataset)	89
Qualitative Inspection (Training Dataset)	90
Internal Consistency (Training Dataset).....	95
Factor Analysis Suitability (Training Dataset).....	101
Preliminary Analysis (Training Dataset).....	105
Data Reduction Technique Selection (Training Dataset)	108
Initial Dimensionality Assessment (Training Dataset).....	113
Initial Exploratory Factor Analysis and Interpretation (Training Dataset)	115
JEPR Model Revision (Based on Initial Factor Analysis Findings)	125
Final Dimensionality Assessment (Training Dataset).....	137
Final Exploratory Factor Analysis and Interpretation (Training Dataset).....	138
JEPR Decision Support System Tool Revision.....	143
V. Multivariate Analysis and Results	155
Chapter Overview.....	155
Data Solicitation Process (Test Dataset)	157
Qualitative Inspection (Test Dataset)	158
Internal Consistency (Test Dataset).....	162

Factor Analysis Suitability (Test Dataset).....	164
Preliminary Analysis (Test Dataset).....	166
Data Reduction Technique Selection (Test Dataset).....	167
Dimensionality Assessment (Test Dataset)	170
Exploratory Factor Analysis and Interpretation (Test Dataset).....	172
Structural Equation Modeling Overview.....	177
Confirmatory Factor Analysis Overview (Test Dataset).....	179
Confirmatory Factor Analysis Data Suitability (Test Dataset)	181
Confirmatory Factor Analysis Evaluation of Fit Criteria (Test Dataset)	194
Confirmatory Factor Analysis and Interpretation (Test Dataset)	201
Qualitative Classification (Test Dataset).....	214
Artificial Neural Network Suitability (Test Dataset)	219
Artificial Neural Network Background (Test Dataset)	221
Artificial Neural Network and Interpretation (Test Dataset)	223
VI. Conclusions and Recommendations	241
Conclusion.....	241
Significant Research Contributions.....	241
Recommendations for Future Research.....	244
Summary.....	245
Appendix I	246
Single Attribute Value Function for Duty Performance.....	246
Single Attribute Value Function for Duty Leadership	247
Single Attribute Value Function for Teamwork and Followership.....	248

Single Attribute Value Function for Respect for Service and Standards	249
Single Attribute Value Function for Discipline and Self-Control.....	250
Single Attribute Value Function for Communication	251
Single Attribute Value Function for Responsibility	252
Single Attribute Value Function for Honesty and Accountability	253
Single Attribute Value Function for Physical Fitness	254
Single Attribute Value Function for Military Awards	255
Single Attribute Value Function for Base and Community Involvement	256
Single Attribute Value Function for Education.....	257
Single Attribute Value Function for Administrative Actions Correction Factor	258
Appendix II	259
Sensitivity Analysis of Weight 1	259
Sensitivity Analysis of Weight 2.....	260
Sensitivity Analysis of Weight 3	261
Sensitivity Analysis of Weight 4.....	262
Sensitivity Analysis of Weight 5	263
Sensitivity Analysis of Weight 6.....	264
Sensitivity Analysis of Weight 7	265
Sensitivity Analysis of Weight 8.....	266
Sensitivity Analysis of Weight 9.....	267
Sensitivity Analysis of Weight 10.....	268
Sensitivity Analysis of Weight 11	269
Sensitivity Analysis of Weight 12.....	270

Appendix III.....	271
Value Breakout Attribute Contribution for Each JEPR Attribute	271
Appendix IV.....	272
Value Breakout Contribution for Each JEPR Fundamental Objective	272
Appendix V.....	273
Value-Gap Strengths and Shortfalls of Notional Airman A.....	273
Value Gap Strengths and Shortfalls of Notional Airman B	274
Value Gap Strengths and Shortfalls of Notional Airman C	275
Value Gap Strengths and Shortfalls of Notional Airman D.....	276
Value Gap Strengths and Shortfalls of Notional Airman E	277
Value Gap Strengths and Shortfalls of Notional Airman F.....	278
Value Gap Strengths and Shortfalls of Notional Airman G.....	279
Value Gap Strengths and Shortfalls of Notional Airman H.....	280
Appendix VI.....	281
Approved Exemption Request from Human Experimentation Requirements	281
Appendix VII	282
Confirmatory Factor Analysis Model Outputs	282
JEPR Test Dataset CFA Model (Baseline).....	282
JEPR Test Dataset CFA Model (Modified Model #1)	288
JEPR Test Dataset CFA Model (Modified Model #2)	293
JEPR Test Dataset CFA Model (Modified Model #3)	299
Appendix VIII.....	305
JEPR Test Dataset Artificial Neural Network (ANN) MATLAB Code	305
Appendix IX.....	307
Artificial Neural Network (ANN) SNR Values and Feature Weights	307

Bibliography	309
Vita.....	323

List of Figures

	Page
Figure 1. Contribution to Promotion Score (EPR Factor Included)	7
Figure 2. Contribution to Promotion Score (EPR Factor Nullified Due to Inflation)	8
Figure 3. Current Junior Enlisted Performance Report (Front Side)	9
Figure 4. Current Junior Enlisted Performance Report (Back Side)	10
Figure 5. Value Model Construction, Validation, and Analysis Process Overview	17
Figure 6. Overview of Value Hierarchy Refinement Methodology	41
Figure 7. Strategic Value Hierarchy	46
Figure 8. Strategic Value Hierarchy Focusing on Evaluations	49
Figure 9. Refined Value Hierarchy Framework	51
Figure 10. Example of Local Weighting Construct	58
Figure 11. Example of Global Weighting Construct	59
Figure 12. JEPR Value Hierarchy with Global Weight Structure	60
Figure 13. Duty Performance SAVF Function Example	63
Figure 14. Overall Scoring Scheme Comprised of Value Hierarchy Framework	66
Figure 15. Comparison of Exponential vs. Piecewise Function Fit	67
Figure 16. Value Breakout by Attribute of Value Function Scores	74
Figure 17. Value Breakout by Fundamental Objective of Value Function Scores	78
Figure 18. Value Gap Graph (Scores for Notional Airmen B Shown)	80
Figure 19. Overview of the Model Validation Chapter	82
Figure 20. Sensitivity of Duty Performance Weight $W_{DP}<32%$ and $W_{DP}>78%$	85
Figure 21. Sensitivity of Duty Leadership Weight $W_{DL}>13%$	86

Figure 22. Sensitivity of Physical Fitness Weight $W_{PF}<22\%$ and $W_{PF}>77\%$	87
Figure 23. Sensitivity of Communication Weight $W_{COMM}>63\%$	88
Figure 24. Distribution of 71 Performance Ratings (Current EPR System).....	91
Figure 25. Distribution of 71 Performance Ratings (JEPR System)	92
Figure 26. JEPR Distribution Ratings for Subjects Rated “5” (Current EPR System).....	94
Figure 27. JEPR Bidirectional Scaling Scheme (Increasing Value to the Right).....	98
Figure 28. Scree Plot of Initial Eigenvalues from the Initial Correlation Matrix [R].....	108
Figure 29. Data Reduction Techniques Tree (EFA Branch Highlighted)	110
Figure 30. Scree Plot of Eigenvalues from the Reduced Correlation Matrix [R*].....	115
Figure 31. JEPR Reduced Factors after Promax Oblique Rotation	120
Figure 32. JEPR Reduced Factors after Varimax Orthogonal Rotation	120
Figure 33. JEPR Value Hierarchy (Core Values Aligned on Rotated Factor Loadings)	125
Figure 34. Revised JEPR Value Hierarchy (Based on Core Values)	127
Figure 35. Revised JEPR Value Hierarchy (Theorized Factor Structure Overlay)	130
Figure 36. Scree Plot of Eigenvalues from the Reduced Correlation Matrix [R*].....	137
Figure 37. Factor Loading Plot for Two Factor JEPR Model (Rotated Orthogonally) ..	141
Figure 38. Revised JEPR Value Hierarchy (Validated Two Factor Structure Overlay)	142
Figure 39. JEPR DSS Service Before Self Factor	150
Figure 40. JEPR DSS Integrity Factor	151
Figure 41. JEPR DSS Excellence Factor	152
Figure 42. JEPR DSS Administrative Actions Penalty Function	153
Figure 43. JEPR DSS Career Target Output.....	154
Figure 44. Overview of Multivariate Analysis and Results Chapter	156

Figure 45. Distribution of 159 Performance Ratings (Current EPR System).....	158
Figure 46. Distribution of 159 Performance Ratings (JEPR System)	159
Figure 47. JEPR Distribution Ratings for Subjects Rated “5” (Current EPR System)...	161
Figure 48. Scree Plot of Initial Eigenvalues from the Initial Correlation Matrix [R].....	167
Figure 49. Data Reduction Techniques Tree (EFA Branch Highlighted)	168
Figure 50. Scree Plot of Eigenvalues from the Reduced Correlation Matrix [R*].....	171
Figure 51. Factor Loading Plot for Two Factor JEPR Model (Rotated Orthogonally) ..	175
Figure 52. Causal Model with Measurement and Structural Sub-Models.....	178
Figure 53. Data Reduction Techniques Tree (CFA Branch Highlighted)	180
Figure 54. Possible Outlier Points (Overall JEPR Score vs. Classification Category)...	191
Figure 55. Overidentified SEM Model of JEPR Test Dataset	203
Figure 56. Hypothesized SEM of JEPR Test Dataset (Baseline Model).....	204
Figure 57. Hypothesized SEM of JEPR Test Dataset (Modified Model #1).....	207
Figure 58. Hypothesized SEM of JEPR Test Dataset (Final Model)	212
Figure 59. JEPR Versus EPR Scoring (JEPR Classification Classes Overlaid).....	219
Figure 60 . McCulloch-Pitts Model Neuron or Single Threshold Unit (Krogh, 2008) ..	222
Figure 61. Feed-forward Two-Layer Network Example (Krogh, 2008)	223
Figure 62. ANN JEPR Classifier (JEPR Classes Shown)	225
Figure 63. ANN JEPR Classifier (MATLAB NPR Tool, 2012)	226
Figure 64. ANN EPR Training Confusion Matrix (111 of 159 Randomly Sampled)	228
Figure 65. ANN JEPR Validation Confusion Matrix (24 of 159 Randomly Sampled) .	229
Figure 66. ANN JEPR Test Confusion Matrix (24 of 159 Randomly Sampled)	230
Figure 67. ANN JEPR Combined Confusion Matrix (159 of 159 Randomly Sampled)	230

Figure 68. JEPR vs. EPR Scoring (JEPR Classification Classes Overlaid)	231
Figure 69. ANN EPR Classifier (EPR Classes Shown).....	232
Figure 70. ANN EPR Training Confusion Matrix (111 of 159 Randomly Sampled)	234
Figure 71. ANN EPR Validation Confusion Matrix (24 of 159 Randomly Sampled) ...	235
Figure 72. ANN EPR Test Confusion Matrix (24 of 159 Randomly Sampled).....	236
Figure 73. ANN EPR Combined Confusion Matrix (159 of 159 Randomly Sampled) .	236
Figure 74. JEPR vs. EPR Scoring (Translated EPR Classification Classes Overlaid) ...	237
Figure 75. JEPR vs. EPR Scoring (JEPR Classification Classes Overlaid)	238
Figure 76. Duty Performance Single Attribute Value Function	246
Figure 77. Duty Leadership Single Attribute Value Function.....	247
Figure 78. Teamwork and Followership Single Attribute Value Function	248
Figure 79. Respect for Service and Standards Single Attribute Value Function.....	249
Figure 80. Discipline and Self Control Single Attribute Value Function.....	250
Figure 81. Communication Single Attribute Value Function.....	251
Figure 82. Responsibility Single Attribute Value Function.....	252
Figure 83. Honesty and Accountability Single Attribute Value Function.....	253
Figure 84. Physical Fitness Single Attribute Value Function.....	254
Figure 85. Military Awards Single Attribute Value Function	255
Figure 86. Base and Community Involvement Single Attribute Value Function.....	256
Figure 87. Education Single Attribute Value Function	257
Figure 88. Administrative Actions Independent External Function	258
Figure 89. Scores For Eight Notional Airmen Using Provided Weights.....	259
Figure 90. Score Changes From Performance Weight Change	259

Figure 91. Scores For Eight Notional Airmen Using Provided Weights.....	260
Figure 92. Score Changes From Duty Leadership Weight Change.....	260
Figure 93. Scores For Eight Notional Airmen Using Provided Weights.....	261
Figure 94. Score Changes From Teamwork and Followership Weight Change	261
Figure 95. Scores For Eight Notional Airmen Using Provided Weights.....	262
Figure 96. Score Changes From Respect for Service and Standards Weight Change....	262
Figure 97. Scores For Eight Notional Airmen Using Provided Weights.....	263
Figure 98. Score Changes From Discipline and Self-Control Weight Change	263
Figure 99. Scores For Eight Notional Airmen Using Provided Weights.....	264
Figure 100. Score Changes From Communication Weight Change.....	264
Figure 101. Scores For Eight Notional Airmen Using Provided Weights.....	265
Figure 102. Score Changes From Responsibility Weight Change	265
Figure 103. Scores For Eight Notional Airmen Using Provided Weights.....	266
Figure 104. Score Changes From Honesty and Accountability Weight Change.....	266
Figure 105. Scores For Eight Notional Airmen Using Provided Weights.....	267
Figure 106. Score Changes From Physical Fitness Weight Change.....	267
Figure 107. Scores For Eight Notional Airmen Using Provided Weights.....	268
Figure 108. Score Changes From Military Awards Weight Change	268
Figure 109. Scores For Eight Notional Airmen Using Provided Weights.....	269
Figure 110. Score Changes From Base and Community Involvement Weight Change.	269
Figure 111. Scores For Eight Notional Airmen Using Provided Weights.....	270
Figure 112. Score Changes From Education Level Weight Change	270
Figure 113. Scores for Eight Notional Airmen Using Provided Weights.....	271

Figure 114. Contribution to Overall Score by Value Type for Eight Notional Airmen .	271
Figure 115. Scores for Eight Notional Airmen Using Provided Weights.....	272
Figure 116. Contribution to Overall Score by Value Type for Eight Notional Airmen .	272
Figure 117. Scores for Eight Notional Airmen (Airman A Highlighted).....	273
Figure 118. Value Gap Feedback for Notional Airman A.....	273
Figure 119. Scores for Eight Notional Airmen (Airman B Highlighted)	274
Figure 120. Value Gap Feedback for Notional Airman B	274
Figure 121. Scores for Eight Notional Airmen (Airman C Highlighted)	275
Figure 122. Value Gap Feedback for Notional Airman C	275
Figure 123. Scores for Eight Notional Airmen (Airman D Highlighted).....	276
Figure 124. Value Gap Feedback for Notional Airman D.....	276
Figure 125. Scores for Eight Notional Airmen (Airman E Highlighted)	277
Figure 126. Value Gap Feedback for Notional Airman E	277
Figure 127. Scores for Eight Notional Airmen (Airman F Highlighted).....	278
Figure 128. Value Gap Feedback for Notional Airman F	278
Figure 129. Scores for Eight Notional Airmen (Airman G Highlighted).....	279
Figure 130. Value Gap Feedback for Notional Airman G.....	279
Figure 131. Scores for Eight Notional Airmen (Airman H Highlighted).....	280
Figure 132. Value Gap Feedback for Notional Airman H.....	280

List of Tables

	Page
Table 1. Contribution to Promotion Score (EPR Factor Included).....	6
Table 2. Contribution to Promotion Score (EPR Factor Nullified Due to Inflation).....	7
Table 3. Rating Categories for Each Attribute.....	52
Table 4. Leadership/Performance and Values/Responsibilities Ratings Categories	52
Table 5. Physical Fitness Ratings Category.....	52
Table 6. Awards Ratings Category	53
Table 7. Education Level Ratings Category	53
Table 8. Base and Community Involvement Ratings Category.....	53
Table 9. Initial Rating Category Definitions for Duty Performance, Duty Leadership, and Communication in the Leadership & Performance Fundamental Objective	54
Table 10. Initial Rating Category Definitions for Leadership & Performance Fundamental Objective	54
Table 11. Initial Rating Category Definitions for Values & Responsibilities Fundamental Objective	55
Table 12. Initial Rating Category Definitions for Awards Sub-Category in Professional Qualities Fundamental Objective.....	55
Table 13. Initial Rating Category Definitions for Education Level Sub-Category in Professional Qualities Fundamental Objective	55
Table 14. Initial Rating Category Definitions for base and Community Involvement Sub- Category in Professional Qualities Fundamental Objective	55
Table 15. SME Ranking of Importance of Objectives and Weight Assignments	57
Table 16. Gamma Shaping Component for SAVFs Used in VFT Function	62
Table 17. Initial Rating Category Definitions for Penalty Function.....	65
Table 18. Piecewise Sectional Ranges and Slopes for Physical Fitness SAVF.....	69

Table 19. Piecewise Ranges and Slopes for Teamwork and Followership SAVF	69
Table 20. Piecewise Ranges and Slopes for Revised Education SAVF	69
Table 21. Piecewise Ranges and Slopes for Revised Penalty Function	70
Table 22. SAVF Scores for Notional Personnel A through D and an Ideal Airman	73
Table 23. SAVF Scores for Notional Personnel E through H and an Ideal Airman.....	73
Table 24. Fundamental Objective Hierarchy	77
Table 25. Scoring by Fundamental Objective.....	78
Table 26. Value Gap Computations (Scores for Notional Airmen B Shown).....	80
Table 27. JEPR Weight Assignments Based on SME Importance	83
Table 28. Cronbach's Alpha Value Quality for Internal Consistency.....	96
Table 29. JMP Generated Statistics for JEPR Data	99
Table 30. Raw Cronbach's Alpha Measures (Overall and with Excluded Attributes) ...	100
Table 31. Initial Correlation Matrix [R] Eigenvalues (JEPR Training Dataset).....	107
Table 32. SMC Based Prior Communality Estimates (Training Dataset)	111
Table 33. Final Communality Estimates (Training Dataset)	112
Table 34. Reduced Correlation Matrix [R*] Eigenvalues (JEPR Training Dataset)	113
Table 35. Unrotated Factor Loadings of the JEPR Training Dataset	117
Table 37. Orthogonal Rotation Results of JEPR Training Data	119
Table 36. Oblique Rotation Results of JEPR Training Data	119
Table 38. Service Before Self Core Value Relationship to JEPR Common Factor One	122
Table 39. Integrity Core Value Relationship to JEPR Common Factor Two.....	123
Table 40. Excellence Core Value Relationship to JEPR Common Factor Three	124
Table 41. JEPR Attributes Related to Standards	129

Table 42. JEPR Attributes Related to Professional Expectations.....	129
Table 43. Final JEPR Service Before Self Fundamental Objective Categories	131
Table 44. Final JEPR Excellence Fundamental Objective Categories	132
Table 45. Final JEPR Integrity Fundamental Objective Categories.....	133
Table 46. Final JEPR Administrative Actions Independent Penalty Function.....	133
Table 47. Final Gamma Shaping Components for SAVFs Used in VFT Function.....	134
Table 48. Final Ranges and Slopes for Piecewise Physical Fitness SAVF	135
Table 49. Final Ranges and Slopes for Piecewise Teamwork and Followership SAVF	135
Table 50. Final Ranges and Slopes for Piecewise Education SAVF	135
Table 51. Final Ranges and Slopes for Piecewise JEPR Penalty Function	136
Table 52. Two Factor JEPR Model Unrotated Factor Loadings	139
Table 53. Two Factor JEPR Model Orthogonally Rotated Factor Loadings.....	139
Table 54. Two Factor JEPR Model Oblique Rotated Factor Loadings	140
Table 55. JEPR Classification Classes and Class Descriptions	148
Table 56. Cronbach's Alpha Value Quality for Internal Consistency.....	162
Table 57. Raw Cronbach's Alpha Measures (Overall and with Excluded Attributes) ...	163
Table 58. Initial Correlation Matrix[R] Eigenvalues (JEPR Test Dataset)	166
Table 59. Final Communality Estimates from Factor Analysis (Test Dataset)	169
Table 60. Reduced Correlation Matrix [R*] Eigenvalues (Test Dataset).....	170
Table 61. Unrotated Factor Loadings from the JEPR Test Dataset.....	172
Table 62. Orthogonally Rotated Factor Loadings.....	173
Table 63. Two Factor JEPR Model Oblique Rotated Factor Loadings	174
Table 64. Empirical Mixed Distributions and Parameters (JEPR Test Dataset)	183

Table 65. Empirical Johnson SL Distributions and Parameters (JEPR Test Dataset)	183
Table 66. Assessment of Normality Data for JEPR Test Dataset.....	186
Table 67. Outlier Test of JEPR Test Dataset Using (<i>MDist2</i>) Distances	190
Table 68. Absolute and Incremental Fit Indices Used for CFA (JEPR Test Dataset)	196
Table 69. Statistical Tests for CFA of JEPR Test Dataset (Baseline Model).....	205
Table 70. Recommended MIs for CFA of JEPR Test Dataset (Baseline Model).....	206
Table 71. Covariances Added for JEPR Test Dataset CFA Modification	211
Table 72. Statistical Tests Summary for CFA of JEPR Test Dataset	211
Table 73. EFA Loadings vs. CFA Regression Weights (JEPR Test Dataset)	213
Table 74. JEPR and Translated EPR Classification Classes.....	214
Table 75. Pivot Table of Translated EPR Classes and JEPR Classification Classes.....	215
Table 76. Classification Discrepancies (JEPR Below Standards Classification)	216
Table 77. Classification Discrepancies (JEPR Meets Standards Classification).....	217
Table 78. JEPR and Translated EPR Classification Classes.....	220
Table 79. JEPR Classification Classes.....	225
Table 80. Translated EPR Classification Classes	232

List of Equations

	Page
Equation 1. Single Attribute Value Function for VFT Function Sub-Objectives.....	62
Equation 2. VFT Weighted Value Model.....	64
Equation 3. Initial Penalty Function Weight Assignment	65
Equation 4. Initial Penalty Function	65
Equation 5. Initial Penalty Function Weighted Model	65
Equation 6. JEPR Overall Score	66
Equation 7. Revised SAVF for Piecewise Sub-Objectives.....	68
Equation 8. Revised Piecewise Penalty Function	70
Equation 9. Cronbach's Alpha Equation (Raw)	96
Equation 10. Sum Squares Equation for Computing Correlation Matrix Elements	101
Equation 11. Initial Correlation Matrix Design (JEPR Training Dataset).....	102
Equation 12. Kaiser-Meyer-Olkin (KMO) Index Measure of Sampling Adequacy.....	102
Equation 13. Bartlett's Test of Sphericity (JEPR Training Dataset).....	105
Equation 14. Characteristic Equation for Eigenvalue and Eigenvector Extraction	106
Equation 15. Prior Community Variance Estimation (SMC Method).....	111
Equation 16. Reduced Correlation Matrix Design (Initial Communalities).....	112
Equation 17. Reduced Correlation Matrix (Final Communalities).....	113
Equation 18. Loadings Computation from the Reduced Correlation Matrix.....	116
Equation 19. Final SAVF for Exponential Sub-Objectives	133
Equation 20. Final SAVF for Piecewise Sub-Objectives	134
Equation 21. Final Piecewise JEPR Penalty Function.....	136

Equation 22. Bartlett's Test of Sphericity (JEPR Test Dataset).....	165
Equation 23. Reduced Correlation Matrix (Final Communalities).....	170
Equation 24. Centroid Matrix for Mardia's Test Statistic.....	187
Equation 25. Centroid Matrix for Mardia's Test Statistic (Transposed).....	187
Equation 26. Inverse Covariance Matrix for Mardia's Test Statistic.....	187
Equation 27. Matrix of Squared Mahalanobis Distances.....	187
Equation 28. Mardia's Measure	187
Equation 29. Mardia's Multivariate Kurtosis Test	188
Equation 30. AGFI Absolute Fit Index.....	197
Equation 31. RMSEA Absolute Fix Index	198
Equation 32. SRMR Absolute Fix Index	198
Equation 33. TLI Incremental Fit Index	199
Equation 34. CFI Incremental Fix Index	200
Equation 35. SEM Regression Equations (Standards Factor)	202
Equation 36. SEM Regression Equations (Professional Expectations Factor).....	202
Equation 37. SEM Regression Equation (Covariance Between Factors).....	202
Equation 38. ANN Mean Square Error Function.....	227
Equation 39. ANN Back-Propagation Algorithm.....	227

VALUE FOCUSED THINKING APPROACH USING MULTIVARIATE VALIDATION FOR JUNIOR ENLISTED PERFORMANCE REPORTING IN THE UNITED STATES AIR FORCE

I. Introduction

General Issue

In the commercial business world, the topic of performance appraisals has long been a controversial topic for managers and employees alike (Jafari, Bourouni, & Amiri, 2009; Meyer, Kay, & French, 1965). Organizations use appraisal systems to let employees gage how well their performance compares with the expectations of the supervisor and the company. Performance appraisal systems are also used by organizations to identify areas where employees may require additional training or development to reach full potential in their assigned positions (Bae, 2006; Boice & Kleiner, 1997). When employee performance expectations are met or exceeded, the company benefits from increased productivity or efficiency, incur a financial savings, or increases in profit. Employees meeting or exceeding expectations may be rewarded with bonuses, promotions, and/or future leadership opportunities. However, in situations where employee performance is deficient, companies may experience a loss in productivity, or even worse, incur catastrophic disasters (including financial losses and/or loss of life). Therefore, unsatisfactory performance must be conveyed to the employee and documented to provide a record for charting improvement, demotion, or termination (Gizaw, 2010).

From an employee's standpoint, performance appraisals provide employees insight into how their performance is viewed by supervisors or their organizational

leadership, and provide an avenue for job progression and increased responsibilities and salaries. Regardless of whether the employee needs to improve performance or needs to continue to sustain their current level of performance, employees must know areas of strength in their work habits and weakness in their duty performance. Because the consequences of performance appraisal systems have the potential to significantly impact both the organization and the employees, it is vital that the performance appraisal framework be systematic and ensure appraisals are conducted in a fair and consistent manner (Boice & Kleiner, 1997).

Military organizations are no different to their civilian counterparts, as they use performance appraisals to reward high performing employees with promotions and increased leadership opportunities. Military organizations, like civilian entities, also use performance appraisals to provide feedback to employees, and if an individual is under-performing, appraisals are used to provide a training roadmap to enable the employee to meet expectations. If the employee cannot meet expectations, performance appraisals, just as they do in civilian companies, provide military organizations an avenue for demotion or termination of under-performing employees. The consequences of performance appraisal systems can be significant for both military organizations and military members. Therefore, any organization which relies on appraisal systems to determine employee progression or censure must use a performance appraisal framework that is systematic, fair, and consistent (Boice & Kleiner, 1997; Roberts, 2003).

Historically, performance appraisal systems used by the United States military have been a topic of discussion when concerning the design of a systematic, consistent, and fair system. As cited by D.J. Jackson and Ward, studies conducted by the Air Force

Military Personnel Center in 1988, concluded that the enlisted evaluation system (EES) was ineffective (D. J. Jackson & Ward, 1992). Efforts were undertaken in 1990 and again in 2006 and 2009 in an attempt to improve the Air Force enlist evaluation process; resulting in the system we have today (Air Force Pamphlet 36-2241, 2013, p. 197, p. 200).

The Air Force enlisted appraisal system of year 2014 strives to provide the rater a means to assess and document the ratees' performance, quantify performance, and to provide constructive feedback based on the supervisors observations of work habits (Air Force Pamphlet 36-2241, 2013, p. 252). The evaluation is intended to measure the ratees' performance versus the standards conveyed by the supervisor at the beginning of the rating period. The process is also intended to provide an avenue for the supervisor to evaluate the ratees' future potential to meet the standards and expectations (Air Force Pamphlet 36-2241, 2013, p. 252).

However, since 2008, there has been increasing pressure to reevaluate the fairness and equity of the Air Force enlisted performance appraisal system. A 2008 Project Air Force study by the RAND Corporation noted the Air Force enlisted promotion system is not generating consistent and deliberate results and is not meeting the intended goal for promotions (Schiefer, Robbert, Crown, Manacapilli, & Wong, 2008). The RAND study went on to determine that the current system is failing to meet the intent of Air Force Policy Directive (AFPD) 36-25, which requires the enlisted promotion system to "identify those people with the highest potential to fill positions of increased grade and responsibility" (Schiefer et al., 2008). Additionally, there has been a recent flurry of Opposite the Editorial page articles (Larter, Dec 2011; Losey, Sep 16, 2013; Losey, Sep

18, 2013; Schogol, Jan 2013; Schogol, Feb 2013; Schogol, Mar 2013; Schogol, Sep 2013) published concerning problems with the Air Force EPR system. What was surprising about these articles was that the members voicing dissatisfaction with the current system ranged from Air Force senior leaders to the most junior of airmen. Finally, a recent military War College research paper by a senior Air Force leader contemplated the effectiveness of the Air Force personnel evaluation system (Yates, 2011, p. 7). So, one might ask, is there really a problem with the current Air Force appraisal system?

Air Force Colonel Brian Yates addressed the topic of appraisal rating inflation in great detail in 2011. According to Colonel Yates, during the 2009 E-7 Air Force promotion board, there were 1,269 members selected to the rank of Master Sergeant, all of whom had perfect evaluation scores. Yet there were 11,502 other E-6 airmen who were also rated “Truly Among the Best”, who were not selected for promotion (Yates, 2011, p. 7). So if the EPR does not appear to be delineating airmen performance, then what role is the EPR serving? In February 2013, the Chief Master Sergeant of the Air Force (CMSAF), CMSgt James Cody, addressed the issue of EPRs. Speaking to a group of deployed Airmen from the 386th Expeditionary Wing, Chief Cody stated:

“When you talk about the EPR specifically and our performance assessment, we have a responsibility to give our airmen fair and honest feedback. Those performance assessments need to be fair and we need to delineate who is the very best, who has met the standards and we need to clearly show those who have not met the standards. So as we move forward -- and I've talked with General Welsh about this several times -- we are going to look at EPRs. We promise you that. But we are going to begin in a very thoughtful way and that is to go back and look at what we have already looked at to make sure we reevaluated and reviewed those

things we were thinking of in the past before we move forward. But really the commitment we have to this is we are going to take a serious look at EPRs and the entire system to ensure they are doing what we want them to do for us as an Air Force” (Thompson, Feb 2013).

Since February, the Air Force has been engaged in an exhaustive review of the current enlisted performance appraisal system. Upon conclusion of the investigative research effort, Chief Cody again addressed the media on 18 September 2013. Chief Cody was quoted as saying “It’s an inflated system. It’s clear in the numbers” (Losey, Sep 2013).

During part of the review process mandated by Chief Cody, the data revealed that from a high point in 2009 where 85.3 percent of airmen received perfect performance ratings, the percentage has dropped only 2.4 percent to 82.9 percent as of 2011 (Schogol, Mar 2013). Chief Cody also confirmed that performance appraisals, the most heavily weighted component in promotion consideration for enlisted airmen, has largely become a non factor due to over-inflated scores, with other factors such as specialty knowledge test scores, time in grade, time in service, or medals being the deciding factors (Losey, Sep 2013). If the inflation of performance reports is nullifying the EPR component in promotion determinations, then what factors are dominant in determining promotion fitness? A quick overview of the Weighted Airmen Promotion System (WAPS) may be able to provide some insight as to what issues were discovered during the Air Force level review.

From a management science standpoint, the use of performance appraisals for promotion consideration and delineating performance of civilian employees has long

been an established method (Hubbell & Chory-Assad, 2005; Mayer & Davis, 1999). The Air Force is no different. The Air Force meets this objective through AFPD 36-25, which requires that the enlisted promotion system “identify those people with the highest potential to fill positions of increased grade and responsibility.” To meet the intent of AFPD 36-25, the Air Force created the WAPS system which is used to determine which enlisted airmen are suitable for promotion to the next higher rank. The WAPS system consists of six weighted factors which sum together into an overall score, which is used for promotion determination. The first component is comprised of weighted EPRs, with the more current reports having increased impact on the point’s total. The second factor is the specialty knowledge test (SKT), which is a test of an individual’s specific career field knowledge, while the third factor is the promotion fitness examination (PFE), and is based on general Air Force knowledge. The remaining factors are time in service (TIS), time in grade (TIG), and number and type of decorations awarded. Each factor is assigned points based on its importance, with 460 points being the maximum that can be earned overall. Looking at the contribution of each component based on the maximum number of points available, Table 1 and Figure 1 provide clear evidence that the EPR was designed to be the most dominant factor in deciding junior enlisted promotions.

Table 1. Contribution to Promotion Score (EPR Factor Included)

Promotion Factor	Maximum Score Possible	% Contribution to Overall Promotion Score
EPRs	135	29%
Specialty Knowledge Test	100	22%
Promotion Fitness Examination	100	22%
Time In Grade	60	13%
Time In Service	40	9%
Decorations	25	5%

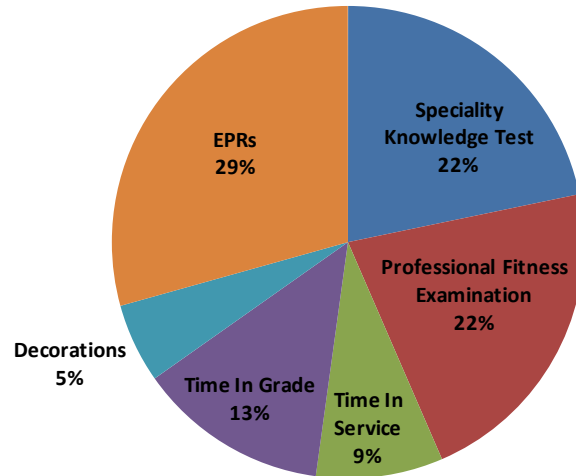


Figure 1. Contribution to Promotion Score (EPR Factor Included)

However, if the appraisal system is truly experiencing inflation, and most members are maximizing the EPR component, then the EPR factor, the most heavily weighted factor by design and doctrine, is effectively nullified from the computation. Looking at the component contributions from this vantage point, it is apparent that in an inflated appraisal system, the SKT and PFE components dominate the remaining portions of the overall score. In an inflated system, the EPR, which is intended to be the most heavily weighted component, is effectively nullified, with 62% of the promotion determination coming from the SKT and PFE written test examinations. This can be seen explicitly in Table 2 and Figure 2.

Table 2. Contribution to Promotion Score (EPR Factor Nullified Due to Inflation)

Promotion Factor	Maximum Score Possible	% Contribution to Overall Promotion Score
Specialty Knowledge Test	100	31%
Promotion Fitness Examination	100	31%
Time In Grade	60	18%
Time In Service	40	12%
Decorations	25	8%

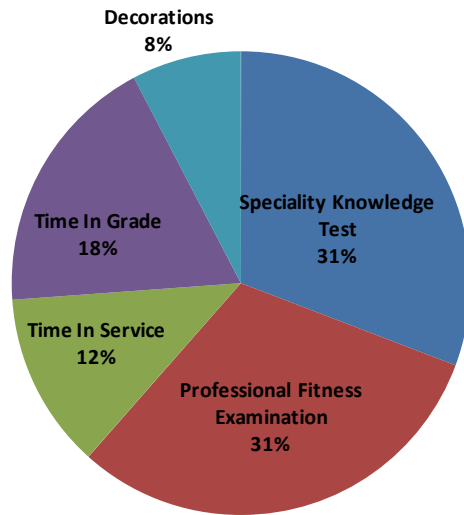


Figure 2. Contribution to Promotion Score (EPR Factor Nullified Due to Inflation)

The quick look at the promotion system seems to support Chief Cody's and Colonel Yates conclusions that the primary problem with the promotion system is inflation. However, because several promotion factors are interdependent, inflation of EPR ratings can also impact the ability to delineate airmen and may also affect the consistency of the appraisal system.

Part of the concern with the current enlisted appraisal system voiced by senior leaders, users, and by independent analysts such as RAND may be attributed to the current system's design construct. Motivation suffers when employees believe that their behaviors will not be rewarded (Hubbell & Chory-Assad, 2005; Mayer & Davis, 1999; Noe, Hollenbeck, Gerhart, & Wright, 1997, p. 236). If users feel that performance is marginalized, or inflated, they may feel there is a lack of consistency with ratings. In the civilian sector, analyses by psychologists such as Greenberg (1986) support the concerns of consistency. Greenberg's research indicates that subordinates' beliefs about a fair performance evaluation may be based on the procedures by which the evaluation process

was constructed, regardless of the ratings received. When considering the existing Air Force appraisal design, Figure 3 and Figure 4 illustrate the current performance assessment form construct, the Air Force form AF910.

ENLISTED PERFORMANCE REPORT (AB thru TSgt)			
I. RATEE IDENTIFICATION DATA (Refer to AFJ 36-2406 for instructions on completing this form)			
1. NAME (Last, First, Middle Initial)	2. SSN	3. GRADE	4. DAFSC
5. ORGANIZATION, COMMAND, LOCATION, AND COMPONENT		6. PAS CODE	7. SRID
8. PERIOD OF REPORT From: Thru:		9. NO. DAYS SUPERVISION	10. REASON FOR REPORT
II. JOB DESCRIPTION			
1. DUTY TITLE		2. SIGNIFICANT ADDITIONAL DUTY(S)	
3. KEY DUTIES, TASKS, AND RESPONSIBILITIES (Limit text to 4 lines)			
III. PERFORMANCE ASSESSMENT			
1. PRIMARY/ADDITIONAL DUTIES (For SSgt/TSgt also consider Supervisory, Leadership and Technical Abilities) Consider Adapting, Learning, Quality, Timeliness, Professional Growth and Communication Skills (Limit text to 4 lines)			
<input type="checkbox"/> Does Not Meet <input type="checkbox"/> Meets <input type="checkbox"/> Above Average <input type="checkbox"/> Clearly Exceeds			
2. STANDARDS, CONDUCT, CHARACTER & MILITARY BEARING (For SSgt/TSgt also consider Enforcement of Standards and Customs & Courtesies) Consider Dress & Appearance, Personal/Professional Conduct On/Off Duty (Limit text to 2 lines)			
<input type="checkbox"/> Does Not Meet <input type="checkbox"/> Meets <input type="checkbox"/> Above Average <input type="checkbox"/> Clearly Exceeds			
3. FITNESS (Maintains Air Force Physical Fitness Standards) (For referrals, limit text to 1 line)			
<input type="checkbox"/> Does Not Meet <input type="checkbox"/> Meets <input type="checkbox"/> Exempt			
4. TRAINING REQUIREMENTS (For SSgt/TSgt also consider PME, Off-duty Education, Technical Growth, Upgrade Training) Consider Upgrade, Ancillary, OJT and Readiness (Limit text to 2 lines)			
<input type="checkbox"/> Does Not Meet <input type="checkbox"/> Meets <input type="checkbox"/> Above Average <input type="checkbox"/> Clearly Exceeds			
5. TEAMWORK/FOLLOWERSHIP (For SSgt/TSgt also consider Leadership, Team Accomplishments, Recognition/Reward Others) Consider Team Building, Support of Team, Followership (Limit text to 2 lines)			
<input type="checkbox"/> Does Not Meet <input type="checkbox"/> Meets <input type="checkbox"/> Above Average <input type="checkbox"/> Clearly Exceeds			
6. OTHER COMMENTS Consider Promotion, Future Duty/Assignment/Education Recommendations and Safety, Security & Human Relations (Limit text to 2 lines)			
IV. RATER INFORMATION			
NAME, GRADE, BR OF SVC, ORGN, COMMAND AND LOCATION		DUTY TITLE	DATE
SSN		SIGNATURE	
		Click here to sign	

AF FORM 910, 20080618

PREVIOUS EDITIONS ARE OBSOLETE

PRIVACY ACT INFORMATION: The information in this form is FOR OFFICIAL USE ONLY. Do not release to the public. Act of 1974

Figure 3. Current Junior Enlisted Performance Report (Front Side)

V. OVERALL PERFORMANCE ASSESSMENT		RATEE NAME: _____				
Overall Performance During Reporting Period						
ASSESSMENT	POOR (1)	NEEDS IMPROVEMENT (2)	AVERAGE (3)	ABOVE AVERAGE (4)	TRULY AMONG THE BEST (5)	
RATER'S ASSESSMENT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
ADDITIONAL RATER'S ASSESSMENT	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Last feedback was performed on: _____ If feedback was not accomplished in accordance with AFI 36-2406, state the reason.						
VI. ADDITIONAL RATER'S COMMENTS (Limit text to 3 lines) <input type="checkbox"/> CONCUR <input type="checkbox"/> NON-CONCUR						
NAME, GRADE, BR OF SVC, ORGN, COMMAND AND LOCATION		DUTY TITLE			DATE	
		SSN _____ SIGNATURE _____				
					Click here to sign	
VII. FUNCTIONAL EXAMINER/AIR FORCE ADVISOR <input type="checkbox"/> FUNCTIONAL EXAMINER <input type="checkbox"/> AIR FORCE ADVISOR						
(Indicate applicable review by marking the appropriate box.)						
NAME, GRADE, BR OF SVC, ORGN, COMMAND AND LOCATION		DUTY TITLE			DATE	
		SSN _____ SIGNATURE _____				
					Click here to sign	
VIII. UNIT COMMANDER/CIVILIAN DIRECTOR/OTHER AUTHORIZED REVIEWER <input type="checkbox"/> CONCUR <input type="checkbox"/> NON-CONCUR						
NAME, GRADE, BR OF SVC, ORGN, COMMAND AND LOCATION		DUTY TITLE			DATE	
		SSN _____ SIGNATURE _____				
					Click here to sign	
IX. RATEE'S ACKNOWLEDGEMENT						
I understand my signature does not constitute agreement or disagreement. I acknowledge all required feedback was accomplished during the reporting period and upon receipt of this report.						
<input type="checkbox"/> Yes <input type="checkbox"/> No						
SIGNATURE _____		DATE _____				
					Click here to sign	
INSTRUCTIONS						
Complete this report IAW AFI 36-2406. Reports written by Colonels or civilians (GS-15 or higher, or Supervisory Pay Band 3), do not require an additional rater; however, endorsement by the rater's rater is permitted unless the report is written by a senior rater or the Chief Master Sergeant of the Air Force. When the rater's rater is not at least a MSgt or civilian (GS-07 or higher, or Supervisory Pay Band 1), the additional rater is the next official in the rating chain meeting grade requirements. An overall rating of 2 or negative comments require the EPR to be referred IAW AFI 36-2406. Rationale for any additional evaluator nonconcurring with an overall rating must be included. Section VIII Reviewer nonconcurrency must be included on an AF Form 77, Letter of Evaluation. If ratee is deployed, provide copy and feedback via e-mail/telecon.						
PRIVACY ACT STATEMENT						
AUTHORITY: Title 10 United States Code, Section 8013 and Executive Order 9397, 22 November 1943.						
PURPOSE: Information is needed for verification of the individual's name and Social Security Number (SSN) as captured on the form at the time of rating.						
ROUTINE USES: May specifically be disclosed outside the DoD as a routine use pursuant to 5 U.S.C. 552a(b)(3).						
DISCLOSURE: Disclosure is mandatory; SSN is used for positive identification.						
AF FORM 910, 20080618		PREVIOUS EDITIONS ARE OBSOLETE			PRIVACY ACT INFORMATION: The information in this form is FOR OFFICIAL USE ONLY. Protect IAW the Privacy Act of 1974.	

Figure 4. Current Junior Enlisted Performance Report (Back Side)

Of particular note, notice that the performance feedback section markings (section III) on the front side of the form are mathematically independent of the overall rating section

(section V) located on the backside of the form. This lack of connectivity is further illustrated when considering the doctrine that governs the Air Force Officer and Enlisted Evaluation System. Paragraph 3.1.10.1.4 of AFI 36-2406, of the Officer and Enlisted Evaluation Systems Instruction, states the following:

3.1.10.1.4. Above Average (4): Performs beyond established standards and expectations, performs at higher level than many of their peers. A rates' performance assessments on the front of the AF Form 910 or AF Form 911 may, or may not, all be marked "Clearly Exceeds" with a fitness assessment of "Meets" or "Exempt" and still receive this rating (Air Force Instruction 36-2406, 2013, p. 83).

Therefore, the design of the appraisal system and doctrine appear to contribute to the current systems perceived deficiencies.

According to Chief Cody, the current enlisted appraisal system appears to be at the root of this problem and has created a climate where inaccurate evaluations mask the true performance of airmen. In an 18 Sep 2013 interview, Chief Cody stated "Today, it is the other factors that we evaluate that discriminate...Performance is not the great discriminator." (Losey, Sep 2013). Chief Cody's assessment is further supported by the quick look of the WAPS system, where SKT and PFE testing were shown to be the dominant factors for promotion and progression in a suspected inflated EPR environment. Finally, the findings of the 2008 RAND Project Air Force study supports the belief that the Air Force enlisted promotion system is not generating consistent and deliberate results and is not meeting the intended promotion goals, which is to "identify those

people with the highest potential to fill positions of increased grade and responsibility” (Schiefer et al., 2008).

Problem Statement

The civilian community values a performance appraisal framework which is systematic and ensures appraisals are conducted in a fair and consistent manner (Boice & Kleiner, 1997; Heslin & Don VandeWalle, 2011). Looking back at the statement given by the CMSAF, Air Force leadership appears to share the values of the civilian sector when it comes to junior enlisted appraisals. The Air Force desires an appraisal framework that is fair, can delineate the best airmen, and is consistent (Thompson, 2013). The performance appraisal framework should incorporate leadership values, provide an avenue to translate qualitative measurements of performance to quantitative values, and should quantitatively highlight areas of performance feedback for the airman in relation to organizational goals and standards.

Research Approach

The purpose of this project is to develop a model framework which revamps the junior enlisted EPR system. This revision will seek to provide consistency, control ratings inflation, and provide the ability to delineate airmen. The vision is to provide a framework for a new performance evaluation system which qualitatively captures the performance of the individual over the evaluation period in meaningful areas of performance for both the Air Force and to the individual.

This new method seeks to identify superior performers for future leadership opportunities and promotion while also providing constructive feedback to the individual

concerning both areas of strength and weakness. The new system is expected to also reduce the administrative footprint of report generation for supervisors and senior unit leaders through the use of secure CAC encrypted web-based technologies. The use of Decision Theory, by the way of Value-Focused Thinking (VFT), will transform the evaluation process by translating qualitative inputs into quantitative output, which is focused on the performance areas Air Force leadership value the most. From a management science prospective, the systems underlying construct will be linked doctrinally to a set common factors at the heart of the Air Force value structure. Finally, the system will use management science techniques to assist in the control of bias, to control design inflation, to invoke trust, and to ensure internal consistency of the design.

Research Goals

The first goal of this research project is to illustrate how a VFT approach could be utilized to more accurately capture the true performance on junior enlisted personnel in the United States Air Force (USAF). By using a VFT approach, personnel who exhibit the traits most desired by the USAF can be recognized for their stellar performance, selected for promotion, and identified for future leadership opportunities. Conversely, substandard personnel could also be clearly identified as incongruent with the USAF value structure. The VFT methodology also creates a medium to provide improved feedback to members by detailing areas of strength and areas requiring improvement. Airmen will be presented quantitative metrics on their observed performance and will also be provided quantitative data on what is required to maximize performance. Finally, adopting the evaluation framework seeks to reduce the administrative demands on unit

senior leaders through the use of a web-based application, the incorporation of a more streamlined appraisal routing process, and through the use of a revised signatory process in performing junior enlisted performance appraisals. With reduced administrative demands, leaders will be able to increase their focus more onto ‘hands-on’ leadership and mentoring of junior members.

A second goal of this research is to use established management statistical techniques to validate the VFT framework is congruent with Air Force values, organizational goals, and doctrine. One of these established methods is to use Cronbach’s alpha (Cronbach, 1951) to measure the internal consistency of the VFT framework. Another technique from the management science community to be applied is to use Exploratory Factor Analysis (EFA). Factor analysis is a multivariate statistical procedure that is commonly used in the fields of psychology and education for the development, refinement, and evaluation of tests, scales, and measures (Williams, Brown, & Onsmann, 2012). For this research, EFA will be used on an initial validation data set to determine what the underlying unobserved factors (values) are that comprise the Air Force appraisal system and will examine the suitability of the initial VFT hierarchy structure. Once this underlying structure is confirmed with EFA, the VFT framework will be adjusted if necessary to ensure that the construct is congruent with Air Force doctrine, goals, and values.

Finally, the third goal of this project is to confirm the analytic capabilities of the model and to validate the results. First, using a larger sample from the Air Force population, Confirmatory Factor Analysis (CFA) will be applied to confirm that the VFT framework remained consistent with the factors (values) uncovered during EFA, and that

the larger population model accurately captured the performance of airmen during the appraisal process, and that the process remained congruent with Air Force doctrine, goals, and values during the appraisal. Secondly, an Artificial Neural Network (ANN) classifier will be applied to the large sample data to confirm that the values solicited to construct the VFT framework can accurately classify appraisals as Exceeds Standards, Meets Standards, or Below Standards in accordance with Air Force values and doctrine. The values from the VFT Framework will also be studied for classification success versus the current EPR system method of classification of ratees' as Exceeds Standards, Meets Standards, or Below Standards.

Preview

Chapter two discusses the literature review that was compiled in researching this problem. This research focuses on the desired traits of a personnel appraisal system, mitigation techniques for appraisal system concerns, and then explains why a valued focused approach is appropriate for performing personnel appraisals.

Chapter three focuses on the VFT and management science techniques used for model development and validation of the system. Chapter three discusses the methods used for solicitation and development of an initial value hierarchy, the development of Single Attribute Value Functions (SAVFs) for each attribute, and the creation of a Multi-Attribute Value Function (MAVF) that fully describes the desired performance attributes for junior enlisted airmen. Chapter three also discusses how the use of Decision Analysis techniques were used to study how each attribute contributed to the overall design of the framework using deterministic analysis techniques.

Chapter four explores how sensitive the model is to changes in the weighting schemes that were solicited from a group of Subject Matter Experts (SMEs). From a management science perspective, chapter four details the use of Cronbach's alpha on a training data set to validate the internal consistency of the framework and measurement scales, while also discussing the suitability and Exploratory Factor Analysis (EFA) to confirm how the value hierarchy is related to doctrine. Chapter four concludes by discussing the modifications to the hierarchy and framework based on the discoveries revealed during the factor analysis and variable rotation.

Chapter five details the multivariate analysis and results after introducing a statistically relevant real world data set from an Air Force sample population. Using this real world population, chapter five verifies the revised framework's consistency, again through the use of grounded management science techniques such as Cronbach's alpha, Confirmatory Factor Analysis (CFA), and variable rotation, and relates how the framework is directly derived from Air Force doctrine. Chapter five also explores Artificial Neural Networks (ANNs) are used to verify the classification effectiveness of both the current EPR system and JEPR system based on the VFT Framework solicited in chapter three and validated in chapter four. Finally, chapter six details the conclusions that were arrived at from the research and analytical effort while also providing insight into the modeling effort.

Chapter six concludes with how the model mitigated several of the common shortcomings of appraisal systems, and where this type of model could be incorporated in future efforts or research. An overview of the entire analytical process encompassed by this research is illustrated in Figure 5. The green dashed line Figure 5 highlights the VFT

processes that occurred during chapter III, the JEPR Value Model Construction, the red dashed line encapsulates the multiple EFA efforts performed during chapter IV, Validation, while the gray dashed lines illustrates the CFA and ANN analysis that occurred during chapter V, Multivariate Analysis and Results.

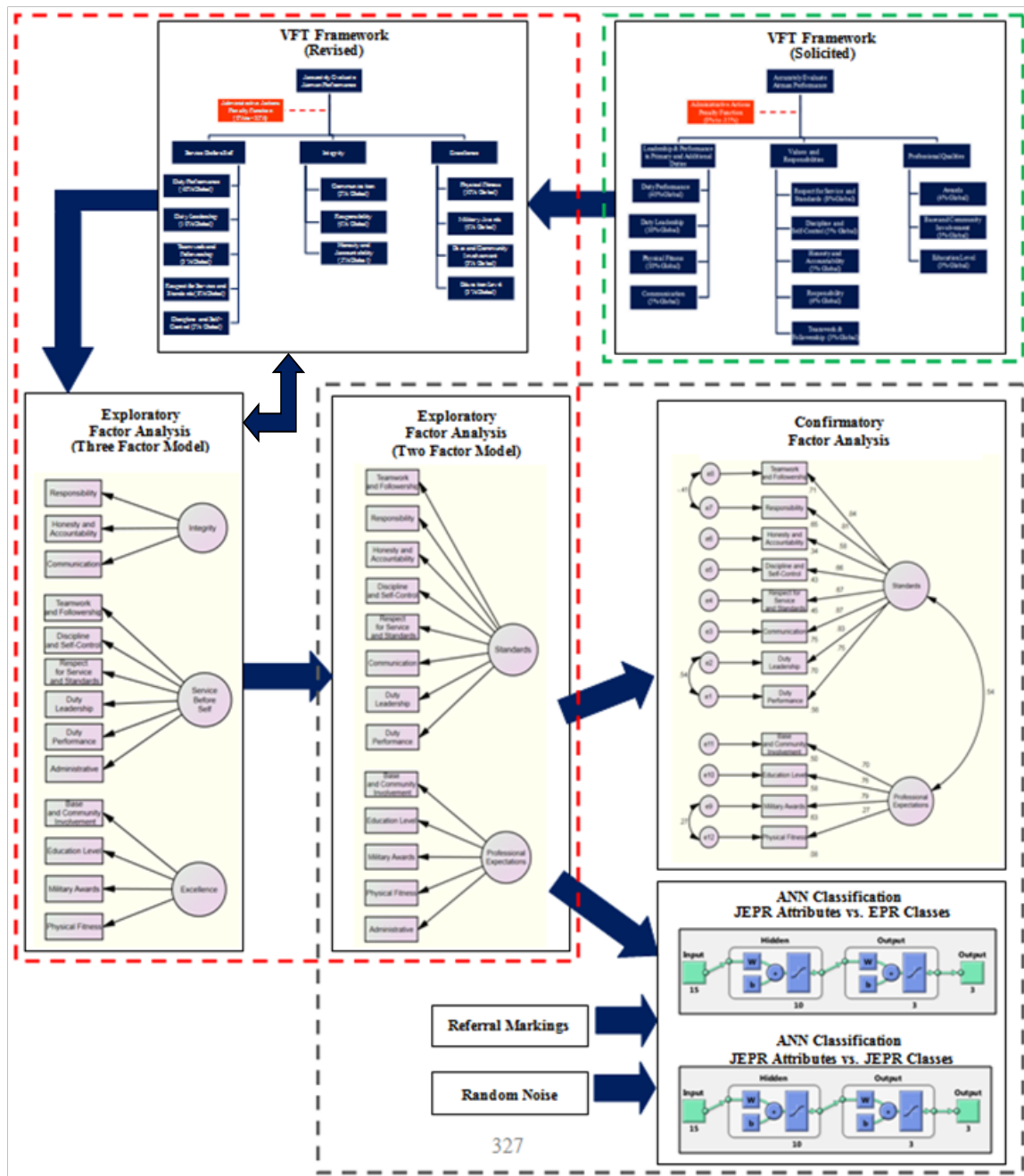


Figure 5. Value Model Construction, Validation, and Analysis Process Overview

II. Literature Review

Chapter Overview

This chapter focuses on the need for effective performance appraisal systems that effectively capture the value of the organization. In the area of Performance Management, organizations where employer, supervisor, and employee relationships exist use appraisal systems as measurement tools by leadership to assess the amount of contribution provided by the specific behaviors and results from employees in achieving the overall objectives of the organization (Bae, 2006). Appraisal systems not only provide the results of worker performance, they also provide performance feedback to workers, which in turn, significantly influence the productivity of an organization (Bae, 2006; Lee, 1989, p. 91). Kernan, as cited by (D. J. Jackson & Ward, 1992, p. 6), believed that reliable and timely feedback is essential to preserving elevated levels of achievement. Despite the importance of this topic, measurement and management systems and techniques seldom receive the attention they deserve, given the potential risks involved in doing them poorly (Noe et al., 1997, p. 233). This research in this chapter will first focus on the desired traits of a personnel appraisal system, and then will detail several mitigation techniques used to address appraisal system concerns. Finally, the chapter will explain why a Value-Focused Thinking approach that is validated using established management Science multivariate statistical techniques is the best suited approach for performing personnel performance appraisals.

Desired Traits of an Appraisal System

According to Yee and Chen (2009), maintaining a talented and knowledgeable workforce is vital in the workplace of today. In an effort to better manage the vital resource of human capital, organizations have increasingly relied on performance appraisal processes to support managerial decisions. For the organization, performance appraisals are crucial in identifying and promoting the most qualified candidates and are essential to maintaining a competitive advantage (Yee & Chen, 2009). This process is also important to the ratee. Subordinates have become increasingly aware that performance appraisal data is used to determine organizational rewards such as bonuses and promotions, and that appraisal data is also used to determine current and future career opportunities within the organization (Yee & Chen, 2009). However, as Higgins and Bargh noted, as cited in (Bol, 2011), supervisors are more often concerned with completion of the subordinates performance evaluation versus ensuring that the ratees' performance is in-line with the organizations goals. Managers often view performance appraisals as burden (Bol, 2011).

Evaluating the performance of an employee is difficult for many reasons (Moers, 2005). First, the simple classification of an employee as “poor”, “average”, or “outstanding” is not an easy decision (Yee & Chen, 2009). Poor appraisal design may unintentionally bias the employees' appraisal rating without accurately representing true job performance (Bae, 2006). Poor design of rating categories and definitions may cause inter-category correlations, which in effect, leads to “Halo Error” (Murphy, Jako, & Anhalt, 1993). This particular type of “Halo Error” may result in the rater inflating ratings due to mental correlations between less well defined categories, and better defined

or observed categories (Anusic, Schimmack, Pinkus, & Lockwood, 2009, Murphy et al., 1993). Many managers prefer to be non-confrontational, and find it easier to provide sterile nondescript feedback and evaluations rather than record the true observations (Gizaw, 2010). Next, many managers feel that applying excessively accurate ratings to an employee will cause problems in the organization after the fact, thus effecting their own standing in the organization (Gizaw, 2010; Longenecker, Sims, & Gioia, 1987). Finally, managers dread administering performance appraisals due to the long-term ramifications that a poor appraisal can have on the employees' career (Gizaw, 2010; Longenecker et al., 1987). Therefore, they choose to inflate ratings rather than accurately capture performance.

Consistency of an appraisal system is paramount for an organization striving for efficiency, effectiveness, and fairness. To create a consistent appraisal system, the system must be tied to clearly defined organizational goals or values, which are deemed key for successful operation (Gagné, 2009). Both Aguinis & Joo (2012) and Noe et al. (1997, p.234) define performance management as the means through which managers ensure their employees' activities and outputs are congruent with the organizations goals. Employees must be aware of these key organization goals or values, and be aware of how their performance contributes to the overall success of the organization (Gagné, 2009). Military organizations are no different. For military organizations, these values are rooted in doctrine. For the Air Force, Air Force Instruction 1-1, The Air Force Culture, captures these values, and details that airmen, "...whether at home station or forward deployed, encompasses the actions, values and standards we live by each and every day, whether on or off-duty. From defined missions to force structure, each of us must understand not

only where we fit, but why” (Air Force Instruction 1-1, 2012, pg. 4). For the enlisted corps of the United States Air Force, much of the value structure is outlined in Air Force Instruction (AFI) 36-2618, The Enlisted Force Structure, AFI 1-1 and in The Air Force Core Values Manual Air Force Directive (AFD) 070906-003. The core values manual and AFI 1-1 provides a basic value structure for all personnel to adhere to, while the Enlisted Force Structure outlines expectations for enlisted personnel at each level of rank. Specific career-field expectations are also outlined in doctrine in the Career-Field Education and Training Plan (CFETP).

Whenever human beings are involved in a decision making process, bias will always be present, and will affect the consistency of the decision (P.M. Podsakoff, MacKenzie, Lee, & N.P. Podsakoff, 2003). Appraisal systems are no different. An effective appraisal system must strive to control inconsistency and bias through the use of sound structural design (Aguinis et al., 2012; Bae, 2006). Bias, both method bias and rater bias, must be minimized in the design of a performance appraisal system to ensure consistency and fairness (Aguinis et al., 2012; Bae, 2006). When perceived or actual bias is encountered, the employee may respond in a fashion that results in inefficiencies on many levels for the organization (Moers, 2005; Prendergast & Topel, 1993). An employee who feels discriminated against may quit (Prendergast & Topel, 1993), or at the very least the employee may withdraw from productive activities (Moers, 2005), or may even begin to engage in counterproductive behavior. There is also a reciprocal effect to bias from employees who were favored due to a manager’s centrality bias and leniency bias (Bol, 2011). Workers who were favored during evaluations often may expend less effort during subsequent evaluation periods (Bol, 2011), as the individual may perceive a

sense of entitlement, with no fear of consequences for underperformance. If bias is present, it becomes extremely difficult to differentiate good performance from favoritism (Moers, 2005; Prendergast & Topel, 1993). In organizations where employee incentives are used and subjective performance measurements exist, the development of interpersonal relationships between managers and employees can form biases where workers attempt to influence the performance appraisals for personal gain (Bol, 2011; Prendergast & Topel, 1993). In summary, appraisal systems should be researched thoroughly to eliminate biases, as biases result in higher compensation costs, generate complexity in making personnel decisions, inject difficulties in determining incentives, and can create losses in motivation from employees (Moers, 2005).

Annison & Wilford, Fukuyama, Mishra, Shaw, and Mayer & Davis, as cited in (Mayer & Gavin, 2005), all noted that organizations have begun to realize the importance of trust in the organization by their employees. Robinson, as cited in (Hopkins & Weathington, 2006), noted that there is a reciprocal trust between organizations and employees, where Argyris, as cited in (Mayer & Davis, 1999), theorized that trust creates an environment where common goals are envisioned and strived for. For organizations, trust in management is directly tied to productivity output of employees (Mayer & Gavin, 2005). Organizations must trust that employees will act in a manner that is most beneficial to the organization (Hopkins & Weathington, 2006). Employees on the other hand must trust that the organization will act in good faith and reward their activities with additional opportunities or promotions (Hopkins & Weathington, 2006). In an effort to accomplish these goals, organizations utilize performance appraisals to delineate employee performance (Yee & Chen, 2009). Mayer, Davis, and Schoorman, as cited in

(Mayer & Gavin, 2005), conveyed that, in performing their jobs, employees make themselves vulnerable to the organization when they expend effort.

If extra effort is expended to reduce errors or defects, or the employee suggests methods to improve quality, the employee is then dependent on the appraisal system to capture this increased effort and contribution (Hubbell & Chory-Assad, 2005; Mayer & Davis, 1999). If an appraisal system fails to reward employees who have contributed to the organization with “above and beyond” effort, the employees’ level of trust in the appraisal system and in the organization will erode (Hubbell & Chory-Assad, 2005; Mayer & Davis, 1999). However, if the appraisal system does delineate between employees based on the level of performance, the level of trust and confidence employees place in the appraisal system and in the organization will increase (Yang 2005, p. 16; Mayer & Davis, 1999).

Quality driven organizations must also break away from constructs where managers exclusively control appraisal systems (Bae, 2006; Ghorpade, Chen, & Caggiano, 1995). In large Multi-National Corporations (MNCs) and matrixed organizations, where appraisers are physically separated from the employees, the appraisers struggle to make an objective assessment of an employee’s daily task performance and grapple to delineate performance between “near peer” employees (Appelbaum, Roy, & Gilliland, 2011). Ideally, eliminating or reducing the proximity between the employee and the appraiser can improve the accuracy of the appraisal due better communication familiarity, and trust in the relationship (Appelbaum et al., 2011). However, when the reduction in the physical gap between employees and appraisers are not possible in MNCs or matrixed organizations, then communication becomes paramount between the manager that the

employee actually works for, and the manager who is the appraiser (Appelbaum et al., 2011).

Good communication aids the appraiser in accurately evaluating the employee, and reduces “Halo Effect”, where the appraiser view of the employee does not cloud his/her appraisal of the employees’ true performance (Appelbaum, Nadeau, & Cyr, 2008). In organizations where appraisers are physically separated from the employees, managers must strive to build relationships through regular contact, and if possible regular face-to-face contact, to mitigate the absence of the formal and informal communications that occur with the daily interactions of other organizational designs (Appelbaum et al., 2011). Finally, to prevent loss of information on the employees’ accolades and difficulties, managers and direct supervisors should engage in systematically gathering information concerning the employees’ performance, and communicate the observations to the appraiser while the information is recent to improve accuracy of the appraisal (Bol, 2011; Ghorpade et al., 1995).

Organizational and industrial psychologists have long felt that job performance is central to the work psychology construct (Viswesvaran & Ones, 2000). As psychologists evolved the field of performance appraisal, many multi-attribute models have been applied in an effort to capture better measure job performance (Yee & Chen, 2009). To properly delineate between employees, methods and criteria must be used to measure and quantify observations. Barrick and Mount discovered that the "Big Five" personality dimensions (Extraversion, Emotional Stability, Agreeableness, Conscientiousness, and Openness to Experience) were statistically related to three job performance criteria (Barrick & Mount, 1991; Mount, Ilies, & Johnson, 2006). These job criteria (job

proficiency, training proficiency, and personnel data) were specific for five unique occupational groups (professionals, police, managers, sales, and skilled/semi-skilled) (Barrick & Mount, 1991). Using the correlations of data compiled through observation, their results illustrated the benefits of using personality models to accumulate, communicate, and quantify empirical findings especially for use in performance appraisal. Therefore, any appraisal system must be able to translate observed personality dimension data to statistically sound information that can be used to quantify inter-relationships (Mount et al., 2006).

Building on the five factor research by Barrick & Mount, Bae, and Guion detailed that assessments must be utilized to help managers identify the strengths and weaknesses of employees to improve training shortcomings and for optimal placement decisions for the organization (Bae 2006; Guion, 1998). To do this in appraising job performance, the appraisal must be able to scale actions, behaviors, and outcomes that an employee engages in which support and contribute to the overall organizational goals (Viswesvaran & Ones, 2000). In designing an appraisal, the actions, behaviors, and outcomes should measure the task performance of the employee, the citizenship behavior of the employee, and the counterproductive behaviors of the employee (Viswesvaran & Ones, 2000). Viswesvaran & Ones noted that specialized jobs such as the military fall under this design construct (Viswesvaran & Ones, 2000).

Military Research in Appraisal System Design

The US Army has studied job performance in great detail and has developed several models for determining work effectiveness (Campbell, 1990). One such model

researched by the Army is known as Project A (Campbell, 1990). In an effort to better delineate soldiers, Project A sought to generate criterion variables, predictor measures, analytical methods, and validation data for selecting and classifying entry-level positions in the US Army (Campbell, 1990). Although not specifically used for performance reporting, many of the techniques and measures could be applied to measuring job performance. In researching the Project A study that was conducted for the US Army, Campbell found that there were five job performance criteria for entry-level jobs (Campbell, 1990). The five criteria identified by Campbell were core technical proficiency, general soldiering proficiency, effort and leadership, personal discipline, and physical fitness and military bearing (Viswesvaran & Ones, 2000). In addition to job performance, Borman, Motowidlo, Rose, & Hanser, as cited in (Viswesvaran & Ones, 2000), furthered Campbell's research and discovered that allegiance, teamwork, and determination were also vital performance dimensions for unit effectiveness.

A second US Army related study concerning delineation of soldiers through job performance was created to measure and appraise the "WholeSoldier Performance" of a soldier, quantifying moral, cognitive, and physical domains during the evaluation (Dees, Nestler, & Kewley, 2013). This study relied on Value-Focused Thinking (VFT) techniques from the Operations Research (OR) field, and was reinforced by the management science technique of factor analysis. According to Keeney, "Valued-Focused Thinking is a way to channel a critical resource-hard thinking-in order to make better decisions" (Keeney, 1994). In applying VFT, inputs from Subject Matter Experts (SMEs) were solicited in constructing a value hierarchy. These inputs better known as attributes or objectives were then quantified using several single-attribute value functions

(Keeney, 1992, p. 141-144). These functions were then weighted based on stakeholder inputs, and then combined into a multiattribute value function (Keeney, 1992, p. 327-331). This multiattribute function captures the contribution of an attribute in the entire decision space (Kirkwood, 1996, p. 61). Further, Dees et al. validated the construct of the “WholeSoldier Performance” model by applying standards and measurements of the management science community to the model. Dees et al. utilized Cronbach’s alpha to verify the models measurement scales construction, and then utilized the Principal Axis Factoring method of factor analysis to gain insight as to the underlying construct formed by the correlations among the measured variables (Fabrigar, Wegener, MacCallum, & Strahan, 1999). The “WholeSoldier Performance Appraisal” was not the first proposed usage of a weighted multi-criteria method. In 2009, Yee and Chen proposed a weighted multi-criteria model using Fuzzy Set Theory would be a transparent and fair method to conduct military performance evaluations (Yee & Chen, 2009).

Mitigation Techniques to Address Appraisal System Concerns

Preventing or reducing inflation of any performance appraisal system is a difficult challenge, as rating leniency and inflation are consequences of workplace politics, image management, organizational norms, discomfort with performance appraisals, and or aversion to interpersonal conflicts (Spence & Keeping, 2011). Designing a rating instrument design with descriptive anchored ratings scales is one way that appraisal accuracy can be improved (Lilley & Hinduja, 2007). Raters are more apt to correctly categorize observed behaviors when the appraisal design categorizes behaviors and ties ratings directly to standards, values, and doctrine (Bae, 2006).

A second method to control performance appraisal rating inflation is to utilize a forced distribution in the rating process (Murphy, 2008). Some organizations have adopted forced distribution models for conducting performance appraisals in an effort to mitigate rater biases (Berger, Harbring, & Sliwka, 2013). Appraisal systems that utilize forced distribution models allocate a predetermined distribution of ratings to supervisors, then mandate that the supervisors adhere to the predetermined allocations when assigning employee appraisal ratings (Berger et al., 2013). Forced distribution appraisal models have been shown to be successful in some corporate environments in controlling inflation of appraisal ratings (Murphy, 2008). General Electric is one company that has been successful in the implementation of forced distribution rating methods (Blume, Baldwin, & Rubin, 2009). General Electric leadership touted forced distribution appraisal methods as an efficient method of rewarding performance output by employees, and as a key factor in strengthening the organization (Blume et al., 2009).

However, Roch, Sturnburgh, and Caputo, as cited in (Murphy, 2008), concluded that organizational psychologists generally view forced distribution techniques as a less fair appraisal technique than methods used by other rating systems. Blume also noted that the adoption of forced distribution appraisal systems by several U.S. companies resulted in both an internal and external backlash from employees and the media (Blume et al., 2009). Employees became infuriated, claiming the system was unfair and inequitable, when previously high performing employees were appraised as subpar and dismissed from the organization (Blume et al., 2009). Both Ford and Goodyear backtracked from the forced distribution appraisal systems, but not before sustaining substantial damage to

both public images and the morale of the workforce of both companies (Blume et al., 2009).

Forced distribution rating systems have also been criticized for masking performance differences across organizational divisions and workgroups (Murphy, 2008). The main limiting factor of forced distributions occurs when the percentage of employees forced into the distribution that actually meet the cutoff criteria is greater or less than the cutoff percentage (Almond et al., 2005). High performing employees may be under appraised, while sub-par employees may be inflated to meet rating cutoff criteria requirements (Giangreco, Carugati, Sebastiano, & Tamimi, 2012). Scullen et al., as cited in (Berger et al., 2013) performed a simulation study that illustrated that using a forced distribution for appraisals and personnel management and discovered that although forced distributions can increase organizational performance in the short-run, the effects decay over time as the pool of under performers is exhausted and are forced out of the organization. Additionally, another reason the effects of a forced distribution appraisal system also wane is that employees initially understand that they need to work harder to achieve good evaluations, which are tied to bonuses and promotions, but soon become demotivated, when they realize that they no longer can achieve the appraisal ratings they were accustomed to under the previous appraisal system to earn bonuses and promotions (Berger et al., 2013).

A final method to reduce rating inflation is to communicate the raters rating history to the ratee and to the raters' rater as a part of the appraisal (Dees et al., 2013). This technique was suggested as a method to reduce inflation of appraisals in a newly proposed article detailing a proposed revamp of the enlisted appraisal system for the U.S.

Army. This method not only promotes accuracy, but also conveys a climate of transparency to ratees' in the appraisal process (Dees et al., 2013). This technique also allows senior leaders the ability to observe the rating histories of raters in the organization, and to better manage personnel under their control. If a senior leader observes a widely spread distribution for a raters rating history, the senior leader can feel confident that the rater is differentiating the levels of performance between the employees under his management. However, if the rater's history is skewed left or right, the senior leadership has the ability to engage with the rater to find out why.

For example, if the chain of command identifies that a specific supervisors rating history is skewed to the left, then the unit leaders can investigate whether the supervisor has historically been assigned a large number of underperforming employees, or if the supervisor has been possibly under valuing or improperly accomplishing the performance appraisal ratings (Dees et al., 2013). Conversely, if a second supervisors' historic rating distribution is narrow and excessively high, the unit senior leaders have the valuable historic information to ascertain whether the supervisor has been supervising a large number of high performing employees, to investigate further to determine if the supervisor is over valuing performance, and to further research whether or not the supervisor is properly accomplishing performance appraisal ratings (Dees et al., 2013). In either case, supervision now has additional information and insight to quickly identify trends, and either redistribute the work center personnel to balance skills and performance of employees, improve task training for employees where shortfalls are noted, expand training of employees where positive trends are discovered, or improve or expand supervisory guidance for performing evaluations. Allowing the ratee to view the

appraisers rating history provides transparency, while allowing supervision at the organizational level to view the raters' history allows for better skill set distribution of personnel and supervisors, and also facilitates better mentorship of raters by leadership (Dees et al., 2013).

Bias in employee appraisals is problematic as it increases the difficulty in making the right personnel decisions (Moers, 2005). However, steps can also be taken to reduce bias and improve consistency through the use of sound appraisal systems and organizational designs (Aguinis et al., 2012; Prendergast & Topel, 1993). Boice and Kleiner provided a good example of bias control. Boice and Kleiner remarked that bias could be reduced through the use of multiple rater systems which are computerized allowing for statistical analysis to identify bias both during design and execution (Boice & Kleiner, 1997). Using this technique, designers can mitigate the construct based on the bias discoveries during testing, then after implementation, organizations can address any biases that surface through training, education, or policy (Aguinis et al., 2012). This effectively controls design and implementation bias, thus improving the overall consistency of the appraisal system (Aguinis et al., 2012).

The consistency of an appraisal system is also affected by who was involved in the design of the system. From an organizational standpoint, an appraisal system is more likely to gain acceptance if all levels of the stakeholders were involved in the design or redesign process (Ghorpade et al., 1995; Nankervis & Compton, 2006). From a ratee perspective, employees are more apt to accept a system as fair, and more willingly accept the results generated by an appraisal system as accurate when they have had a voice in the design construct (Bae, 2006; Ghorpade et al., 1995). The inclusion of multiple levels

of stakeholders in the design process provides valuable insight as to the requirements or how a system would react from the perspective of a user, ratee or mid level manager (Bae, 2006). Bobko & Colella, Mohrman, Resnick, & Lawler, and Waldman, as cited in (Bae, 2006), all noted that the failure to include stakeholders in the design of an appraisal system may result in negative reactions which may damage the company, employees careers, or both.

Frequent recording and documentation is crucial to improving the factual content that is often included in performance appraisals (Stone, 1999). Too often, large portions of information are either lost or become muddled when a supervisor waits until the end of an evaluation period to record performance observations, creating an environment of subjectivity, which in hand creates an inability to delineate performance, ultimately resulting in ratings inflation or marginalization (Balzer, 1986; Murphy 2008). Bernardin & Walter, Guion, and Hakel, Appelbaum, Lyness, & Moses; as cited in (Balzer, 1986), all noted that a performance appraisal system that relies on immediate supervisors to collect performance observations of their employees in a timely manner, such as using a “Behavioral Diary”, can reduce subjectivity, improve delineation of employees, and reduce ratings inflation.

Why use a Value-Focused Thinking Approach?

A consistent employee appraisal system measures the contributions of the employee toward attributes that are valued by the organization (Bae, 2006). These attributes or “values” define all that is fundamentally important to the organization (Keeney, 1994). Kirkwood, as cited in (Orwat, 2008, p. 51), noted that a Value Focused

Thinking (VFT) approach enables the model designer the ability to capture the desires of the relevant decision makers and stakeholders in defining what is valued through a formal, repeatable, and defensible process. This approach also incorporates the four axioms of decision analysis, which provide the rationale and theoretical feasibility for the decision maker to “divide and conquer” the problem (Keeney, 1982). The four axioms are as follows. First, a VFT approach allows the decision maker to structure the problem. Second, it allows the decision maker to assess the impact of alternatives. Third, a VFT approach allows the designers to capture the decision makers’ preferences. Finally, the fourth axiom allows the decision maker to evaluate and compare alternatives.

For military decisions, the values of a VFT Framework should be the future values of national-security decisionmakers desire along with the values that the users, and customers of a service, regard as important (Parnell, 2007). The measurement approach for a VFT Framework should also be quantitative, in that, the use of numbers clarifies the elements of the process, and forces explicit reasoning in designing the system (Kirkwood, 1996, pg. 3). Looking at the Air Force enlisted appraisal system as an example, Air Force Instruction (AFI) 1-1, The Air Force Culture, details the values and standards expected of Air Force members (Air Force Instruction 1-1, 2012, pg. 4). Additionally, AFI 36-2406 describes the Enlisted Performance Report (EPR) as the measurement tool for appraising the ability of enlisted airmen to meet the aforementioned standards. From the above descriptions, it appears that the Air Force appraisal system is based on Value Focused Thinking methodologies. However, what are the benefits to using a VFT Framework and what would be the benefits of applying this method to the Air Force appraisal system? Keeney, as cited

in (Parnell, 2007), has identified nine benefits to using a VFT approach for decision opportunity situations.

The first benefit of using a VFT process is that a VFT Framework helps the decision maker and stakeholders apply and translate Strategic Thinking to a specific problem (Keeney, 1992, p. 27-28). Strategic thinking in the VFT process helps the decision maker identify objectives that are the foundation of the organization, and unchanging (Keeney, 1992, p. 27-28). For military organization such as the United States Air Force, doctrine is the fundamental principles by which the military forces guide their actions in support of strategic national objectives. It is authoritative but requires judgment in application (Air Force Pamphlet 36-2241, 2013, p. 500). Air Force doctrine clearly defines three strategic objectives which explain the need for a personnel evaluation system (AFI 36-2406, 2013, p. 8). The first reason that an evaluation system is needed is to provide meaningful organizational and supervisor feedback to airmen (AFI 36-2406, 2013, p. 8). This feedback details to the airmen on how well they are meeting expectations, what is expected of the airman by the supervisor and organization, and provides mentorship and planning for the airmen how to better meet expectations (AFI 36-2406, 2013, p. 8). The doctrine also describes that the second reason to have an evaluation system for airmen is to provide a reliable, long-term, cumulative record of performance and potential (AFI 36-2406, 2013, p. 8). Finally, Air Force doctrine states that the third strategic objective of the evaluation system is to provide sound data for promotion and for other force management decisions to Air Force systems and leaders (AFI 36-2406, 2013, p. 8).

The second benefit of using a VFT process is that it helps the decision maker with consistent decision making by applying the same set of ultimate objectives (Keeney, 1992, p. 26). These consistent objectives must be in-step with the decisionmaker's strategic objectives, and be the driving reason for undertaking the project (Keeney, 1992, p. 26). This interconnection of ideas through the use of a VFT design, allows for consistency and repeatability under the same set of weights and objectives.

The third benefit of using a VFT design is that a VFT facilitates the collection of only information which is important to achieving the values of the organization (Keeney, 1992, p. 24-25). Extraneous information, not explicitly identified as an objective, should not be considered (Keeney, 1992, p. 24-25). Additionally, only data that will contribute to creating a better alternative or wiser choice should be collected.

The fourth benefit of using a VFT process is that a VFT construct facilitates involvement (Keeney, 1992). Lack of consideration of what is valued by stakeholders will erode support from those who have a vested interest in the decision outcome (Keeney, 1992, p. 25-26). By involving those with a vested interest about what is valued in the decision, further discussions can be initiated concerning consequences of a decision, leading to "buy-in", compromise, and conflict resolution (Keeney, 1992, p. 26). Therefore, a VFT framework allows leadership to consider all stakeholders inputs during design, increasing familiarity and acceptance by users (Keeney, 1992, p. 25-26).

The fifth reason that a VFT design is beneficial is that it improves communication (Keeney, 1992, p. 25). Decisions often revolve around complex problems where technical experts have knowledgeable insight that is beneficial in arriving at a solution (Keeney, 1992, p. 25). Use of a VFT design can translate the complex technical concepts of

technical experts into common language that can be easily understood by stakeholders (Keeney, 1992, p.25).

The sixth benefit to using a VFT based system is the ability to evaluate alternatives (Keeney, 1992). Parnell identified that evaluation of alternatives was especially relevant to operational military analysis (Parnell, 2007). Through the use of sensitivity analysis, decisionmaker's can test the "what if" factor of the model, by seeing the ramifications of what weight changes or ratios would have on the program (Keeney, 1992, p. 26). This allows for identification or study of possible unknown or unimagined scenarios (Keeney, 1992, p. 26), prior to incorporation. If logic problems are identified in the solution, the model can be adjusted to create a more robust design with a more accurate output. Since this testing occurs before implementation, the number of changes and the severity of the changes after fielding are greatly reduced, building confidence in the design and reducing retrofit costs.

The seventh benefit of a VFT Framework is that hidden objectives can be uncovered (Keeney, 1992). Often it is difficult to ascertain what values are important, or how to articulate why they are important (Keeney, 1992). Other times you may not be aware of a value, or a set of values, that are relevant to the decision (Keeney, 1992), until analysis or preliminary discovers the objectives. By using a VFT construct, preliminary analysis and testing can be accomplished that may help capture hidden objectives (Keeney, 1992, p. 24), before the system is fielded.

The eighth advantage to utilizing a VFT Framework is the creation of alternatives (Keeney, 1992). VFT, unlike many decision other decision methods which restrict alternative creation, promotes the creation of alternatives (Keeney, 1992, p. 27). The

choices or alternatives are often compared and contrasted in the form of value gaps. Value gaps illustrate the ideal “best possible” score and individual value hierarchy attribute scores of the ideal “best possible” solution versus the overall value score and attribute scores for each alternative being considered (Parnell, 2007). This allows of comparing and contrasting of alternatives by both overall scores and by attribute scores, to assist in determining the overall best solution (Parnell, 2007).

The final advantage to a adopting a VFT design is the ability to identify decision opportunities (Keeney, 1992). A VFT design is not solely constrained to the final evaluation (Keeney, 1992, p. 27). VFT provides the ability to systematically revisit a previous decision and study how well the decision is addressing or has addressed the problem (Keeney, 1992, p. 27). Leveraging this VFT advantage may yield opportunities to improve on current decisions due to increased knowledge and understanding, or may provide additional decision opportunities to pursue (Keeney, 1992, p. 27).

Validating a VFT Framework Using Multivariate Management Science Methods

A VFT Framework is a useful tool that aids decisionmakers in making difficult decisions by translating value structures to mathematical models (Pruitt, 2011, p. iv). VFT models provide a methodology that allows decisionmakers to make tradeoffs between multiple, sometimes conflicting objectives (Keeney, 1992, p. 130). VFT models also provide additional insight that can better prepare the decisionmaker for the next time a decision opportunity arises (Keeney, 1992, p. 27). However, VFT models are rarely statistically validated for accuracy (Pruitt, 2011, p. iv). Pruitt suggested that multivariate

techniques should be used as a method for validating VFT Framework's for statistical relevance and classification consistency (Pruitt, 2011, p. 38).

In applying a VFT Framework for redesigning the Air Force appraisal system, as was the case with the "WholeSoldier" article, Cronbach's alpha could be used to validate the measurement scales of the attributes used in a VFT Framework (Dees et al., 2013). Exploratory Factor Analysis (EFA) techniques could also be used to validate the VFT Framework (Dees et al., 2013); by ensuring the framework is in-line with fundamental objectives such as Air Force and military doctrine. Additionally, Confirmatory Factor Analysis could be used to verify that the factor solutions generated from the EFA construct is statistically correct (Helfrich, Li, Mohr, Meterko, & Sales, 2007). Confirmatory Factor Analysis is a powerful hypothesis test based statistical tool, which has long been used by psychologists and researchers to develop, refine, and assess the validity of behavioral measurement constructs (D.L. Jackson, Gillaspay Jr., & Purc-Stephenson, 2009). Finally, as suggested by Pruitt, Artificial Neural Networks could be utilized to validate the effectiveness of the appraisal system to correctly classify personnel based on the values provided by the VFT Framework (Pruitt, 2011, p. 38). The use of Artificial Neural Networks in Management Science has shown that ANNs perform better than traditional method of classification, without incurring distributional assumptions or linearity (Krycha & Wagner, 1999). This merging of VFT concepts from Operations Research and established Management Science multivariate statistical techniques would provide credibility to the design of a new appraisal system for the Air Force among the work force, managers, and academia, validating that the newly devised

system is a fair and statistically defensible method for accomplishing performance appraisals.

III. Value Model Construction

Chapter Overview

This chapter begins by describing the purpose behind attempting to revise the current junior level enlisted performance appraisal system. The chapter then discusses how values and objectives were solicited from Air Force doctrine, tactical level decisionmakers, and Subject Matter Experts (SMEs) to identify what traits are considered important during the appraisal of junior level enlisted members. Once identified, the chapter details how the values were grouped into a strategic hierarchal framework, then how the framework was continually refined to focus on the specific area of appraisal design modification. The chapter then discloses how the weightings of importance were solicited for each attribute or objective that had been identified by the SMEs, then how those weights were applied to the framework design. The chapter then explains how mathematical functions were derived to accurately represent how the tactical level leadership valued each attribute of the framework. Next, a data collection plan was unveiled that involved the development of a prototype Decision Support System tool to collect data samples from the field for validating the design, then testing the design after analysis. Finally, the chapter concludes with a Deterministic Analysis using computer generated data for eight notional airmen to verify that the weighted attributes of the framework function as intended. Figure 6 provides an overview of the methodology detailed in this chapter, illustrating the development of the strategic hierarchy, the identification of appraisal modification objective for better evaluations, and the development of the tactical level hierarchy to address appraisal modification.

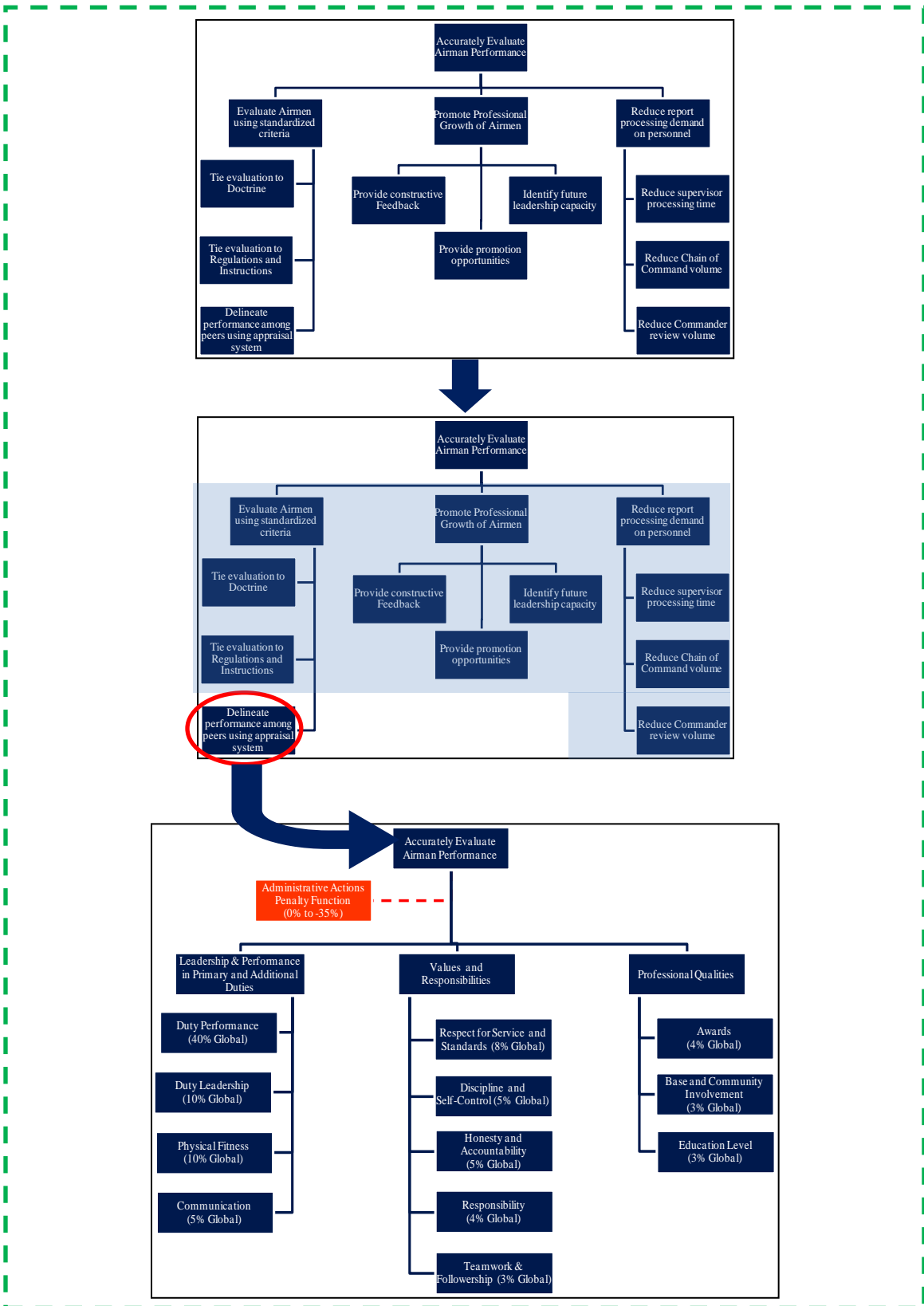


Figure 6. Overview of Value Hierarchy Refinement Methodology

Purpose

The purpose of this research is to develop a prototype model which modifies how junior enlisted EPR appraisals are accomplished and calculated. This project involves more than a simple revision of the form. The vision is to create a new way to evaluate the performance of the junior enlisted force which captures true performance of the individual over the evaluation period in meaningful areas. This new method will also provide constructive feedback to the individual concerning both areas of strength and weakness, and will reduce the administrative footprint of report generation for supervisors. It is believed that use of Value Focused Thinking techniques enhance appraisals by providing a consistent framework to translate qualitative inputs into quantitative output. Success can be determined through this process if non-equal performers which would have received the same overall ratings under the current system, can be delineated from each other under criterion that is generated and adopted by the United States Air Force Senior Non-Commissioned Officer (SNCO) Corps.

Airman performance, appraisals, and promotions are an Air Force wide issue, and not only affect the mission, but the direction of the force, and have far reaching effects on all ranks. As enlisted performance reports directly factor into promotions, we want to ensure the right Airmen are selected for leadership positions. The aim of any revised system should seek to use the correct criterion when evaluating today's junior enlisted airmen, as they will serve as the leaders of tomorrow and support commanders, allies, and citizens. These stakeholders require nothing less than the most highly skilled airmen, who exhibit Integrity, place Service Before Self, and demonstrate Excellence in all endeavors.

A simple revision to a form will not change the norms and cultures of the enlisted force. Air Force instruction changes, along with enforcement and management by the Senior Enlisted force, must occur simultaneously to any change in the rating computations and physical redesign of the evaluation form to address the culture of inflation, which senior leaders have acknowledged has taken root in enlisted the ranks (Losey, Sep 2013). In an effort to develop a solution from the tactical level Air Force stakeholders, a team of 25 Senior Non-Commissioned Officers (SNCOs), led by a CMSgt select maintainer from Barksdale Air Force Base, volunteered to serve as Subject Matter Experts (SMEs) for the development of a new Junior Enlisted Performance Report (JEPR) Framework. This framework would not only involve revising the computation methods and form design, but would also identify the changes needed to the associated Air Force Instructions and doctrine.

This research focused on the junior enlisted appraisal system; however, the framework could be adapted for appraisals at any level for any type of organization. The revised performance report construct will also drive changes to the Weighted Airman Promotion System (WAPS) as outlined in chapter one, and correct the scenario where PFE and SKT test scores comprise 62% of the promotion score in an inflated environment. This system uses a portion of the performance reporting ratings for computations toward promotion selection. If the proposed changes to the junior enlisted EPR system prove successful, then the SNCO evaluation system and the Officer evaluation system should also be considered for revision.

As stated earlier, this is an Air Force wide issue, and not only affects the mission, but also affects the direction of the force and has far reaching effects into all ranks. The

new system must be a sound process that invokes *a cultural change that MUST occur among leadership and commanders* to eliminate over-inflation and accurately capture true performance and provide feedback. The enlisted force is the backbone of our military. Therefore, true success cannot be determined immediately, but will be determined by the quality of the future leader identified for promotion under the revamped method. We must ensure that we develop well rounded future leaders who possess traits that are valued the most by the Air Force, and that true excellence is distinguishable from very good, and that average performance is classified as average.

VFT Values and Objectives

The use of a tactical level SNCO SME team helped bound the problem for the analysis by identifying shortcomings that exist in the current junior enlisted evaluation program. The tactical level SMEs also helped by communicating values that are important at the immediate supervisory level, along with what was valued from a future enlisted force development level. In applying this value framework, the team worked to develop the evaluation criteria and categories for a new prototype evaluation construct. The tactical level SMEs are key stakeholders in this process. They are the subject-matter experts and stakeholder representatives from their respective career fields. Parnell, as cited by Merrick, Parnell, Barnett, and Garcia, deemed the use of this level of expertise for value solicitation in a multiple-objective value model as the “Silver Standard” (Merrick, Parnell, Barnett, & Garcia, 2005).

The team sought to tie the evaluation categories and criteria directly to doctrine such as the Air Force Core Values manual and Air Force Instruction 36-2618, which

outline the responsibilities of Airmen as a whole and of the enlisted force structure. In utilizing regulations and doctrine, we hope to apply the Decision Analysis Gold standard and extract doctrinal values of the Air Force. Once the values were identified, they were used to develop and weight a value hierarchy which provides a framework for the new prototype for junior enlisted performance reporting.

In discussion with SNCO SMEs concerning the Junior Enlisted EPR project, the team determined three alternatives for addressing this decision. The first alternative was to keep the system as it is without any revisions. The second option was to modify the existing construct to include value focused thinking when performing an appraisal. The third option was to completely revamp the system, where new guidance, cultural changes, and new methods for appraisal are introduced. With known alternatives, a “Bottom Up” approach was taken for structuring objectives.

The SMEs identified that the Strategic Objective of an appraisal system is the ability to “accurately evaluate performance of junior enlisted airmen”. This objective was supported by all other underlying objectives, and thus is the overall goal of the project.

Three fundamental objectives support achievement of the strategic objective. In developing objectives, the team of Subject Matter Experts (SMEs) used their practical experience, discussed the US Army Whole Soldier Performance Appraisal Study (Dees et al., 2013), and reviewed the USAF Core Values Manual and AFI 36-2618, The Enlisted Force Structure. The first fundamental objective the team decided on was the “need to evaluate airmen using a standardized criterion.” This would, in essence, change a subjective process into a quantifiable process that is standardized across the junior enlisted tier. The second fundamental objective the team decided on was that “the system

must promote professional career growth of Airmen”. This would emphasize the development of professionalism and leadership, and provide feedback to members seeking opportunities to improve and advance. The third fundamental objective decided on was to “reduce the administrative footprint of the current process.” Currently, supervisors and SNCOs within the chain of command spend many hours accomplishing administrative tasks such as writing, rewriting and defending the EPR ratings of personnel. This is time that could be better spent mentoring, training, and sampling the work of junior enlisted members. Using the fundamental objectives, the team developed a value hierarchy as illustrated below in Figure 7.

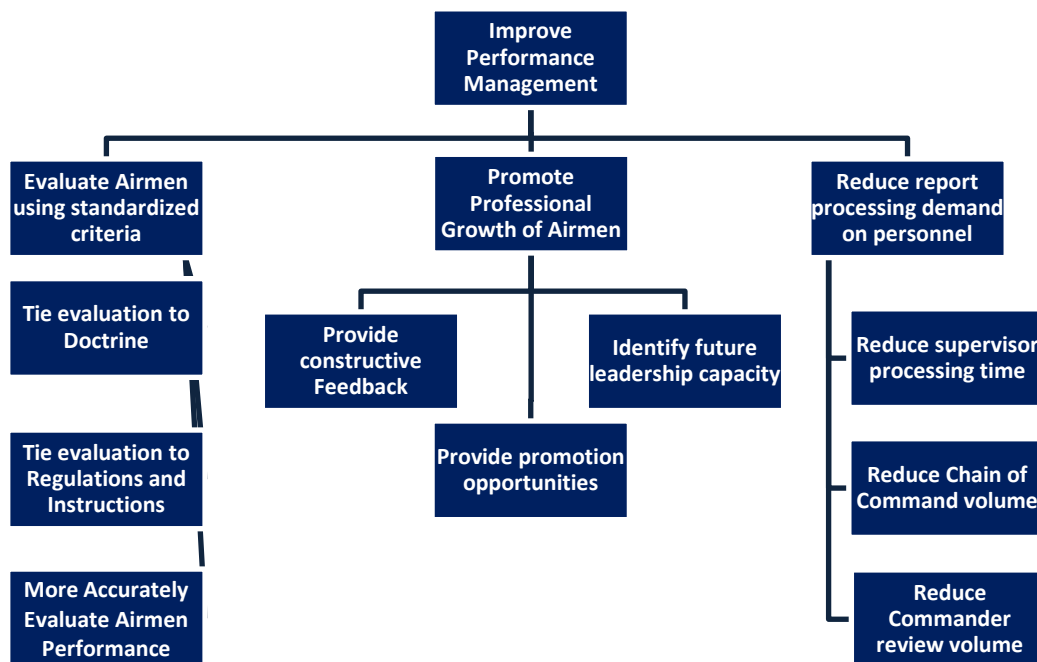


Figure 7. Strategic Value Hierarchy

The SMEs evaluated each of the third tier objectives for attributes. These attributes are in essence an individual airman’s Measures of Performance or Measures of

Effectiveness. For the fundamental objective, “to evaluate using a standardized criteria”, the attribute, “to Incorporate Core Values into Performance Report evaluation criteria”, relies on using the three Air Force Core Values as the doctrine to tie back to the reporting process. This is a natural extension as the core values manual details many of the desired traits that the SNCOs felt defined the standard of what an airman should adhere to. Since the rateses’ performance cannot be measured directly against these three main traits, Integrity, Service before Self, and Excellence, a proxy and constructed scale would be used to evaluate the Airman’s ability to meet these criteria. For the second attribute, “to tie evaluations to regulations and instructions”, AFI 36-2618, The Enlisted Force Structure Instruction, was chosen for use as it specifically details responsibilities by rank and skill-level. Again, this is measured with a proxy and constructed scale as many of the factors cannot be measured directly. Finally, for the attribute “to delineate performance among peers”, a proxy and constructed scale will be used as performance standing could be measured against peers within a section.

For the fundamental objective “to promote professional growth”, the attribute to “Identify Future Leadership Capacity” could be scored by the use of narrow sub-categories that could identify areas of strength. This would be a proxy and constructed scale, as the supervisor’s observations could be included into an overall value function. For the attribute of “Providing Constructive Feedback”, uninflated evaluations could provide the member quantifiable strengths and weaknesses in areas as a roadmap to success and growth. This would be a proxy and constructed measure, as actions evaluated by the supervisor would factor into an overall value function as a contribution. Finally,

for the attribute to “Provide Promotion Opportunities”, the WAPS test scores of the airman would provide a direct and natural scale for measuring promotion opportunities.

For the third fundamental objective, “Accurately Evaluate Airman Performance”, the attribute, “reduce supervisor processing time” can be directly measured from the number of hours that each supervisor will expend completing reports. For the attribute, “reduce chain of command volume”, again this can be directly measured by the number of EPRs that are handled by individuals in the chain. Finally, the third attribute, “reduce commander volume”, is a direct and natural measure, as the number of EPRs handled by the commander can be directly computed.

There are several value judgment implications related to the decision to revise or replace an evaluation system, including the current junior enlisted evaluation system. If alternative two (revise current system) or alternative three (develop a new system) were selected, new criteria must be developed for the supervisor to consider when evaluating the ratee. The supervisor would experience value changes corresponding to the new standard. The ratee would also experience value changes in an attempt to conform to the new standard. From a macro level, the enlisted corps as a whole would experience value changes in aligning to the new standard. Finally, Commanders would experience value changes, as they adjust to how they view the quality of their personnel based on the new standard. Therefore, we have chosen alternative three and will develop a new rating system that will utilize a Value Focused approach.

Development of the Strategic Value Hierarchy was necessary for identifying potential approaches to change the appraisal system. This change requires more than just a new computational method and a new form. The Air Force culture, doctrine, and Air

Force Instructions must change to fully implement any new evaluation process. However, for the purpose of this research, Figure 8 illustrates the intent of this project is to narrow the scope on the development of the new evaluation process. Therefore we intend to focus on the strategic attribute “More Accurately Evaluate Airmen Performance”.

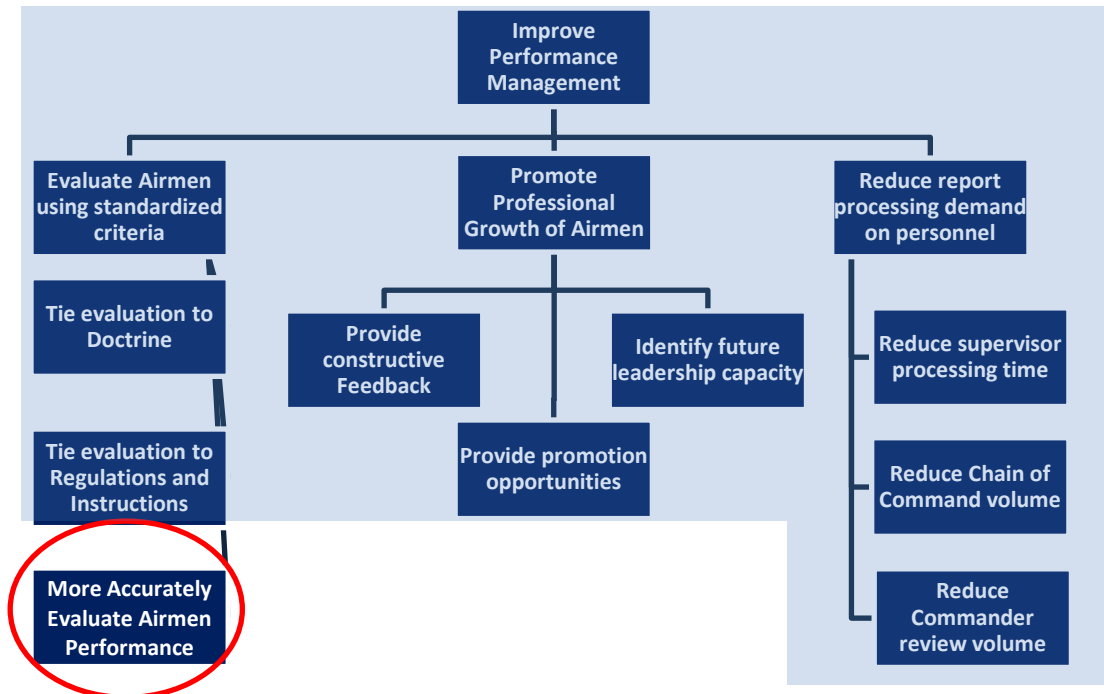


Figure 8. Strategic Value Hierarchy Focusing on Evaluations

VFT Evaluation Hierarchy

Focusing on the strategic objective to “More Accurately Evaluate Airman performance” in the strategic value hierarchy, the SNCO SMEs developed a more specific value hierarchy that provided clear and concise objectives which would allow supervisors to be able to more accurately evaluate Airmen performance. The pool of

objectives yielded three key fundamental objectives, which SMEs felt more accurately captured the desired performance traits of Airmen. The first Fundamental Objective identified was Leadership and Performance in Primary and Additional Duties. This was the most important objective to the SMEs, as they felt the intent of the EPR is to not only capture the performance of an Airman, but to also quantify leadership. This objective is a key principle outlined doctrinally by rank and position in Air Force Instruction 36-2618, The Enlisted Force Structure.

The next fundamental objective identified was Values and Responsibilities. This objective captures a myriad of traits which are detailed in the Air Force Core Values. Both on and off-duty actions are captured here.

The third category decided upon was the Professional Qualities objective. Currently, it is very difficult to accurately delineate factors among airmen that are simply doing their job. This category would capture the efforts of airmen who attempt to better themselves in the profession of arms, support unit activities, and who also support the local community. The SMEs felt that inclusion of this objective would create a more competitive environment among airmen trying to separate themselves from their peers for promotion and open doors to eventual leadership opportunities.

Underneath these three fundamental objectives, 12 attributes were identified. These 12 objectives all were able to be tied back to the fundamental objectives, with each attribute describing a portion of a specific fundamental objective. Reviewing the fundamental objectives and attributes, it became apparent that the current junior enlisted EPR system could not meet the objectives that the team had established. This was primarily due to form design and lack of connectivity of the categories to doctrine.

Therefore, it was decided that the junior enlisted EPR form should be redesigned using a Value Focused Thinking approach as described in chapter 2, where an additive multi-attribute value function would be used to quantitatively score the performance of an Airmen. Doctrine and SME inputs were essential to developing constructed proxy measures for each of these 12 attributes. The value Hierarchy can be seen in Figure 9.

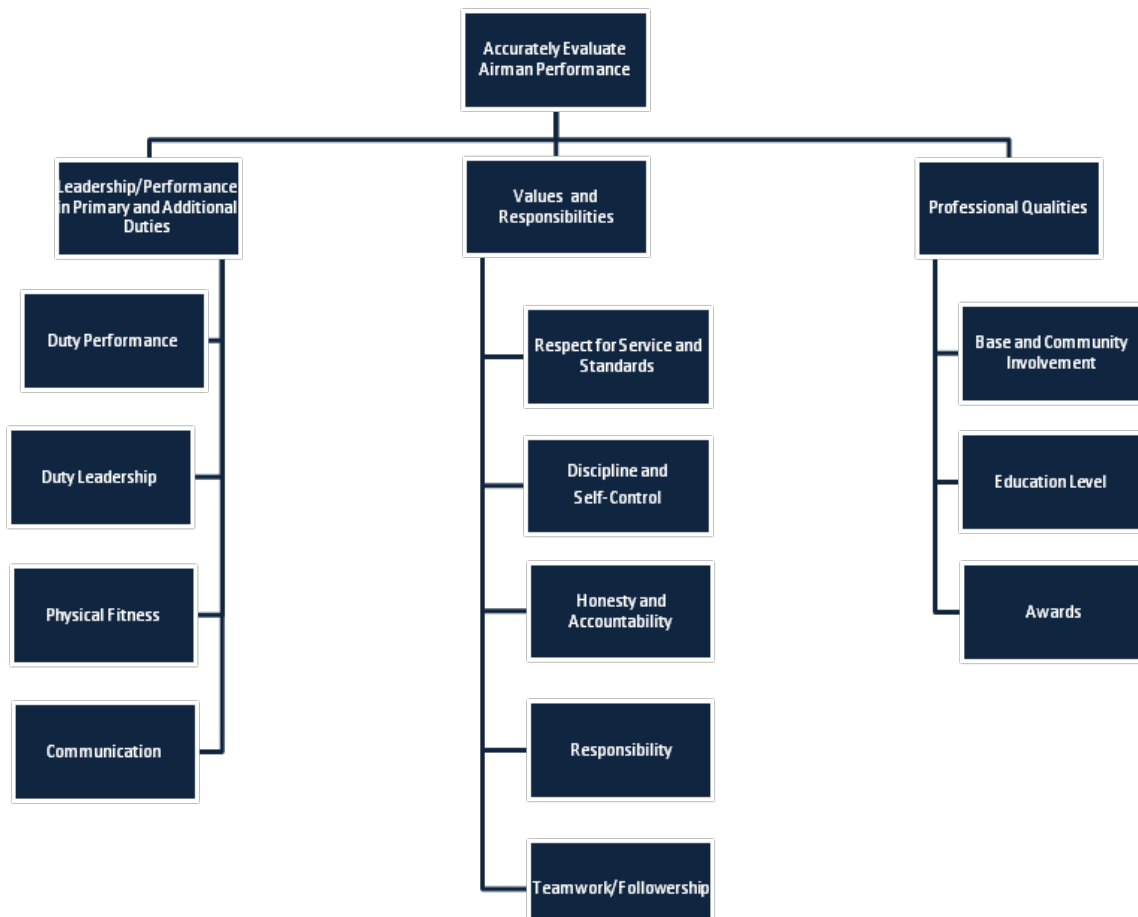


Figure 9. Refined Value Hierarchy Framework

The SMEs developed four rating categories or blocks for each of the attributes. Each of these categories was assigned a definition in an effort to categorize the performance of the Airmen. These categories are shown in Table 3.

Table 3. Rating Categories for Each Attribute

Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
Below Standard	Potential	At Standard	Exceeds Standard

However, the rating categories were broad, and were not numerically defined. Therefore, the SMEs needed to further define the above categories using published Air Force Doctrine. For the Leadership and Performance in Primary and Additional Duties Objective and for the Values and Responsibilities Objective, the use of rank, skill-level, and duty position helped further define the four rating categories. The refined rating categories are shown in Table 4 and Table 5.

Table 4. Leadership/Performance and Values/Responsibilities Ratings Categories

	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
Leadership/Performance in Primary/Additional Duties And Values/Responsibilities	Below Standard	Potential	At Standard	Exceeds Standard
	Meets Minimal Objectives Not Consummate With Rank and Duty Position	Meets Some Objectives Consummate With Rank and Duty Position	Meets All Objectives Consummate With Rank and Duty Position	Meets Objectives For Next Higher Rank and Duty Position

Table 5. Physical Fitness Ratings Category

	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
Physical Fitness	Exempt in All Components	Below Standard	At Standard	Exceeds Standard
	Current with Min Passing Score Applied for Full PT Test Exemption	Non-Current or Current Failure in Overall Score or 1+ Components	Current and Meets Standards for Overall Score and all Components	Current and Exceeds Standards for Overall Score and Meets all Components

Since each career field is unique, the SMEs felt the specific Career-Field Education and Training Plan, the Enlisted Force Structure, and the Core Values Manual provided common ground and clarity to the rater and ratee in defining the rating categories.

For the Professional Qualities objective, each of the three attributes was decidedly different, and thus required different definitions for each of the four rating categories.

The SMEs developed unique definitions for each of the rating categories, for each of the attributes. Using this method, the team was able to more easily quantify each of the three attributes. The rating definitions of each of the four categories are listed below in Table 6 through Table 8 for each of the three attributes which comprise the professional qualities fundamental objective.

Table 6. Awards Ratings Category

Awards	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	No Awards	Consider Squadron, Group, and Wing Nominee	Consider Squadron, Group, and Wing Awards	Consider NAF/MAJCOM/HQ USAF/ Joint Level Awards

Table 7. Education Level Ratings Category

Education Level	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Not Pursuing Education Opportunities	Currently Pursuing a Degree or Certification	Possesses CCAF and/or Associate Degree	Possesses Bachelors or Graduate Degree

Table 8. Base and Community Involvement Ratings Category

Base and Community Involvement	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Does not Participate in Base or Community Events	Participates in 1 Base or Community Event	Participates in 2+ Base or Community Events	Active in 4+ Base or Community Events with Leadership Role in 1+ Event

In further refining the rating categories, the SMEs created variable ranges for scoring inside each ratings category. The structure was similar to the ratings blocks used in the

current design, but provided better delineation among performers by providing the rater flexibility within a ratings category, thus allowing the rater to better quantify the observed qualitative measurements as quantitative values, and to not simply score the attribute by placing a rating in a “bin”, where one size fits all. Each attribute was designed to be scored by the rater on a 0 to 100 point scale. Within each of the four ratings categories, the SMEs determined what portion of the 100 point scale applied to each particular category for each particular attribute. Table 9 through Table 14 captures the completed rating categories.

Table 9. Initial Rating Category Definitions for Duty Performance, Duty Leadership, and Communication in the Leadership & Performance Fundamental Objective

Leadership/Performance in Primary/Additional Duties	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Meets Minimal Objectives Not Consummate With Rank and Duty Position	Meets Some Objectives Consummate With Rank and Duty Position	Meets All Objectives Consummate With Rank and Duty Position	Meets Objectives For Next Higher Rank and Duty Position
Duty Performance	0 to 14	15 to 39	40 to 64	65 to 100
Duty Leadership	0 to 19	20 to 39	40 to 59	60 to 100
Communication	0 to 19	20 to 39	40 to 59	60 to 100

Table 10. Initial Rating Category Definitions for Leadership & Performance Fundamental Objective

Physical Fitness	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Exempt in All Components	Below Standard	At Standard	Exceeds Standard
	Current with Min Passing Score Applied for Full PT Test Exemption	Non-Current or Current Failure in Overall Score or 1+ Components	Current and Meets Standards for Overall Score and all Components	Current and Exceeds Standards for Overall Score and Meets all Components
Physical Fitness	75	0 to 100 0% Awarded for Raw Score	75 to 89	90 to 100

**Table 11. Initial Rating Category Definitions for Values & Responsibilities
Fundamental Objective**

Values and Responsibilities	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Meets Minimal Objectives Not Consummate With Rank and Duty Position	Meets Some Objectives Consummate With Rank and Duty Position	Meets All Objectives Consummate With Rank and Duty Position	Meets Objectives For Next Higher Rank and Duty Position
Respect for Service & Standards	0 to 24	25 to 49	50 to 74	75 to 100
Discipline & Self-Control	0 to 19	20 to 39	40 to 59	60 to 100
Honesty & Accountability	0 to 19	20 to 39	40 to 59	60 to 100
Responsibility	0 to 14	15 to 29	30 to 49	50 to 100
Teamwork & Followership	0 to 29	30 to 44	45 to 64	65 to 100

**Table 12. Initial Rating Category Definitions for Awards Sub-Category in
Professional Qualities Fundamental Objective**

Awards (Sub-Category of Professional Qualities Fundamental Objective)	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	No Awards	Consider Squadron, Group, and Wing Nominee	Consider Squadron, Group, and Wing Awards	Consider NAF/MAJCOM/HQ USAF/ Joint Level Awards
	0 to 14	15 to 29	30 to 49	50 to 100

**Table 13. Initial Rating Category Definitions for Education Level Sub-Category in
Professional Qualities Fundamental Objective**

Education Level (Sub-Category of Professional Qualities Fundamental Objective)	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Not Pursuing Education Opportunities	Currently Pursuing a Degree or Certification	Possesses CCAF and/or Associate Degree	Possesses Bachelors or Graduate Degree
	0 to 39	40 to 49	50 to 69	70 to 100

**Table 14. Initial Rating Category Definitions for base and Community Involvement
Sub-Category in Professional Qualities Fundamental Objective**

Base and Community Involvement (Sub-Category of Professional Qualities Fundamental Objective)	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Does not Participate in Base or Community Events	Participates in 1 Base or Community Event	Participates in 2+ Base or Community Events	Active in 4+ Base or Community Events with Leadership Role in 1+ Event
	0 to 29	30 to 49	50 to 79	80 to 100

VFT Weight Solicitation

With the evaluation categories now defined, the SMEs rank ordered the fundamental objectives. They ranked Leadership and Performance in Primary/Additional Duties as the most important objective, followed by Values and Responsibilities, with Professional Qualities as the third most important fundamental objective in performance. This same method was used for each of the attributes inside the Fundamental Objective categories. With the categories ranked, swing weighting was utilized for determining the appropriate weights for the new appraisal VFT Framework (von Winterfeldt & Edwards, 1986). The SMEs felt that swing weighting techniques would best capture the level of importance and impact to the Airmen, the unit, and the Air Force as a whole.

Swing weighting determines a weighting scheme by querying the decision makers and/or key stakeholders using a series of questions (Pöyhönen & Hämäläinen, 2001). For the new junior enlisted appraisal project, the SNCO SMEs were utilized as key stakeholders for the weight determinations per the Decision Analysis “Silver Standard” (Merrick et al., 2005). Initially during the swing weighting process, all weights for all attributes were moved to the lowest possible level (Pöyhönen & Hämäläinen, 2001). Using a 0 to 100 point scale, the SMEs were asked which attribute they felt was the most important (Pöyhönen & Hämäläinen, 2001). Unanimously, the SMEs felt that Duty Performance was the most important attribute. Duty Performance was assigned the maximum value of 100 points (Pöyhönen & Hämäläinen, 2001). Next, the SMEs were asked which attribute was the second most important to move from the lowest to the highest weighting level (Pöyhönen & Hämäläinen, 2001). Duty Leadership was chosen by the SMEs, and after much discussion, the SMEs felt that Duty Leadership had

possessed one fourth of the importance in the performance of a junior enlisted airman than did Duty Performance. Therefore, a lower portion of points than Duty Performance, 25 of a possible 100 points, were assigned to Duty Performance (Pöyhönen & Hämäläinen, 2001). Physical Fitness was the next attribute in importance as determined by the SNCO SMEs. The SMEs again assigned 25 points to the Physical Fitness attribute, as they felt that in today's Air Force climate, Air Force leadership and Air Force Instructions highly value Physical Fitness. This process continued with the remaining attributes. Trade spaces and value differences were discussed, until finally the SMEs agreed on a ranking and weighting scheme. After all attributes were considered, the SMEs had allocated a total of 250 points for all the attributes. The weights were then normalized so that all of the weights summed to one. Table 15 reflects the final rank ordering of the attributes and the determined weights.

Table 15. SME Ranking of Importance of Objectives and Weight Assignments

SNCO SME				
Ranking of Importance of Value Function Objectives				
Attribute Importance to SMEs	Objective Number	Description	Raw Swing Weight Points Score	Normalized Weight Assignments
1	1	Duty Performance	100	0.40
2	2	Duty Leadership	25	0.10
3	3	Physical Fitness	25	0.10
4	5	Respect for Service and Standards	20	0.08
5	4	Communication	12.5	0.05
6	6	Discipline and Self-Control	12.5	0.05
7	7	Honesty and Accountability	12.5	0.05
8	8	Responsibility	10	0.04
9	10	Awards	10	0.04
10	9	Teamwork and Followership	7.5	0.03
11	11	Education	7.5	0.03
12	12	Base and Community Involvement	7.5	0.03

The next task in the weighting process was determining whether to use a local or global weighting scheme. A local weighting scheme partitions the weights among Fundamental Objectives, then partitions the weight assigned to that specific Fundamental Objective to the attributes located underneath the respective objectives. Suppose 75% of the weighting is assigned to Fundamental Objective 1 and 25% of the weighting is assigned to Fundamental Objective 2. If attribute 1A, underneath Fundamental Objective 1 has 65% of the importance of Fundamental Objective 1, then attribute 1A actually contributes only 48.75% of the weight to the overall model. Figure 10 illustrates how a local weighting scheme is derived.

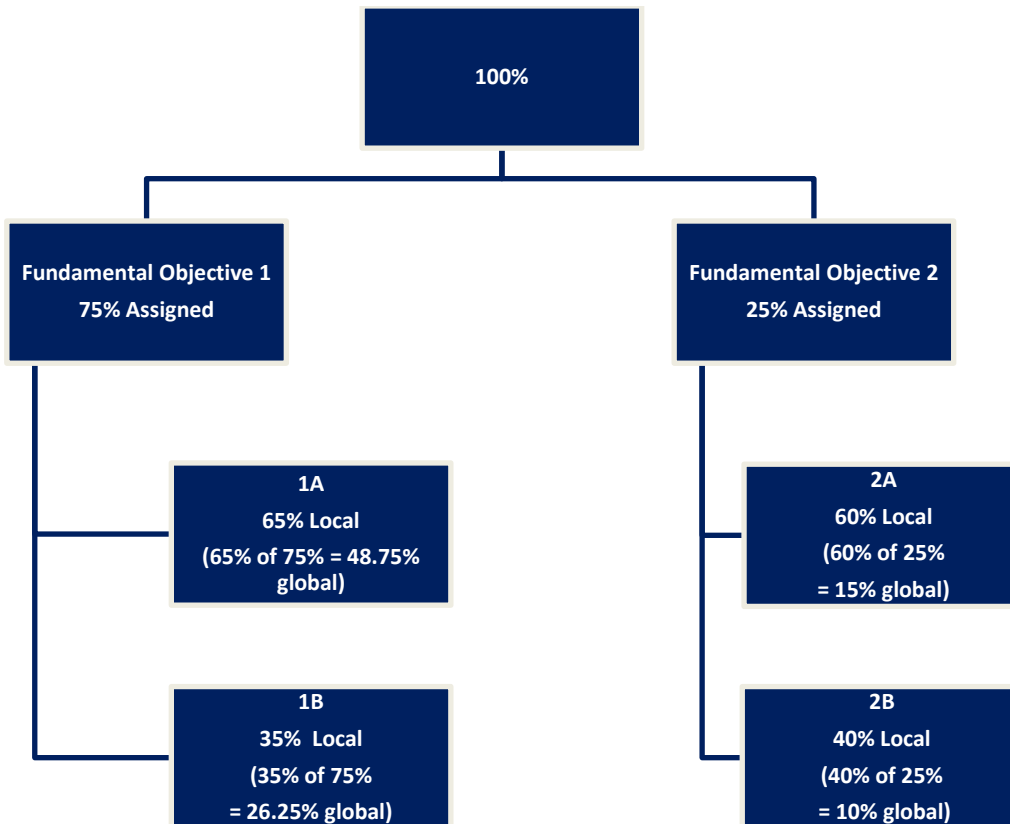


Figure 10. Example of Local Weighting Construct

In a global weighting scheme, the weights are partitioned among the attributes, not the Fundamental objectives. Each attribute weight contributes directly to the overall 100% of the weighting allocation. In the global weighting scheme, attribute 1A, weighted at 65%, contributes 65% to the overall weighting. This can be seen explicitly in Figure 11.

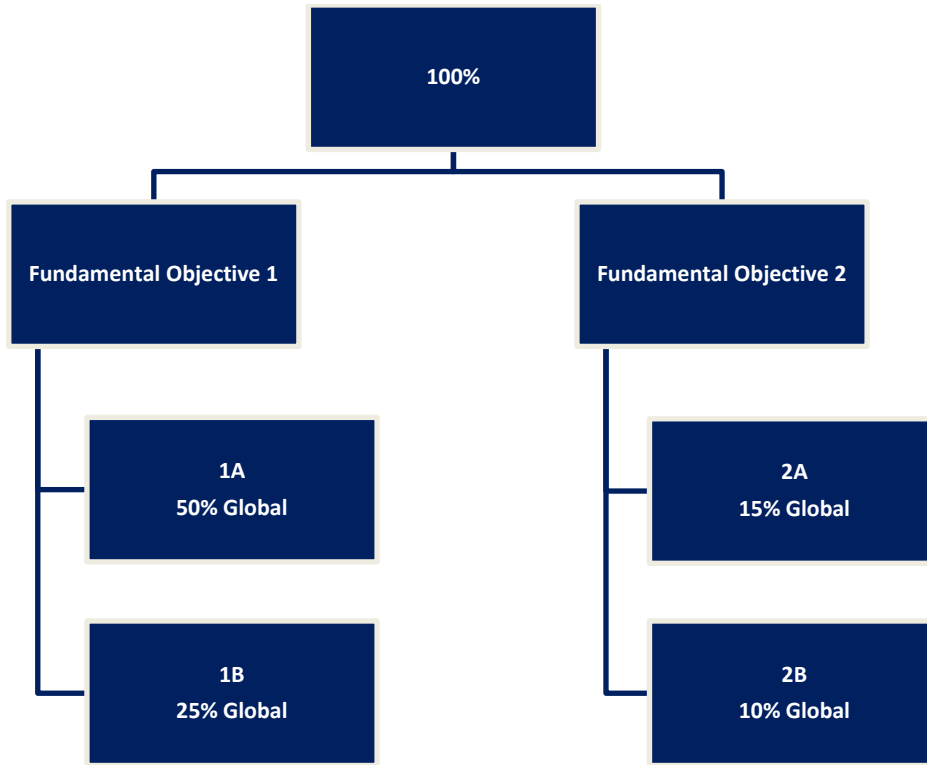


Figure 11. Example of Global Weighting Construct

Although global weighting structures are easier to understand, when a VFT Framework involves a diverse and broad group of stakeholders, local weighting schemes are usually superior. In large stakeholder models, the local decision maker at the Fundamental Objective level is usually more knowledgeable in their specific areas of control when partitioning the weights. Had a larger hierarchy had been used, with several

hundreds of attributes, the solicitation and assignment of a global weighting scheme would have been impractical. However, for the JEPR project, the small number of attributes used in this model made obtaining and assignment of global weights possible. Figure 12 illustrates how the derived weighting scheme was applied globally to the JEPR VFT Framework. This weights associated with the VFT Framework will later be utilized in computing the additive value functions for each attribute used by the proposed JEPR appraisal system.

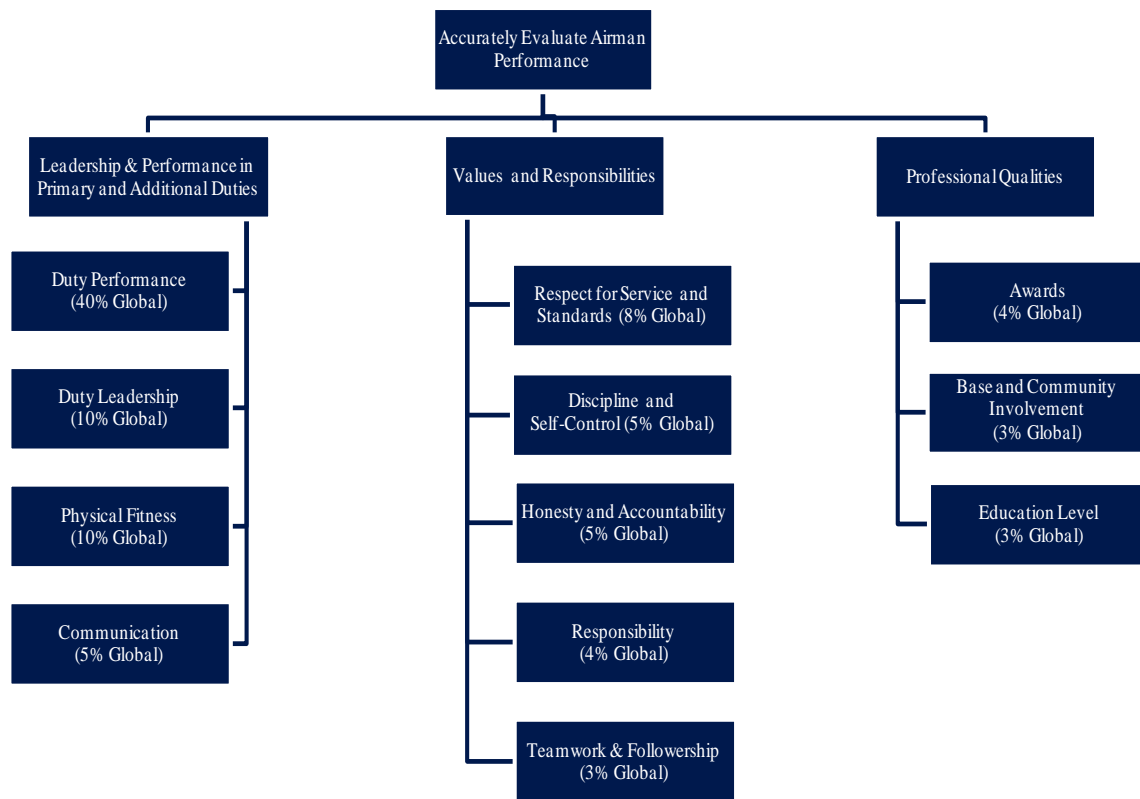


Figure 12. JEPR Value Hierarchy with Global Weight Structure

VFT Attribute Function Development

With the rating categories developed and the swing weights created, the next portion of the analysis was to develop the functions based on data solicited from the SMEs. We asked for three data points for each of the 12 attributes to be able to construct a unique Single Attribute Value Function (SAVFs) for each of the attributes. Known as the bisection method, the purpose of this process is to solicit points of performance for each attribute from the SNCO SMEs, then generate a curve that generates the lowest sum of squares computations between the solicited points (Watson, 1987). The curve for each attribute then will reflect the value of the function at all locations between the minimum and maximum values for the attribute.

For the possible scores Airmen could receive for an attribute, the top possible data point was set at 100, meaning the best score that could be earned in the category would be 100. The bottom data point was also fixed with the minimum score that could be earned in the category determined as 0. In addition to these minimum and maximums, for each attribute, we asked the Subject Matter Experts to provide the following:

- 1. *What score would you apply to someone meeting 25% of the attribute standard?***
- 2. *What score would you apply to someone meeting 50% of the attribute standard?***
- 3. *What score would you apply to someone meeting 75% of the attribute standard?***

Using these solicited data points, SAVFs were constructed for each attribute using an Exponential Single Dimensional Value Function (Kirkwood, 1996). The SAVF function initially used for this study is shown in Equation 1. Looking closer at Equation 1 in

determining the specific value of a function at a given point, x_i is the point of interest along the curve, x_i^0 is the minimum possible value of the curve, while x_i^* is the maximum possible value of the curve. Finally, the γ_i (Gamma) value is the unique shaping component for the specific attributes curve. Table 16 illustrates the specific γ_i (Gamma) values for each of the VFT Framework SAVF functions.

$$f_i = \frac{1 - e^{-\gamma_i(x_i - x_i^0)}}{1 - e^{-\gamma_i(x_i^* - x_i^0)}} \quad (1)$$

Table 16. Gamma Shaping Component for SAVFs Used in VFT Function

Gamma Shaping Component for Value Function Objectives		
Attribute Number	Attribute	Gamma Value Used
1	Duty Performance	0.009679388
2	Duty Leadership	0.009386208
3	Physical Fitness	0.009679388
4	Communication	0.009386208
5	Respect for Service and Standards	0.0000000001
6	Discipline and Self-Control	0.00938621
7	Honesty and Accountability	0.00938621
8	Responsibility	0.018435884
9	Teamwork and Followership	0.002990016
10	Awards	0.018435884
11	Education	-0.00295596
12	Base and Community Involvement	-0.00281841

Figure 13 illustrates the Duty Performance SAVF fitted between the performance data points solicited from the SMEs during the function design. Notice how the curve has been fitted between the solicited points to minimize the sum of squares total between the solicited points.

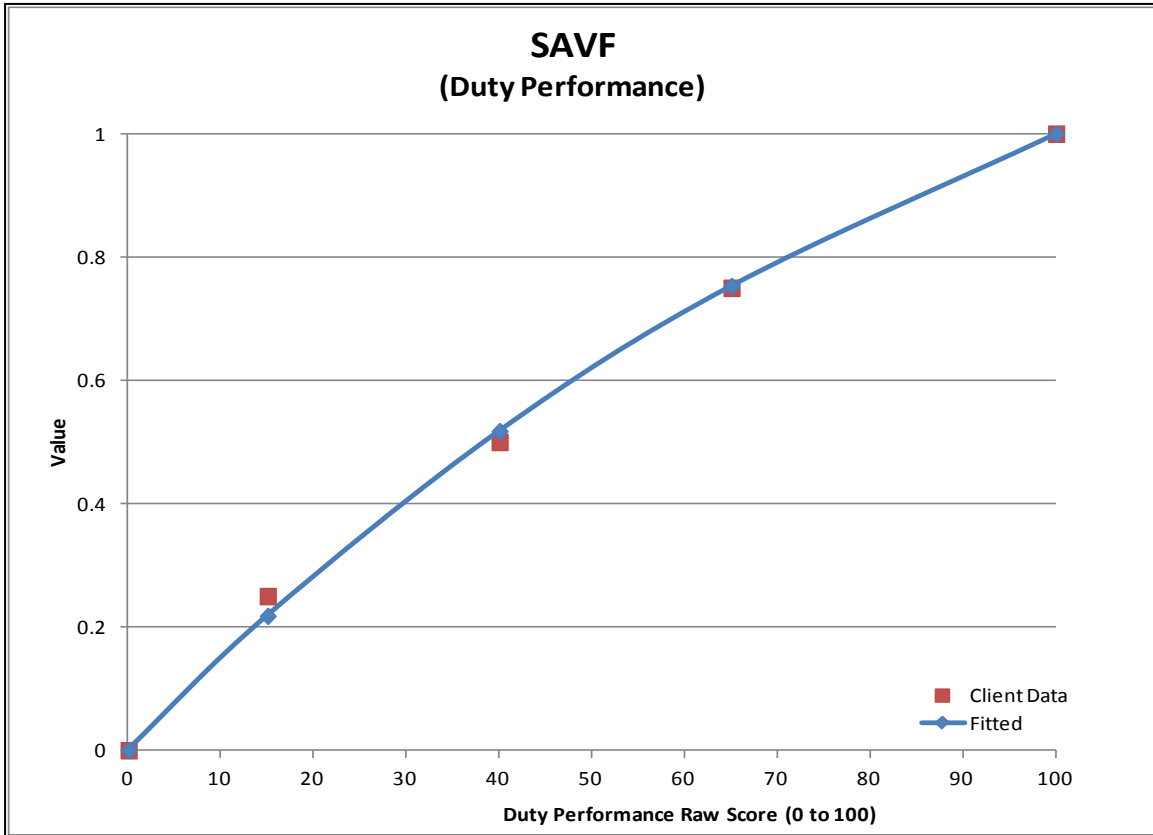


Figure 13. Duty Performance SAVF Function Example

With the SAVF functions now developed, the functions and weights could be combined to form an additive Multi-Attribute Value Function (MAVF). The model for the revised performance report would work as follows. The supervisor would enter the raw scores (0 to 100) for the ratee for each of the 12 attributes. The scores would then have the shaping functions from Table 16 applied (these functions were based on data solicited from the SMEs for each particular attribute). The weights would then be applied for each particular attribute, and then all 12 components would be summed together using an additive MAVF. This MAVF would yield the final performance report score for the Airman of interest. The mathematical model is reflected in Equation 2.

$$\begin{aligned}
v(x) &= \sum_{i=1}^{12} w_i f_i \\
&= w_1 f_1 + w_2 f_2 + w_3 f_3 + w_4 f_4 + w_5 f_5 + w_6 f_6 \\
&+ w_7 f_7 + w_8 f_8 + w_9 f_9 + w_{10} f_{10} + w_{11} f_{11} + w_{12} f_{12}
\end{aligned} \tag{2}$$

However, at this point a problem arose. The SNCOs wanted to be able to deduct points away from an individual when Administrative Actions had to be taken to correct repeated poor behavior or repeated gross negligence. These activities are well above and beyond the normal counseling and mentoring sessions between supervisors and rates', and are formally documented in the individuals Personal Information File. In an effort prevent marginalizing the effects or disrupting the weight structure of the VFT function, an external Penalty Function was created as a correction factor, to capture the negative impacts of Administrative Actions. The Penalty Function is not part of the Value Hierarchy, as it is a correction factor after the value score had been generated. If Administrative Actions had occurred for a particular Airman, the Penalty Function corrects the additive VFT Function score after the fact, by deducting a penalty to yield an overall JEPR score. The purpose of this was to capture the impact and ramifications of the Administrative Actions. If no Administrative Actions occurred, only the additive VFT function would determine the JEPR overall score. In essence, the Administrative Action function would be treated as an independent variable, similarly to how cost is treated in an acquisitions decision where when cost is deemed as a Cost As an Independent Variable, and is introduced after computation of the value of the system. The thought behind this from the SMEs was that the EPR, regardless of score, should

definitely reflect the fact that Administrative Actions had been documented during the rating period. Using the same techniques as before, the penalty function rating categories were developed along with a weighting scheme. Table 17 and Equation 3 below reflect the rating categories and the weight of the penalty function.

$$pw = 0.35 \tag{3}$$

Table 17. Initial Rating Category Definitions for Penalty Function

Administrative Actions (Correction Factor)	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Article 15/UCMJ	LOC/LOA/LOR	LOC/LOA/LOR	Min/No Negative Indicators
	Documented Article 15 or UCMJ Actions	Reoccurring disciplinary issues with multiple LOCs/LOAs/LORs in PIF	Documented disciplinary issue with single LOC/LOA/LOR in PIF	Minimal to no disciplinary issues. Consider PT failures in Period if now Passing
	-100 to -81	-80 to -61	-60 to -31	-30 to 0

The penalty function did not follow was computed the same shape as the additive multi-attribute value functions did. The structure was negative, with a Gamma shaping component of -0.00673012. Equation 4 shows the initial function used in building the penalty function while Equation 5 shows how the independent penalty function was integrated into the value hierarchy.

$$pf = -1 \left[\frac{1 - e^{-\gamma_i(x_i + x_i^*)}}{1 - e^{-\gamma_i(x_i^* + x_i^0)}} \right] \tag{4}$$

Mathematically the completed penalty function with weights included s as follows, where pw is the penalty weight and x is the value of the function acting on the raw penalty score provided by the supervisor:

$$p(x) = (pw)(pf) \tag{5}$$

Figure 14 illustrates the value hierarchy with the weights applied.

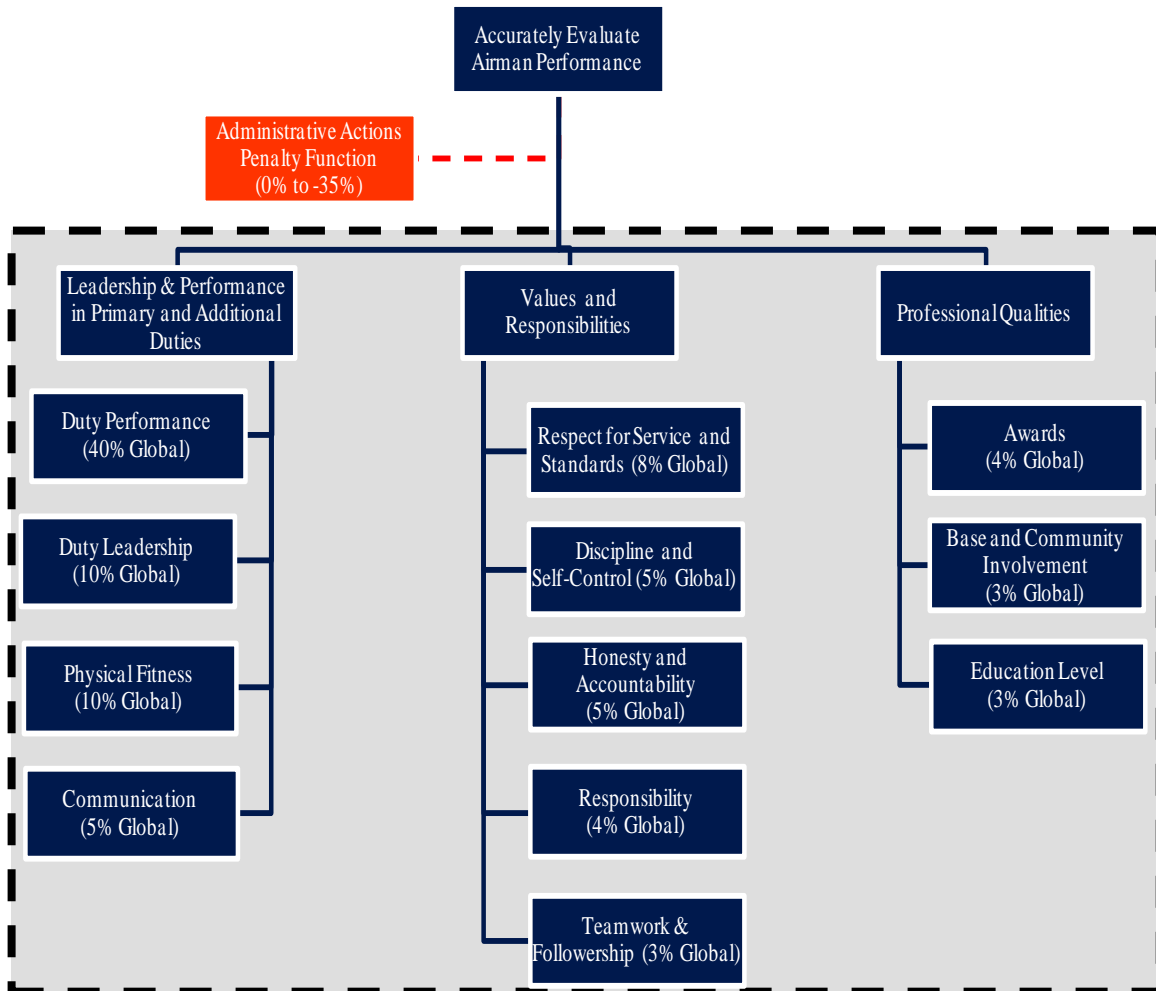


Figure 14. Overall Scoring Scheme Comprised of Value Hierarchy Framework

Therefore the completed JEPR overall score is computed as shown in Equation 6, with the penalty function having a negative value:

$$z(x) = \begin{cases} v(x) + p(x), & \text{if } pf < 0 \\ v(x), & \text{if } pf = 0 \end{cases} \quad (6)$$

VFT Attribute Function Revisions

An in-depth review of the Single Attribute Value Functions (SAVFs) in this project revealed that the exponential functions did not fit well when the sum of squares was evaluated for the Physical Fitness, Teamwork and Followership, and the Educational Activity attributes when compared against the Subject Matter Experts (SMEs) provided data. The lack of fit for the exponential function was also noted for the independent penalty function for the Administrative Actions correction factor. Therefore, the functions for these four attributes were redesigned incorporating a piecewise design. Figure 15 contrasts the lack of fit experienced with the exponential function versus the Piecewise SAVF function for the Teamwork and Followership attribute.

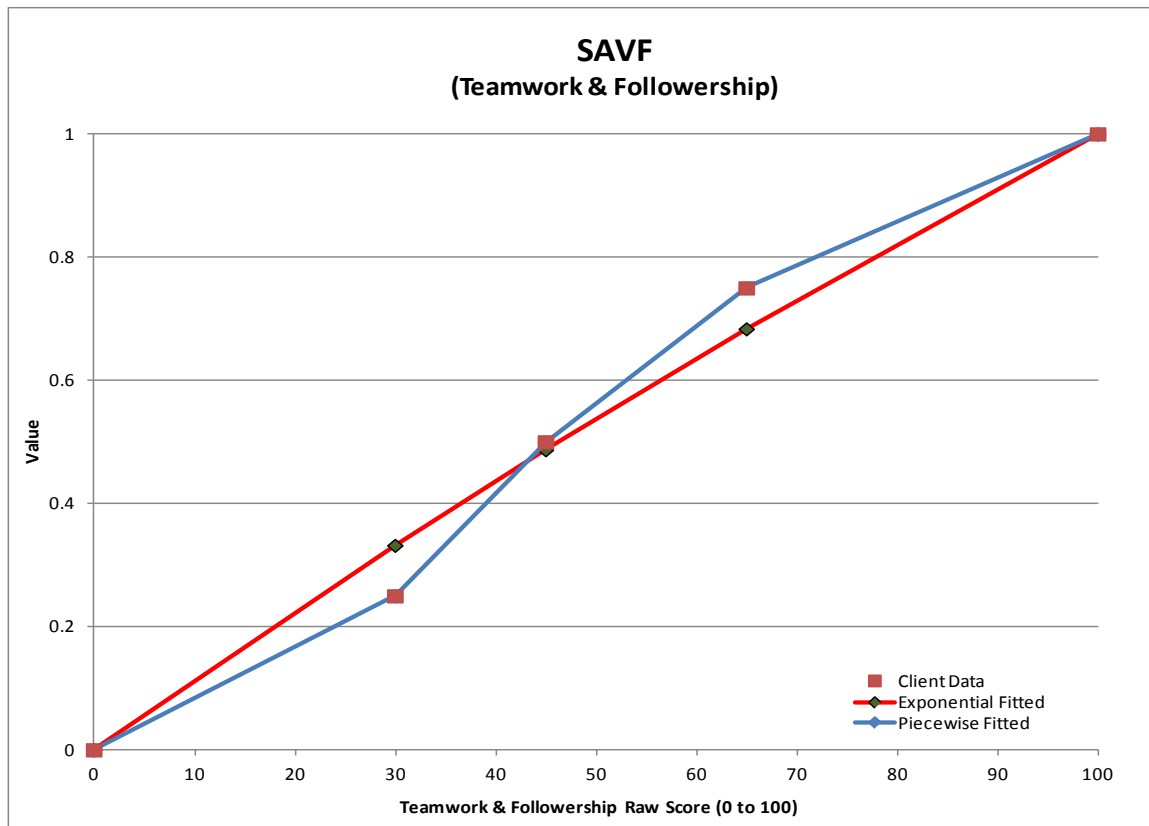


Figure 15. Comparison of Exponential vs. Piecewise Function Fit

For the Physical Fitness attribute, compromises were made to align with the current system. A method for exemptions was developed along with a method to capture failures where the overall fitness score was satisfactory, yet a minimum passing score in one of the test components was not achieved by the member. Therefore, the team determined that a failure had no value in the function, while a fully exempted member would receive a minimal passing score. The team hoped to capture the lack of readiness by awarding the minimum passing score without dramatically affecting the overall value score. The team felt this would promote physical fitness testing versus reliance on full fitness test exemptions, as more promotion points would be available. The revised attribute functions are shown in Equation 7 through Equation 8 for all Single Value Attribute Functions (SAVFs), where i is the attribute number using the function, j is the additive sum of the function before slope k , and k is the current section of the function. Each piecewise function used in the VFT Framework was comprised of four sections. The Piecewise function sectional ranges and slopes are also provided and are compiled in Table 18 through Table 21.

$$f_i = \begin{cases} \frac{\left(\frac{RAW}{SLOPE_k}\right)}{100}, & k = 1 \text{ and } RAW \leq MAX_k \\ \frac{\left(\sum_{j=2}^k \frac{(MAX_{j-1} - MAX_{j-2})}{SLOPE_{j-1}} + \frac{(RAW - MAX_{k-1})}{SLOPE_k}\right)}{100}, & 2 \leq k \leq 4 \text{ and } RAW \leq MAX_k \end{cases} \quad (7)$$

Table 18. Piecewise Sectional Ranges and Slopes for Physical Fitness SAVF

Objective 3		
Physical Fitness		
Percentage of What an Ideal Employee Provides	Raw Score Ranges Solicited	Calculated Piecewise Slopes
0%	0	0
25%	1 to 74	2.96
65%	75 to 75	0.025
95%	76 to 90	0.50
100%	91 to 100	2.00
NOTE		
Function Values are artificially terminated for overall PT scores below 75% or for a failure in 1 or more components regardless of score. For these scenarios, 0% value is awarded for the SAVF. This is due to Air Force Instruction 36-2905 Guidance.		

Table 19. Piecewise Ranges and Slopes for Teamwork and Followership SAVF

Objective 9		
Teamwork and Followership		
Percentage of What an Ideal Employee Provides	Raw Score Ranges Solicited	Calculated Piecewise Slopes
0%	0	0
25%	1 to 30	1.20
50%	31 to 45	0.60
75%	46 to 65	0.80
100%	66 to 100	1.40

Table 20. Piecewise Ranges and Slopes for Revised Education SAVF

Objective 11		
Education		
Percentage of What an Ideal Employee Provides	Raw Score Ranges Solicited	Calculated Piecewise Slopes
0%	0	0
25%	1 to 40	1.60
50%	41 to 50	0.40
75%	51 to 70	0.80
100%	71 to 100	1.20

$$pf = \begin{cases} \frac{\left(\left(\sum_{j=k}^4 \frac{(MAX_{j-1} - MAX_j)}{SLOPE_j} \right) + \frac{RAW - MAX_{k-1}}{SLOPE_k} \right)}{100}, & 1 \leq k \leq 3 \text{ and } RAW \leq MAX_k \\ \frac{\left(\frac{MAX_{k-1} - MAX_k}{SLOPE_k} + \frac{RAW - MAX_{k-1}}{SLOPE_k} \right)}{100}, & k = 4 \text{ and } RAW \leq MAX_k \end{cases} \quad (8)$$

Table 21. Piecewise Ranges and Slopes for Revised Penalty Function

Negative Value Contribution		
Independent Penalty Function		
Percentage of What an Ideal Employee Provides	Raw Score Ranges Solicited	Calculated Piecewise Slopes
0%	-100 to -81	1.00
25%	-80 to -61	0.57142286714
50%	-60 to -31	1.00
75%	-30 to -1	2.00
100%	0	0

VFT Data Collection

The data collection effort was an iterative process and was conducted in two phases. The first phase was the training phase. This phase was used to validate the accuracy of the JEPR model's numerical output versus the qualitative performance observations of the tactical level supervisors. The training phase was also sought to verify that the VFT framework was consistent with Air Force strategic values and doctrine. To prevent inadvertently influencing the ratings of the current system and to also accurately capture the tactical level supervisors observations in a timely manner without loss of data, JEPR system appraisals were completed immediately following the completion of the official EPR for a test subject. By using the JEPR system as a shadow system, the intent

was to record near parallel data under both systems to better understand and capture the values of the rater, the organization, and the enlisted force structure as a whole. For the training phase, 71 test subjects, across eight unique AFSCs, had their overall appraisal ratings recorded using the current EPR system. Upon immediate completion of the formal report, the test subjects were then appraised using the JEPR system construct. The initial findings were presented to the work group, the Barksdale Top Three SNCO organization and the Barksdale Chiefs Group for discussion, consideration, and refinement. This iterative process will allow a myriad of different enlisted perspectives, career field expectations, and training to further define the categories for an accurate evaluation. The analytical intent was to use these initial 71 data points as training data, where the JEPR model could be adjusted or corrected based on observations noted by the raters during the initial effort.

The second phase of data collection was the test phase. This phase was used to verify that the JEPR models numerical output was consistent with the qualitative performance observations of the tactical level supervisors and from the previous training effort. Additionally, this phase sought to verify that the VFT framework was consistent with Air Force strategic values and doctrine and did not deviate from the underlying construct discovered during the training phase of data collection. Again, to prevent inadvertently influencing the ratings of the current system and to also accurately capture the tactical level supervisor's observations in a timely manner without loss of data, the JEPR system appraisals were completed immediately following the completion of each official EPR for each test subject. For the test phase, 159 test subjects, across 24 unique AFSCs were involved in the JEPR test effort.

VFT Deterministic Analysis (Notional Dataset)

Once the SAVFs and the MAVF were redesigned, a Deterministic analysis of the VFT Weighted Value Model was performed. Due to the rapid departure of this proposal from the current Junior Enlisted Performance Reporting (EPR) structure, translation of historical EPR scoring could not be accomplished. In particular, it is impossible to translate the banded discrete rating categories of the historical EPR format to the expanded and narrowly defined JEPR categories. Therefore, before field testing the prototype, notional JEPR data was generated to ensure the model design was sound, and to validate that the scoring outputs generated by the JEPR model fall within the expectations of the SMEs based on their inputs that were solicited during the design.

For this project, scores were generated for each JEPR attribute for eight notional junior enlisted personnel using a random number generator in Microsoft Excel, with the random attribute scores ranging between 0.00 and 1.00. The Administrative penalty function was not considered at any point during the analysis, as it is independent of the VFT framework, and is not a part of the VFT Weighted Value Model. Once the independent attribute scores were generated, the overall value score for each notional Airman was computed by summing the attribute scores. Additionally, an “Ideal” employee was also included in the analysis as a baseline. The “Ideal” employee is considered “The Best of the Best” and reflected the maximum possible score for each category across all attributes. The independent randomly generated weighted SAVF scores along with the VFT Weighted Value Model overall scores are shown in Table 22 and Table 23.

Table 22. SAVF Scores for Notional Personnel A through D and an Ideal Airman

SAVF Scores for Ideal Airman and Notional Personnel A through D (Overall Score and Ranking Included)					
Attribute	Ideal	A	B	C	D
Duty Performance	0.4000	0.3600	0.3120	0.1120	0.3280
Duty Leadership	0.1000	0.0440	0.0960	0.0650	0.0040
Teamwork and Followership	0.0300	0.0018	0.0264	0.0222	0.0003
Respect for Service and Standards	0.0800	0.0088	0.0056	0.0784	0.0224
Discipline and Self-Control	0.0500	0.0315	0.0450	0.0290	0.0300
Communication	0.0500	0.0400	0.0335	0.0005	0.0245
Responsibility	0.0400	0.0068	0.0356	0.0396	0.0144
Honesty and Accountability	0.0500	0.0325	0.0030	0.0000	0.0110
Physical Fitness	0.1000	0.0910	0.0770	0.0840	0.0000
Awards	0.0400	0.0324	0.0036	0.0208	0.0092
Base and Community Involvement	0.0300	0.0147	0.0051	0.0063	0.0078
Education	0.0300	0.0159	0.0237	0.0207	0.0180
Overall Score	1.0000	0.6794	0.6665	0.4785	0.4696
Rank		1	2	6	7

Table 23. SAVF Scores for Notional Personnel E through H and an Ideal Airman

SAVF Scores for Ideal Airman and Notional Personnel E through H (Overall Score and Ranking Included)					
Attribute	Ideal	E	F	G	H
Duty Performance	0.4000	0.3080	0.3840	0.2040	0.1640
Duty Leadership	0.1000	0.0430	0.0190	0.0260	0.0610
Teamwork and Followership	0.0300	0.0153	0.0036	0.0000	0.0075
Respect for Service and Standards	0.0800	0.0096	0.0032	0.0184	0.0640
Discipline and Self-Control	0.0500	0.0375	0.0225	0.0075	0.0090
Communication	0.0500	0.0245	0.0290	0.0470	0.0345
Responsibility	0.0400	0.0012	0.0324	0.0348	0.0112
Honesty and Accountability	0.0500	0.0010	0.0330	0.0440	0.0055
Physical Fitness	0.1000	0.0000	0.0000	0.0810	0.0650
Awards	0.0400	0.0124	0.0280	0.0036	0.0260
Base and Community Involvement	0.0300	0.0207	0.0105	0.0045	0.0057
Education	0.0300	0.0264	0.0198	0.0093	0.0144
Overall Score	1.0000	0.4996	0.5850	0.4801	0.4678
Rank		4	3	5	8

Looking at Table 22 and Table 23, the first thing noted was that personnel D, E, F received a zero score for the Physical Fitness SAVF. This was because the randomly

generated raw scores for these Airmen were less than 75. Although the Physical Fitness SAVF function does generate values below the raw score of 75, Air Force Instruction 36-2905 considers a fitness score less than 75 as unsatisfactory, and thus a failure to meet an established standard. Therefore, a score of zero is assigned for the Physical Fitness attribute in the JEPR VFT Weighted Value Model when the randomly generated raw score was below 75. Looking at the data in graphical form, Figure 16 created a Value Breakout which shows the contribution of each attribute to the overall JEPR VFT Weighted Value Model score. Figure 16 graphically shows that the Duty Performance attribute dominated all other attributes when looking at the contribution percentage of each attribute to each employee's overall score.

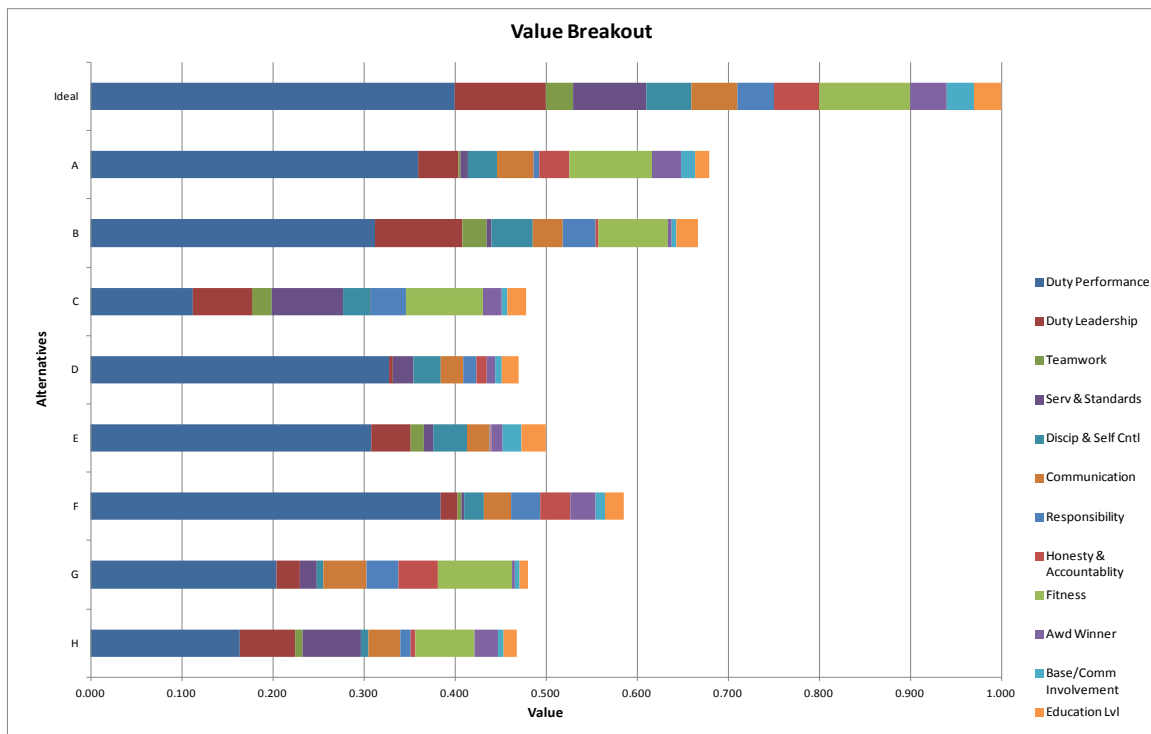


Figure 16. Value Breakout by Attribute of Value Function Scores

This was as anticipated as Duty Performance possessed the largest overall contribution weight (40%) to the overall score. In developing the value function, Duty Performance dominance was a trait that was consistently advocated by the SMEs, as the primary intent of this appraisal is to be able to more accurately capture on the job performance. Poor Duty Performance value scores directly reflected the overall score of the individual, whereas higher scores in other categories simply could not overcome poor Duty Performance. This is directly reflected in personnel C as shown in Table 22, Table 23, and Figure 16, where a relatively high fitness score of 8.4% out of 10% could not overcome the poor Duty Performance score of 11.2% out of 40%.

Review of the next highest weighted attributes of Table 22, Table 23, and Figure 16, Duty Leadership and Fitness, each weighted at 10% of the overall score, reflect a somewhat different pattern. Strong performances in lesser or equivalent categories allowed the employee to overcome a weak score in another area. This can be explicitly seen in the scores of employee B, who had a Physical Fitness score of 7.7% out of 10%, which equates to a score of 81 out of 100 on the Air Force Physical Fitness Test. However, this low score was partially compensated for by the Duty Leadership attribute with a score of 9.6% out of 10.0%. This was due to the construction of the VFT Weighted Value Model (MAVF), where a higher score in one attribute may be able to partially offset a lower score in another attribute if the weightings of the two attributes were approximately equivalent, without inflating the overall score. This type of detailed information concerning strengths and shortcomings in specific attributes has great potential as quantitative feedback for the ratee. A good example of this phenomenon can be seen in personnel D as shown in Table 22, Table 23, and Figure 16; where a strong

score in Duty Performance was ultimately impacted by the accumulation of lower scores in the remaining attribute areas. For example, the low score of 0.4% earned in the Duty Leadership category, although weighted only at 10%, did impact the overall score for personnel D. Had personnel D achieved a marginally better score in this category, for instance a score >3.5%, personnel D would have been rated fourth among the population versus seventh. Again, comparison of the remaining attributes of the Value Breakout followed this pattern, where higher scoring attributes weighted approximately the same could compensate for lower scoring attributes. However many high scoring low weight attributes (i.e. Communication, Education Level, and Responsibility) could not overcome a poor score in a heavily weighted attribute such as Duty Performance.

Next we looked at the Fundamental Objective level Value Breakout in Table 24 and Table 25. The Fundamental Objectives are the major areas which tie all the attributes that were solicited from the SNCO SMEs back to what the SMEs felt was valued by the Air Force at a strategic level. Inspection of the Fundamental Objectives was important step of the analysis, as we needed to ensure that the accumulated attributes of higher valued Fundamental Objectives dominated the accumulated attributes of lesser valued Fundamental Objectives in the VFT Weighted Value Model score. Table 24 illustrates the Fundamental Objective hierarchy.

Table 24. Fundamental Objective Hierarchy

Leadership/ Performance in Primary and Additional Duties		Values and Responsibilities		Professional Qualities	
Duty Performance	40%	Respect for Service and Standards	8%	Military Award Winner	4%
Duty Leadership	10%	Discipline and Self-Control	5%	Education Level	3%
Physical Fitness	10%	Honesty and Accountability	5%	Base and Community Involvement	3%
Communication	5%	Responsibility	4%		
		Teamwork and Followership	3%		
Total	65%	Total	25%	Total	10%

Looking at Table 25, the VFT Weighted Value Model scores at the Fundamental Objective level reveal that the heavily weighted Fundamental Objective of Leadership/Performance in Primary and Additional Duties (65% of total 100% of weighted areas) dominated the scoring. The high scores in the lesser weighted Fundamental Objectives of Values and Responsibilities (25%) and Professional Qualities (10%) were unable to offset a poor score in the Leadership/Performance in Primary and Additional Duties. The scores for the notional Airman G, as shown in Table 25 and Figure 17, are a good example of this behavior. Airman G had the highest Values and Responsibilities score and the 5th rated Professional Qualities score. Yet, the weak Leadership/Performance in Primary and Additional Duties score of 0.256 could not be overcome by the high scores in the lower weighted Fundamental Objectives.

Table 25. Scoring by Fundamental Objective

Notional Airman	Leadership/ Performance in Primary and Additional Duties	Values and Responsibilities	Professional Qualities	VFT Weighted Value Model Score
Ideal	0.660	0.140	0.200	1.000
A	0.446	0.079	0.154	0.679
B	0.485	0.072	0.109	0.667
C	0.307	0.040	0.132	0.479
D	0.385	0.050	0.035	0.470
E	0.413	0.027	0.060	0.500
F	0.432	0.094	0.058	0.585
G	0.256	0.126	0.098	0.480
H	0.306	0.051	0.111	0.468

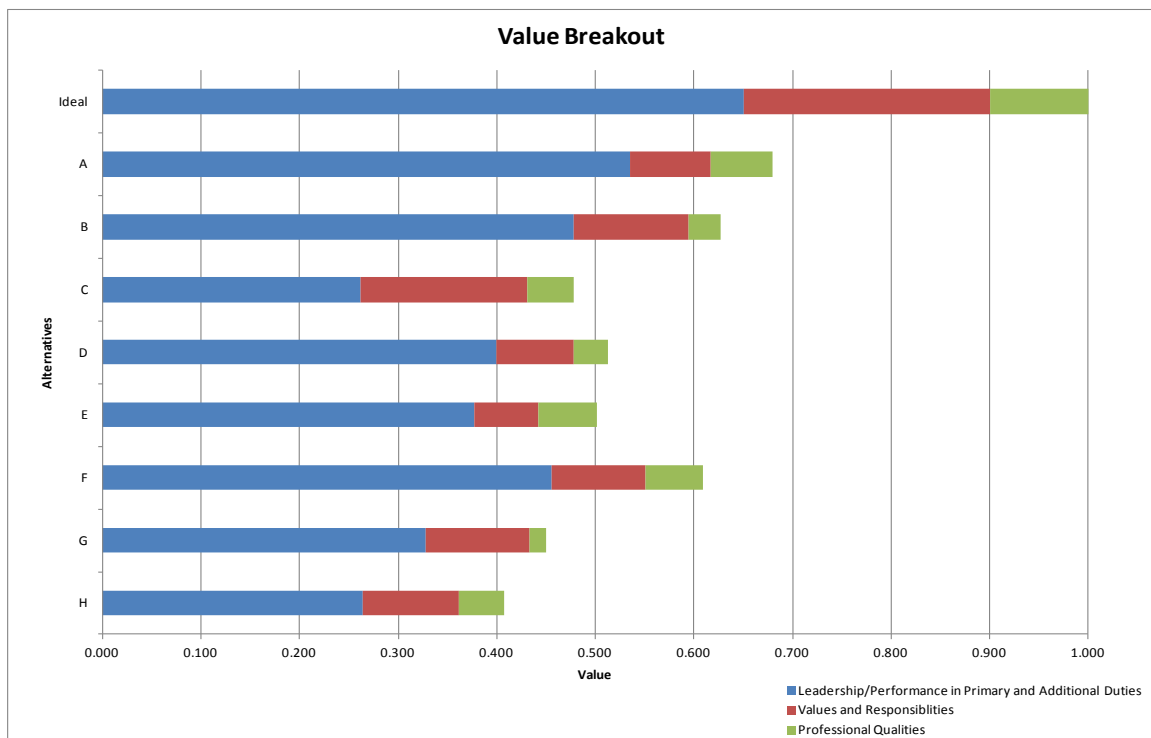


Figure 17. Value Breakout by Fundamental Objective of Value Function Scores

After review of the data in Table 25 and Figure 17, the SMEs felt the model accurately captured their value structure. As Leadership/Performance in Primary and Additional Duties was deemed to be the most important value for the Air Force by the SMEs, the JEPR model mirrored this importance as Leadership/Performance in Primary and Additional Duties was shown to be the most dominant feature in the JEPR model. After analyzing the Value Breakout tables and charts, a Value Gap analysis was performed. The purpose of the Value Gap analysis was to numerically and graphically capture the detailed qualitative feedback that the JEPR model was capable of generating from each attribute. For the analysis, the value scores for each of the 12 attributes were recorded and charted for the eight notional employees. Additionally, the difference between each individual's value score in each attribute area and the "Ideal" Airman who is "The Ideal Best of the Best" was also recorded and charted. The Value Gap Graph provided in-depth insight, both numerically and visually concerning the notional Airman's performance. For a real evaluation, this type of information would be invaluable to both the rater and to the ratee in illustrating graphically and numerically on where the ratees' performance stands in relation to the best rating that could have been achieved, for each attribute measured. The Value Gap also provides a vector to both the supervisor and to the employee on areas of strength, and for areas that need further training and mentorship. Finally, the ratee can see how a particular attribute impacts their overall score and ranking. An example of the Value Gap data and graph for personnel B can be seen in Table 26 and Figure 18. Again, the SMEs felt the simulated data from the eight notional airmen reflected in the Value Gap analysis was an accurate reflection of their Value Hierarchy.

Table 26. Value Gap Computations (Scores for Notional Airmen B Shown)

Value Gap From Notional Airman B		
Attribute	Attribute Score	Value Gap From Ideal Airmen
Duty Performance	0.3120	0.0880
Duty Leadership	0.0960	0.0040
Teamwork and Followership	0.0036	0.0364
Respect for Service and Standards	0.0056	0.0744
Discipline and Self-Control	0.0335	0.0165
Communication	0.0450	0.0050
Responsibility	0.0030	0.0470
Honesty and Accountability	0.0356	0.0044
Physical Fitness	0.0264	0.0036
Awards	0.0770	0.0230
Base and Community Involvement	0.0237	0.0063
Education	0.0051	0.0249

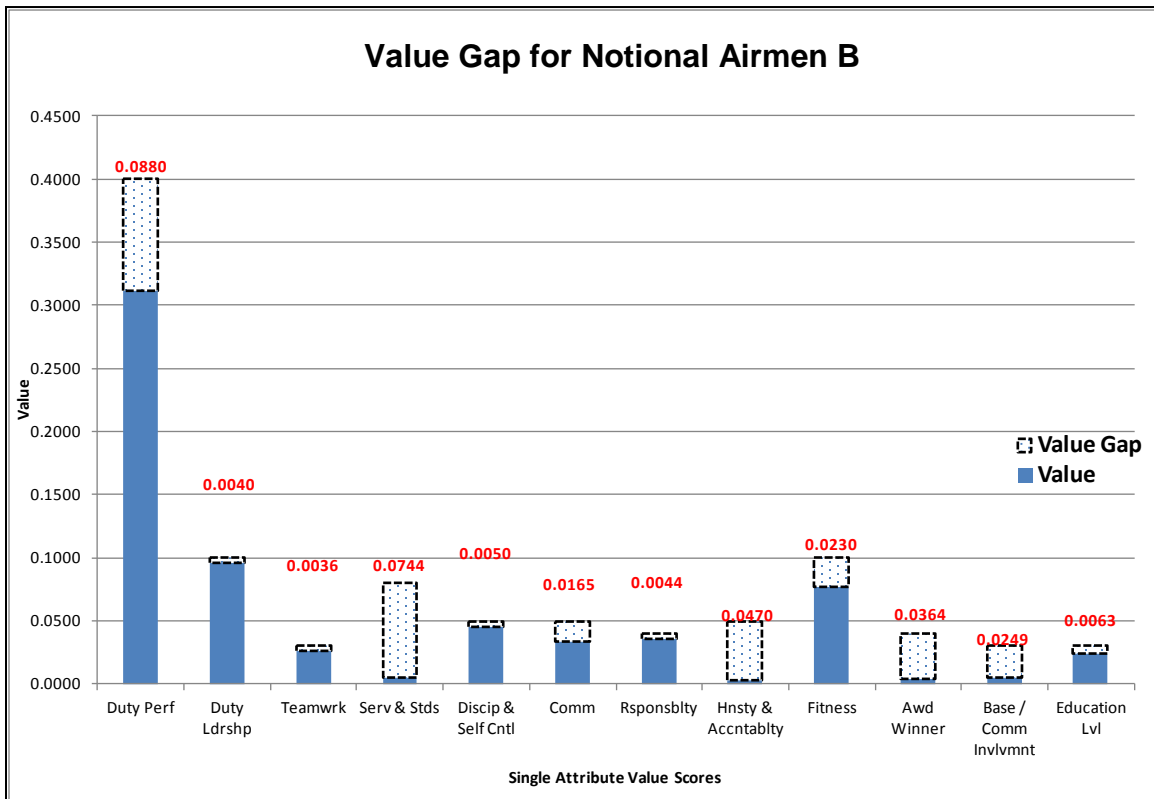


Figure 18. Value Gap Graph (Scores for Notional Airmen B Shown)

IV. Model Validation

Chapter Overview

This chapter focuses on validating the proposed model of JEPR framework. First, a Sensitivity Analysis was performed on the weights assigned to each of the JEPR attributes to determine if the members rating would change with minor changes in weightings. For the Sensitivity Analysis, the effects on the overall JEPR scores for each of the eight notional airmen were studied as the weights of each attribute were maximized incrementally. Any drastic change in the overall JEPR scores and ranking order for the notional airmen were discussed, and the weighting scheme reassessed.

A small sample of 71 JEPR reports was solicited from the Air Force population using a representative JEPR model. The representative model captured the scores for each JEPR attribute in addition to the independent Administrative Action correction factor and the overall JEPR score. Each attribute as well as the overall score from this small data sample were qualitatively inspected for behavior, shape, and statistical relationships. After the qualitative inspection, the small sample of JEPR data used as training data analyzed the consistency of the JEPR measurement scale constructs for each of the JEPR attributes. The JEPR training data was then subjected to several tests to verify suitability for factor analysis, with Exploratory Factor Analysis techniques next being applied. Finally, minor revisions were made to the model based on the observations from the JEPR Training Data analysis and discussion with the SMEs, yielding a final two factor JEPR model. This two factor model will be used for Confirmatory Factor Analysis

and Artificial Neural Networks classification analysis in Chapter V. Figure 19 provides an overview of this chapter.

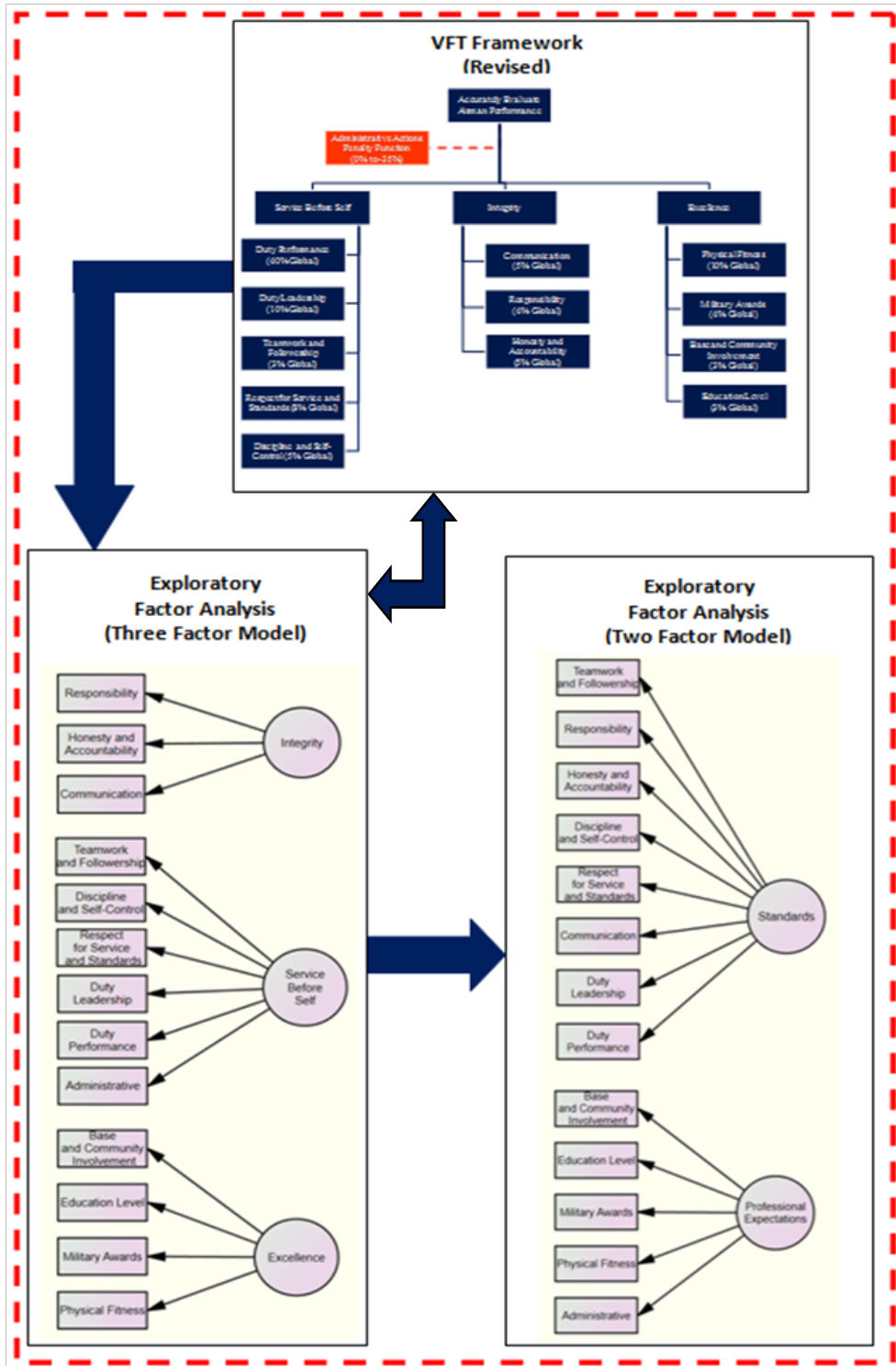


Figure 19. Overview of the Model Validation Chapter

Sensitivity Analysis (Notional Dataset)

The Sensitivity Analysis studied the effects on the overall scores for each of the notional airmen based on incremental changes to each of the individual weights which comprised the VFT Weighted Value Model. Since the Administrative Action correction factor is a penalty, independent of the VFT Weighted Value Model, the attribute was not included in this portion of the analysis. The JEPR weighting construct that was determined by the SNCO SMEs is reflected in Table 27.

Table 27. JEPR Weight Assignments Based on SME Importance

SNCO SME JEPR Weighting Assignments		
Attribute Importance to SMEs	Description	Normalized Weight Assignments
1	Duty Performance	0.40
2	Duty Leadership	0.10
3	Physical Fitness	0.10
4	Respect for Service and Standards	0.08
5	Communication	0.05
6	Discipline and Self-Control	0.05
7	Honesty and Accountability	0.05
8	Responsibility	0.04
9	Awards	0.04
10	Teamwork and Followership	0.03
11	Education	0.03
12	Base and Community Involvement	0.03

The goal of the JEPR sensitivity analysis was to verify the accuracy of the VFT Framework by performing small incremental changes to the weighting scheme (Kirkwood, 1996). If the JEPR Framework could consistently yield the same rankings of the notional, airmen, regardless of the value of the particular weight, then the model

would be deemed as an accurate representation of their value structure. However, if the initial value solicitations and swing weighting proved to be too sensitive, where minor changes to the weighting scheme resulted in changes to the notional airmen's ranking, then further work with the SMEs would have to be done to better define the functions and weights of the JEPR VFT Framework.

The team created a Microsoft Excel tool to assist with the sensitivity analysis. The tool provided the ability to study the effects that weight changes had on the overall JEPR scores for the eight notional airmen by changing each weight one at a time. For each particular weight of interest, the tool graphically illustrated how the scoring would change as the weight was increased or decreased throughout the entire range from 0% to 100%. The proportions for all other weights with the model remained within their solicited ratios, as the weight of interest was increased or decreased (Kirkwood, 1996). The use of sensitivity analysis, and the development of the Microsoft Excel Weight Sensitivity Analysis tool, proved invaluable in being able to visually communicate the ramifications of weight changes to the VFT Framework. The SMEs were able to see how weight changes affected the overall results of the scoring, and how the ranking of the notional airmen changed as the weighting scheme was adjusted (Kirkwood, 1996).

Figure 20 shows that personnel A (ranked #1 initially with the weight $W_{DP} = 40\%$) dominated through the majority of the weighted range. Only if the weight of Duty Performance was changed to $W_{DP} < 32\%$, would the overall best performer change from personnel A to personnel B. In the upper end of the weighting range, if the weight of Duty Performance was $W_{DP} > 78\%$, the best performer would change from personnel A to personnel F who was ranked #3 overall initially. This behavior confirmed the intuitions

of the team that a heavily weighted category such as Duty Performance would dominate the overall scoring as the weight was increased.

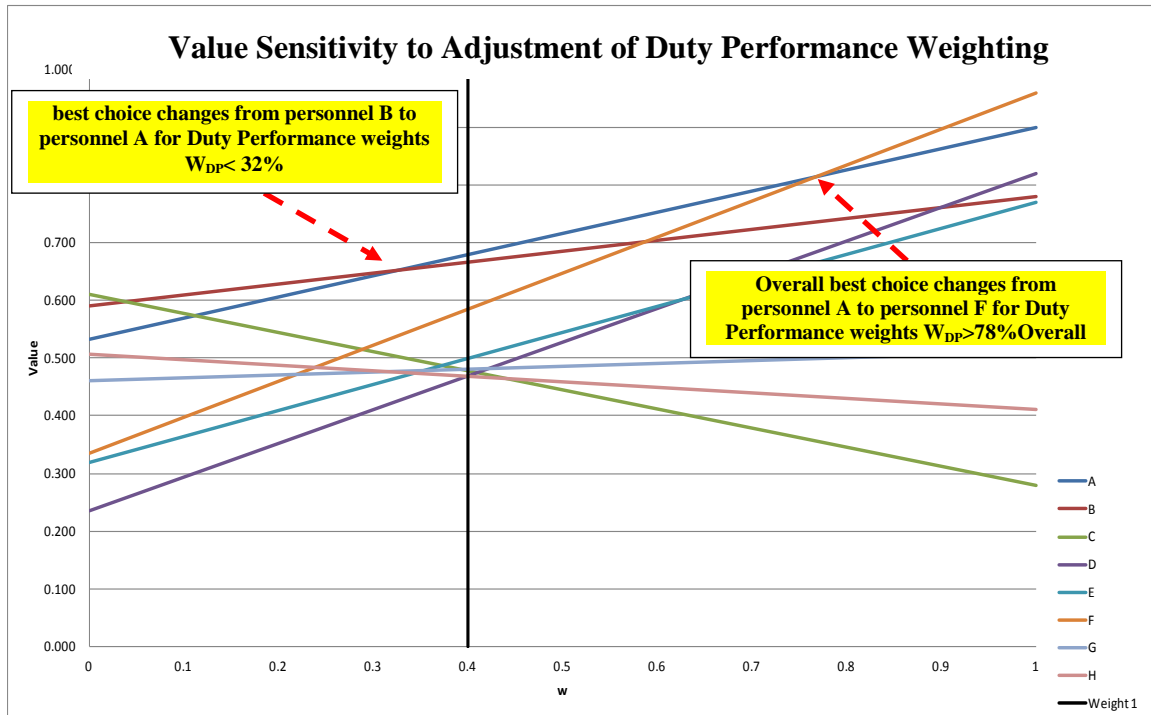


Figure 20. Sensitivity of Duty Performance Weight $W_{DP} < 32\%$ and $W_{DP} > 78\%$

For Duty Leadership, shown in Figure 21, personnel A maintained the best overall performer status through the early portion of the range until the weight was raised to $W_{DL} > 13\%$. After this point, personnel A was supplanted by personnel B, with personnel B being deemed the best overall performer at all Duty Leadership weightings above 13%. This behavior was also witnessed in personnel C, as an increase in weighting of importance of Duty Leadership $W_{DL} > 55\%$ raised personnel C from an initial overall rating of 6th to the second best performing airmen. In the Duty Performance attribute, the minimum change Δ in the weightings construct that would result in a change in the

overall ranking of the notional airmen was 8%. However, for the Duty leadership attribute, the minimum weight change Δ which would change the overall rankings of the notional airmen was only 3%.

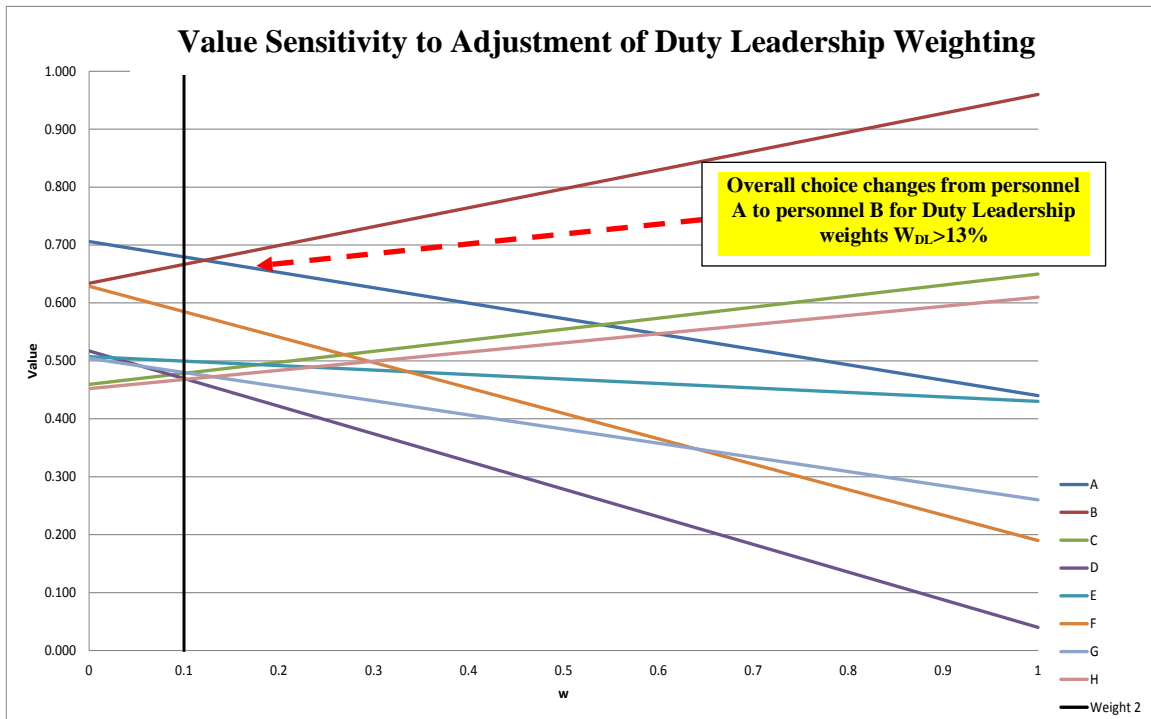


Figure 21. Sensitivity of Duty Leadership Weight $W_{DL} > 13\%$

Although the Duty Leadership was more sensitive than Duty Performance, this sensitivity occurred only above the 13% threshold that had been established by the SMEs. Below the 13% weighting mark, the results were consistent throughout the weighting range, with no changes occurring in the overall rankings for the notional airmen. In discussing the Duty leadership attribute weight, the SMEs conveyed that leadership is considered one of the institutional competencies of the enlisted force structure (Air Force Instruction 36-2618, 2012, p. 3). Because of this, the SMEs felt that a Duty Leadership weighting of less than

10% was unrealistic. However, the SMEs also noted that Air Force Instruction 36-2618 described leadership responsibilities as tiered process, and that junior enlisted members are expected to operate at the tactical level, where primary occupational skills perfected and knowledge of Air Force institutional competencies are developed (Air Force Instruction 36-2618, 2012, p. 3). Therefore, the SMEs felt it was highly unlikely that senior leadership would desire to weight Duty Leadership greater than 13% for junior enlisted airmen, who are expected to operate at the tactical level.

Looking at the Physical Fitness attribute weighting as illustrated in Figure 22, personnel A dominated throughout the entire weight range, with personnel B falling to the 4th best overall best performer at weighting values $W_{PF} > 86\%$.

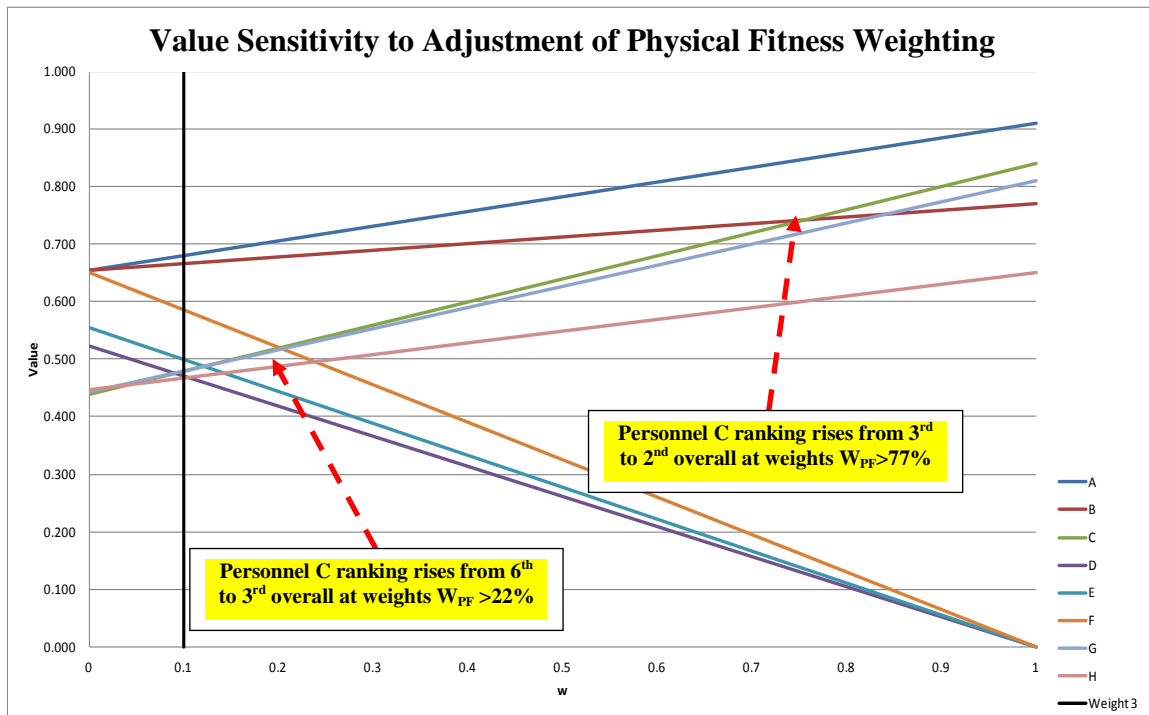


Figure 22. Sensitivity of Physical Fitness Weight $W_{PF} < 22\%$ and $W_{PF} > 77\%$

This was due to a Physical Fitness score that met standards but was at the lower end of the Physical Fitness scoring range with a raw Physical Fitness score of 81. At weights $W_{PF} > 22\%$, personnel C moved from the 6th overall best score to the 3rd overall best score. At even higher weighting values for Physical Fitness where $W_{PF} > 77\%$, personnel B became the 2nd overall best performer among the eight notional airmen.

Looking at the weighted Communication attribute shown in Figure 23, personnel A maintained overall dominance in scoring until the weight was $W_{COMM} > 63\%$, where personnel G overtook personnel A, and became overall the best performer.

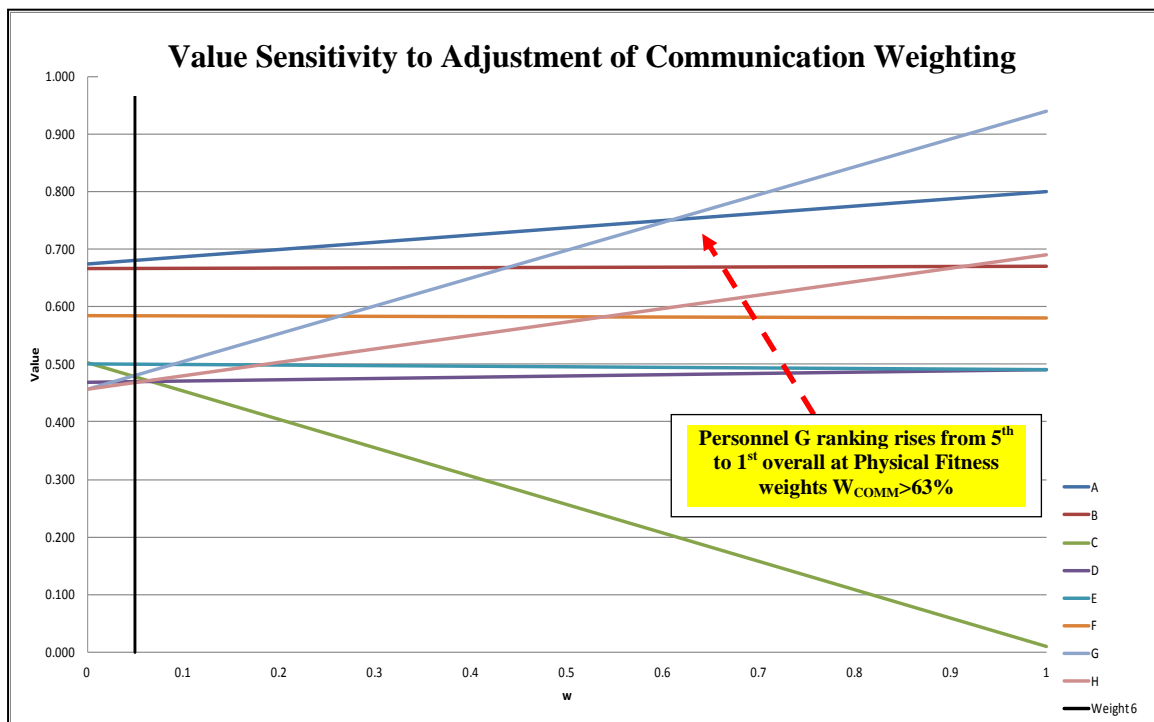


Figure 23. Sensitivity of Communication Weight $W_{COMM} > 63\%$

For the remaining attributes the distance from the baseline weight to the position where a change in the best performer occurred ranged from approximately $\pm 10\%$ to never

changing throughout the range. Sensitivity Analysis for each attribute can be seen in Appendix II of this document.

Although there was some weight sensitivity noticed in the JEPR model, after a lengthy discussion with the SMEs, it was believed that the JEPR accurately captured the desired Value Hierarchy and the stated goals of senior leadership concerning the evaluations. A prime example is the Duty Performance attribute. Although the weight was sensitive at values less than $W_{DP} < 32\%$, it was insensitive at values between $W_{DP} \Rightarrow 32\%$ and $W_{DP} < 78\%$. For Duty Performance, the SMEs felt that the 40% weighting of the attribute accurately reflected Air Force senior leadership goals, and that weighting of importance would likely not to change by more than 5%, regardless of which senior leaders were queried. A strong Duty Performance JEPR weighting is directly in-line with the current Air Force goals, as Air Force senior leadership has continued to state that they desire that Duty Performance be the dominant and discriminating factor in performance appraisals (Losey, Sep 2013).

Data Solicitation Process (Training Dataset)

Using a prototype JEPR database system that has been developed, the team sought to accrue a small sample of data for further refinement and analysis of the proposed JEPR Framework. The group of SMEs generated JEPR reports using the prototype system after closeout of actual performance reports using the current EPR system. The data compiled by using this case by case method was used to further validate the JEPR prototype. This test bed also served as a feedback mechanism to modify the value function and/or weighting schemes of the JEPR model.

The results of the preliminary analysis included 71 preliminary EPRs chosen across eight career fields to serve as a validation, or training dataset for the JEPR model. SNCOs from the eight participating career-fields were asked to score EPRs as usual, and after EPR completion, score the airman using the JEPR program. This was done to prevent bias from entering the actual report. The supervisors also recorded the overall score using the current EPR system after the fact for later comparison with the JEPR outputs. During data collection, no personnel identifying information was collected, only the JEPR scoring results and a record number identifying the career field for the ratee. Supervisors were assigned a pseudo block of phantom identification numbers for creating the case files for analysis. Upon completion of the data collection effort, supervisors sent the data back for compilation and analysis. The goal was to use this training data set to support the primary objectives of this research which were to more accurately capture the true performance for junior enlisted personnel using established management statistical techniques and to confirm the JEPR Framework was congruent with Air Force values, organizational goals, and doctrine. Success would be determined if the JEPR Framework illustrated the ability to delineate between near peers, and the Framework could be aligned with doctrine. Secondary effects such as impacts to promotions and impacts to the future force structure could not be measured at this time.

Qualitative Inspection (Training Dataset)

Using the JEPR Training Dataset that was collected from the eight different participating career fields, the data was studied qualitatively for trends and distribution. The data was exported from the Microsoft Access to Microsoft Excel for analysis. First,

the overall ratings of the 71 test subjects scored under the current EPR system were studied using a histogram. Immediately, it was noticed that 56 of the 71, or 79% of the airman received the maximum score possible, an overall “5” rating, which was described as “Truly Among the Best”. Only 9 of the 71, approximately 12.6% of the airmen were given an overall rating of “4” which equated to “Above Average”. The distribution showing the 71 test subjects evaluated under the current system is shown in Figure 24.

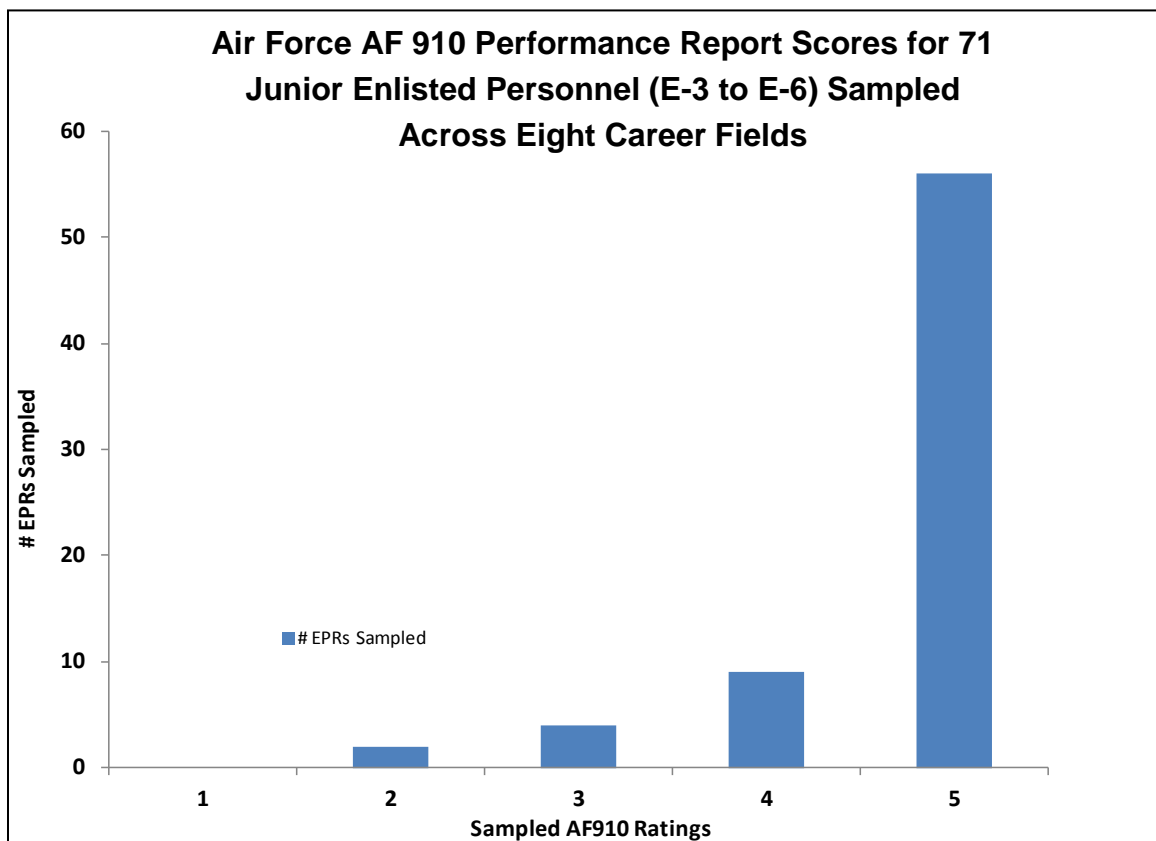


Figure 24. Distribution of 71 Performance Ratings (Current EPR System)

However, looking at a histogram of the same 71 personnel evaluated using the JEPR system in Figure 25; there was clearer delineation among the population. The histogram showed a right skewed mound distribution, with two distinct tails. The right

skewed distribution indicated that the Air Force values high quality personnel who exhibit the traits of leadership, values, and professional qualities, which happen to be same Fundamental Objectives the SMEs had identified for the JEPR model. The mean JEPR score of the population was found to be 72 (out of 100), with a standard deviation of approximately 21. With an alpha of 0.05, with 95% confidence, the mean JEPR score of the population falls between 67 (out of 100) and 77 (out of 100). Again this indicates Air Force’s desire for a junior enlisted core of higher performing individuals.

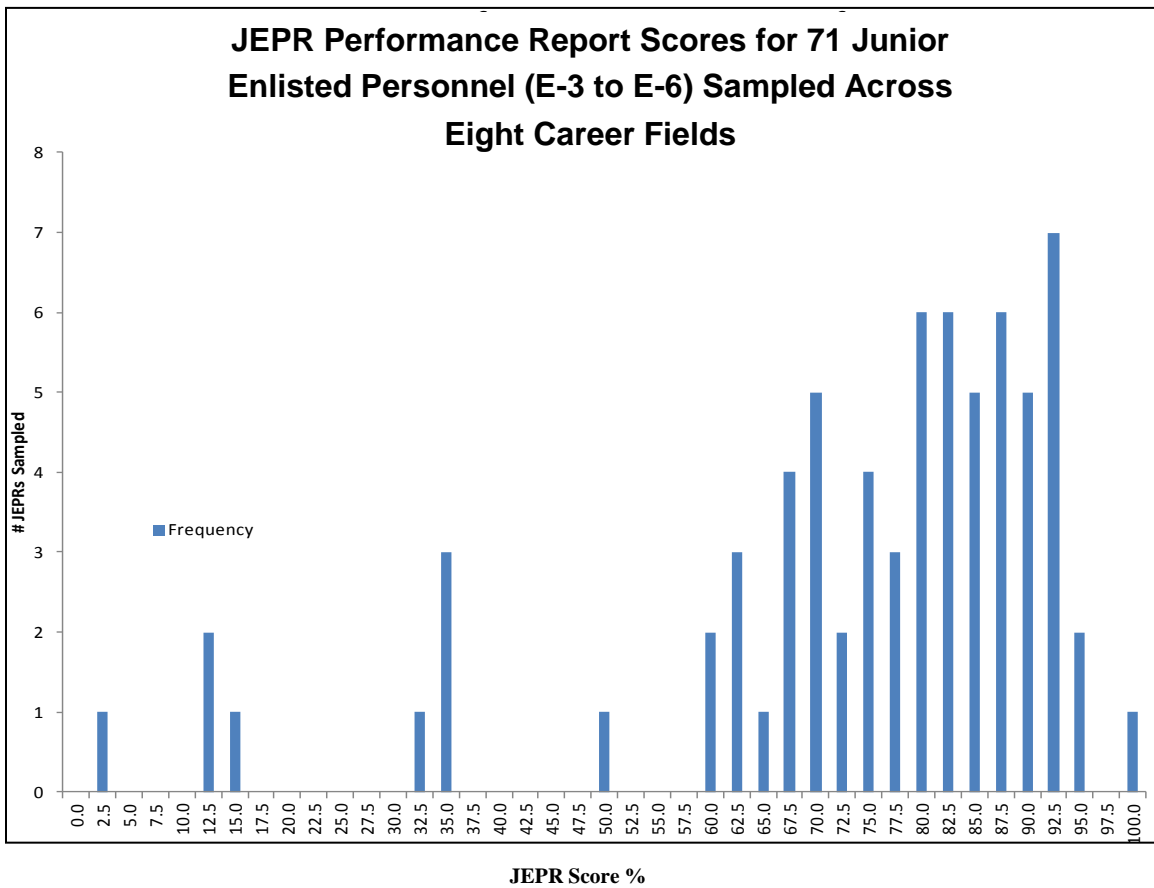


Figure 25. Distribution of 71 Performance Ratings (JEPR System)

The left tail of the distribution was very long and gradual, while the right tail was short and abrupt due to the truncation of the scores at 100. This shape is indicative of the

Johnson SL distribution, which is an empirical logarithmic distribution that is closely related to a normal distribution (Kaplan & Knowles, 2004). The Johnson SL distribution is used for modeling real world data for valuation of commodities (Kaplan & Knowles, 2004). The non-normal behavior of the JEPR training data will have greater importance in Chapter V.

The Shapiro-Wilk W test failed to reject the hypothesis that the distribution was from a Johnson SL distribution with a p-value of 0.2339, meaning we can assume the Johnson SL distribution is suitable for the data. The long left tail, right skewed distribution indicated a wide dispersal for the airman who scored lower than the population concentration by the JEPR. Examination of the scoring and JEPR comment bullets for these test subjects indicated disciplinary actions had occurred and been recorded; the test subject had failed to meet standards, or had exhibited low evaluation numbers in the heavily weighted categories of Performance in Primary Duties or Duty Leadership as observed and recorded by the supervisor. The short right tail indicated that for performers above the concentration of the population, lesser weighted factors provided delineation of outstanding performers. This was confirmed after review of the individual category scores and supervisor performance comments. Therefore, from a qualitative standpoint, delineation can be achieved using the JEPR program with the ability to separate near-peers based on all factors considered under the value hierarchy.

Further qualitative analysis narrowed the scope of the study and looked only at 56 test subjects who were rated as overall “5s”, “Truly One of The Best” under the current system. Study of this sub-population using the JEPR program again illustrated a Johnson SL distribution with a long left tail and a short right tail. This sub-population that had

been scored as “Truly Among the Best” under the current system had JEPR scores that were concentrated between 60 to 95, with a mean of 79. This was approximately 7% higher than the mean of the JEPR scores for overall population, indicating that the “Truly One of The Best” sub-population as a whole appeared to be better performers. Delineation occurred in this sub-population, and it was possible to delineate performance between near-peer test subjects. The observations are shown in Figure 26.

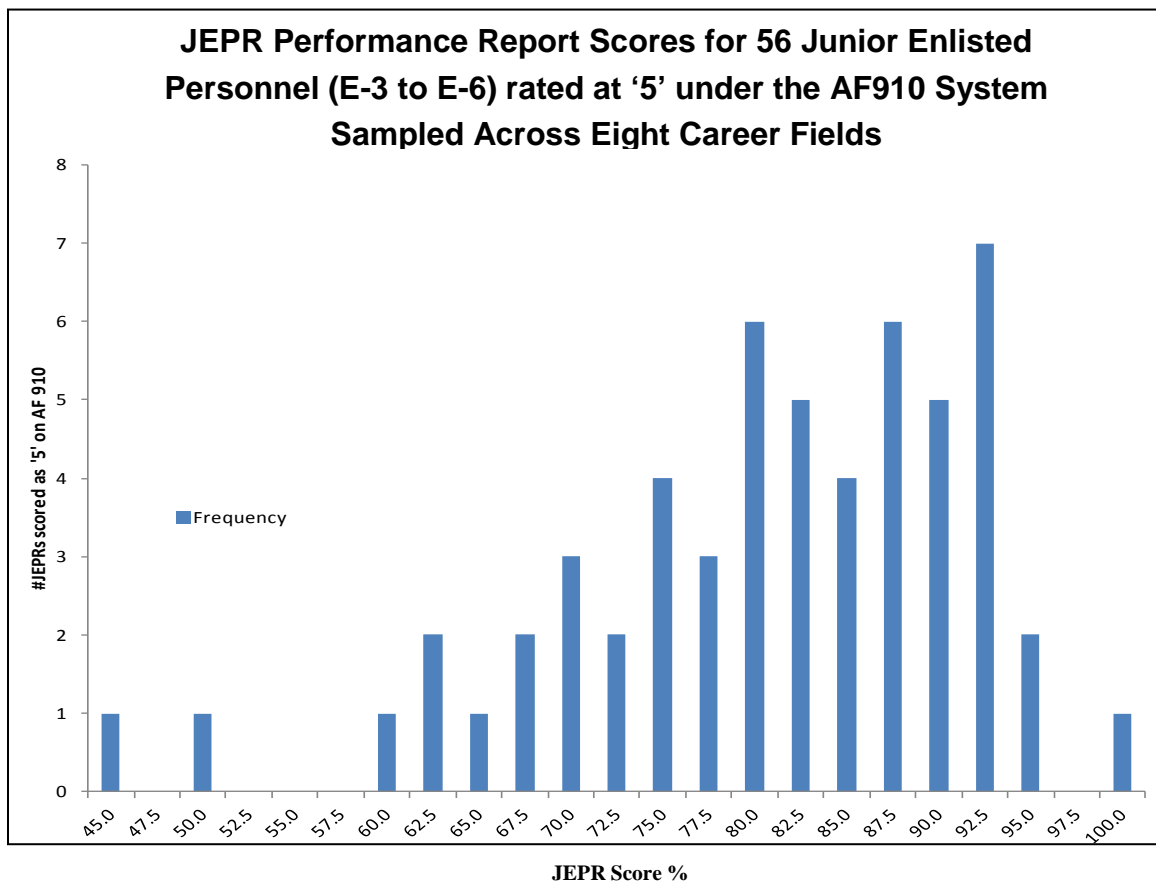


Figure 26. JEPR Distribution Ratings for Subjects Rated “5” (Current EPR System)

The standard deviation was found to be 12.31, which was very high. There were two low scoring data points in the left tail noticed when inspecting the distribution. Further

analysis of these test subjects revealed that although they were stellar performers in almost all categories, the supervisor had assigned very low scores to the heavily weighted Duty Performance and Duty leadership categories, thus impacting the score. Supervisor comments of the report confirmed the accuracy of the markings as the individuals had issues with upgrade training and on the job performance. After exclusion of these two points, the mean was determined to be approximately 80.5, with a standard deviation of 9.78. From further study of ratings versus comments, it was concluded that the JEPR ranges, being weighted, were capturing the value structure that the SNCO SMEs had developed as to what qualities they thought were more important in defining a high performing airman. These initial results highlight the ability to delineate among near-peer performers, consistent with doctrine and SME values.

Internal Consistency (Training Dataset)

In line with current psychometric trends, Cronbach's Alpha was used for testing the internal consistency of the JEPR model (Tavako et al., 2011). Internal consistency, in psychometric terms, means that when items are used to form a measurement scale, such as a JEPR attribute, the items should be correlated with each other, and should all measure the same thing (Bland & Altman, 1997). The rationale for the selection of Cronbach's alpha for model validation was that the JEPR program was developed using Likert-Type scales, with four defined ratings categories, with each possessing bounded internal ranges for scoring an individual in each attribute category. Cronbach's alpha, when used to measure internal consistency, verifies the quality of a Likert-Type scale by evaluating the internal consistency between the scale or test attributes (J. Gliem & R.

Gliem, 2003). A scale exhibiting a high Cronbach's alpha score ensures that all items are measuring the same metric, and therefore should be correlated to one another (Bland & Altman, 1997). The closer Cronbach's alpha coefficient is to 1.0, on a measurement scale from 0 to 1.0, the greater the internal consistency of the items in the scale. Equation 9 illustrates the raw Cronbach's alpha formula for computing internal consistency. Looking closer at Equation 9, K represents the total number of attributes in the JEPR model ($K = 13$), i is the number of the attribute being summed, $\sum_{i=1}^K \sigma_{attribute_Scores_i}^2$ represents the sum of the variance in the scores for i JEPR attributes, and $\sigma_{JEPR_Overall_Scores}^2$ represents the variance of all K JEPR overall scores.

$$\alpha = \frac{K}{K - 1} \left(1 - \frac{\sum_{i=1}^K \sigma_{attribute_Scores_i}^2}{\sigma_{JEPR_Overall_Scores}^2} \right) \quad (9)$$

According to George and Mallery, as cited by (J. Gliem & R. Gliem, 2003), Table 28 provides the basic rules for determining the quality of the Cronbach's alpha value.

Table 28. Cronbach's Alpha Value Quality for Internal Consistency

Cronbach's α Value	Description
≥ 0.9	Excellent
≥ 0.8	Good
≥ 0.7	Acceptable
≥ 0.6	Questionable
≥ 0.5	Poor
< 0.5	Unacceptable

Because each JEPR attribute consists of a scale, and the entire VFT hierarchy of the JEPR model consists of a series of scales, Cronbach's alpha coefficient was deemed an

appropriate measure for validating the internal consistency of the JEPR model rating scales and attributes (J. Gliem & R. Gliem, 2003). For the JEPR training set data, the raw Cronbach's alpha was 0.7864. This value was deemed as an "acceptable" alpha value for measuring internal consistency, and approached the "good" range as defined by George and Mallery with only 71 test points. In 2006, Helms et al., as cited by Spiliotopoulou, noted that increasing the number of participants measured by a scale can increase the value of Cronbach's alpha, as adding participants increases the amount of covariance among responses (Spiliotopoulou, 2009). Therefore, it is expected that the Cronbach's alpha value will increase during the analysis of the JEPR Test Dataset in Chapter V, where approximately 150 test subjects will be appraised.

Because the Administrative Actions correction factor is highly correlated to several other attributes within the JEPR program, it had to be included in the test for internal consistency. Although the Administrative Actions correction factor uses a different numeric scale than the other attributes (-100 to 0), the orientation remains the same, as it counts upward. The Administrative Actions correction factor is not a negatively scaled (inverted values). This attribute is bidirectional, just as the other JEPR attributes, except that the scale resides on the negative side of the value axis. As with the other JEPR attributes, as the supervisors value increases, the ratings categories of the Administrative Actions correction factor also increase in value from left to right, with the numerical values that can be assigned in the categories also increasing. The JEPR bidirectional scaling scheme is illustrated in Figure 27.

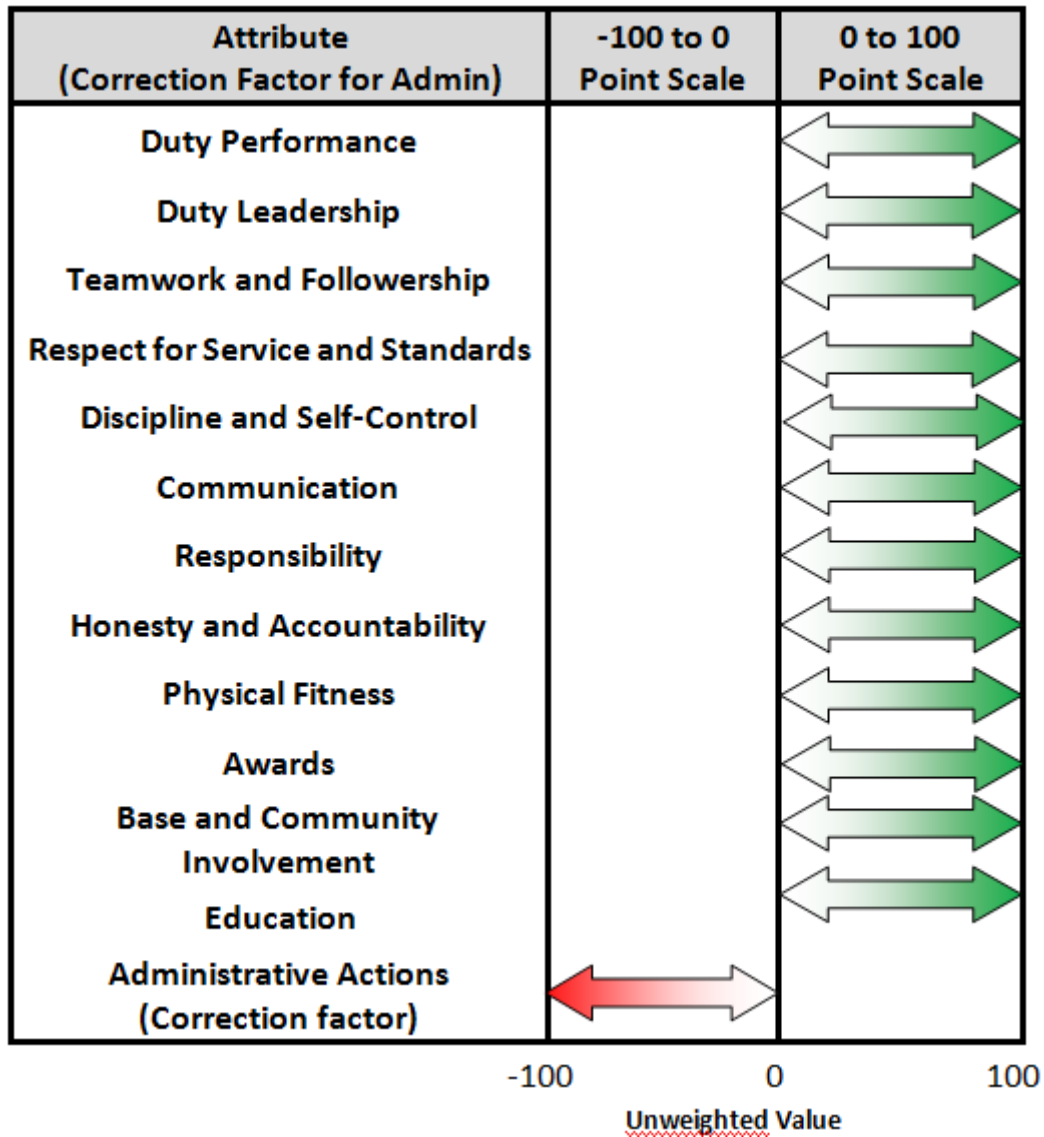


Figure 27. JEPR Bidirectional Scaling Scheme (Increasing Value to the Right)

The simple statistics computed by the JMP Software (JMP 11.0, 2013) for the JEPR Training Dataset, shown in Table 29, illustrated the negative mean generated by the Administrative Actions correction factor.

Table 29. JMP Generated Statistics for JEPR Data

JEPR Training Data Multivariate Simple Statistics							
Column	N	DF	Mean	Std Dev	Sum	Minimum	Maximum
Duty Performance	71	70	0.3159	0.0753	22.4262	0.0817	0.4000
Duty Leadership	71	70	0.0735	0.0230	5.2168	0.0000	0.1000
Physical Fitness	71	70	0.0859	0.0212	6.0970	0.0000	0.1000
Communication	71	70	0.0369	0.0109	2.6234	0.0000	0.0500
Respect for Service and Standards	71	70	0.0606	0.0156	4.3024	0.0120	0.0800
Discipline and Self-Control	71	70	0.0375	0.0115	2.6619	0.0000	0.0500
Honesty and Accountability	71	70	0.0393	0.0133	2.7916	0.0000	0.0500
Responsibility	71	70	0.0314	0.0095	2.2272	0.0000	0.0400
Teamwork and Followership	71	70	0.0241	0.0063	1.7102	0.0023	0.0300
Military Awards	71	70	0.0202	0.0120	1.4367	0.0000	0.0400
Education Level	71	70	0.0145	0.0094	1.0302	0.0000	0.0300
Base and Community Involvement	71	70	0.0151	0.0080	1.0713	0.0000	0.0300
Administrative (Correction Factor)	71	70	-0.0293	0.0701	-2.0773	-0.2835	0.0000

The negative mean was expected, because the Administrative Actions correction factor is a negative quality indicator, and resided on the negative side of the value axis. As Cronbach’s alpha is effectively a variance determined measure, the negative mean of the Administrative Actions did not affect the Cronbach’s alpha computation

Different variants of Cronbach’s alpha were considered for reporting internal consistency. However, they were rejected after closely examining the JEPR construct. First, the JEPR model design relies on summed attribute scores to yield an overall JEPR score. These scores are raw and are not standardized. Second, within each JEPR attribute, a unique sub-scale is utilized to measure only that specific trait that has a unique variance

and unique standard deviation (Cortina, 1993). Therefore it was determined for proper estimation of the internal consistency for the JEPR model, the raw Cronbach's alpha was the best suited measure for reporting internal consistency, as the raw Cronbach's measure accounts for differences in variance between items, and is appropriate for non-standardized data (Cortina, 1993; J. Gliem & R. Gliem, 2003).

Looking at the raw Cronbach's alpha outputs by attribute in Table 30, revealed that when an attribute is excluded, the overall Cronbach's alpha value changed only by a minimum of 0.0007, or a maximum of 0.0445. This not only confirmed that internal consistency existed for all the measures in the entire JEPR model, but that internal consistency of the measures also existed between attributes, with very little variation in the overall alpha value if one attribute was excluded.

Table 30. Raw Cronbach's Alpha Measures (Overall and with Excluded Attributes)

JEPR Model Cronbach's α	
Entire Set	α Value
Overall	0.7864
Excluded Column	
	α
Duty Performance	0.7660
Duty Leadership	0.7419
Physical Fitness	0.7820
Communication	0.7746
Respect for Service and Standards	0.7608
Discipline and Self-Control	0.7721
Honesty and Accountability	0.7806
Responsibility	0.7749
Teamwork and Followership	0.7795
Military Awards	0.7746
Education Level	0.7778
Base and Community Involvement	0.7871
Administrative (Correction factor)	0.7573

Although Cronbach's alpha is a good indicator of internal consistency for the items within a scale, it does not necessarily indicate that the measurement scale is unidimensional (J. Gliem & R. Gliem, 2003). Having unidimensionality means that the scale is measuring the same underlying concept (J. Gliem & R. Gliem, 2003). Factor analysis is one technique that can be used to help determine the dimensionality of a scale (J. Gliem & R. Gliem, 2003). The use of factor analysis is a logical step in the validation process for the JEPR model, as factor analysis has long been used in validation exploration and validation in psychological research (Fabrigar et al., 1999; Worthington & Whittaker, 2006). However, before factor analysis techniques can be applied, suitability tests must be performed on the JEPR data to ensure the model construct is sound and acceptable for further analysis.

Factor Analysis Suitability (Training Dataset)

To begin the suitability tests, an initial correlation matrix was generated using the data matrix generated from the 13 JEPR attributes of all 71 JEPR training data observations. The correlation matrix was chosen for the analysis instead of the covariance matrix because the Administrative correction factor data had been negatively scaled while all other JEPR attributes were positively scaled. To create the correlation matrix, a Sum of Squares for each of the attribute columns was generated from the data matrix columns to create the elements $SS_{(j,k)}$ of the correlation matrix. Equation 10 shows the Sum of Squares computation formula for the correlation matrix.

$$SS_{(j,k)} = \sum_{i=1}^n (x_{i,j} - \bar{x}_j)(x_{i,k} - \bar{x}_k) \quad (10)$$

Using Equation 10, the elements for correlation matrix R were generated for the JEPR Training Dataset. The initial correlation matrix structure is illustrated in Equation 11.

$$R = \begin{bmatrix} \frac{SS_{(1,1)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(1,1)}})]} & \frac{SS_{(1,2)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(2,2)}})]} & \dots & \frac{SS_{(1,13)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(13,13)}})]} \\ \frac{SS_{(2,1)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(1,1)}})]} & \frac{SS_{(2,2)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(2,2)}})]} & \dots & \frac{SS_{(2,13)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(13,13)}})]} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{SS_{(13,1)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(1,1)}})]} & \frac{SS_{(13,2)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(2,2)}})]} & \dots & \frac{SS_{(13,13)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(13,13)}})]} \end{bmatrix} \quad (11)$$

After generating the correlation matrix, the first test for factor analysis suitability that was performed was the Kaiser-Meyer-Olkin (Kaiser, 1970) test. The Kaiser-Meyer-Olkin (KMO) test was used as an index to measure sampling adequacy. In essence KMO is a measure of the strength of the relationship among variables (Williams et al., 2012). The KMO formula is shown in Equation 12.

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} u_{ij}^2} \quad (12)$$

The $\sum_{i \neq j} r_{ij}^2$ term is the sum of the squares, not including the diagonal elements, for all attributes from the initial correlation matrix. The $\sum_{i \neq j} u_{ij}^2$ term is the sum of the squares, not including the diagonal elements, for all attributes from the partial correlation matrix. The R matrix is inverted to yield the R^{-1} inverse correlation matrix, which is then used to compute the partial correlation matrix. The individual partial correlations reflect a

measure of the strength of the relationship between two variables, with the effects of other variables controlled (JMP 11, 2013).

The KMO index ranges from values of 0 to 1.0 and compares the magnitudes of the observed correlation coefficients to the magnitudes of the partial correlation coefficients (Williams et al., 2012). If the sum of the squared partial correlations, $\sum_{i \neq j} u_{ij}^2$ are large when compared to the sum of the squared correlations, $\sum_{i \neq j} r_{ij}^2$, then the KMO index value will be near 0, indicating the correlations are widely spread across many variables, and are not clustering on a small number of variables (Leung, Wong, Ko, Lam, & Fok, 2005; A. Trappey, C. Trappey, Wu, & Lin, 2012). If the sum of the squared partial correlations, $\sum_{i \neq j} u_{ij}^2$, are small when compared to the sum of the squared correlations, $\sum_{i \neq j} r_{ij}^2$, then the KMO index value will be near 1, indicating the correlations are clustering on a small number of variables and that the data is suitable for factor analysis (Leung et al., 2005; Trappey et al., 2012; Williams et al., 2012). For factor analysis, a value of 0.5 or greater is considered suitable for factor analysis (Williams et al., 2012). For the JEPR model training data, the KMO index value was generated using the SPSS software (SPSS 18.0, 2009). SPSS computed a KMO index value of 0.862, which was categorized as “meritorious”, far exceeding the 0.5 threshold for factor analysis consideration as detailed by Kaiser (Hutcheson & Sofroniou, 1999, p. 225).

The second suitability test to be performed on the JEPR training data was the Bartlett’s Test of Sphericity (Bartlett, 1950). This test verified that the correlation matrix of the JEPR data was not an identity matrix, and that correlation existed between the attributes (Maciel et al., 2013; Merkle, Layne, Bloomberg, & Zhang, 1998). If correlation

was not present between the variables, then attributes are completely unrelated, and factor analysis is not possible (Maciel et al., 2013; Merkle et al., 1998). To perform the Bartlett's Test of Sphericity, a hypothesis test was used to determine the probability that the JEPR training data is an identity matrix and is completely uncorrelated. The hypothesis test used a Bartlett's value which was an approximation of the Chi-Square distribution. The Bartlett's value was computed using the number of observations in the JEPR data, the number of attributes (variables) that comprised the data, and the determinant of the correlation matrix for the data (Maciel et al., 2013). The Bartlett's value was then compared against a Chi-Square test statistic value which was based on a predetermined alpha level for hypothesis for acceptance or rejection (Maciel et al., 2013). For the JEPR Training Dataset, the null hypothesis was that the data was completely uncorrelated and unsuitable for factor analysis. The significance level for acceptance of the null hypothesis was set at $\alpha = 0.05$. The hypothesis test for the Bartlett's Test of Sphericity is shown in Equation 13. For the JEPR data, the significance p-value for the Bartlett's Test of Sphericity was very small ($7.06394E - 82$) and was well below the significance threshold of 0.05, indicating that the JEPR training data was not completely uncorrelated, and was suitable for factor analysis (Merkle et al., 1998; Williams et al., 2012).

Alternatives

H_0 : JEPR Data is an Uncorrelated Identity Matrix
 H_a : JEPR Data IS NOT Uncorrelated Identity Matrix

Assumptions

$$\alpha = 0.05$$

$$df = \frac{(\#Attributes^2 - \#Attributes)}{2} = \frac{(13^2 - 13)}{2} = 78$$

Test Statistics

$$\begin{aligned} \text{Bartlett's } s &= \left[-1 * \left[(\#Observations - 1) - \frac{((2 * \#Attributes) + 5)}{6} \right] * \text{Ln}|R| \right] \\ &= \left[-1 * \left[(70) - \frac{((2 * 13) + 5)}{6} \right] * -9.28158462 \right] \\ &= 601.771 \approx \chi^2 \end{aligned}$$

$$\chi^2_{(1-\alpha,df)} = \chi^2_{(0.95,78)} \approx 99.616$$

Decision Rule

if Bartlett's $s \leq \chi^2_{(0.95,78)}$ conclude H_0

if Bartlett's $s > \chi^2_{(0.95,78)}$, conclude H_a

Conclusion

$$601.771 > 99.616$$

\therefore conclude H_a with P - Value of $7.06394E - 82$ (13)

Preliminary Analysis (Training Dataset)

With the data now deemed suitable for preliminary analysis, the next step in the analysis process was to extract the eigenvalues and the eigenvectors from the correlation matrix that was previously generated. Once extracted, the eigenvalues were formed into a single diagonal matrix, while the eigenvectors were captured in a separate matrix. The

characteristic equation shown in Equation 14 was used to extract the eigenvalues and eigenvectors from the correlation matrix

$$\det(R - \lambda I_{13}) = \begin{bmatrix} R_{(1,1)} - \lambda & R_{(1,2)} & \cdots & R_{(1,13)} \\ R_{(2,1)} & R_{(2,2)} - \lambda & \cdots & R_{(2,13)} \\ \vdots & \vdots & \ddots & \vdots \\ R_{(13,1)} & R_{(13,2)} & \cdots & R_{(13,13)} - \lambda \end{bmatrix} \quad (14)$$

Using Equation 14, JMP generated the initial eigenvalues and eigenvectors from the JEPR program training data sample using the JMP software. Kaiser's method was used in the preliminary analysis to initially study how many factors to retain. Kaiser's method, as cited by Zwick & Velicer recommended that the number of components or factors for retention should be equivalent to the number of all eigenvalues that are greater than 1.0 (Zwick & Velicer, 1986). Kaiser, as cited by (Zwick & Velicer, 1986), further explained that the retention of eigenvalues greater than one ensured that nonnegative component reliability existed, as eigenvalues greater than 1.0 possess more summing power in accounting for variance than a single variable. Looking at the JEPR Training Dataset eigenvalues generated from the correlation matrix using Kaiser's method, the first three eigenvectors yielded eigenvalues of 6.5005, 1.7488, and 1.1248. These three vectors accounted for approximately 72.1% of the variation associated with the training model data. The eigenvalues and variance accounted for can be seen explicitly in Table 31.

Table 31. Initial Correlation Matrix [R] Eigenvalues (JEPR Training Dataset)

Eigenvalues of the Initial Correlation Matrix			
Number	Eigenvalue	Percent	Cumulative Percent
1	6.005	50.004	50.004
2	1.7488	13.453	63.456
3	1.1248	8.652	72.108
4	0.7394	5.688	77.796
5	0.5543	4.264	82.060
6	0.5153	3.964	86.024
7	0.4474	3.441	89.465
8	0.3768	2.899	92.364
9	0.3348	2.576	94.940
10	0.2113	1.625	96.565
11	0.2013	1.549	98.114
12	0.1486	1.143	99.257
13	0.0966	0.743	100.000

The Scree test was also studied during the preliminary analysis of the JEPR training data. This is a graphical approach used to confirm the correct number of components or factors that should be retained in a model (Cattell, 1966). The Scree test relies on inspection and interpretation by the analyst to determine the correct number of components or factors to retain, using a graphical plot of the eigenvalues from the initial correlation matrix. The shape of the graph illustrates an area where the eigenvalues begin to equalize and the graph begins to flatten out. This “elbow” area is the point where the variance explanation provided by the eigenvalues decreases dramatically, and provides little benefit for inclusion. For the JEPR model, two elbows were noted, one occurred at the line segment between the second and third eigenvalues, with the second elbow occurring between the third and fourth eigenvalues. These “elbows” can be seen explicitly in Figure 28. If the first three components or factors are retained for the JEPR model, then approximately 72.1% of the variance could be accounted for. Therefore, based on the preliminary

analysis, retention of three components or factors for model seemed intuitive, as the JEPR value hierarchy was constructed from three Fundamental Objectives.

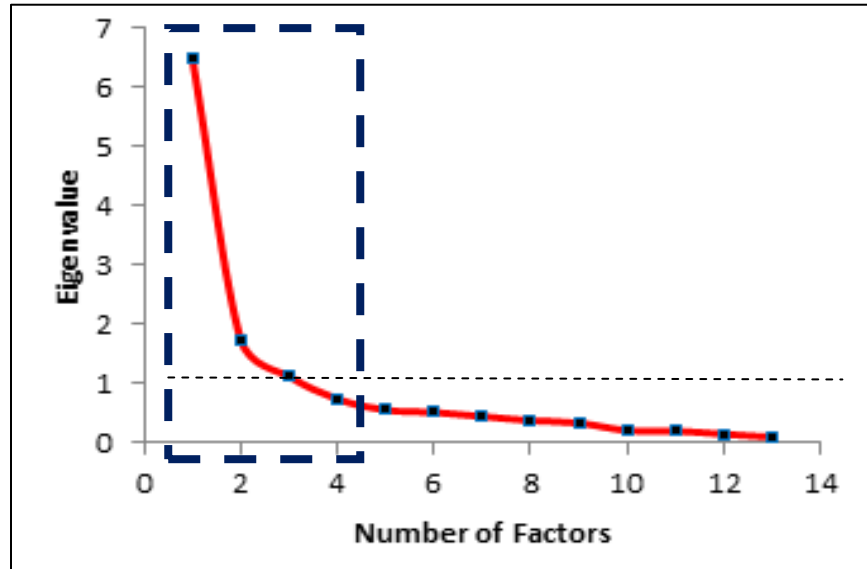


Figure 28. Scree Plot of Initial Eigenvalues from the Initial Correlation Matrix [R]

Data Reduction Technique Selection (Training Dataset)

However, because the goal of this analysis was to validate the underlying construct of the VFT Framework was correct, the correct data reduction technique had to be selected before proceeding. Performing the factor analysis using the Principal Component Method (PCM) was considered inappropriate, as PCM simply strives to explain the variables in a lesser number of factors (Henson & Roberts, 2006). Additionally, PCM tries to maximize the variance explained by the factors, and does not attempt to separate common and unique variances within the attribute (Conway & Huffcutt, 2003).

The Maximum Likelihood Estimation (MLE) method for factor analysis was also considered for the EFA effort. MLE is often used for EFA due to the numerous goodness of indices available and the ability to apply significance testing and confidence intervals to the results (Fabrigar et al., 1999). The downside of using MLE is that it requires that the input data be normally distributed, and if used on non-normal data, will generate distorted results (Fabrigar et al., 1999). However, as Micceri noted, as cited in (Curran, West, & Finch, 1996), the majority of behavioral research data collected is not normally distributed. Since the JEPR construct is founded on measuring observed behavioral data, the use of MLE was deemed inappropriate.

To better explain the underlying construct, the Principal Axis Factoring (PAF) method of factor analysis was chosen for the JEPR project. The PAF method is focused on discovering hidden structures through the explanation of common variance between the variables (Henson & Roberts, 2006). The use of the PAF technique required the ones located on the diagonals of the original correlation matrix to be replaced with estimates of the common variance, which is also known as communality. These estimates represent the proportion of variance in each input variable that is shared with other input variables in the dataset (Henson & Roberts, 2006). The use of communalities more accurately reflects the true variance between variables than does the principal component method (ones on the correlation matrix diagonal) of factor analysis (Conway & Huffcutt, 2003). Figure 29 illustrates the different data reduction techniques available and their relationship.

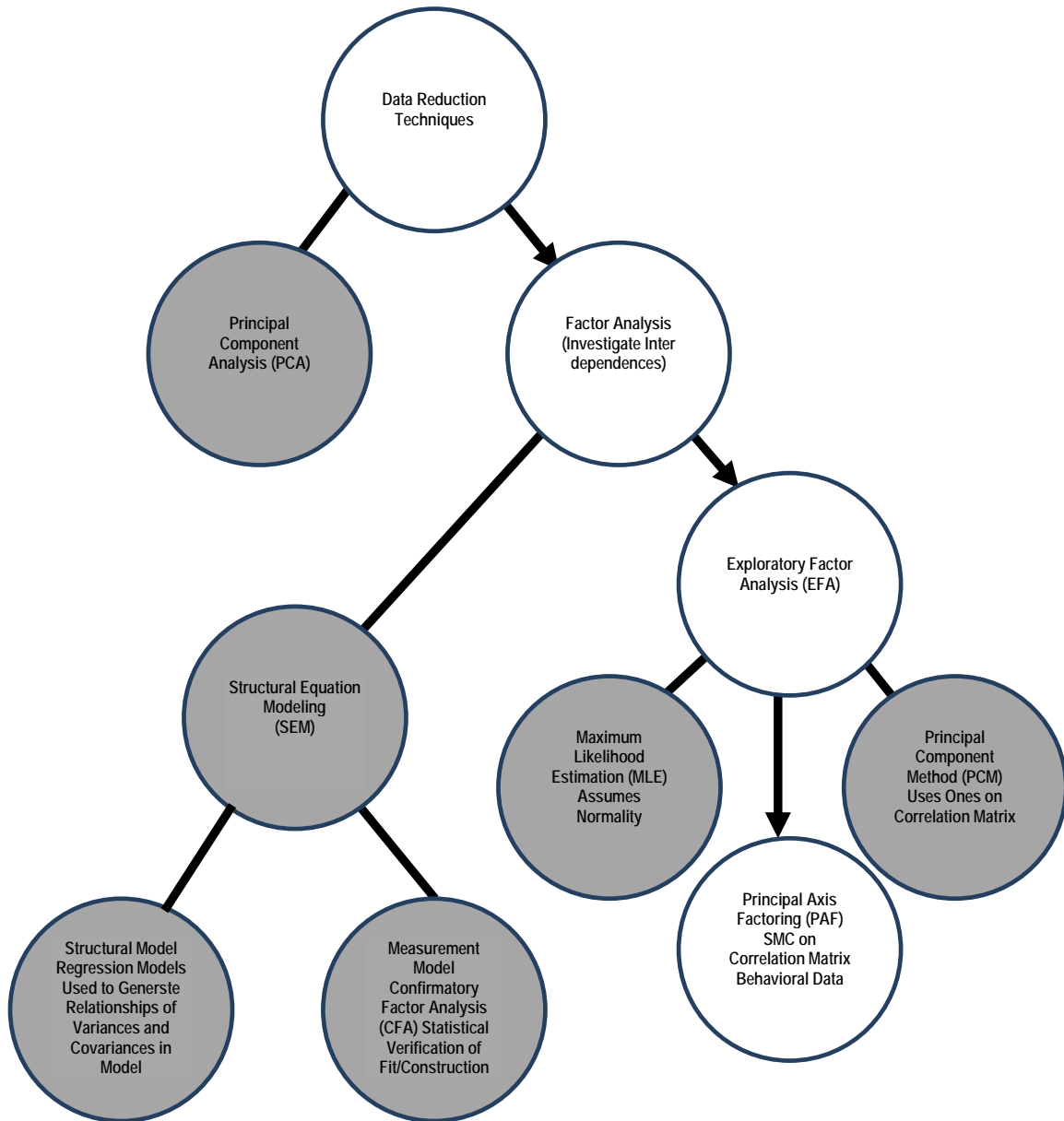


Figure 29. Data Reduction Techniques Tree (EFA Branch Highlighted)

JMP used iterated estimates of the communalities, starting with the Squared Multiple Correlations (SMCs) for each attribute. Iterative methods for estimating communalities are better at fitting the data, and usually stabilize at a consistent value regardless of the starting value (Widaman & Herringer, 1985). The SMC based prior communality

estimates for each attribute were computed using the diagonal elements of the inverse of the initial correlation matrix. The equation for computing the SMC based prior communality estimates for the i th attribute is shown in Equation 15.

$$\hat{h}_i^2 = 1 - \left(\frac{1}{\{R^{-1}\}_{i,i}} \right) \quad (15)$$

Table 32 lists the SMC based prior communality estimates generated by the JMP Software using Equation 15 for the JEPR training data set.

Table 32. SMC Based Prior Communality Estimates (Training Dataset)

Prior Communality Estimates (SMC)	
Attribute	Communality Value
Duty Performance	0.80549
Duty Leadership	0.84591
Physical Fitness	0.36247
Communication	0.73910
Respect for Service and Standards	0.73516
Discipline and Self-Control	0.63411
Honesty and Accountability	0.42433
Responsibility	0.66142
Teamwork and Followership	0.76669
Military Awards	0.56332
Education Level	0.52284
Base and Community Involvement	0.41556
Administrative (Correction Factor)	0.63670

The modified correlation matrix, with the SMC prior communalities on the diagonals, became the reduced correlation matrix R^* as shown in Equation 16.

$$R^* = \begin{bmatrix} 0.80549 & \frac{SS_{(1,2)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(2,2)}})]} & \dots & \frac{SS_{(1,13)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(13,13)}})]} \\ \frac{SS_{(2,1)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(1,1)}})]} & 0.84591 & \dots & \frac{SS_{(2,13)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(13,13)}})]} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{SS_{(13,1)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(1,1)}})]} & \frac{SS_{(13,2)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(2,2)}})]} & \dots & 0.63670 \end{bmatrix} \quad (16)$$

After the SMCs replaced the main diagonals of the initial correlation matrix, JMP iterated back through the modified correlation matrix, extracted new factors, recomputed the communalities again, and placed the recomputed communalities back onto the main diagonal using regression (Floyd & Widaman, 1995). This process continued until the communality estimates stabilized, yielding a final reduced correlation matrix (Floyd & Widaman, 1995). The final communalities are shown in Table 33.

Table 33. Final Communality Estimates (Training Dataset)

Final Communality Estimates	
Attribute	Communality Value
Duty Performance	0.76758
Duty Leadership	0.82973
Physical Fitness	0.31730
Communication	0.78334
Respect for Service and Standards	0.74040
Discipline and Self-Control	0.59436
Honesty and Accountability	0.43005
Responsibility	0.66932
Teamwork and Followership	0.76833
Military Awards	0.60297
Education Level	0.56270
Base and Community Involvement	0.42215
Administrative (Correction Factor)	0.66675

The final reduced correlation matrix is represented in Equation 17 and was used for the remainder of the JEPR training data factor analysis.

$$R^* = \begin{bmatrix} 0.76758 & \frac{SS_{(1,2)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(2,2)}})]} & \dots & \frac{SS_{(1,13)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(13,13)}})]} \\ \frac{SS_{(2,1)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(1,1)}})]} & 0.82973 & \dots & \frac{SS_{(2,13)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(13,13)}})]} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{SS_{(13,1)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(1,1)}})]} & \frac{SS_{(13,2)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(2,2)}})]} & \dots & 0.66675 \end{bmatrix} \quad (17)$$

The new eigenvalues generated from the final reduced correlation matrix which utilized the final communality estimates as the main diagonal entries are reflected in Table 34.

Table 34. Reduced Correlation Matrix [R*] Eigenvalues (JEPR Training Dataset)

Eigenvalues of the Reduced Correlation Matrix	
Number	Eigenvalue
1	6.1907
2	1.2696
3	0.6947
4	0.2723
5	0.1696
6	0.0833
7	0.0516
8	0.0321
9	-0.0640
10	-0.1038
11	-0.1291
12	-0.1490
13	-0.2048

Initial Dimensionality Assessment (Training Dataset)

The dimensionality assessment that had been performed during earlier the preliminary analysis was now no longer valid since the correlation matrix had been

modified into the reduced correlation matrix, which possessed different eigenvalues. As cited by Fabrigar et al., Gorsuch and Horn noted that Kaiser's rule cannot be used to determine the number of factors to retain when communalities are placed on the diagonals of a reduced correlation matrix (Fabrigar et al., 1999). Although Kaiser's rule could not be applied in this situation, it was possible to reduce the dimensionality to some extent by inspecting the eigenvalues of the reduced correlation matrix. Looking closer at Table 34, it was noticed that the eigenvalues of the reduced correlation matrix for factors 9 through 13 were negative. Dillon and Goldstein noted that any factor with a negative eigenvalue also has a corresponding imaginary eigenvector, and cannot contribute to factor analysis (Dillon & Goldstein, 1984, p. 74). Therefore the dimensionality could be reduced from 13 to eight factors, simply by inspection. However, dimensionality could be further reduced.

When the goal is to examine factors that pertain to the study of common variance, a Scree Test generated from the reduced correlation matrix eigenvalues is a viable method for the assessing dimensionality needed for factor analysis (Fabrigar et al., 1999). The Scree Plot of the reduced correlation matrix eigenvalues as shown in Figure 30 illustrated graphically that a drastic difference in contribution existed between eigenvalues three and four. Therefore, factors one, two, and three from the reduced correlation matrix were selected for retention and eigenvalues four through eight were not retained due to their small contributions to the JEPR model in explaining variance (Dillon & Goldstein, 1984, p. 74).

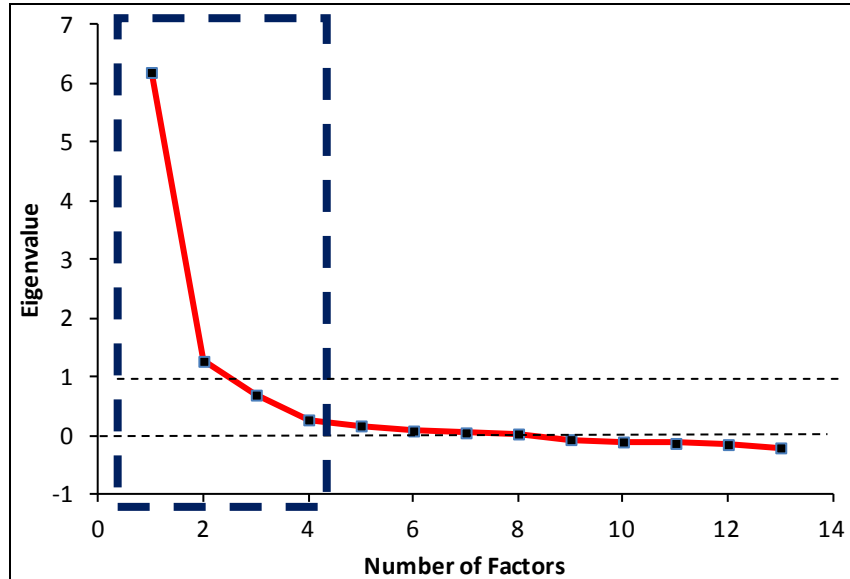


Figure 30. Scree Plot of Eigenvalues from the Reduced Correlation Matrix [R*]

Initial Exploratory Factor Analysis and Interpretation (Training Dataset)

For almost a century, the psychological research community has been using factor analysis as a method for examining interrelationship, data reduction, classification, description of data, data transformation, and hypothesis testing, and mapping construct space (Ford, MacCallum, & Tait, 1986). In an effort to discover the hidden structures explained by common variance, the factor loadings generated during factor analysis are studied and manipulated to provide insight (Henson & Roberts, 2006). Therefore the use of factor analysis procedures, such as loadings analysis and rotation, would allow for a systematic assessment of the JEPR as prescribed by the Applied Psychology field (Ford et al., 1986).

Factor loadings are regression weights generated in a matrix form and reflect the correlations between each original variable and the underlying related factor (DeCoster,

1998). The higher the strength of each loading value, the more relevant the variable is in defining the factor's dimensionality (DeCoster, 1998). The JEPR loadings matrix was created in JMP from a matrix of eigenvectors $\tilde{e}_i'^*$ and a diagonal matrix of eigenvalues λ_i^* from the reduced correlation matrix, where i was the number of factors that were retained. For the JEPR training data, $i = 1, 2, 3$ as only the first three factors were chosen for retention during the dimensionality assessment. Equation 18 illustrates the formula used by JMP for computing the unrotated loadings matrix from the JEPR Training Dataset.

$$A^* = \sqrt{\lambda_i^*} \tilde{e}_i'^* \quad (18)$$

Inspection of the initial unrotated factor loadings indicated that a majority of the variables were heavily loaded on one factor. The groupings of the variables were not intuitive, and did not resemble any recognizable structure tied to Air Force doctrine or otherwise. The original unrotated factor loadings matrix is shown in Table 35.

In an attempt to better interpret the underlying factor structure of the JEPR model, the loadings matrix was rotated. The belief was that after rotation, the attributes would realign under the three factors and reveal a structure that was akin to the Fundamental Objectives of the JEPR Framework. There are two types of rotation methods for factor analysis: oblique rotations and orthogonal rotations.

Table 35. Unrotated Factor Loadings of the JEPR Training Dataset

Unrotated Factor Loading Matrix			
Objective	Factor 1	Factor 2	Factor 3
Duty Performance	0.825072	-0.155601	-0.250256
Duty Leadership	0.889890	-0.051500	-0.187545
Physical Fitness	0.264630	0.454923	-0.200783
Communication	0.771665	-0.251238	0.353202
Respect for Service and Standards	0.831920	-0.121190	-0.183368
Discipline and Self-Control	0.728016	-0.176955	-0.181771
Honesty and Accountability	0.500435	-0.255872	0.337851
Responsibility	0.774218	-0.049021	0.259807
Teamwork and Followership	0.814135	-0.324765	-0.006491
Military Awards	0.581608	0.505750	0.094465
Education Level	0.624878	0.373564	0.180758
Base and Community Involvement	0.301467	0.503163	0.279452
Administrative (Correction Factor)	0.705828	0.323704	-0.252526

An orthogonal rotation redistributes the variance between factors, forcing uncorrelated factor structure (Williams et al., 2012). Oblique rotations, on the other hand, allow correlation to exist between the factors, and is often considered more realistic for behavioral research, (Williams et al., 2012). Ford, Fabrigar and Gorsuch, as cited in (Conway & Huffcutt, 2003), all agreed that an oblique rotation is preferred if the factors are truly are correlated. The use of an orthogonal rotation, where true correlation exists between the factors, can generate an unrealistic factor loadings structure, creating a false interpretation of the factor relationships (Conway & Huffcutt, 2003). However, Floyd and Widaman noted, as cited in (Conway & Huffcutt, 2003), that if there is little to no correlation between the factors, then an orthogonal rotation and an oblique rotation will yield very similar results. Therefore, in conducting the JEPR analysis, both an Oblique Promax rotation and an Orthogonal Varimax rotation were studied for suitability.

During the analysis, all variables with loadings greater than or equal to 0.40 were considered statistically significant. The results for both the oblique and the orthogonal rotations are shown in Table 36 and Table 37.

For each rotation type, the highest loading value for each variable is shown in bold. Surprisingly, both the oblique and the orthogonal rotations aligned the same variables under the three factors used for the factor analysis, with both methods identifying almost the same variables in each factor as significant. The only difference being that the orthogonal rotation identified the Responsibility as relating to both factors one and two, and Teamwork and Followership which was also dispersed between factors one and two. The relaxation of the orthogonality requirement in the oblique rotation allowed for dispersion of the loadings to better align responsibility to only factor two and to also align Teamwork and Followership only to factor one.

The orthogonal rotation was able to account for 62.73% of the variance using only three factors. The variance for the oblique rotation was not computed as the variance cannot be partitioned among factors after an oblique rotation has been applied (Macallum, 1983). Regardless of the rotation method chosen, the loadings of JEPR Training Data variables clearly aligned with a specific factor in a set of three common factors. This supported the intuition that the factors were genuinely uncorrelated, as the orthogonal and the oblique rotations produce almost identical results (Costello & Osborne, 2005). Since both rotations revealed the same loading structure, the orthogonal rotation will be studied first from this point forward for simplicity for all other factor analysis efforts, and then verified against the oblique rotation to ensure consistency.

Table 36. Oblique Rotation Results of JEPR Training Data

Factor Analysis Settings Technique #1 (Oblique)			
Factoring Method	Principal Axis Factoring		
Prior Communality	Common Factor Analysis (SMC)		
Factors Selected	3		
Rotation Method	Oblique Promax		
Significance Threshold	=> 0.4		
Objective	Factor 1	Factor 2	Factor 3
Duty Performance	0.858749	0.066335	-0.044105
Duty Leadership	0.796382	0.106469	0.101831
Physical Fitness	0.209869	-0.330890	0.486384
Communication	0.231483	0.728063	0.027771
Respect for Service and Standards	0.780527	0.123639	0.014894
Discipline and Self-Control	0.732072	0.117369	-0.069088
Honesty and Accountability	0.071979	0.628263	-0.047352
Responsibility	0.253886	0.540255	0.214926
Teamwork and Followership	0.661727	0.395218	-0.151562
Military Awards	0.090240	0.057724	0.708325
Education Level	0.078670	0.221290	0.604635
Base and Community Involvement	-0.284536	0.165319	0.695196
Administrative (Correction Factor)	0.601755	-0.187507	0.437038

Table 37. Orthogonal Rotation Results of JEPR Training Data

Factor Analysis Settings Technique #2 (Orthogonal)			
Factoring Method	Principal Axis Factoring		
Prior Communality	Common Factor Analysis (SMC)		
Factors Selected	3		
Rotation Method	Orthogonal Varimax		
Significance Threshold	=> 0.4		
Objective	Factor 1	Factor 2	Factor 3
Duty Performance	0.811289	0.281011	0.174431
Duty Leadership	0.795622	0.322030	0.304980
Physical Fitness	0.208025	-0.197588	0.484749
Communication	0.425886	0.758026	0.165398
Respect for Service and Standards	0.766929	0.322965	0.218895
Discipline and Self-Control	0.701859	0.293449	0.125071
Honesty and Accountability	0.232819	0.611586	0.042504
Responsibility	0.434501	0.609141	0.330866
Teamwork and Followership	0.696147	0.529325	0.059365
Military Awards	0.252689	0.170673	0.714138
Education Level	0.265352	0.310122	0.629372
Base and Community Involvement	-0.070135	0.174009	0.622054
Administrative (Correction Factor)	0.603617	0.034695	0.548809

This can be seen graphically in Figure 31 and Figure 32. However, further analysis was needed to interpret what these three latent constructs were.

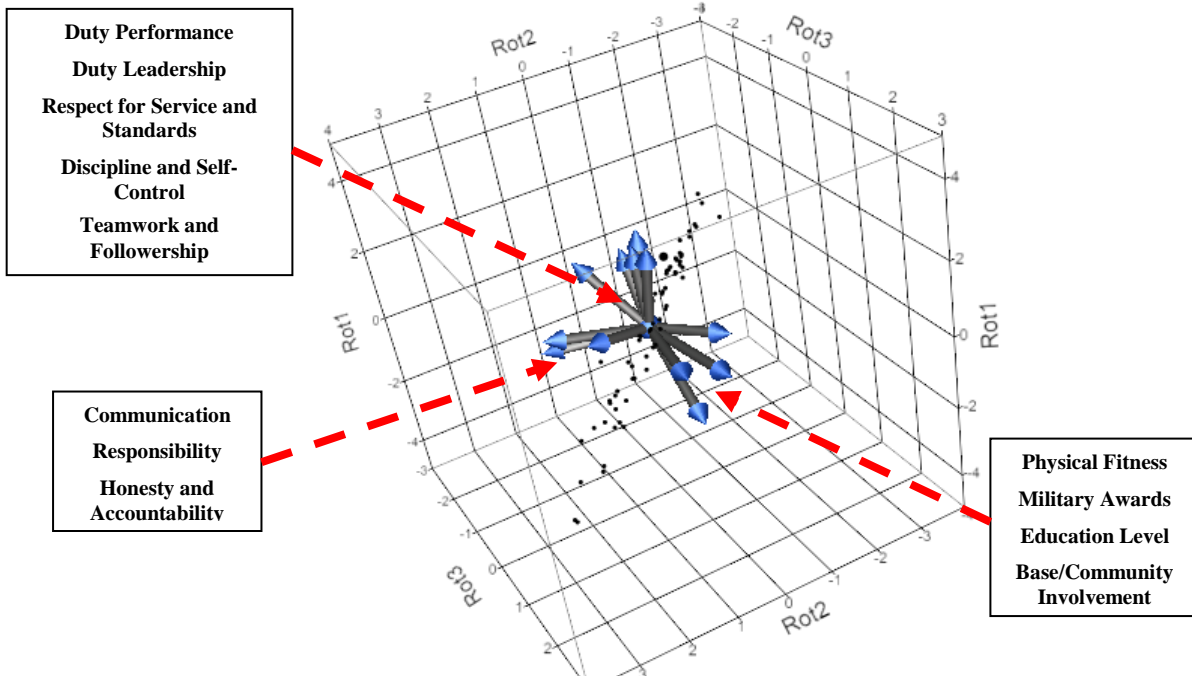


Figure 31. JEPR Reduced Factors after Promax Oblique Rotation

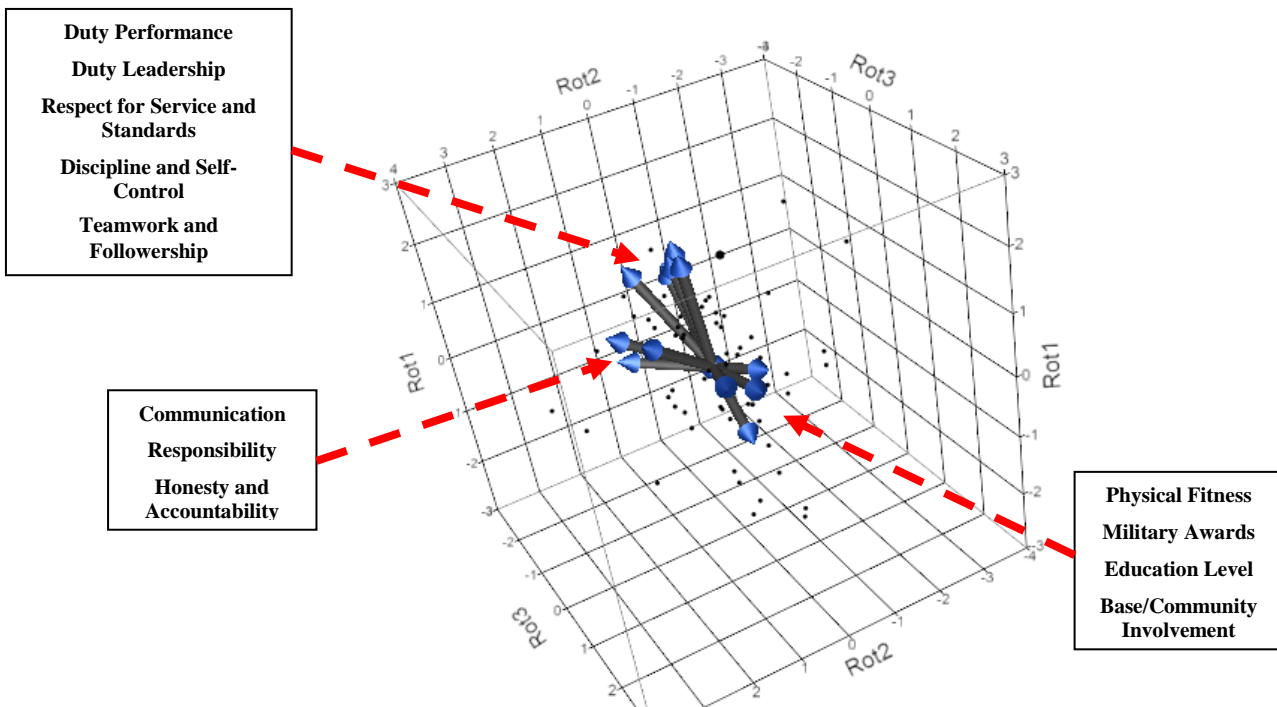


Figure 32. JEPR Reduced Factors after Varimax Orthogonal Rotation

Initially, it was hypothesized that these three factors would be Leadership /Performance in Primary and Additional Duties, Values and Responsibilities, and Professional Qualities, which were the fundamental objectives of the value hierarchy. It was also believed the attributes would align underneath the appropriate Fundamental Objective, as seen in the value hierarchy. However, if factor one was indeed Leadership/Performance in Primary and Additional Duties, then the Duty Performance and Duty Leadership variables were properly associated. Yet the Respect for Service and Standards, Discipline and Self-Control, and Teamwork and Followership variables also aligned underneath factor one, but in the hierarchy, they were associated with Values and Responsibilities, not Leadership/ Performance in Primary and Additional Duties. The incongruency between the variables and common factors continued through factors two and three. A resinspection of Air Force doctrine provided insight to the apparent misalignment of factors and variables with the value hierarchy.

The common factors and variables were indeed not describing the constructed value hierarchy. The factors were found to more closely align with the Air Force core values. This can be intuitively seen by observing that the large factor loading values and factor alignments coincide with specific core value traits in AFD-070906-003, the Air Force Core Values doctrine. Table 38, Table 39, and Table 40 show by factor, the loadings, the factor alignment, and the doctrinal alignment. For this comparison, the orthogonal rotated data was used; however, the oblique rotated data produces the same result as the largest factors identified for each variable are the same as well as the variable alignment with the specific factors.

Table 38. Service Before Self Core Value Relationship to JEPR Common Factor One

JEPR Training Data Rotated Factor Loading (Orthogonal)		
Service Before Self Core Value		
Objective	Factor 1 Loading	Doctrine
Duty Performance	0.811289	<i>"Service before self tells us that <u>professional duties take precedence</u> over personal desires."</i>
Duty Leadership	0.795622	<i>"While it may be the case that professionals are expected to exercise judgment in the performance of their duties, good professionals understand that rules have a reason for being, and the default position must be to follow those rules unless there is a clear, operational reason for refusing to do so." <u>"...if a leader resists the temptation to doubt 'the system', then subordinates might follow suit."</u></i>
Respect for Service and Standards	0.766929	<i>"<u>To lose faith in the system is to adopt the view that you know better than those above you in the chain of command</u> what should or should not be done. In other words, to lose faith in the system is to place self before service."</i>
Discipline and Self-Control	0.701859	<i>"Discipline and self-control. <u>Professionals cannot indulge themselves in self-pity, discouragement, anger, frustration, or defeatism.</u> They have a fundamental moral obligation to the persons they lead to strike a tone of confidence and forward-looking optimism."</i>
Teamwork and Followership	0.696147	<i>"Respect for others. Service before self tells us also that a <u>good leader places the troops ahead of his/her personal comfort.</u> We must always act in the certain knowledge that all persons possess fundamental worth as human beings"</i>

Table 39. Integrity Core Value Relationship to JEPR Common Factor Two

JEPR Training Data Rotated Factor Loading (Orthogonal)		
Integrity Core Value		
Objective	Factor 2 Loading	Doctrine
Communication	0.758026	"Openness. <i>Professionals of integrity encourage a free flow of information</i> within the organization. They seek feedback from all directions to ensure they are fulfilling key responsibilities, and they are never afraid to allow anyone at any time to examine how they do business."
Honesty and Accountability	0.611586	"Honesty. <i>Honesty is the hallmark of the military professional because in the military</i> , our word must be our bond. We don't pencil-whip reports, we don't cover up tech data violations, we don't falsify documents, and we don't write misleading operational readiness messages. The bottom line is we don't lie, and we can't justify any deviation." ... " <i>Accountability. No person of integrity tries to shift the blame</i> to others or take credit for the work of others; "the buck stops here" says it best."
Responsibility	0.609141	" <i>Responsibility. No person of integrity is irresponsible</i> ; a person of true integrity acknowledges his or her duties and acts accordingly."

Table 40. Excellence Core Value Relationship to JEPR Common Factor Three

JEPR Training Data Rotated Factor Loading (Orthogonal)		
Excellence In All We Do Core Value		
Objective	Factor 3 Loading	Doctrine
Physical Fitness	0.484749	"Personal Excellence. <i>Military professionals must...stay in physical and mental shape...</i> "
Military Awards	0.714138	"Excellence in all we do directs us to develop a sustained passion for the <i>continuous improvement and innovation</i> that will propel the Air Force into a long-term, <i>upward spiral of accomplishment and performance.</i> "
Education Level	0.629372	"Personal Excellence. Military professionals must... <i>continue to refresh their general educational backgrounds.</i> "
Base and Community Involvement	0.622054	"Product/service excellence. We must focus on providing services and generating products that fully respond to customer wants and anticipate customer needs, and we must do so within the boundaries established by the taxpaying public."

After reviewing the doctrinal relationships that were uncovered by the factor analysis, it is clear to see that the JEPR value hierarchy is sound in that all Air Force Core Values are covered and each of the JEPR Fundamental Objectives is comprised of at least two Core Values. The overlap of the core values can be explicitly seen in Figure 33.

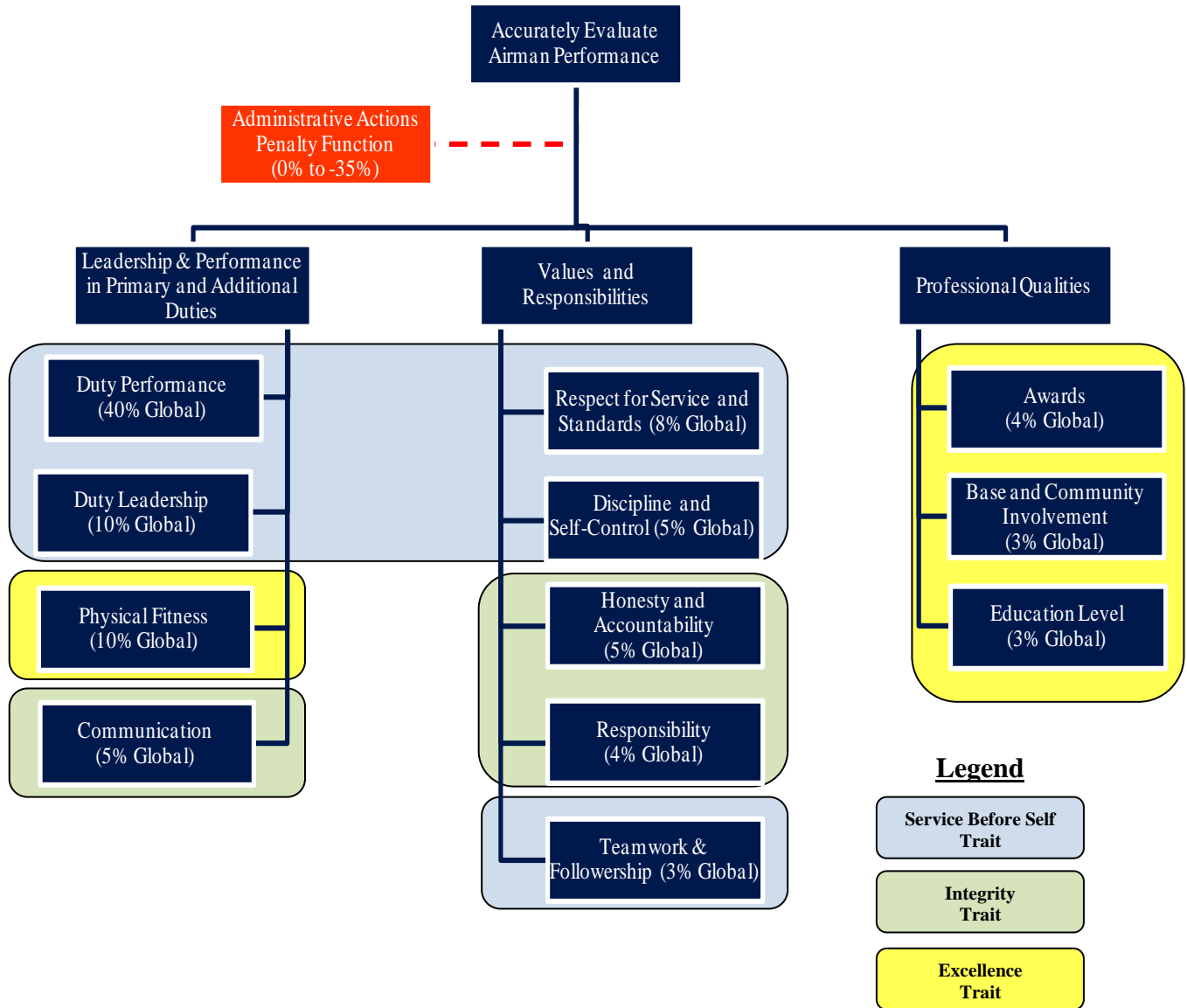


Figure 33. JEPR Value Hierarchy (Core Values Aligned on Rotated Factor Loadings)

JEPR Model Revision (Based on Initial Factor Analysis Findings)

Based on the insight provided by the factor analysis and rotation, the value hierarchy was reconstructed and aligned under the Air Force Core Values doctrine, Air Force Directive 070906-003. This was possible due to the global weighting scheme used by the JEPR VFT Framework. The global weighting scheme provided flexibility to the

SMEs during the redesign, as the attributes of a globally weighted construct independently assess the importance of the attribute to the overall VFT Framework, rather than requiring the SMEs to make tradeoffs among different categories using local scales (Monat, 2009). If local scales had been used, the local weighting values that were assigned to each Fundamental Objective would had to have been redistributed for each weight moved. Only the fundamental objectives were renamed and components realigned based on the results of the loadings matrices. No weights were changed from the initial global weighting scheme originally solicited from the SNCO subject matter experts. Figure 34 shows the revised value hierarchy. Appendix III through Appendix V show the value breakouts and value gaps for the eight notional airmen after attribute reorganization.

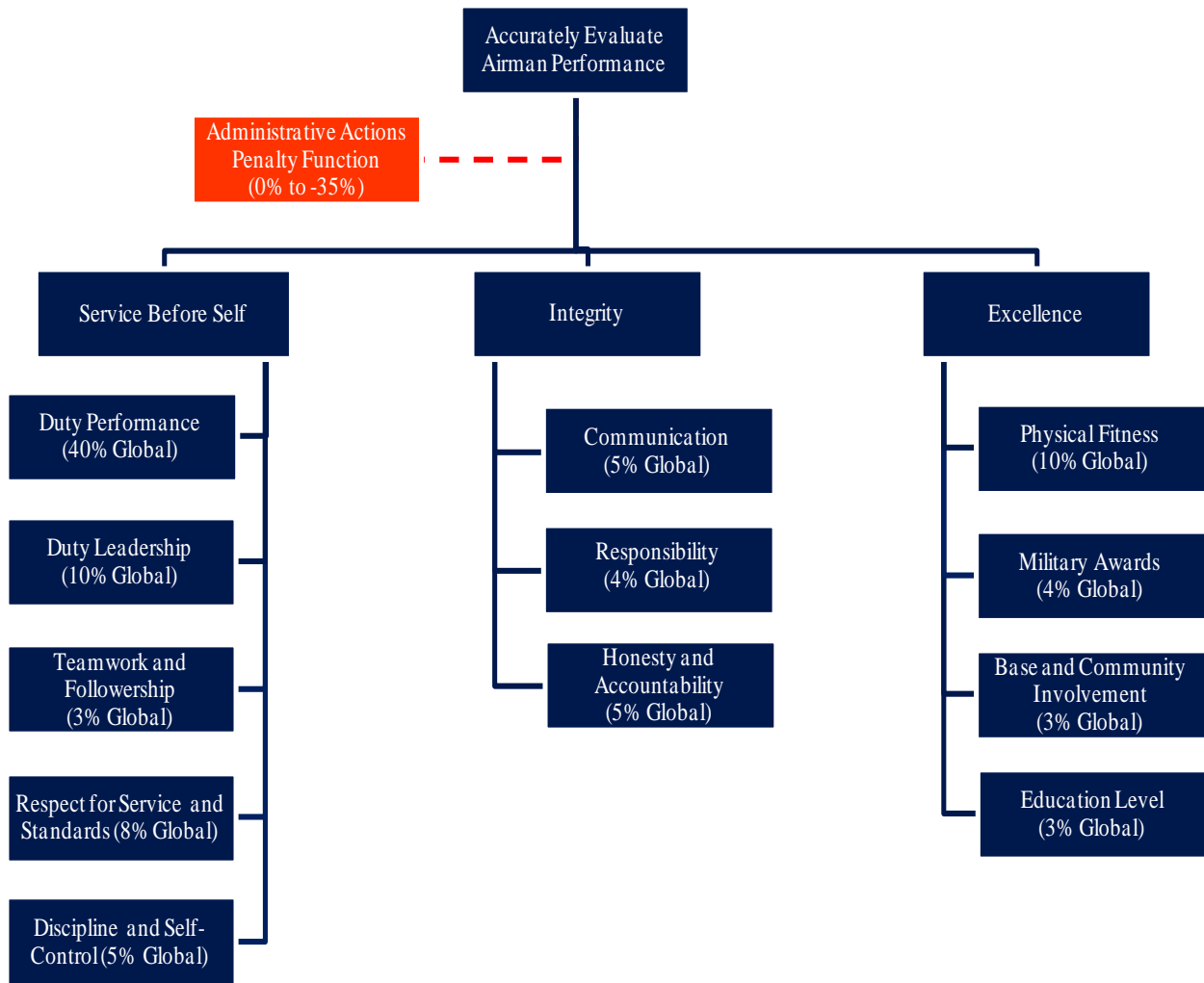


Figure 34. Revised JEPR Value Hierarchy (Based on Core Values)

In addition to the realignment of the attributes under the underlying core value structure, the SNCO SMEs provided additional recommendations after factor analysis of JEPR Training Dataset. First, they felt that more detailed definitions of the ratings categories would better help the rater classify individuals during appraisals. Although the duty and rank centric rating categories fared well in describing job performance categories, they were inadequate in categorizing behavioral observations. The SMEs felt that the category descriptions for attributes under the Service Before Self category

objective and the integrity objective were directly tied to standards. The SMEs also believed that attributes under the excellence objective, with the exception of Physical Fitness, were tied to professionalism and professional growth as it is described in AFI 36-2618, The Enlisted Force Structure. The SMEs felt that a left marking in Military Awards, Base and Community Involvement, or Education Level were not violations of standards, but did indicate the individual was not maximizing their abilities to the fullest extent for professional growth into becoming a well-rounded airman. Therefore, the attributes were divided into two distinct groups. One group was identified as a Standard, while the other group was Professional Expectations. A Standard was defined as a category for attributes that were tied to meeting a military standard. A failure of a ratee to meet a standard would drive a referral EPR and severely impact the ratees' overall JEPR score. The Professional Expectation group was defined for attributes that quantify the ratees' effort to maximize their professional growth and airmanship. If a ratee was appraised to be "Below Professional Expectations" for a Professional Expectation attribute, the ratee would be considered below the expectations in this area for professional growth. However, a "Below Professional Expectations" rating for an attribute would not generate a referral EPR for the ratee, but would impact the ratee by not contributing any points from this attribute to the ratees' overall JEPR score

However, the Physical Fitness attribute was problematic to define. In the current Air Force ratings appraisal system, Physical Fitness is a binary rating, where no value or an increase in rating is given for exceeding the standards. In the JEPR model, the Physical Fitness attribute is deemed a standard up to the point of a passing score, then transitions to reward the ratee for better Physical Fitness performance through

incremental increases in the overall JEPR rating as the Physical Fitness performance rises. Physical Fitness was believed to bridge the two categories in a piecewise fashion. The SMEs felt that initially, Physical Fitness should be treated as a Standard up to the point that a passing score is achieved, then should transfer to a Professional Expectation after the ratee was above the minimum standard. This would allow the category to meet the intent of a standard, yet provide increased value (captured as incremental increases in the overall JEPR score) to the ratee at points above the minimum standard. Further testing with factor analysis later in this research will provide better insight as to which group that this attribute is more closely aligned with. Table 41 and Table 42 illustrate the two groups of attributes, while Figure 35 illustrates the two theorized groupings of the JEPR attributes.

Table 41. JEPR Attributes Related to Standards

Attribute	Type
Duty Performance	Standard
Duty Leadership	Standard
Teamwork and Followership	Standard
Respect for Standards	Standard
Discipline and Self-Control	Standard
Communication	Standard
Responsibility	Standard
Honesty and Accountability	Standard
Physical Fitness	Standards *Bridges both groupings

Table 42. JEPR Attributes Related to Professional Expectations

Attribute	Type
Military Awards	Professional Expectation
Base and Community	Professional Expectation
Involvement	Professional Expectation
Education Level	Professional Expectation *Bridges both groupings

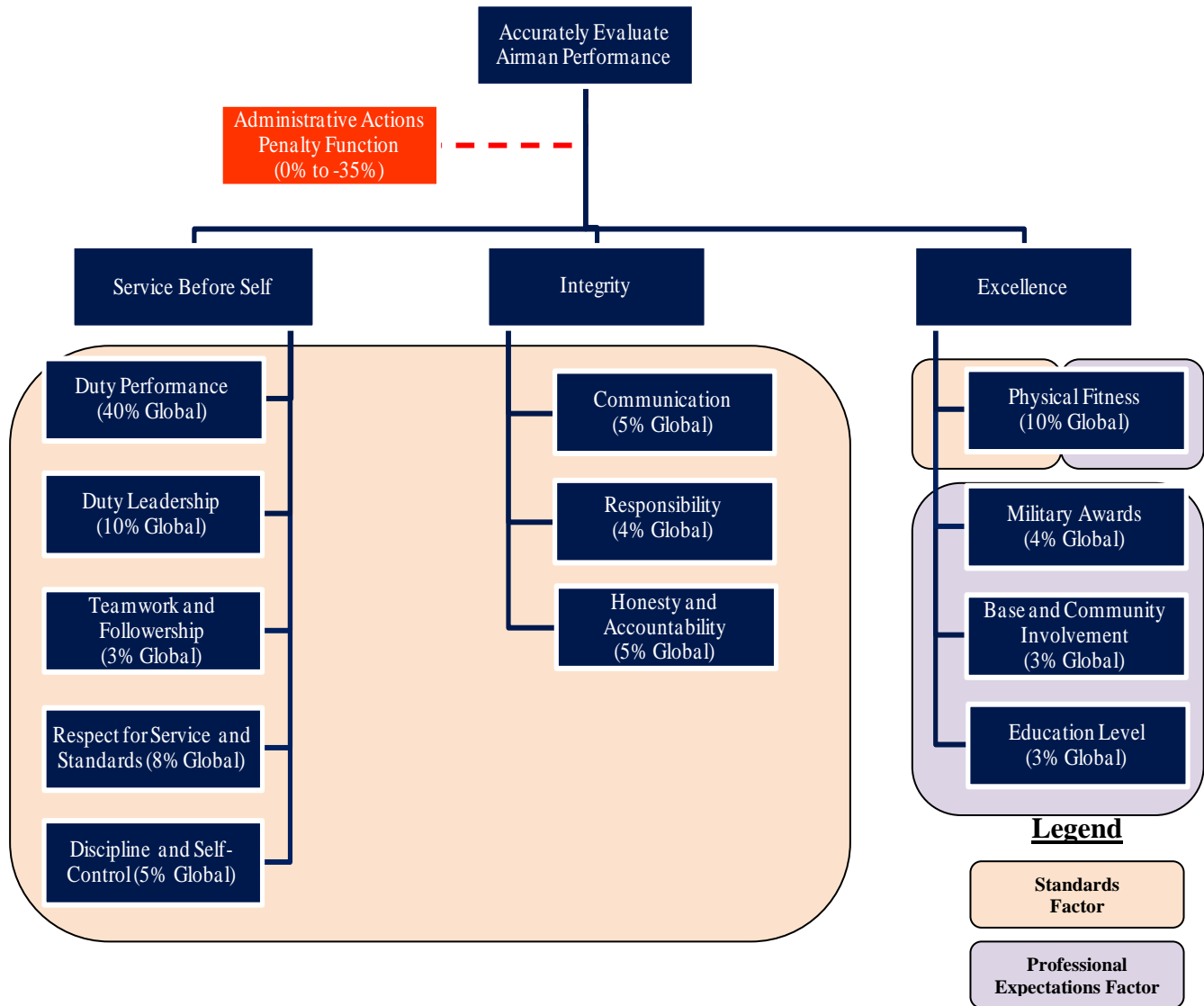


Figure 35. Revised JEPR Value Hierarchy (Theorized Factor Structure Overlay)

Several SMEs felt that where an attribute described an individual’s personal values, discrete (all or none) markings were needed. In determining Honesty and Accountability for example, some SMEs felt the individual either exhibits or does not exhibit the trait. However, others felt that there were instances where a ratee may be honest when confronted. Therefore Table 43 through Table 46 through show the final revised rating

categories. They capture inputs from both groups of SMEs, with better definitions, better category descriptions, discrete markings at the upper and lower bounds, and variable settings in the middle of the categories where the individuals' personal values are captured.

Table 43. Final JEPR Service Before Self Fundamental Objective Categories

Service Before Self	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Meets Minimal Objectives Not Consummate With Rank and Duty Position	Meets Some Objectives Consummate With Rank and Duty Position	Meets All Objectives Consummate With Rank and Duty Position	Meets Objectives For Next Higher Rank and Duty Position
Duty Performance	0 to 14	15 to 39	40 to 64	65 to 100
Duty Leadership	0 to 19	20 to 39	40 to 59	60 to 100
Teamwork and Followership	0 to 29	30 to 44	45 to 64	65 to 100

Service Before Self	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Consistently Does Not Demonstrate Respect for Service and Standards	Frequent Mentorship Needed to Maintain Respect for Service and Standards	Minimal Mentorship Needed to Maintain Respect for Service and Standards	Exhibits Respect for Service and Standards at all Times
Respect for Service and Standards	0	1 to 49	50 to 99	100

Service Before Self	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Consistently Does Not Demonstrate Discipline and Self-Control	Frequent Mentorship Needed to Maintain Discipline and Self-Control	Minimal Mentorship Needed to Maintain Discipline and Self-Control	Exhibits Discipline and Self-Control at all Times
Discipline and Self-Control	0	1 to 39	40 to 99	100

Table 44. Final JEPR Excellence Fundamental Objective Categories

Excellence	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Exempt in All Components	Below Standard	At Standard	Exceeds Standard
	Current with Min Passing Score Applied for Full PT Test Exemption	Non-Current or Current Failure in Overall Score or 1+ Components	Current and Meets Standards for Overall Score and all Components	Current and Exceeds Standards for Overall Score and Meets all Components
Physical Fitness	75	0 to 100 0% Awarded for Raw Score	75 to 89	90 to 100

Excellence	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Professional Expectation	Broadening Professionalism	At Professional Expectation	Exceeds Professional Expectation
	Consider No Awards or Nominations at Any Level	Consider Section/Squadron /Group/Wing Nominee	Consider Squadron/Group /Wing Awards	Consider NAF/MAJCOM/H Q USAF/Joint Level Awards
Military Awards	0	1 to 29	30 to 49	50 to 100

Excellence	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Professional Expectation	Broadening Professionalism	At Professional Expectation	Exceeds Professional Expectation
	Does Not Participate in Base/Community Events	Participates in 1 Base or Community Event	Participates in 2+ Base or Community Events	Active in 4+ Base or Community Events with Leadership Role in 1+ Event
Base and Community Involvement	0	1 to 39	40 to 59	60 to 100

Excellence	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Professional Expectation	Broadening Professionalism	At Professional Expectation	Exceeds Professional Expectation
	Not Pursuing Educational Opportunities	Currently Pursuing Degree/Certification or Enrolled in CDCs	Possesses CCAF and/or Associate Degree	Possesses Bachelors or Graduate Degree
Education Level	0	1 to 49	50 to 69	70 to 100

Table 45. Final JEPR Integrity Fundamental Objective Categories

Integrity	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Meets Minimal Objectives Not Consummate With Rank and Duty Position	Meets Some Objectives Consummate With Rank and Duty Position	Meets All Objectives Consummate With Rank and Duty Position	Meets Objectives For Next Higher Rank and Duty Position
Communication	0 to 19	20 to 39	40 to 59	60 to 100
Responsibility	0 to 14	15 to 29	30 to 49	50 to 100
Integrity	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Below Standard	Potential	At Standard	Exceeds Standard
	Consistently Does Not Demonstrate Honesty and Accountability	Exhibits Honesty & Accountability in Adverse Situations When Confronted	Exhibits Honesty & Accountability in Adverse Situations Voluntarily	Exhibits Honesty and Accountability at all Times
Honesty and Accountability	0	1 to 39	40 to 99	100

Table 46. Final JEPR Administrative Actions Independent Penalty Function

Administrative Actions (Correction Factor)	Rating Category 1	Rating Category 2	Rating Category 3	Rating Category 4
	Article 15/UCMJ	LOC/LOA/LOR	LOC/LOA/LOR	Min/No Negative Indicators
	Documented Article 15 or UCMJ Actions	Reoccurring disciplinary issues with multiple LOCs/LOAs/LORs in PIF	Documented disciplinary issue with single LOC/LOA/LOR in PIF	Minimal to no disciplinary issues. Consider PT failures in Period if now Passing
	-100 to -81	-80 to -61	-60 to -31	-30 to 0

After incorporating the SNO SME recommendations, the final VFT attributes slated to use the exponential function were determined, and the associated Gamma values were computed. The exponential function used is shown in Equation 19 with the associated Gamma values in Table 47 for the applicable JEPR attributes.

$$f_i = \frac{1 - e^{-\gamma_i(x_i - x_i^0)}}{1 - e^{-\gamma_i(x_i^* - x_i^0)}} \tag{19}$$

Table 47. Final Gamma Shaping Components for SAVFs Used in VFT Function

Gamma Shaping Component for Value Function Objectives		
Objective Number	Attribute	Gamma Value Used
1	Duty Performance	0.009679388
2	Duty Leadership	0.009386208
4	Communication	0.009386208
5	Respect for Service and Standards	0.0000000001
6	Discipline and Self-Control	0.00938621
7	Honesty and Accountability	0.00938621
8	Responsibility	0.018435884
10	Awards	0.018435884
12	Base and Community Involvement	-0.00281841

The final attributes for the VFT Framework using a Piecewise function were also determined, and the associated slope values were computed. The Piecewise function used for the VFT framework is shown in Equation 20, where i is the attribute number using the function, j is the additive sum of the function before slope k , and k is the current section of the function. Each piecewise function used in the VFT Framework was comprised of four sections. The associated slope values for the applicable JEPR attributes are shown in Table 48 through Table 50.

$$f_i = \begin{cases} \frac{\left(\frac{RAW}{SLOPE_k}\right)}{100}, & k = 1 \text{ and } RAW \leq MAX_k \\ \frac{\left(\sum_{j=2}^k \frac{(MAX_{j-1} - MAX_{j-2})}{SLOPE_{j-1}} + \frac{(RAW - MAX_{k-1})}{SLOPE_k}\right)}{100}, & 2 \leq k \leq 4 \text{ and } RAW \leq MAX_k \end{cases} \quad (20)$$

Table 48. Final Ranges and Slopes for Piecewise Physical Fitness SAVF

Objective 3		
Physical Fitness		
Percentage of What an Ideal Employee Provides	Raw Score Ranges Solicited	Calculated Piecewise Slopes
0%	0	0
25%	1 to 74	2.96
65%	75 to 75	0.025
95%	76 to 90	0.50
100%	91 to 100	2.00
NOTE		
Function Values are artificially terminated for overall PT scores below 75% or for a failure in 1 or more components regardless of score. For these scenarios, 0% value is awarded for the SAVF. This is due to Air Force Instruction 36-2905 Guidance.		

Table 49. Final Ranges and Slopes for Piecewise Teamwork and Followership SAVF

Objective 9		
Teamwork and Followership		
Percentage of What an Ideal Employee Provides	Raw Score Ranges Solicited	Calculated Piecewise Slopes
0%	0	0
25%	1 to 30	1.20
50%	31 to 45	0.60
75%	46 to 65	0.80
100%	66 to 100	1.40

Table 50. Final Ranges and Slopes for Piecewise Education SAVF

Objective 11		
Education		
Percentage of What an Ideal Employee Provides	Raw Score Ranges Solicited	Calculated Piecewise Slopes
0%	0	0
25%	1	0.04
50%	2 to 50	1.96
75%	51 to 70	0.80
100%	71 to 100	1.20

Finally, the final Piecewise penalty function is shown in Equation 21 and the associated slope values in Table 51 for the JEPR Penalty Function.

For Equation 21, j is the additive sum of the function before slope k , and k is the current section of the function. Each piecewise function used in the VFT Framework was comprised of four sections. The associated slope values

$$pf = \begin{cases} \left(\frac{\left(\sum_{j=k}^4 \frac{(MAX_{j-1} - MAX_j)}{SLOPE_j} \right) + \frac{RAW - MAX_{k-1}}{SLOPE_k}}{100} \right), & 1 \leq k \leq 3 \text{ and } RAW \leq MAX_k \\ \left(\frac{\frac{MAX_{k-1} - MAX_k}{SLOPE_k} + \frac{RAW - MAX_{k-1}}{SLOPE_k}}{100} \right), & k = 4 \text{ and } RAW \leq MAX_k \end{cases} \quad (21)$$

Table 51. Final Ranges and Slopes for Piecewise JEPR Penalty Function

Negative Value Contribution		
Independent Penalty Function		
Percentage of What an Ideal Employee Provides	Raw Score Ranges Solicited	Calculated Piecewise Slopes
0%	-100 to -81	1.00
25%	-80 to -61	0.57142286714
50%	-60 to -31	1.00
75%	-30 to -1	2.00
100%	0	0

With the JEPR functions redesigned based on the findings from the initial factor analysis, the analysis effort shifted to see if the theorized two factor structure of the JEPR Training Dataset truly existed, and to test whether the factor structure could still describe the Air Force Core Values.

Final Dimensionality Assessment (Training Dataset)

The SMEs input indicated the attributes could be regrouped into two distinct categories due to the modification. The SMEs theorized that the three Fundamental Objectives of Service Before Self, Integrity, and Excellence from the VFT Framework could really be reduced to just two latent factors: Standards and Professional Expectations. To test this assumption, a second EFA analysis was performed using only two factors to describe the VFT Framework.

For the second EFA effort, the eigenvalues from the reduced correlation matrix used in the initial dimensionality assessment were again used for the final dimensionality assessment. As identified earlier, the negative eigenvalues for factors nine through 13 were immediately eliminated, as they corresponded to negative eigenvectors, and could not contribute to the factor analysis (Dillon & Goldstein., 1984, p. 74). With only eight factors remaining, a Scree Plot of the reduced correlation matrix eigenvalues was generated as shown in Figure 36.

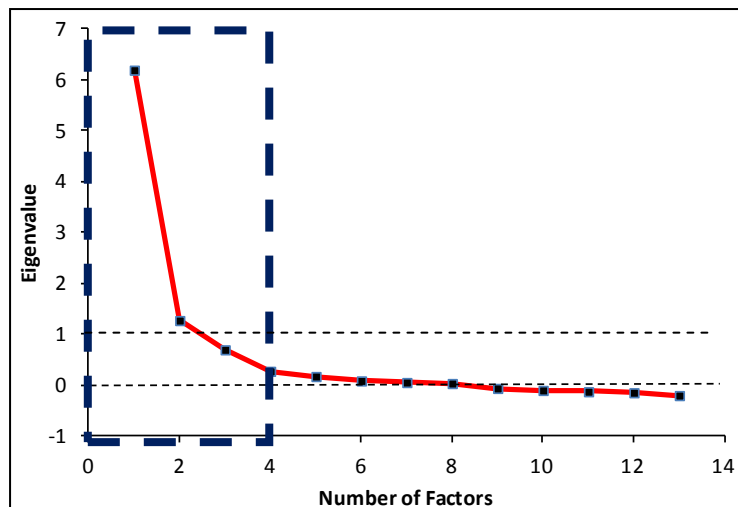


Figure 36. Scree Plot of Eigenvalues from the Reduced Correlation Matrix [R*]

The Scree Plot illustrated that the JEPR model received only a minimal contribution from eigenvalues four through eight. The Scree Plot also graphically highlighted a noticeable difference between the slopes of eigenvalues two and three. Therefore, it was decided to retain only eigenvalues one and two for the final EFA model, as the eigenvalues for factors three through eight were so small that their contributions to explaining variance in the JEPR model would have been minimal (Dillon & Goldstein., 1984, p. 74).

Final Exploratory Factor Analysis and Interpretation (Training Dataset)

With the dimensionality of the final model now determined, the factor loadings were again generated from the reduced correlation matrix using only two latent factors. Inspection of the final unrotated factor loadings from the JEPR training data indicated that a majority of the variables were heavily loaded on one factor. The groupings of the variables were not intuitive, and did not resemble the two factor latent structure of Standards and Professional Expectations that were identified after the JEPR model was redesigned. The original unrotated factor loadings matrix is shown in Table 52.

As was the case with the initial three factor model, the unrotated loadings for the two factor model were rotated orthogonally in an attempt to test whether the VFT Framework of the JEPR model could be interpreted as Standards and Professional Expectations as defined by the SMEs. An orthogonal Varimax rotation was applied to the two factor model, with all variables with factor loadings greater than or equal to 0.40 being considered statistically significant. The results of the rotated loadings for the two factor model are shown in Table 53, with the highest loading value for each variable shown in bold.

Table 52. Two Factor JEPR Model Unrotated Factor Loadings

Unrotated Factor Loading Matrix		
Objective	Factor 1	Factor 2
Duty Performance	0.825072	-0.155601
Duty Leadership	0.889890	-0.051500
Physical Fitness	0.264630	0.454923
Communication	0.771665	-0.251238
Respect for Service and Standards	0.831920	-0.121190
Discipline and Self-Control	0.728016	-0.176955
Honesty and Accountability	0.500435	-0.255872
Responsibility	0.774218	-0.049021
Teamwork and Followership	0.814135	-0.324765
Military Awards	0.581608	0.505750
Education Level	0.624878	0.373564
Base and Community Involvement	0.301467	0.503163
Administrative (Correction Factor)	0.705828	0.323704

Table 53. Two Factor JEPR Model Orthogonally Rotated Factor Loadings

Factor Analysis Settings Technique #1 (Orthogonal)		
Factoring Method	Principal Axis Factoring	
Prior Communality	Common Factor Analysis (SMC)	
Factors Selected	2	
Rotation Method	Varimax	
Significance Threshold	=> 0.4	
Objective	Standards	Professional Expectations
Duty Performance	0.801157	0.251203
Duty Leadership	0.809326	0.373562
Physical Fitness	0.019266	0.525940
Communication	0.799070	0.141684
Respect for Service and Standards	0.790996	0.284785
Discipline and Self-Control	0.725587	0.186666
Honesty and Accountability	0.561968	0.009890
Responsibility	0.706111	0.321285
Teamwork and Followership	0.871157	0.096814
Military Awards	0.274978	0.720026
Education Level	0.375390	0.623783
Base and Community Involvement	0.029050	0.585843
Administrative (Correction Factor)	0.470282	0.617911

As expected by the SMEs, the variables of the VFT Framework could be furthered reduced to two factors, explained by the latent factors of Standards and Professional Expectations. The orthogonal rotation was able to account for 57.39% of the common variance between using only two factors. The decision to use the better defined two factor model versus the three factor model resulted in only a 5.34% loss in variance explanation. Additionally, an oblique rotation was performed on the JEPR Training Dataset as shown in Table 54. As was the case with the three factor EFA model, the oblique rotated loadings of the two factor EFA model aligned on the same variables and under the same two factors that the orthogonal rotation did. Both methods identified the same variables in each factor as significant, with only minor differences in loading values.

Table 54. Two Factor JEPR Model Oblique Rotated Factor Loadings

Factor Analysis Settings Technique #2 (Oblique)		
Factoring Method	Principal Axis Factoring	
Prior Communality	Common Factor Analysis (SMC)	
Factors Selected	2	
Rotation Method	Promax	
Significance Threshold	=> 0.4	
Objective	Standards	Professional Expectations
Duty Performance	0.817162	0.041456
Duty Leadership	0.785358	0.177713
Physical Fitness	-0.154533	0.590492
Communication	0.851507	-0.082090
Respect for Service and Standards	0.794480	0.082620
Discipline and Self-Control	0.753765	-0.008763
Honesty and Accountability	0.628916	-0.159500
Responsibility	0.686752	0.149793
Teamwork and Followership	0.947639	-0.154865
Military Awards	0.068125	0.732881
Education Level	0.213336	0.593251
Base and Community Involvement	-0.163595	0.655442
Administrative (Correction Factor)	0.322059	0.557766

Regardless of the rotation method chosen, the loadings of JEPR training data variables clearly aligned with a specific factor in the two factor EFA model. This can be seen graphically in Figure 37 where the dashed dividing line shows the separation of the Standards factor and the Professional Qualities factor.

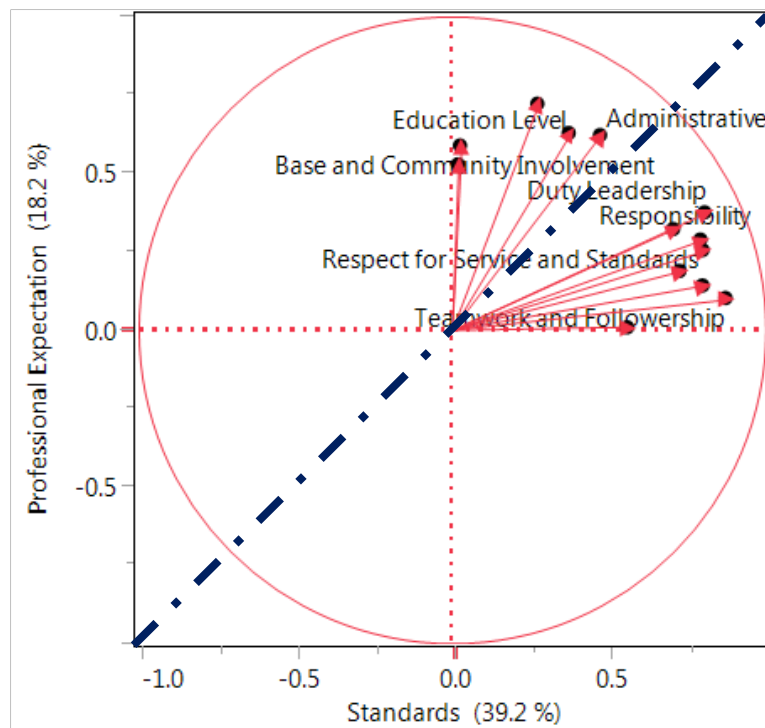


Figure 37. Factor Loading Plot for Two Factor JEPR Model (Rotated Orthogonally)

The two factor EFA model did indeed show that the SMEs were correct in their assumption that there were two latent factors underneath the VFT Framework. Although Physical Fitness was originally theorized to reside in both the Standards and Professional Expectations groupings, the two-factor EFA model clearly illustrated that this attribute belonged in the Professional Expectations factor. This is intuitive as the small loadings shown in column one of Table 53 and Table 54 indicate the correlation between meeting the standards and Physical Fitness, while the much larger loadings shown in column two

of Table 53 and Table 54 indicate the correlation between Physical Fitness and Professional Expectations. From a value standpoint, during the overall scoring of the JEPR, it was clear that the Physical Fitness attribute belonged in the Professional Expectations categories, as the JEPR identified that the attribute provided increased value to both the Air Force and to the ratee, as higher Physical Fitness scores were attained. Figure 38 shows the overlay of the two factor structure onto the Value Hierarchy with Physical Fitness solely represented by the Professional Expectations factor.

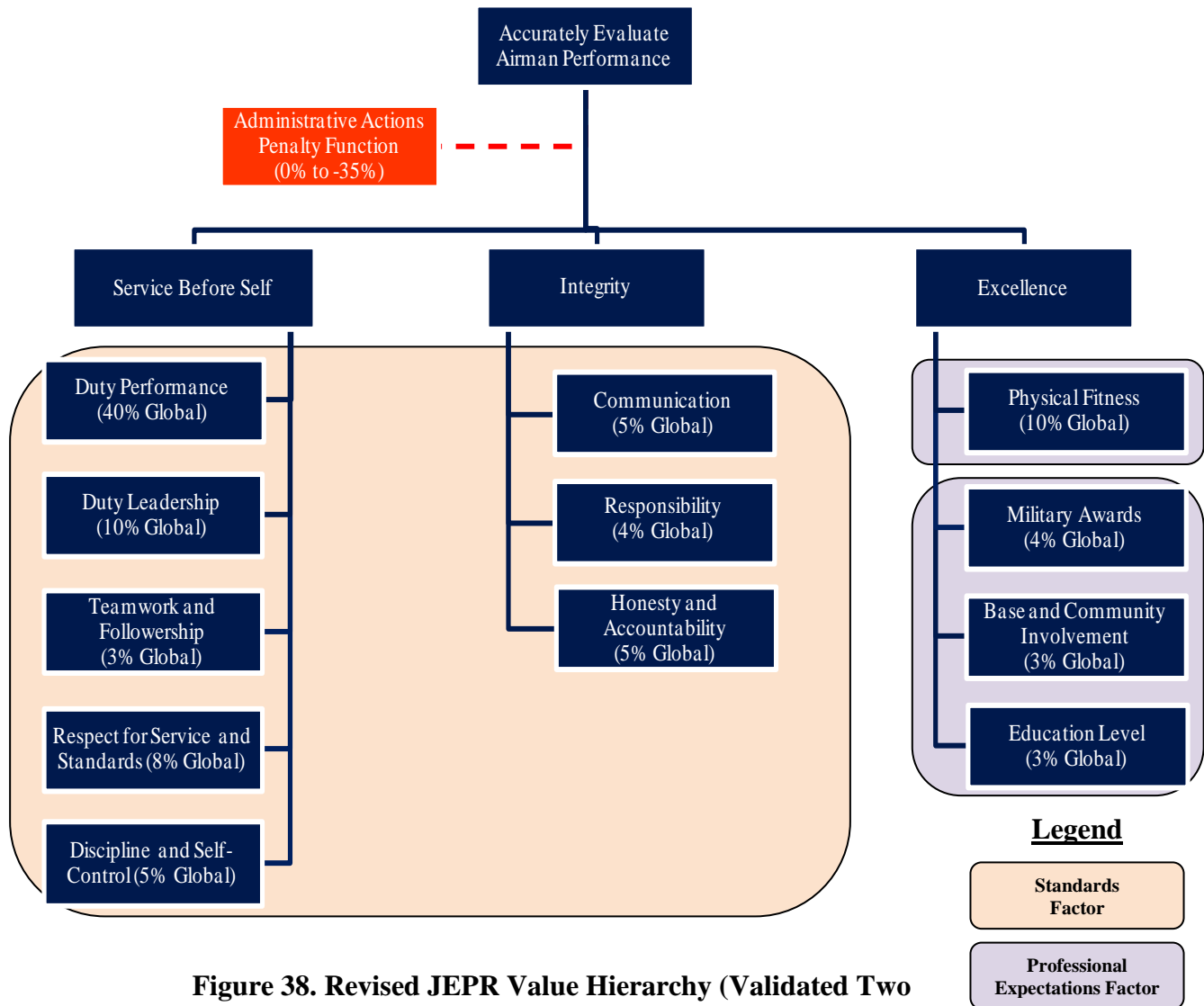


Figure 38. Revised JEPR Value Hierarchy (Validated Two Factor Structure Overlay)

JEPR Decision Support System Tool Revision

With the VFT framework redesigned, the SNO SMEs asked that the prototype Decision Support System (DSS) tool also be redesigned. This redesign incorporated all the changes made to the VFT Framework, and was intended to provide a more accurate representation of what the envisioned web-based user interface would look and act like. The revised DSS also provided a more accurate method of data collection for the supervisors involved in collecting the JEPR Test Dataset samples for the next phase of the analysis. The SNCO SMEs also requested that the DSS be redesigned to include three additional features to improve the appraisal process. First, and most important, the SMEs also asked for additional features to be included into the DSS to help reduce inflation of appraisal ratings. Second, the SNCO SMEs asked that the DSS be able to classify the ratee based on their ability of the ratee to meet Air Force Standards as detailed by doctrine. Finally, the SMEs requested that the DSS be able to provide quantitative feedback to the ratee and the rater. The SMEs requested that the DSS provide areas of strength in performance, areas where improvement in performance was needed, the average score among all AFSCs of the same rank as the ratee within the unit, and the average score among peers of the same rank, in the same career field, Air Force wide. By providing this feedback, a roadmap could be developed between the rater and ratee to achieve clearly defined goals to improve performance for the unit and for the ratee to meet professional goals.

Ratings inflation is a recognized problem in many performance appraisal systems (Murphy, 2008). The SMEs felt that, although a redesign of the JEPR DSS could assist

with appraisal rating inflation, the onus for accurately appraising members truly falls onto the application of doctrine by senior leaders of an organization. In discussing the topic of ratings inflation with the SMEs, several controls methods of controlling inflation were discussed. These methods ranged from the well defined bands and weighted attribute design that the JEPR model used, to using a forced ratings distribution range, to providing a breakdown of the raters rating history for the ratee, rater, and raters chain of command.

From the initial Qualitative Analysis provided by this research, it appeared that the current JEPR design did a very good job of controlling inflation using the clearly defined and consistent attribute categories, with direct ties to doctrine and standards, for appraising airmen. In discussing inflation with the SMEs, the SMEs thought that the use of a weighted attribute scheme in the JEPR model also helped control inflation by providing increased importance and focus on primary duties. The SMEs also believed that the weighted JEPR construct better communicates to the population what attributes are the most important to the Air Force from a strategic vantage point. For example, an airman, who had performed strong in heavily weighted areas associated with primary duties, would accrue more points for their overall JEPR score than an airman who had underperformed in heavily weighted attribute such as Duty Performance. The SMEs felt that this design clearly communicated to both the supervisor and the ratee which attributes are important to the Air Force. The SME also believed the weighted attribute design also conveyed the message to the rater and ratee that all attributes are not equally valued, thus providing delineation, and thus inflation control. With these methods incorporated, the SMEs discussed other possible ways to further control ratings inflation.

The use of a forced distribution to assist in ratings inflation control was discussed in-depth with the SMEs. After much research and discussion, the SMEs felt that this method did not allow delineation of performers within categories, and would unfairly, and artificially, effect organizations and personnel where the number of employees either exceeded or was determined to be below the mandated cutoff level. The SMEs perception of forced distributions was supported by organizational psychologist literature where Roch, Sturnburgh, and Caputo, as cited by (Murphy, 2008), conveyed that organizational psychologists view the use of forced distributions as a less fair appraisal technique than other methods for inflation control. Several large companies such as Ford Motor Company and Goodyear Tire and Rubber Company have in the past experimented with forced distribution appraisal systems (Blume et al., 2009). Both companies experienced unsuccessful results with forced distribution appraisals, and experienced both an internal and external backlash to their use and inconsistent application (Blume et al., 2009). In the case of Ford, many employees who had consistently received positive feedback from their supervisors earlier were suddenly rated as underperformers (Blume et al., 2009). Employees viewed the labeling and dismissal of sub-par performers as unfair and inequitable, and damaging both the workforce morale and the public images of both companies (Blume et al., 2009). Further supporting the SMEs stance on forced distributions, Murphy noted that forced distribution rating systems often mask performance differences across organizations (Murphy, 2008).

Providing the raters appraisal rating history was another method that was discussed for inclusion into the JEPPR DSS in an effort to reduce ratings inflation. A recent research effort in 2008 initiated by the U.S. Army Recruiting Command revealed

that providing the raters rating history to both the ratee and to the rater's supervision chain could significantly reduce appraisal ratings inflation (Dees et al., 2013). Dees et al. elaborated that organizational senior leaders need to be able quickly identify both positive and negative evaluation trends, identify weak and strong workgroups, and recognize training deficiencies for correction, or efficiencies for implementation (Dees et al., 2013). This type of insight also enables managers to better allocate experience level, to correct deficient behaviors quickly, and to propagate positive behaviors by both raters and ratees, improving the organizations quality (Dees et al., 2013). Not only could the ratee, rater, and the supervisors in the chain of command benefit from this capability, career field managers at the Air Force Personnel Center could also benefit from this capability, as they could immediate deduce ratings trends from within enlisted ranks, AFSCs, or locations. The centralized database construct of the JEPR was ideal for this type of analysis, as the DSS relied on Standard Query Language (SQL) queries form grouping of data. Therefore, the JEPR DSS was redesigned to include a graphical representation of the raters' ratings history on the appraisal to provide transparency to the ratee, the rater, and the raters' chain of command. In addition, the prototype JEPR DSS was also modified to allow the raters' chain of command to query the raters' rating history.

To better describe the overall JEPR score and results to both the rater and ratee, the SMEs asked that three distinct classification classes be created to help classify whether or not the ratees' performance had met Air Force standards as defined by doctrine. The SMEs felt that misclassification was a definite shortcoming of the current EPR system, and contributed to inflation. The development of these three classification classes improved the JEPR construct by meeting three distinct goals: The ability to

classify a referral rating, the ability to translate the JEPR scoring scheme to doctrine, and to translate the JEPR scoring scheme to the current EPR construct.

Before the SMEs could define the classification classes, a method for handling a referral report had to be developed. Under the current EPR construct, a referral EPR is an appraisal rendered when the ratee has failed to meet an established standard (Air Force Instruction 36-2406, 2013, p. 40). For the JEPR, it was determined that a referral would be generated if the rater places the ratee into the lowest rating category (failure to meet a standard) for any of the attributes that is defined as Standard. Additionally, the JEPR was also redesigned to issue a referral appraisal if the ratees' overall JEPR score was 30 or lower. Any JEPR referral report is forwarded directly to the commander for review and signature as the senior rating official. Placing the ratee into a far left rating category for any of the attributes defined as Professional Expectations, such as Military Awards, Base and Community Involvement, and Education Level areas does not create a referral situation. This is because these attributes are deemed areas of professional growth, and not a breach of standards.

After the handling of the referral process had been resolved, the JEPR Training Dataset data was inspected to determine the proper numeric boundaries for defining the three classification classes. After much discussion and comparison of the JEPR overall scores to the EPR scores for the JEPR Training Dataset test subjects, the SMEs felt that there were two distinct break points in the data that were identified. Overall JEPR scores less than or equal to 47.57, or that were deemed as a referral would be classified as "Below Standards". Ratees with overall JEPR scores greater than 47.57 and less than 85,

without a referral, would be classified as “Meets Standards”. Lastly, ratees with JEPR overall scores of 85 to 100, without a referral, would be classified as “Exceeds Standards”. Additionally, if the members overall JEPR score was below a 20, or the member failed to meet standards in all nine attributes described as a Standard (including the Administrative Actions correction factor), the JEPR would recommend to the commander to consider whether the individual should be retained for further military service.

In designing the classification classes, a concerted effort was also made to lessen the administrative workload of senior leaders and commanders. In addition to referrals, the SMEs asked that the JEPR be redesigned to forward only appraisals that were “Below Standards” or “Exceeds Standards” to the commander for signature. This would provide the commander insight and details concerning poor performers, as well as providing details concerning the exceptional performers in the unit. The JEPR classification classes and class descriptions are reflected in Table 55.

Table 55. JEPR Classification Classes and Class Descriptions

JEPR Classification Descriptions	
Classification Class Name	JEPR Classification Class Description
Below Standards	Overall Score ≤ 45.57 and/or Failure to Meet any Standard in the Standards group of attributes
Meets Standards	Overall Score >47.57 and <85 . Must meet Standards in all attributes in Standards group
Exceeds Standards	Overall Score ≥ 85 Must meet Standards in all attributes in Standards group

The classification effectiveness of the value based JEPR Framework will be tested for classification effectiveness versus the current system later in Chapter V.

Finally, the SMEs asked for a mechanism to provide increased feedback clarification of areas of strength, areas of weaknesses, the average score in the unit by rank, and the average score in the AFSC by rank Air Force wide. The value gap analysis is provided graphically showing the ratee areas of strength by attribute (blue bars) and areas where they can improve (red dotted bars) by attribute. This graphical representation can be used to facilitate the discussion during feedback of what the supervisor's, the units', and the Air Forces' expectations are for the ratee, what the ratees' career goals are, and how those goals can be achieved. The ratee and supervisor can also discuss how the ratees' score compare to members of the same rank across all AFSCs in their unit, and how their score compares to their peers in the same AFSC across the Air Force.

Although the final factor analysis showed that the three factor JEPR Framework could be further reduced to two factors, the JEPR DSS retained the three factor design to better relate the appraisal format to Air Force doctrine. Figure 39 through Figure 41 illustrate the redesigned JEPR DSS prototype attributes based on the Air Force Core Values revealed from the initial factor analysis of the JEPR Training Data. Figure 42 illustrates the Administrative Actions Penalty Function. Figure 43 illustrates the revised JEPR Career Targets output in-accordance with the SME recommendations.

Junior Enlisted Performance Report (JEPR)

Last Name	First Name	Initial	SSN	Rank	AFSC	JEPR Case File ID
0300	0300	NMI	000-00-0300	A1C	2W031	03001372423909
Duty Location	Organization	PAS	SRID	MAJCOM	Reason	
BARKSDALE AFB LA	2 MUNITIONS SQ	BBSFLKY	gs002	AFGSC	Annual	
Duty Title	Report Start Date	Report Closeout Date	# Days Supervised			
Conventional Maintenance Crew Member	03-May-11	02-Jan-12	244			

Service before Self | Integrity | Excellence | Administrative Actions | Career Targets

Doctrine Source
[AFI 36-2618 The Enlisted Force Structure](#)

Doctrine Source
[AFD-070906-003 - US Air Force Core Values](#)

	Below Standard	Potential	At Standard	Exceeds Standard	Attribute Score
Duty Performance	Meets Minimal Objectives Not Consummate With Rank and Duty Position (0% to 14%)	Meets Some Objectives Consummate With Rank and Duty Position (15% to 39%)	Meets All Objectives Consummate With Rank and Duty Position (40% to 64%)	Meets Objectives For Next Higher Rank and Duty Position (65% to 100%)	30.45%
Duty Leadership	Meets Minimal Objectives Not Consummate With Rank and Duty Position (0% to 19%)	Meets Some Objectives Consummate With Rank and Duty Position (20% to 39%)	Meets All Objectives Consummate With Rank and Duty Position (40% to 59%)	Meets Objectives For Next Higher Rank and Duty Position (60% to 100%)	7.07%
Teamwork and Followership	Meets Minimal Objectives Not Consummate With Rank and Duty Position (0% to 29%)	Meets Some Objectives Consummate With Rank and Duty Position (30% to 44%)	Meets All Objectives Consummate With Rank and Duty Position (45% to 64%)	Meets Objectives For Next Higher Rank and Duty Position (65% to 100%)	2.91%
Respect for Service and Standards	Consistently Does Not Demonstrate Respect for Service and Standards (0%)	Frequent Mentorship Needed to Maintain Respect for Service and Standards (1% to 49%)	Minimal Mentorship Needed to Maintain Respect for Service and Standards (50% to 99%)	Exhibits Respect for Service and Standards at all Times (100%)	8.00%
Discipline and Self-Control	Consistently Does Not Demonstrate Discipline and Self-Control (0%)	Frequent Mentorship Needed to Maintain Discipline and Self-Control (1% to 39%)	Minimal Mentorship Needed to Maintain Discipline and Self-Control (40% to 99%)	Exhibits Discipline and Self-Control at all Times (100%)	2.57%

Rater Digitally Sign JEPR
Main Menu

Figure 39. JEPR DSS Service Before Self Factor

Junior Enlisted Performance Report (JEPR)

Last Name	First Name	Initial	SSN	Rank	AFSC	JEPR Case File ID
0300	0300	NMI	000-00-0300	A1C	2W031	03001372423909
Duty Location	Organization		PAS	SRID	MAJCOM	Reason
BARKSDALE AFB LA	2 MUNITIONS SQ		BBSFLKY	gs002	AFGSC	Annual
Duty Title	Report Start Date		Report Closeout Date		# Days Supervised	
Conventional Maintenance Crew Member	03-May-11		02-Jan-12		244	

Service before Self | Integrity | Excellence | Administrative Actions | Career Targets

	Doctrine Source				Attribute Score
	AFI 36-2618 The Enlisted Force Structure				
	AFD-070906-003 - US Air Force Core Values				
Communication	Below Standard	Potential	At Standard	Exceeds Standard	5.00%
	Meets Minimal Objectives Not Consummate With Rank and Duty Position	Meets Some Objectives Consummate With Rank and Duty Position	Meets All Objectives Consummate With Rank and Duty Position	Meets Objectives For Next Higher Rank and Duty Position	
	(0% to 19%)	(20% to 39%)	(40% to 59%)	(60% to 100%)	
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	
	<input type="text"/>	<input type="text"/>	<input type="text"/>	100	
Responsibility	Below Standard	Potential	At Standard	Exceeds Standard	1.70%
	Meets Minimal Objectives Not Consummate With Rank and Duty Position	Meets Some Objectives Consummate With Rank and Duty Position	Meets All Objectives Consummate With Rank and Duty Position	Meets Objectives For Next Higher Rank and Duty Position	
	(0% to 14%)	(15% to 29%)	(30% to 49%)	(50% to 100%)	
	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	
	<input type="text"/>	24	<input type="text"/>	<input type="text"/>	
Honesty and Accountability	Below Standard	Potential	At Standard	Exceeds Standard	5.00%
	Consistently Does Not Demonstrate Honesty and Accountability	Exhibits Honesty and Accountability in Adverse Situations When Confronted	Exhibits Honesty and Accountability in Adverse Situations Voluntarily	Exhibits Honesty and Accountability at all Times	
	(0%)	(1% to 39%)	(40% to 99%)	(100%)	
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	
	<input type="text"/>	<input type="text"/>	<input type="text"/>	100	

Figure 40. JEPR DSS Integrity Factor

Junior Enlisted Performance Report (JEPR)

Last Name	First Name	Initial	SSN	Rank	AFSC	JEPR Case File ID
0300	0300	NMI	000-00-0300	A1C	2W031	03001372423909
Duty Location	Organization		PAS	SRID	MAJCOM	Reason
BARKSDALE AFB LA	2 MUNITIONS SQ		BBSFLKY	gs002	AFGSC	Annual
Duty Title	Report Start Date		Report Closeout Date		# Days Supervised	
Conventional Maintenance Crew Member	03-May-11		02-Jan-12		244	

Service before Self | Integrity | Excellence | Administrative Actions | Career Targets

Doctrine Source	Doctrine Source	Doctrine Source
AFI 36-2618 The Enlisted Force Structure	AFI 36-2905 Fitness Program	AFD-070906-003 - US Air Force Core Values

Physical Fitness	Exempt in All Components Current with Min Passing Score Applied for Full PT Test Exemption (Exempt 75% Raw Score)	Below Standard Non-Current or Current Failure in Overall Score or 1+ Components (0% Awarded for Raw Score)	At Standard Current and Meets Standards for Overall Score and all Components (75% to 89%)	Exceeds Standard Current and Exceeds Standards for Overall Score and Meets all Components (90% to 100%)	Attribute Score 6.50%
	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	

Military Awards	Below Professional Expectation Consider No Awards or Nominations at Any Level (0%)	Broadening Professionalism Consider Section/Squadron/Group/Wing Nominees (1% to 29%)	At Professional Expectation Consider Squadron/Group/Wing Awards (30% to 49%)	Exceeds Professional Expectation Consider NAF/MAJCOM/HQ USAF/Joint Level Awards (50% to 100%)	Attribute Score 1.97%
	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	

Base and Community Involvement	Below Professional Expectation Does Not Participate in Base or Community Events (0%)	Broadening Professionalism Participates in 1 Base or Community Event (1% to 39%)	At Professional Expectation Participates in 2+ Base or Community Events (40% to 59%)	Exceeds Professional Expectation Active in 4+ Base or Community Events with Leadership Role in 1+ Event (60% to 100%)	Attribute Score 3.00%
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	

Education Level	Below Professional Expectation Not Pursuing Educational Opportunities (0%)	Broadening Professionalism Currently Pursuing a Degree/Certification or Enrolled in CDCs (1% to 49%)	At Professional Expectation Possesses CCAF and/or Associate Degree (50% to 69%)	Exceeds Professional Expectation Possesses Bachelors or Graduate Degree (70% to 100%)	Attribute Score 3.00%
	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	

Figure 41. JEPR DSS Excellence Factor

Junior Enlisted Performance Report (JEPR)

Last Name	First Name	Initial	SSN	Rank	AFSC	JEPR Case File ID	
0300	0300	NMI	000-00-0300	A1C	2W031	03001372423909	
Duty Location		Organization		PAS	SRID	MAJCOM	Reason
BARKSDALE AFB LA		2 MUNITIONS SQ		BBSFLKY	gs002	AFGSC	Annual
Duty Title				Report Start Date	Report Closeout Date	# Days Supervised	
Conventional Maintenance Crew Member				03-May-11	02-Jan-12	244	

[Service before Self](#)
[Integrity](#)
[Excellence](#)
[Administrative Actions](#)
[Career Targets](#)

Doctrine Source
[AFI 36-2907 - UNFAVORABLE INFORMATION FILE \(UIF\)](#)

Administrative	Article 15/UCMJ Documented Article 15 or UCMJ Actions	LOC/LOA/LOR Recurring disciplinary issues with multiple LOCs/LOAs/LORs in PIF	LOC/LOA/LOR Documented disciplinary issue with single LOC/LOA/LOR in PIF	Min/No Negative Indicators Minimal to no disciplinary issues. Consider PT failures in Period if now Passing	Attribute Score 0.00%
	(-100% to -81%)	(-80% to -61%)	(-60% to -31%)	(-30% to 0%)	
	0	0	0	0	

[Rater Digitally Sign JEPR](#)
[Main Menu](#)

Figure 42. JEPR DSS Administrative Actions Penalty Function

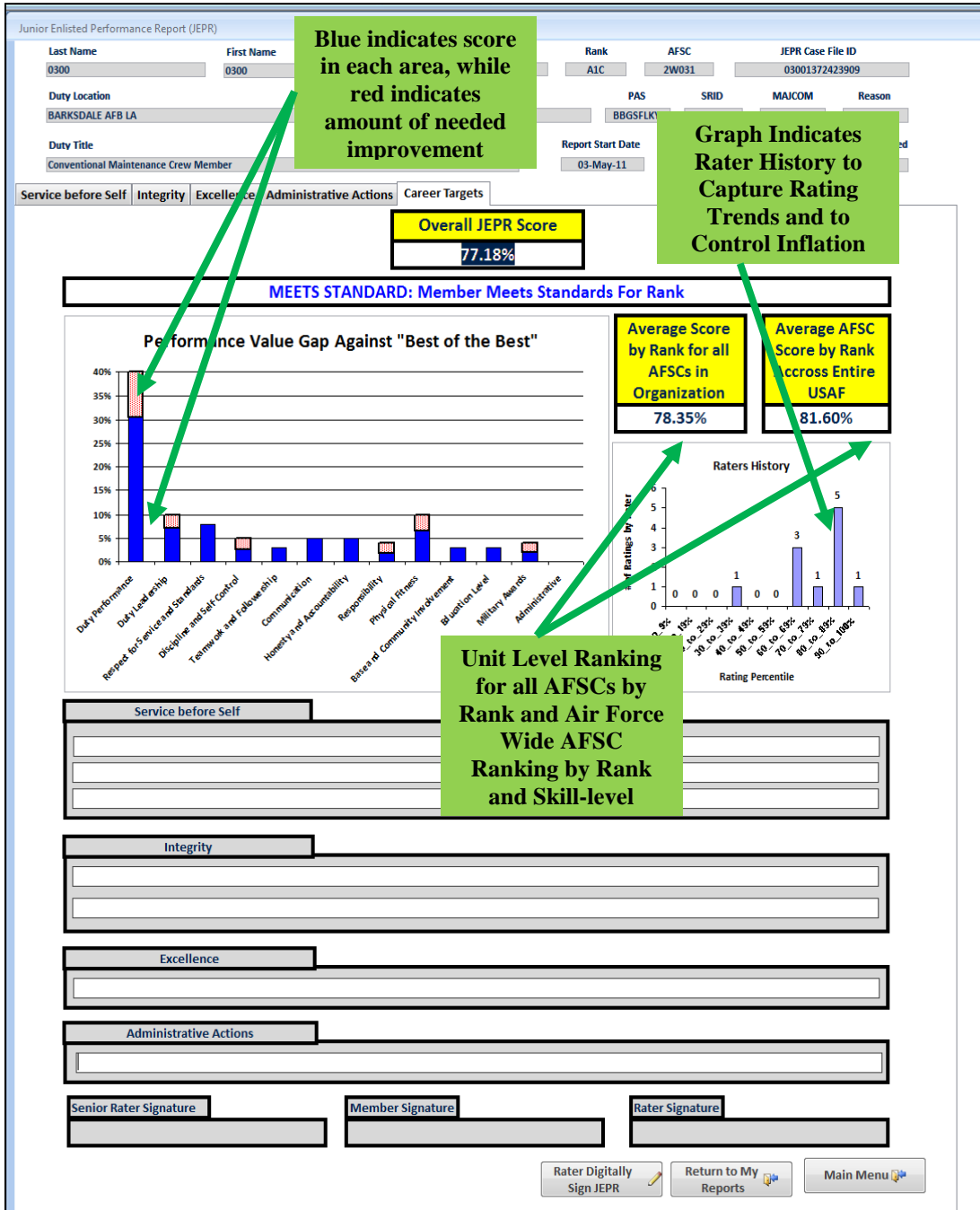


Figure 43. JEPR DSS Career Target Output

V. Multivariate Analysis and Results

Chapter Overview

This chapter focused on confirming that the two-factor structure revealed during Chapter IV was representative of the population. To confirm the structure, a larger dataset was studied to verify that the same factor loading structure existed. This larger dataset, known as the JEPR Test Dataset, was first subjected to factor analysis data suitability tests, without predetermined assumptions, just as had been previously done with the JEPR Training Dataset. Once the suitability tests had been completed, the Exploratory Factor Analysis (EFA) was applied to the dataset, with both orthogonal and oblique rotations utilized for interpretation.

After the two-factor structure of the JEPR framework was verified, Confirmatory Factor Analysis (CFA) was applied to the JEPR Test Dataset, using the same EFA factor structure, to confirm the statistical validity of the JEPR model. The CFA was a hypothesized model built from the EFA loadings construct which included multivariate, multi-equation regression models to create causal relationships among model variables. The regression model weights (factor loadings predicted during the regression) of the hypothesized model were then contrasted to the factor loadings found during the EFA effort from the data sampled from the Air Force population, to support accuracy of the JEPR model.

Finally, two Artificial Neural Networks (ANNs) were applied to further validate the JEPR design. The first ANN tested the classification consistency of the JEPR construct versus the measured attributes of the VFT Framework. The purpose of this

research was to confirm that the JEPR values chosen as cut-off points between the JEPR classifications classes were correct for classifying airmen using the attributes solicited to build the JEPR VFT Framework. A second ANN was also created to for comparative purposes versus the JEPR classifier to contrast how well the test subjects could be classified into the current EPR system using the VFT Framework attributes, given the known overall classification outcome of the current EPR system. Figure 44 illustrates an overview of the EFA, CFA, and the ANN processes detailed in this chapter.

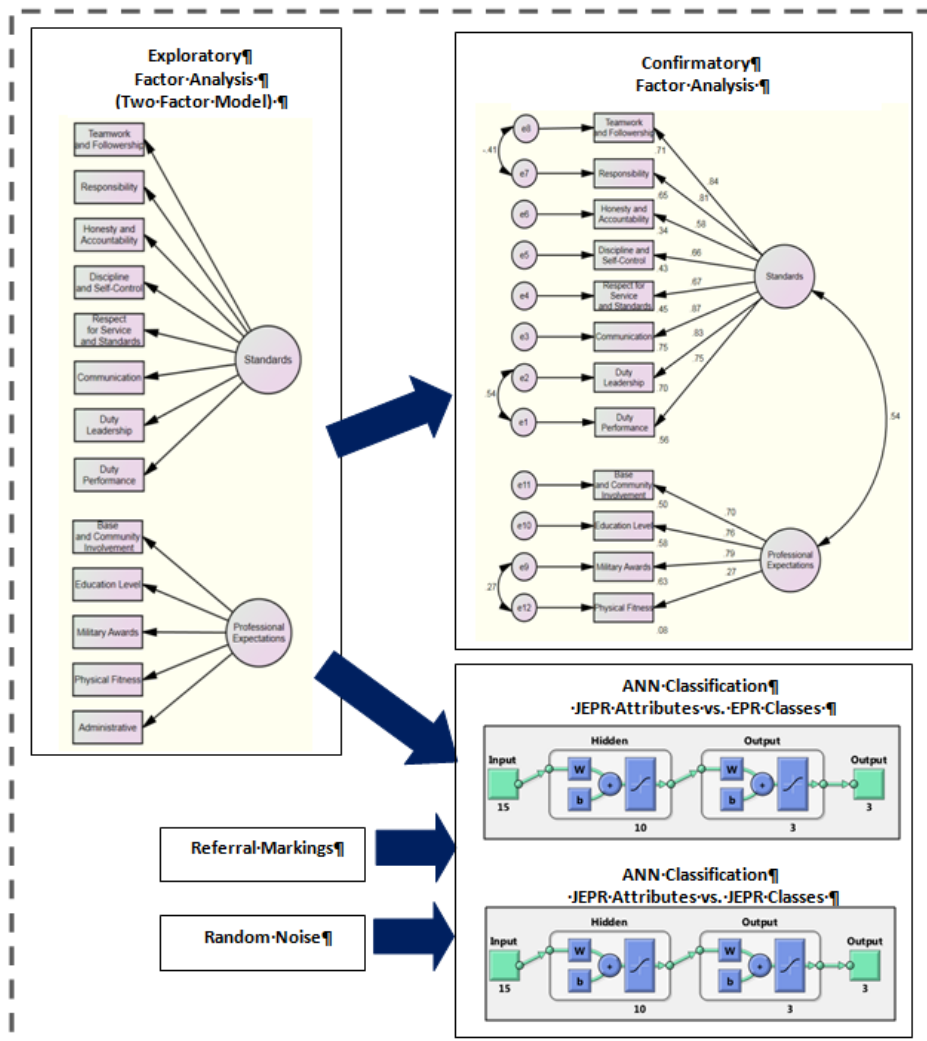


Figure 44. Overview of Multivariate Analysis and Results Chapter

Data Solicitation Process (Test Dataset)

Using the revised prototype JEPR database system, the SMEs collected a larger sample of data known as a test or verification dataset. The purpose of collecting this dataset was to gather a statistically significant sample size that could be used to verify that the two-factor structure that was identified during the JEPR Model Revision (Based on Initial Factor Analysis section in Chapter IV) was correct. A commonly used rule in research to determine an adequate sample is that the sampled data must meet or exceed a 10:1 observation per observed variable ratio (Schreiber, Nora, Stage, Barlow, & King, 2006).

As was done with the JEPR Training dataset, reports were generated after closeout of actual performance reports to prevent unduly influencing the official report. Supervisors utilized pseudo identification numbers for ratee case file creation, with no collection of personnel identifying information. For the JEPR Test Dataset, 159 JEPR samples were collected from 24 participating career-fields. The 159 data samples collected for the 13 observed attributes of the JEPR Test Dataset exceeded the 10:1 ratio rule used for determining an adequate sample (Costello and Osborne, 2005). As was done during the collection of the JEPR training dataset, the raters' gathering the JEPR test dataset were asked to rate official EPRs as usual. Upon completion of the official report, the ratees' overall official EPR rating was recorded using only the pseudo identification number. Next, supervisors evaluated the ratee using the JEPR program. After the JEPR Test Dataset collection effort was completed, the data was compiled for analysis.

Qualitative Inspection (Test Dataset)

The JEPR Test Dataset as next studied qualitatively for trends. Histograms for the 159 data samples were generated and studied. Looking at the histogram, it was immediately noticed that a large portion of the sample were assigned an overall “5” rating. Looking at the histogram table, 115 of the 159 or 72.3% of the airman sampled received the maximum score possible, an overall “5” rating. Doctrinally, this rating is described as “Truly Among the Best”. Only 26 of the 159 test subjects or approximately 16.4% were given an overall rating of “4” which equated to an “Above Average” rating. The distribution of the 159 test subjects measured under the current system is shown in Figure 45.

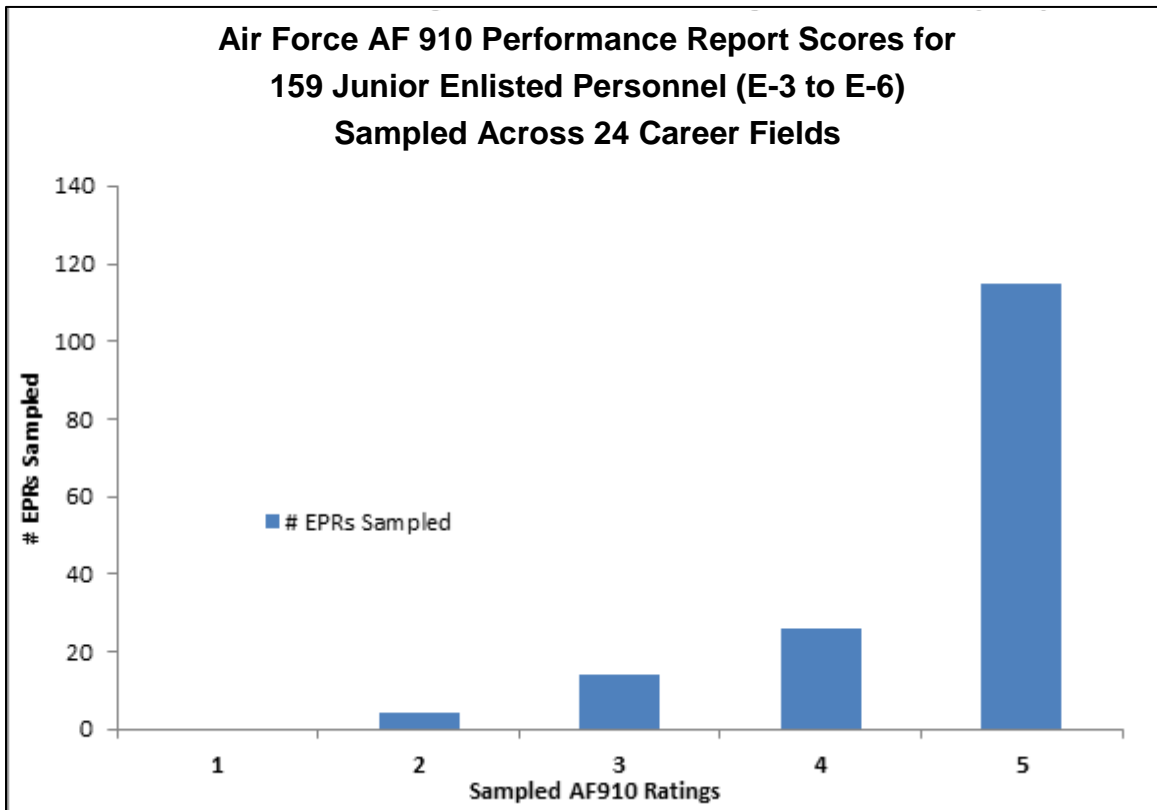


Figure 45. Distribution of 159 Performance Ratings (Current EPR System)

Looking at evaluations of the same 159 personnel using the JEPR system as shown in Figure 46; the JEPR system was able to more clearly delineate the population. The graph illustrated a right skewed mound distribution, with two distinct tails, just as the JEPR Training Dataset has shown. Again, the right skewed distribution was indicative that the Air Force values high quality personnel who possess leadership, values, and professional qualities. The mean JEPR score of the population was found to be 74 (out of 100), with a standard deviation of approximately 21. With an alpha of 0.05, with 95% confidence, the mean JEPR score of the population fell between 70 (out of 100) and 77 (out of 100).

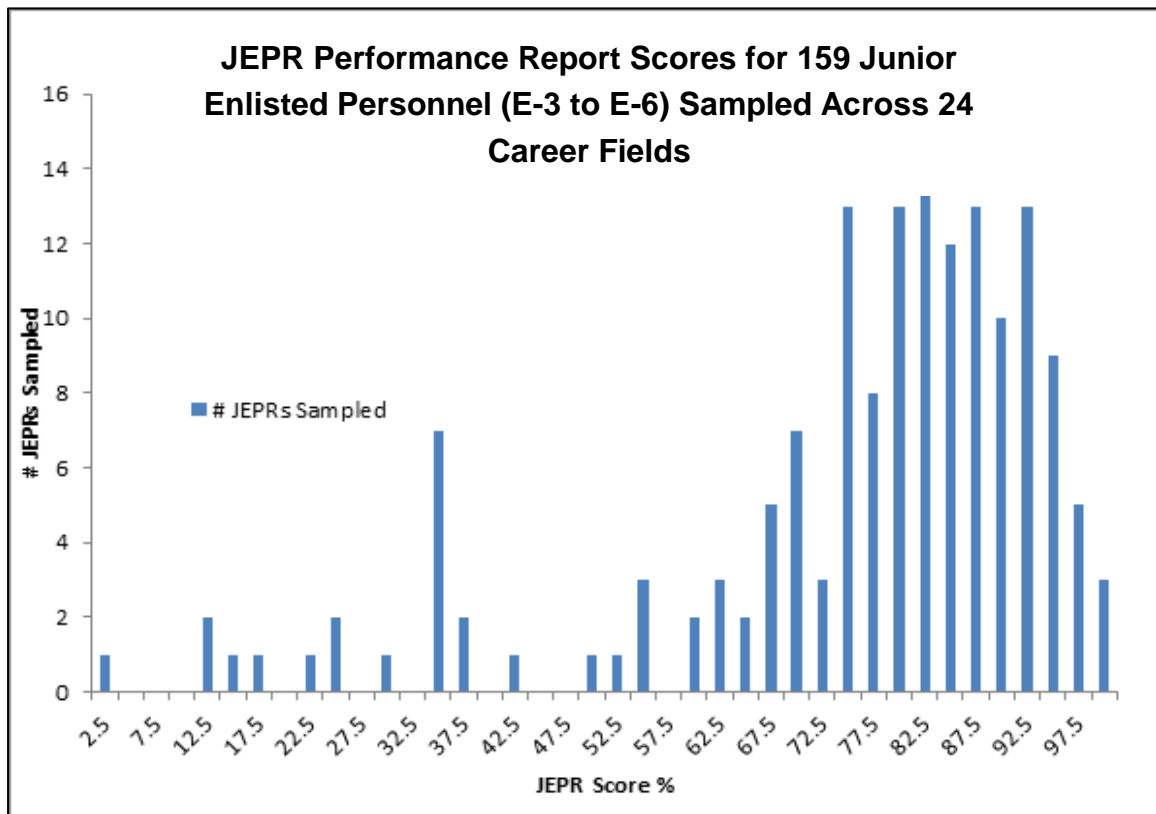


Figure 46. Distribution of 159 Performance Ratings (JEPR System)

Looking further at Figure 46, the long left tail indicated a wide dispersal of airman who scored lower than the population concentration of the JEPR. A review of the scoring of the JEPR attributes and some of the comment bullets that were entered by the supervisor indicated that these test subjects had incurred disciplinary actions, had failed to meet a standard such as Physical Fitness, or had performed poorly in a heavily weighted category such as the Performance in Primary Duties. Review of the attribute scores for a sample of the individuals in the right tail of the distribution indicated strong performances in heavily weighted attributes related to duty performance, with scores in lesser weighted factors providing delineation of outstanding performers. In essence, the right skewed data, the mean, and the confidence intervals indicated the Air Force's desire for a junior enlisted core with high performing individuals (a description of the majority of the Air force junior enlisted population); while the histogram shape indicated the JEPR could delineate these individuals based on the values solicited from the VFT.

For the 115 test subjects who were rated as overall "5"s, or "Truly One of The Best" under the current system EPR system in Figure 47, the histogram illustrated that the mean of JEPR scores from the "Truly One of The Best" was 83 (out of 100), with a standard deviation of approximately 10. With an alpha of 0.05, with 95% confidence, the mean JEPR score for the sub-population of "5" EPRs fell between 80 (out of 100) and 84 (out of 100).

Two low scoring test subjects were observed in the right tail of the distribution. The initial thought was that these subjects did not belong to this population or that the incorrect rating under the current EPR system had been recorded by the supervisor.

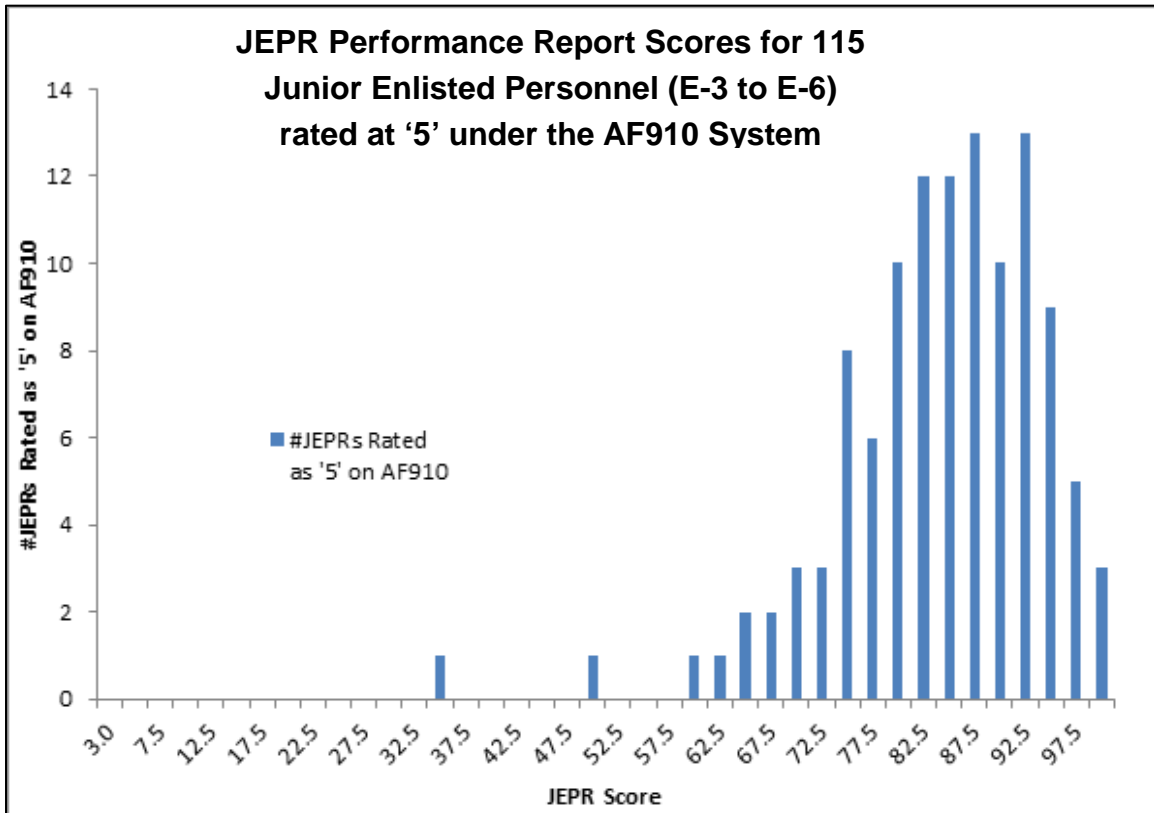


Figure 47. JEPR Distribution Ratings for Subjects Rated “5” (Current EPR System)

However, in discussion with the member’s supervisors, the supervisors revealed that the EPR rating that was recorded was correct. Further research of the JEPR data revealed that the supervisors had evaluated these members as below average in the Duty Performance and Duty Leadership attributes. Additionally, the members also had documented Administrative Actions recorded in their JEPR appraisals. Looking back at these members ratings under the current EPR system, the supervisors stated that they had felt that the rates’ strong performance in other categories had offset weaker performance in duty related areas, justifying the EPR appraisal rating. Had these points been excluded for not belonging to this population, the mean JEPR score for members rated as “Truly One

of The Best” under the current EPR system would have been approximately 83.4 (out of 100), with a standard deviation of 8.81.

Even with the two outliers included, the mean JEPR score for the “Truly One of The Best” sub-population was 9% higher than the mean JEPR scores for the entire sampled population. This indicated that the test subjects who were rated as “Truly One of The Best” appeared to be better performers. The ability of the JEPR to delineate near-peer airmen for this sub-population was clearly evident as illustrated in Figure 47 where the multiple scoring bins of the histogram showed a wide spread, with large counts of test subjects located in bins near the mean with much smaller counts of observations noted in the bins located in the tails of the distribution.

Internal Consistency (Test Dataset)

As had been done with the JEPR Training Data, Cronbach’s alpha was computed to measure the internal consistency. Recall from Chapter IV, that George and Mallery, as cited by (J. Gliem & R. Gliem, 2003), provided the basic rules of thumb shown in Table 56 for classifying the quality of a Cronbach’s alpha value.

Table 56. Cronbach's Alpha Value Quality for Internal Consistency

Cronbach's α Value	Description
≥ 0.9	Excellent
≥ 0.8	Good
≥ 0.7	Acceptable
> 0.6	Questionable
≥ 0.5	Poor
< 0.5	Unacceptable

For the JEPR Test Dataset, the overall Cronbach's alpha was 0.7988 as shown in Table 57. This was an increase of 0.0124 over the Cronbach's alpha calculated earlier using the JEPR Training Dataset. Helms et al., as cited by Spiliotopoulou, noted that increasing the number of participants measured can increase the Cronbach's alpha value, as additional samples increase the covariance (Spiliotopoulou, 2009). Therefore, the increase in the Cronbach's alpha value between the JEPR Training Dataset and the JEPR Test Dataset can be attributed to the increase in sampled population from 71 to 159.

Table 57. Raw Cronbach's Alpha Measures (Overall and with Excluded Attributes)

JEPR Model Cronbach's α	
Entire Set	α Value
Overall	0.7988
Excluded Column	
Duty Performance	0.7983
Duty Leadership	0.7617
Physical Fitness	0.7943
Communication	0.7851
Respect for Service and Standards	0.7741
Discipline and Self-Control	0.7856
Honesty and Accountability	0.7885
Responsibility	0.7878
Teamwork and Followership	0.7920
Military Awards	0.7880
Education Level	0.7935
Base and Community Involvement	0.7963
Administrative (Correction Factor)	0.7613

Looking at Table 57, the systematic exclusion of one attribute at a time showed very little change in the overall Cronbach's alpha value. The largest change occurred when the independent Administrative Actions correction factor was omitted, with the overall Cronbach's alpha value reduced by -0.0375 to 0.7613. The minimal changes

noted in the Cronbach's alpha values as variables were excluded, confirmed that internal consistency existed between the individual measures in the model. Additionally, the lack of change witnessed between the overall Cronbach's alpha, and the alpha values as variables were excluded, indicated the overall measurement methodology was consistent with very little variation. Therefore, the JEPR Test Dataset Cronbach's alpha value computed for the JEPR Test Dataset was deemed as an "acceptable" alpha value range for measuring the JEPRs internal consistency as defined by George and Mallery. A larger sample size should further increase the Cronbach's alpha value.

Factor Analysis Suitability (Test Dataset)

An initial correlation matrix was generated for suitability testing using data from the 13 JEPR attributes obtained for all 159 JEPR Test Dataset observations. The initial correlation matrix was used to first perform the Kaiser-Meyer-Olkin (Kaiser, 1970) test on the JEPR Test Dataset. The Kaiser-Meyer-Olkin (KMO) test was used as a measure of sampling adequacy by measuring the strength of the relationship among variables (Williams et al., 2012). The KMO index values range from 0 to 1.0, with 0.5 considered the minimum threshold for factor analysis consideration (Williams et al., 2012). SPSS computed a KMO index value of 0.888 for the JEPR Test Dataset, indicating that the data was suitable for factor analysis (Hutcheson & Sofroniou, 1999, p. 225).

Bartlett's Test of Sphericity (Bartlett, 1950) was also performed on the JEPR Test Dataset as a part of Factor Analysis Suitability. The purpose of this test was to verify that correlation existed between the attributes (Merkle et al., 1998; Maciel et al., 2013). If correlation did not exist between the attributes, then attributes are completely unrelated,

and factor analysis is not possible (Merkle et al., 1998; Maciel et al., 2013). Equation 22 illustrates the hypothesis test used to perform the Bartlett's Test of Sphericity.

Alternatives

H_0 : JEPR Data is an Uncorrelated Identity Matrix
 H_a : JEPR Data IS NOT Uncorrelated Identity Matrix

Assumptions

$$\alpha = 0.05$$

$$df = \frac{(\#Attributes^2 - \#Attributes)}{2} = \frac{(13^2 - 13)}{2} = 78$$

Test Statistics

$$\begin{aligned} \text{Bartlett's } s &= \left[-1 * \left[(\#Observations - 1) - \frac{((2 * \#Attributes) + 5)}{6} \right] * \text{Ln}|R| \right] \\ &= \left[-1 * \left[(158) - \frac{((2 * 13) + 5)}{6} \right] * -7.731853024 \right] \\ &= 1181.6848 \approx \chi^2 \end{aligned}$$

$$\chi^2_{(1-\alpha, df)} = \chi^2_{(0.95, 78)} \approx 99.616$$

Decision Rule

if Bartlett's $s \leq \chi^2_{(0.95, 78)}$ conclude H_0
 if Bartlett's $s > \chi^2_{(0.95, 78)}$, conclude H_a

Conclusion

1181.6848 > 99.616
 \therefore conclude H_a with P - Value of 1.0632E - 196 (22)

For the JEPR Test Dataset, the significance p-value for the Bartlett's Test of Sphericity was very small (1.0632E - 196) and was well below the significance threshold of 0.05. This indicated that the JEPR Test Dataset data was correlated, and that the data was suitable for factor analysis (Merkle et al., 1998; Williams et al., 2012).

Preliminary Analysis (Test Dataset)

For the preliminary analysis, the correlation matrix was used to extract initial eigenvalues and initial eigenvectors using the JMP software. Kaiser's method was employed for the initial dimensionality assessment to study how many factors might possibly be retained, even through the reduced correlation matrix would later be generated for the actual dimensionality assessment and for Principal Axis Factoring. Any insight gained at this point concerning dimensionality is only an approximation, as Gorsuch and Horn, as (cited by Fabrigar et al., 1999) noted that Kaiser's rule cannot be used to for factor retention decisions for a reduced correlation matrix. Looking at the eigenvalues from the JEPR Test Dataset, the first two eigenvalues accounted for 60.79% of the overall variation associated in the dataset. Table 58 shows a summary of the eigenvalues from the JEPR Test Dataset and the percentages of variance accounted for.

Table 58. Initial Correlation Matrix[R] Eigenvalues (JEPR Test Dataset)

Eigenvalues of the Initial Correlation Matrix			
Number	Eigenvalue	Percent	Cumulative Percent
1	6.3311	48.701	48.701
2	1.5712	12.086	60.787
3	0.9858	7.583	68.370
4	0.7646	5.882	74.252
5	0.6316	4.859	79.111
6	0.5397	4.152	83.263
7	0.4383	3.372	86.634
8	0.4023	3.094	89.728
9	0.3883	2.987	92.716
10	0.3455	2.658	95.374
11	0.2714	2.088	97.462
12	0.1876	1.443	98.904
13	0.1424	1.096	100.000

The Scree Test was also studied during the preliminary analysis for factor retention (Cattell, 1966). Unlike the two “elbows” or drastic changes in slopes that was noted in the initial Scree Plot for the JEPR Training Data, the Scree plot for the JEPR Test Dataset illustrated in Figure 48 shows that only one “elbow” was present where a dramatic slope change had occurred. This “elbow” occurred on the line segment between the first and second eigenvalues. Although a slight slope change was noted between the second and third eigenvalues, it was not as dramatic. Therefore, based on the preliminary analysis, retention of two factors seemed appropriate, especially since the Final EFA model adopted during the JEPR Training Dataset analysis was a two factor model.

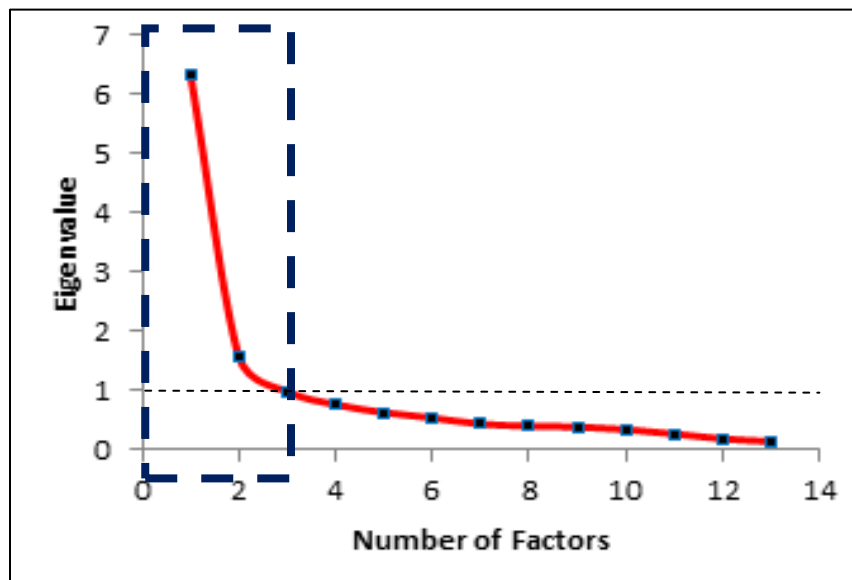


Figure 48. Scree Plot of Initial Eigenvalues from the Initial Correlation Matrix [R]

Data Reduction Technique Selection (Test Dataset)

Since the JEPR Test Dataset was comprised of behavioral science data, the Principal Axis Factoring (PAF) method was selected as the data reduction technique.

Micceri, as cited in (Curran et al., 1996) noted that the preponderance of behavioral research data is not normally distributed. Using PAF provided the ability to focus solely on the common variance portion between factors (D. Tinsley & H. Tinsley, 1987). Figure 49 illustrates where the PAF technique is located on the data reduction techniques tree.

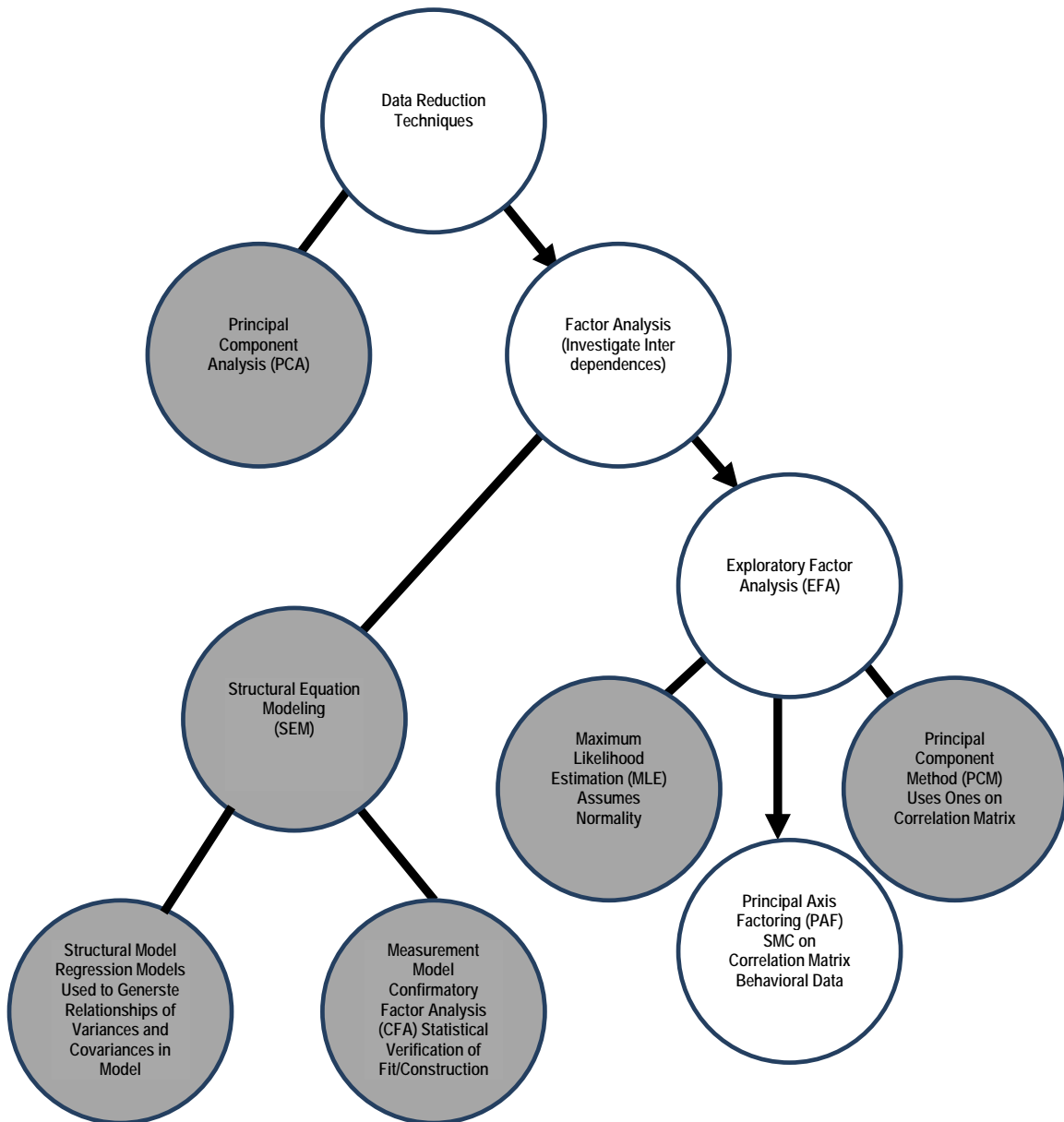


Figure 49. Data Reduction Techniques Tree (EFA Branch Highlighted)

The PAF analysis began by generating initial Squared Multiple Correlations (SMCs) estimates as communalities for the diagonal. The communalities were iteratively recomputed and replaced using regression until the estimates for each attribute converged to a stable value (Floyd & Widaman, 1995). The final estimates were then used as replacements for the variances of the correlation matrix diagonal, to yield the reduced correlation matrix (Henson & Roberts, 2006). The final communalities are reflected in Table 59.

Table 59. Final Community Estimates from Factor Analysis (Test Dataset)

Final Community Estimates	
Attribute	Community Value
Duty Performance	0.6583
Duty Leadership	0.7677
Physical Fitness	0.1711
Communication	0.7023
Respect for Service and Standards	0.4965
Discipline and Self-Control	0.4769
Honesty and Accountability	0.3429
Responsibility	0.5879
Teamwork and Followership	0.6393
Military Awards	0.6106
Education Level	0.5227
Base and Community Involvement	0.4544
Administrative (Correction Factor)	0.5778

The final reduced correlation matrix is represented in Equation 23 and was used for the remainder of the JEPR Training Dataset factor analysis.

$$R^* = \begin{bmatrix} 0.6583 & \frac{SS_{(1,2)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(2,2)}})]} & \dots & \frac{SS_{(1,13)}}{[(\sqrt{SS_{(1,1)}})(\sqrt{SS_{(13,13)}})]} \\ \frac{SS_{(2,1)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(1,1)}})]} & 0.7677 & \dots & \frac{SS_{(2,13)}}{[(\sqrt{SS_{(2,2)}})(\sqrt{SS_{(13,13)}})]} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{SS_{(13,1)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(1,1)}})]} & \frac{SS_{(13,2)}}{[(\sqrt{SS_{(13,13)}})(\sqrt{SS_{(2,2)}})]} & \dots & 0.5778 \end{bmatrix} \quad (23)$$

The new eigenvalues generated from the final reduced correlation matrix which utilized the final communality estimates as the main diagonal entries are reflected in Table 60.

Table 60. Reduced Correlation Matrix [R*] Eigenvalues (Test Dataset)

Eigenvalues of the Reduced Correlation Matrix	
Number	Eigenvalue
1	5.9378
2	1.0707
3	0.4201
4	0.3080
5	0.1717
6	0.0948
7	-0.0349
8	-0.0514
9	-0.0799
10	-0.1027
11	-0.1148
12	-0.1622
13	-0.2304

Dimensionality Assessment (Test Dataset)

Using the eigenvalues in Table 60 from the reduced correlation matrix, it was immediately apparent that only a maximum of six factors could be considered for factor analysis because negative eigenvalues existed for factors seven through 13 (Dillon & Goldstein, 1984, p. 74). As was noted during the analysis of the JEPR Training Dataset,

Gorsuch and Horn, (as cited by Fabrigar et al., 1999), noted that Kaiser’s rule cannot be used to determine dimensionality when communalities are used in a reduced correlation matrix. Therefore, the dimensionality assessment was made using a Scree Plot. The Scree plot is shown in Figure 50.

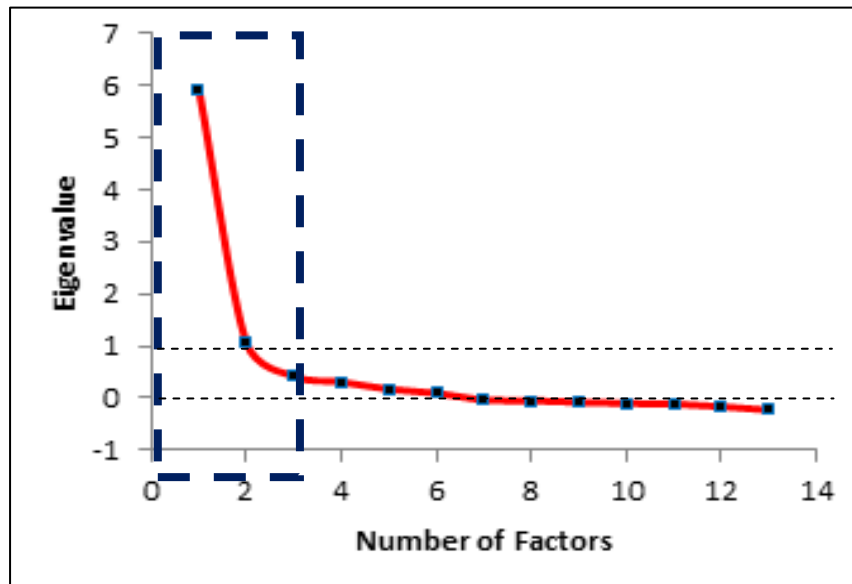


Figure 50. Scree Plot of Eigenvalues from the Reduced Correlation Matrix [R*]

Looking at the Scree Plot in Figure 50, eigenvalues three through six would have provided only minimal additions to the common variance explanation in the JEPR Test Dataset model (Dillon & Goldstein, 1984, p. 74). Therefore, only eigenvalues one and two were retained for the JEPR Test Dataset model, indicating that only two common factors should be retained for EFA (Dillon & Goldstein, 1984, p. 74).

Exploratory Factor Analysis and Interpretation (Test Dataset)

Using the dimensionality assessment, the factor loadings were next generated from the reduced correlation matrix of the JEPR Test Dataset using only two latent factors. The majority of the attributes displayed heavy factor loadings on only the first unrotated factor.

Just as was encountered with the JEPR Training Dataset, the JEPR Test Dataset attribute groupings were not intuitive. The unrotated attribute loadings did not resemble the two factor structure of Standards and Professional Expectations that was theorized during the JEPR Training Data factor analysis that was performed earlier Table 61 reflects the unrotated factor loadings matrix for the JEPR Test Dataset.

Table 61. Unrotated Factor Loadings from the JEPR Test Dataset

Unrotated Factor Loading Matrix		
Objective	Factor 1	Factor 2
Duty Performance	0.79062	-0.18231
Duty Leadership	0.86788	-0.12043
Physical Fitness	0.30320	0.28137
Communication	0.80673	-0.22689
Respect for Service and Standards	0.70073	-0.07409
Discipline and Self-Control	0.65394	-0.22200
Honesty and Accountability	0.57413	-0.11507
Responsibility	0.75172	-0.15119
Teamwork and Followership	0.77075	-0.21273
Military Awards	0.57239	0.53198
Education Level	0.55372	0.46481
Base and Community Involvement	0.46319	0.48977
Administrative (Correction Factor)	0.74870	0.13146

In an effort to better identify the latent constructs described by the JEPR Test Dataset attributes, the unrotated loadings were rotated orthogonally. A Varimax rotation

was utilized for the rotation, with all attributes with factor loadings greater than or equal to 0.40 being considered statistically significant. The orthogonally rotated loadings are shown in Table 62, with the highest loading value for each variable shown in bold.

Table 62. Orthogonally Rotated Factor Loadings

Factor Analysis Settings Technique #1 (Orthogonal)		
Factoring Method	Principal Axis Factoring	
Prior Communality	Common Factor Analysis (SMC)	
Factors Selected	2	
Rotation Method	Varimax	
Significance Threshold	=> 0.4	
Objective	Standards	Professional Expectations
Duty Performance	0.77541	0.23887
Duty Leadership	0.81120	0.33116
Physical Fitness	0.12115	0.39550
Communication	0.81171	0.20839
Respect for Service and Standards	0.64336	0.28740
Discipline and Self-Control	0.67708	0.13597
Honesty and Accountability	0.55440	0.18844
Responsibility	0.72615	0.24629
Teamwork and Followership	0.77347	0.20258
Military Awards	0.22832	0.74733
Education Level	0.24586	0.67986
Base and Community Involvement	0.15502	0.65604
Administrative (Correction Factor)	0.58176	0.48928

As theorized during the JEPR Training Dataset factor analysis effort, the JEPR Test Dataset aligned with the latent factors of Standards and Professional Expectations that had previously been described by the SMEs. Of particular note, the Administrative Actions correction factor crossloaded on both factors when orthogonally rotated. The orthogonal rotation was able to account for 53.91% of the common variance between the

two factors. Additionally, an oblique rotation was performed on the data as shown in Table 63.

Table 63. Two Factor JEPR Model Oblique Rotated Factor Loadings

Factor Analysis Settings Technique #2 (Oblique)		
Factoring Method	Principal Axis Factoring	
Prior Commuality	Common Factor Analysis (SMC)	
Factors Selected	2	
Rotation Method	Promax	
Significance Threshold	=> 0.4	
Objective	Standards	Professional Expectations
Duty Performance	0.80793	0.00625
Duty Leadership	0.81772	0.09953
Physical Fitness	0.00498	0.41089
Communication	0.85986	-0.04114
Respect for Service and Standards	0.64015	0.10728
Discipline and Self-Control	0.73006	-0.07764
Honesty and Accountability	0.57168	0.02467
Responsibility	0.74895	0.03170
Teamwork and Followership	0.81800	-0.03462
Military Awards	0.00871	0.77663
Education Level	0.05166	0.69336
Base and Community Involvement	-0.04440	0.69740
Administrative (Correction Factor)	0.50119	0.35954

An oblique solution creates a simpler, more accurate, and recognizable representation of the relationships between the attributes (Costello & Osborne, 2005; Fabrigar et al., 1999). The oblique rotation identified the same dominant attributes, with only minor differences in loading values. The oblique solution better separated the crossloading correlations in the Administrative Actions correction factor; with factor one inheriting some of the correlation from factor two. However, there was still considerable crossloading noted. With the same simple structure identified in both the oblique and orthogonal rotations, the belief that the factors were uncorrelated was confirmed, as the

orthogonal and the oblique rotations produced nearly identical results (Costello & Osborne, 2005).

Regardless of the rotation method chosen, the loadings of JEPR Test Dataset attributes clearly aligned with the latent factors of Standards and Professional Expectations that had previously been described by the SMEs and discovered during the factor analysis of the JEPR Training Dataset earlier. The delineation of the two latent factors and the alignment of the JEPR attributes can clearly be seen in Figure 51 where the dashed dividing line shows the separation of the Standards factor and the Professional Qualities factor.

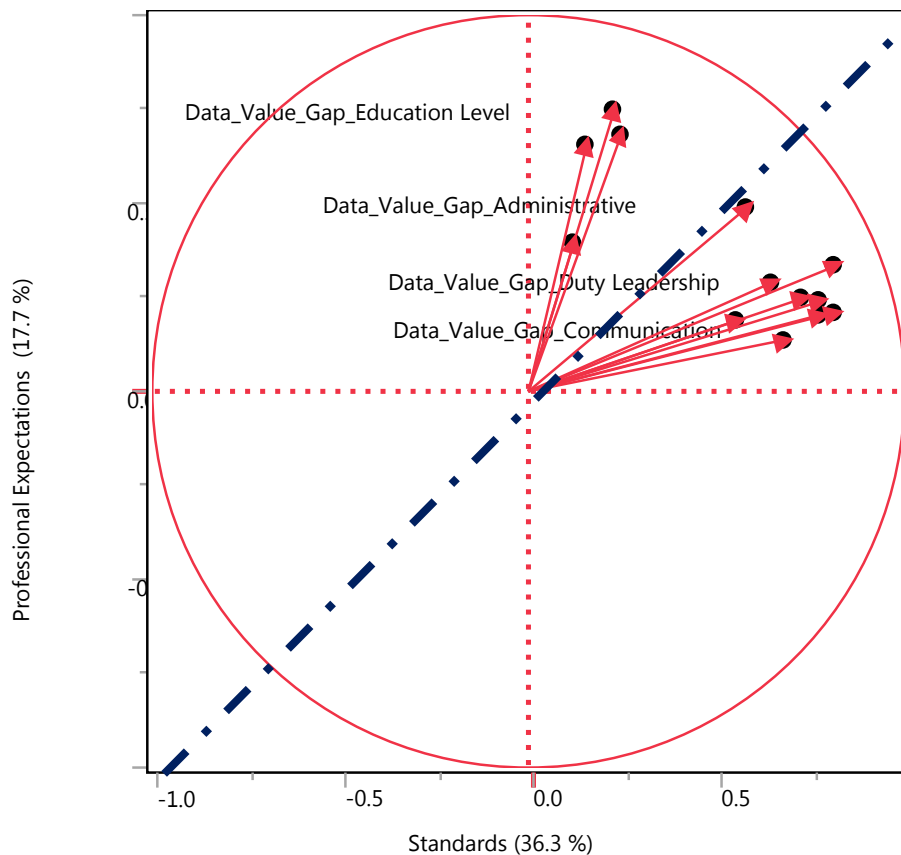


Figure 51. Factor Loading Plot for Two Factor JEPR Model (Rotated Orthogonally)

Although the Physical Fitness attribute did display correlation to the Standards factor, it was more strongly correlated to the Professional Expectations factor. This observation supported the earlier assumption that Physical Fitness belonged in the Professional Expectations factor, due to the JEPR giving incremental scoring increases for Physical Fitness scores that exceed the minimum passing standard. Thus, the two factor EFA model found during the EFA of the JEPR Test Dataset mirrored the two factor model found during the EFA of the smaller sample sized JEPR Training Dataset. This confirmed that the SMEs were correct with their assumption that two latent factors existed underneath the VFT Framework. However, the EFA effort did more than simply identify and confirm the theorized factor construct. The JEPR Test Dataset EFA also identified a potential problem which would have caused problems in the Confirmatory Factor Analysis (CFA) effort (Farrell & Rudd, 2009).

As noted earlier during the EFA on the JEPR Test Dataset, the loadings values identified that the Administrative Actions correction factor was crossloading. Crossloading occurs when a variable loads at a value of 0.32 or higher on two or more factors (Costello and Osborne, 2005). A variable that crossloads is deemed a prime candidate for removal from subsequent analysis, as their retention can adversely affect the fit of the model (Farrell & Rudd, 2009). Looking back at Table 62 for the orthogonal rotation of the two factor model and Table 63 for the oblique rotation solution, both rotation types indicated that the Administrative Actions correction factor was crossloading, with factor loadings well above 0.32 on both factors for this attribute. It was intuitive that the Administrative Actions correction factor would crossload, as this variable was independent of the VFT Framework, and was *applied to the JEPR Overall*

Score after the fact to capture negative quality indicators. Therefore, the Administrative Actions correction factor was not included in the subsequent CFA effort for the JEPR Test Dataset.

Finally, it is worth noting that the same significant factor loadings structure that was identified during the EFA effort for of the larger JEPR Test Dataset that was identified during the JEPR Training Dataset EFA effort. The same results revealed that the design was independent of sample size, career fields sampled, or the supervisor. This validated that the VFT Framework design was consistent in both the computation of appraisal scoring and in the application and interpretation of the appraisal process by supervisors.

Structural Equation Modeling Overview

Structural Equation Models (SEMs) are multivariate, multi-equation regression models where the response variable in one regression equation may be a predictor in another equation; creating causal relationships among variables in the model (Fox, 2002). Within the SEM construct exists two separate models: the structural model and the measurement model (Byrne, 2009, p. 12). A complete SEM model is illustrated in Figure 52.

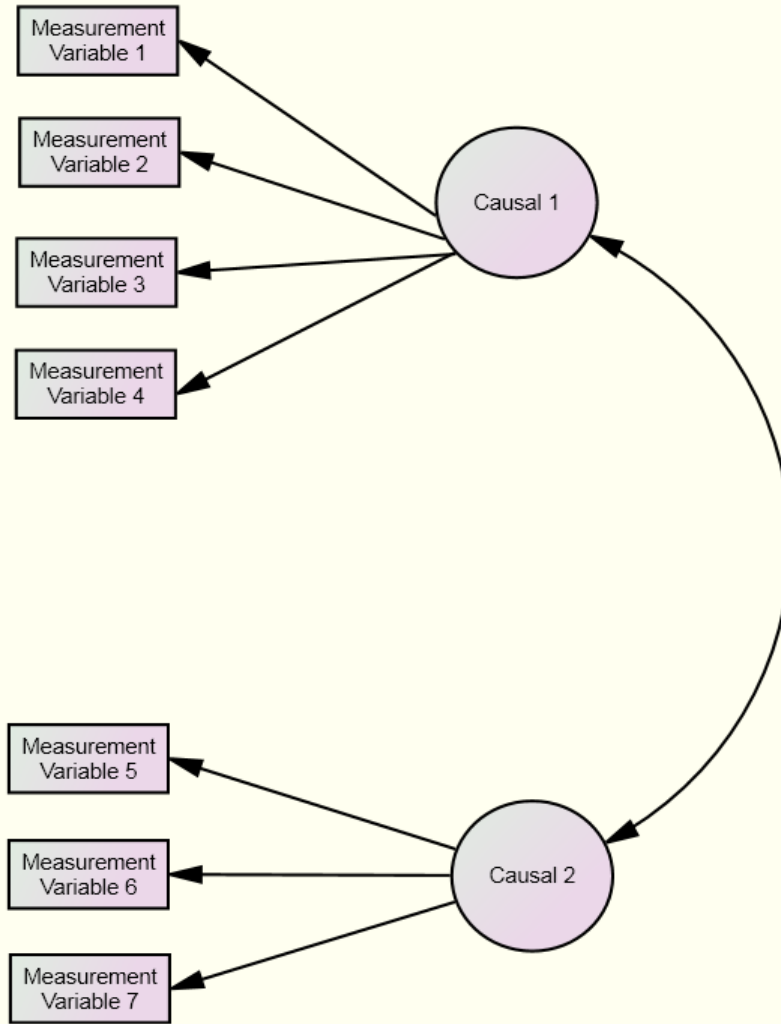


Figure 52. Causal Model with Measurement and Structural Sub-Models

The structural model describes the predicted relationships between latent factors and observed variables of the SEM model, and then compares the results versus the hypothesized model (Byrne, 2009, p. 13; Hatcher, 1996, p. 256; Schreiber et al., 2006). To perform this comparison, the previously mentioned regression models are used to generate directional arcs reflecting relationships of variances and covariances between the measurement models variables and the latent (causal) factors (Hatcher, 1996, p. 256). The arcs shown in Figure 52 between the causal factors and the measurement variables

reflect variances, while the arc between the causal factors reflects the covariance between the factors.

The measurement model describes the relationships and patterns between the observed and unobserved variables. Researchers utilize the measurement model to verify that the variables and structural relationships accurately reflect the desired structure (Hackett, 1996, p. 256; Jackson et al., 2009; Schreiber et al., 2006). During the analysis of the measurement model, researchers study the factor loadings, variances, and modification indices, in an attempt to generate a model that better describes the observed construct statistically (Schreiber et al., 2006). Confirmatory Factor Analysis (CFA) is the measurement model of a SEM (Schreiber et al., 2006).

Confirmatory Factor Analysis Overview (Test Dataset)

Confirmatory Factor Analysis is a special type of hypothesis driven statistical process (Albright & Park, 2006, p. 3). CFA is used to verify the goodness of fit of a hypothesized model which was previously identified during EFA effort. CFA is an iterative process where the factor loadings, variances, covariances, and residual variances, of the original EFA model are constrained or relaxed in search of a statistically valid and intuitively relevant model. Each of the model iterations are evaluated for goodness of fit using a myriad of statistical tests to test for validity. In the field of Psychology, CFA is used to study the relationships between underlying hidden factors and measurable observed attributes such as attitudes, traits, intelligence, clinical disorders (Jackson et al., 2009). Brown noted, as cited in (Jackson et al., 2009) that CFA is often used by the psychological research community to create, validate, or refine measurement

tool constructs and effects discovered during an EFA. With that in mind, the application of CFA was the next logical step in verifying the statistical accuracy of the two factor JEPR construct, as the JEPR also utilizes psychological and social science measurement scales in performing appraisals. In essence, CFA provided a linkage between the management science of conducting appraisals and the behavioral science of measuring psychological, behavioral and social science data. Figure 53 diagrams where SEM and CFA are located on the data reduction techniques tree.

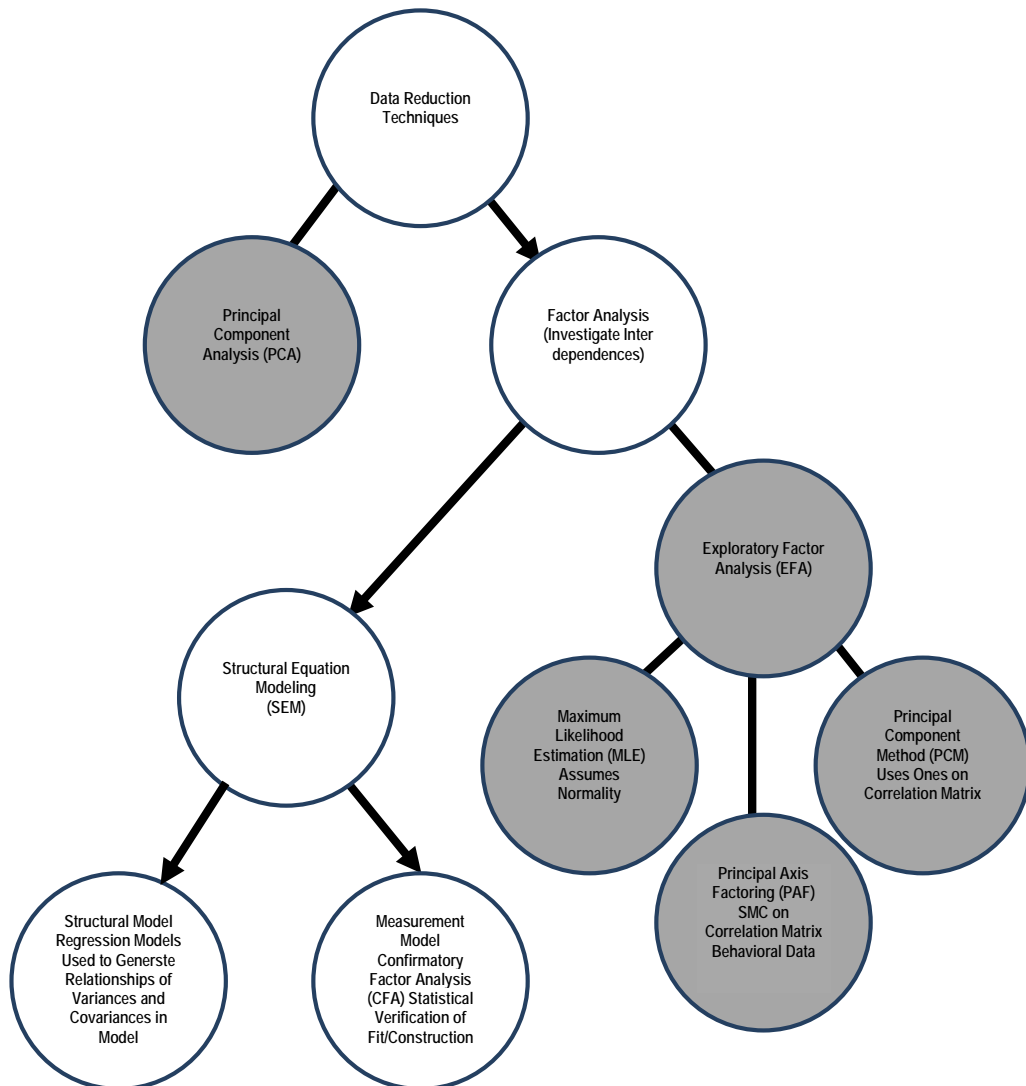


Figure 53. Data Reduction Techniques Tree (CFA Branch Highlighted)

Confirmatory Factor Analysis Data Suitability (Test Dataset)

The goal of Exploratory Factor Analysis is to learn which variables are related, how the variables are related, and to what extent the variables are related (Byrne, 2009, p. 5). In contrast, Confirmatory Factor Analysis (CFA) seeks to test a hypothesized relationship between the variables and latent factors statistically, based on a priori hypothesis about the relationship (Jackson et al., 2009; Byrne, 2009, p. 6). However, before undertaking a Confirmatory Factor Analysis effort, there are several prerequisites for the dataset of interest that must be met (Hatcher, 1994, p. 259-260).

The first prerequisite required for a CFA effort, is that all observed variables in the dataset must be populated (Jackson et al., 2009; Hatcher, 1994, p. 259). McKnight et al. and Schaefer & Graham noted, as cited in (Jackson et al., 2009), that the most common method for dealing with missing data points in preparation of a CFA effort is to use listwise deletion or available case analysis. However, this was not an issue for the JEPR Test Dataset, as all observed variable data fields were populated.

Second, the dataset should be comprised of only continuous data, and the model should contain at least three observed variables per factor, with no more than 20 to 30 observed variables in the model (Hatcher, 1994, p. 259-260). The JEPR Test Dataset met this requirement as well, as the two-factor EFA model was comprised of 12 total observed attributes (variables). Factor one (Standards) was described by eight observed variables, while factor two (Professional Expectations) was described by four observed variables. The Administrative Actions correction factor was not considered for the CFA as it was dropped from the model during the JEPR Test Dataset EFA due to considerable crossloading between factors.

Third, a minimum number of observations must be met (Hatcher, 1994, p. 259). As was stated earlier, a commonly used rule for determining an adequate sample size in research is that the sample must meet or exceed a 10 to 1 sample to observed variable ratio (Schreiber et al., 2006). The JEPR Test Dataset exceeded this rule, with a sample to variable ratio 53 to 4 (13.25 to 1) ratio.

The final prerequisite prior to undertaking a CFA effort is that the complete dataset of all observed variables must exhibit multivariate normality (Byrne, 2009, p. 102; Hatcher, 1994, p. 260). Multivariate normality is vital in applying Structural Equation Modeling (SEM) techniques, of which Confirmatory Factor Analysis is a part of. SEM models rely on Maximum Likelihood Estimate (MLE) or Generalized Least Squares (GLS) for estimations for performing statistical goodness of fit tests for the hypothesized model (Curran et al., 1996). The distortion caused by non-normal data can inflate chi-square test statistics and bias the estimates of the factor loadings that are computed during the CFA regression (Lubke & Muthén, 2004).

One method to determine if multivariate normality is even feasible beforehand is to individually test the data of all the observed variables for univariate normality (Baldwin & Caldwell, 2003). Univariate normality is a prerequisite for the existence multivariate normality (Baldwin & Caldwell, 2003). If any observed variable data field is found to be non-normally distributed, then multivariate normality of the dataset is not possible without transformation. Inspection of the JEPR test data set JMP software revealed that ***none of the JEPR attribute data fields were normally distributed.*** All the empirical distributions from data fields possessed traits of normality, but either demonstrated bi-normal or tri-normal groupings within the attribute or lognormal

behaviors. This was not unexpected, as Micceri, as cited in (Curran et al., 1996), noted the preponderance of behavioral research data does not exhibit multivariate normality nor do the variables follow univariate normality. Looking at the observed empirical distributions for the JEPR Test Dataset using the JMP software, the distributions and parameters for each attribute were identified and summarized in Table 64 and Table 65.

Table 64. Empirical Mixed Distributions and Parameters (JEPR Test Dataset)

JEPR Test Dataset Univariate Mixed Distributions Determinations and Parameters					
Attribute	Duty Leadership	Physical Fitness	Communication	Military Awards	Education
Empirical Distribution Observed	Normal 3 Mixture	Normal 3 Mixture	Normal 3 Mixture	Normal 3 Mixture	Normal 2 Mixture
μ_1	0.055737	0.000166	0.016989	0.000029	0.000406
μ_2	0.069479	0.065152	0.032249	0.014005	0.017346
μ_3	0.093540	0.090198	0.046566	0.030366	
σ_1	0.016467	0.003173	0.006209	0.000744	0.001070
σ_2	0.001910	0.001305	0.003954	0.004812	0.006463
σ_3	0.003974	0.008787	0.002609	0.005902	
π_1	0.340773	0.031447	0.102195	0.225648	0.166476
π_2	0.130410	0.170837	0.330582	0.323492	0.833524
π_3	0.528817	0.797717	0.567223	0.450860	

Table 65. Empirical Johnson SL Distributions and Parameters (JEPR Test Dataset)

Attribute	Duty Perf	Respect for Service and Standards	Discipline and Self Control	Honesty and Acntability	Respsnbly	Teamwrk and Follwrshp
Empirical Distribution Observed	Johnson SL	Johnson SL	Johnson SL	Johnson SL	Johnson SL	Johnson SL
γ	3.007015	0.910318	0.922524	1.155745	0.910732	0.89497
δ	1.055054	0.06739	0.069874	0.06174	0.076369	0.08004
θ	0.408973	0.08	0.05	0.05	0.04	0.03
σ	-1	-1	-1	-1	-1	-1

Although the data appears to be qualitatively non-normal, recent research has shown that some methods of estimation used in CFA are fairly robust to departures from normality (Iacobucci, 2010).

The majority of CFA analysis that has been published has relied on Maximum Likelihood (ML) or Generalized Least Squares (GLS) for estimation (Curran et al., 1996). However, both of these methods are normal theory estimators, with both utilizing the Chi-Square statistic to generate goodness of fit indices (Jackson et al., 2009; Curran et al., 1996). In addition to the Chi-Square goodness of fit test, there are a myriad of additional goodness of fit indices ranging from the Goodness-of-Fit Index (GFI; Jöreskog & Sörbom, 1986), the Root-Mean-Square Error of Approximation (RMSEA; Steiger & Lind, 1980), the Comparative Fit Index (CFI; Bentler, 1990), and the Tucker-Lewis Index (TLI; 1973) to appraise the statistical fit of the hypothesized CFA model (Jackson et al., 2009). Chou et al., Fan & Wang, and Hu all noted, as cited in (Jackson et. al, 2009), that the ML is fairly robust and may be tolerant of mild violations of normality. However, when distributional assumptions are severely violated, both ML and GLS generate inflated Chi-Square values and can potentially generate misleading results concerning the fit of the hypothesized CFA model (Curran et al., 1996).

To statistically evaluate the severity of non-normality, the JEPR Test Dataset was tested for normality using the Analysis of MOment Structures version 18 (AMOS 18; Arbuckle, 2009) software. Using the AMOS software, a representative model was constructed from the JEPR Test Dataset using the orthogonal loadings matrix generated during the EFA effort of the JEPR Test Dataset. The orthogonal matrix was used as both rotation methods had highlighted the same loadings and factor relationships as

significant. As stated earlier, the independent Administration Actions correction factor was omitted from the CFA effort all together due to crossloading (Farrell & Rudd, 2009).

Since univariate normality among attributes is a precursor for multivariate normality, inspection of the individual univariate kurtosis indexes can provide insight into datasets suitability for multivariate normality. As DeCarlo noted, as cited in (Byrne, 2009, p. 103), kurtosis severely impacts variances and covariances. Therefore, kurtotic behavior is of particular concern in SEM analyses such as CFA, as the basis of SEM relies on variance and covariance structures (Byrne, 2009, p. 103). A normal distribution has a standardized kurtosis index value of 3.0 (Byrne, 2009, p. 103). Kline, as cited in (Quilty, Sellbom, Tackett, & Bagby, 2009), indicated that unadjusted univariate skew values greater than 3.0 and kurtosis values greater than 8.0 are indicative of univariate non-normality, and thus multivariate normality. DeCarlo; Kline; West, Finch, & Curran, as cited in (Byrne, 2009, p. 103), noted that most software programs such as AMOS, report a rescaled kurtosis value by subtracting 3.0 from the true kurtosis index making 0.0 as the value indicating normality. Considering the rescaled kurtosis index, West et al., as cited by (Byrne, 2009, p. 103), considered rescaled kurtosis values equal to or greater than 7 as an indicator of departure from normality. Looking at the JEPR Test Dataset assessment of normality data in Table 66, the univariate kurtosis index for the Physical Fitness appears to indicate a departure from normality with a kurtosis index of 7.841, however, it is relatively close to the acceptance threshold, and so further assessment of normality testing was required.

Table 66. Assessment of Normality Data for JEPR Test Dataset

JEPR Test Dataset Assessment of Normality						
Variable	min	max	skew	c.r.	kurtosis	c.r.
Base and Community Involvement	0	0.03	0.419	2.155	-0.717	-1.845
Education Level	0	0.03	-0.184	-0.949	-0.816	-2.1
Military Awards	0	0.04	-0.029	-0.148	-1.228	-3.16
Physical Fitness	0	0.1	-2.426	-12.488	7.841	20.182
Teamwork and Followership	0	0.03	-1.402	-7.217	1.505	3.874
Responsibility	0	0.04	-1.497	-7.704	1.817	4.676
Honesty and Accountability	0	0.05	-1.727	-8.888	1.904	4.901
Discipline and Self-Control	0	0.05	-1.393	-7.171	1.35	3.475
Respect for Service and Standards	0	0.08	-1.368	-7.043	1.243	3.199
Communication	0	0.05	-1.081	-5.564	0.722	1.858
Duty Leadership	0	0.1	-1.031	-5.307	0.669	1.722
Duty Performance	0.082	0.4	-1.266	-6.515	1.098	2.827

The AMOS software also provides a method of testing multivariate normality, through the application of the Mardia's (1970) Multivariate Kurtosis Test (Byrne, 2009, p. 104). Mardia's Multivariate Kurtosis Test is based on standardized fourth moments (Kankainen, Taskinen, & Oja, 2007). To perform Mardia's Multivariate Kurtosis Test, the Mardia's measure of kurtosis had to first be calculated. The measure is generated from the matrix of the centroid distances of the affected data, the inverse covariance matrix from the data, and the transpose of the centroid distance matrix to generate a matrix of Squared Mahalanobis distances (M_{Dist}^2). Each M_{Dist}^2 distance is the squared distances between the vector of an observation and the vector of sample means for all variables, measured in standard deviation units (Byrne, 2009, p. 106; Gao, Mokhtarian, & Johnston, 2008). The Mardia's measure is then generated by summing the squared diagonal entries of the M_{Dist}^2 distances, divided by the number of observations N , yielding $\left(\frac{\sum(M_{Dist}^2)_{ii}^2}{N}\right)$,

and then subtracting $\left(\frac{k(k+2)(n-1)}{(N+1)}\right)$ from the value for N observations and k attributes.

The larger an individual observations $(M_{Dist}^2)^2$ distance is, the greater the contribution to Mardia's measure, and thus the larger the contribution in the departure from multivariate normality (Gao et. al, 2008). Equation 24 through Equation 28 illustrates how the Mardia's measure is calculated.

$$CENTROID = \begin{bmatrix} (Raw_{1,1} - \overline{Raw}_1) & \cdots & (Raw_{1,k} - \overline{Raw}_k) \\ \vdots & \ddots & \vdots \\ (Raw_{k,1} - \overline{Raw}_1) & \cdots & (Raw_{k,k} - \overline{Raw}_k) \end{bmatrix} \quad (24)$$

$$CENTROID^T = \begin{bmatrix} (Raw_{1,1} - \overline{Raw}_1) & \cdots & (Raw_{k,1} - \overline{Raw}_1) \\ \vdots & \ddots & \vdots \\ (Raw_{1,k} - \overline{Raw}_k) & \cdots & (Raw_{k,k} - \overline{Raw}_k) \end{bmatrix} \quad (25)$$

$$\hat{S}^{-1} = \begin{bmatrix} \frac{1}{N} \sum_{i=1}^N ((Raw_{i,1})(Raw_{i,1})) & \cdots & \frac{1}{N} \sum_{i=1}^N ((Raw_{i,1})(Raw_{i,k})) \\ \vdots & \ddots & \vdots \\ \frac{1}{N} \sum_{i=1}^N ((Raw_{i,k})(Raw_{i,1})) & \cdots & \frac{1}{N} \sum_{i=1}^N ((Raw_{i,k})(Raw_{i,k})) \end{bmatrix}^{-1} \quad (26)$$

$$M_{Dist}^2 = ((CENTROID)(\hat{S}^{-1})) CENTROID^T \quad (27)$$

$$Mardia's = \left(\frac{1}{N} * \sum_{i=1}^N [M_{Dist}^2] \right) - \left(\frac{k(k+2)(N-1)}{(N+1)} \right) \quad (28)$$

Using the Mardia's measure, a hypothesis test was performed to determine if the JEPR Test Dataset was multivariate normal by applying Mardia's Multivariate Kurtosis Test.

The null hypothesis was that the data is distributed multivariate normal, while the alternate hypothesis was that the data was not distributed multivariate normal. To determine multivariate normality, Bentler, as cited in (Byrne, 2009, p. 104), suggested that Mardia's measure of kurtosis values greater than 5.00 indicate that the dataset is non-normally distributed. For the hypothesis test, there were $N = 159$ data samples from the $k = 12$ attributes. The hypothesis test for multivariate normality is shown in Equation 29.

Alternatives

H_0 : JEPR Data is asymptotically normally distributed
 H_a : JEPR Data IS NOT asymptotically normally distributed

Assumptions

$\alpha = 0.05$

Test Statistics

$$\begin{aligned} \text{Mardia's } s &= \left(\frac{1}{N} * \sum_{i=1}^N [M_{Dist\ ii}^2]^2 \right) - \left(\frac{k(k+2)(N-1)}{(N+1)} \right) \\ &= \left(\frac{1}{159} 40366.29417 \right) - \frac{12(12+2)(159-1)}{(159+1)} \approx 87.976 \end{aligned}$$

$$\begin{aligned} \text{Mardia's Critical Ratio} &= \frac{\text{Mardia's } s}{\sqrt{\left(\frac{8(k)(k+2)}{N} \right)}} \\ &= \frac{87.97606}{\sqrt{\left(\frac{8(12)(12+2)}{159} \right)}} \approx 30.260 \end{aligned}$$

Decision Rule

if Mardia's Critical Ratio ≤ 5.00 conclude H_0
if Mardia's Critical Ratio > 5.00 conclude H_a

Conclusion

$30.260 > 5.00$
 \therefore conclude H_a

(29)

With a Mardia's measure of kurtosis value of 87.97606, the critical ratio was determined to be 30.260. Based on insight provided by Bentler, as cited in (Byrne, 2009, p. 104), the critical ratio value of 30.260 for the JEPR Test Dataset severely exceeded the 5.00 threshold, indicating that the data was non-normally distributed.

The lack of multivariate normality was problematic in trying to perform the CFA effort. As stated earlier, the majority of goodness of fit measures associated with SEM and CFA rely on multivariate normality of the data (Jackson et al., 2009; Curran et al., 1996). However, when multivariate normality is violated, the inflated Chi-Square values of goodness of fit measures and indices can overestimate the fit of the hypothesized model (Curran et al., 1996). Therefore other methods were researched in an attempt to reduce the kurtosis and improve the multivariate normality of the data.

One method of reducing the kurtosis associated with multivariate normality is to identify and remove outliers from the hypothesized model (Gao et al., 2008). A common approach to identifying potential outliers is to use the diagonal values from the Squared Mahalanobis distances M_{Dist}^2 matrix (Byrne, 2009, p.104-105; DeCarlo, 1997). Just as the case was for the $(M_{Dist}^2)^2$ values of the individual observations during the Mardia's measure calculations, larger M_{Dist}^2 distances also increase multivariate kurtosis, and thus also increase Mardia's measurement value, adversely impacting multivariate normality (Gao et al., 2008). Using the AMOs software, the M_{Dist}^2 diagonal values were computed for all 159 data samples from the JEPR Test Dataset. AMOS generated a table of the M_{Dist}^2 diagonal values by size in decreasing order with two separate p-values to evaluate outliers. The output of the AMOS outlier table is shown in Table 67.

Table 67. Outlier Test of JEPR Test Dataset Using (M_{Dist}^2) Distances

Order (Largest M_{Dist}^2 Size)	Sample# (N)	Mahalanobis Distance Squared (M_{Dist}^2)	p1 value	p2 value
1	135	55.258	0.000	0.000
2	11	49.088	0.000	0.000
3	44	44.781	0.000	0.000
4	159	41.997	0.000	0.000
5	33	39.623	0.000	0.000
6	18	36.778	0.000	0.000
7	157	36.152	0.000	0.000
8	133	34.769	0.001	0.000
9	119	34.561	0.001	0.000
10	158	33.987	0.001	0.000
11	36	32.893	0.001	0.000
12	23	31.817	0.001	0.000
13	129	31.700	0.002	0.000
14	8	30.010	0.003	0.000
15	120	29.058	0.004	0.000
16	138	28.740	0.004	0.000
17	156	28.651	0.004	0.000
18	60	27.784	0.006	0.000
19	34	26.751	0.008	0.000
20	59	25.154	0.014	0.000
21	150	24.731	0.016	0.000
22	153	24.397	0.018	0.000
23	146	23.706	0.022	0.000
24	100	23.703	0.022	0.000
25	105	22.866	0.029	0.000
26	142	21.960	0.038	0.000
27	144	21.684	0.041	0.000
28	97	21.214	0.047	0.000
29	132	20.925	0.051	0.000
30	149	20.889	0.052	0.000
31	15	20.580	0.057	0.000
32	123	17.350	0.137	0.016
33	31	16.455	0.171	0.135
...
158	45	6.593	0.883	1.000
159	38	6.559	0.885	1.000

Looking at Table 67, the $p1$ value is the probability that the point of interest or any other point exceeded the M_{Dist}^2 value for that particular sampled point assuming normality (Arbuckle, 2009). The $p2$ column is the probability that the largest M_{Dist}^2 would exceed the M_{Dist}^2 value computed for particular data point sampled (Arbuckle, 2009). Small $p1$ values are anticipated, however, a small $p2$ value indicates that the sampled point is improbably far from the centroid of the dataset under the assumption of normality (Arbuckle, 2009). Based on the $p2$ values generated by AMOS, 32 of the 159 (approximately 20 percent) of the JEPR Test Dataset could be possible outliers and may be greatly impacting multivariate normality. A graph of the possible outliers with overall JEPR scores plotted against the JEPR classification categories is shown in Figure 54.

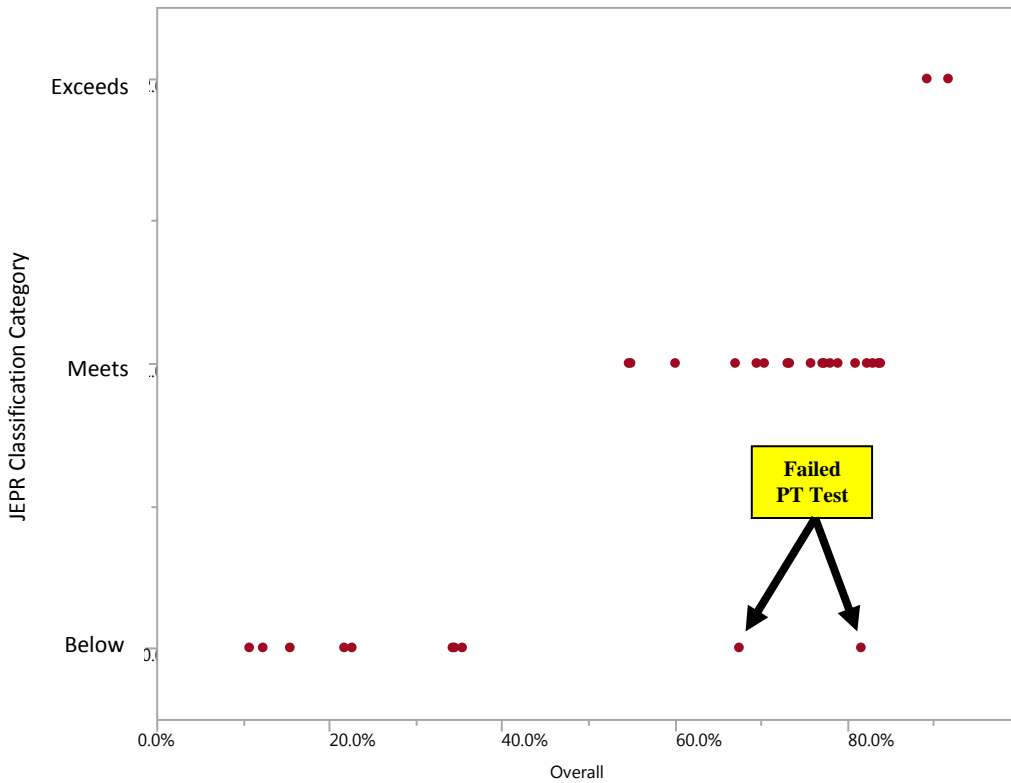


Figure 54. Possible Outlier Points (Overall JEPR Score vs. Classification Category)

Inspecting the individual data elements for each of these 32 possible outliers, no distinguishing abnormalities were identified. The only item of note was that 30 of the 32 identified possible outliers had been classified by the JEPR model as “Below Standards” or “Meets Standards”. Looking at the overall scores versus the JEPR Classification category graphically, this seems to indicate that possibly these points may have been in the tails of the distributions within the sub-populations “Below Standards” or “Meets Standards”. Gao et al. noted that a disadvantage of deleting outliers or possible outliers is that a loss of information and model power occur (Gao et al., 2008). Therefore, without a viable reason to exclude the 32 possible points identified as outliers, the 32 points were retained in the JEPR Test Dataset. Since outlier removal from the JEPR Test Dataset was not possible, other methods were studied to correct the univariate and subsequent multivariate normality issues.

The second method researched for correcting normality of the dataset was transformations. Transformations often can substantially correct univariate skewness and kurtosis when non-normality is severe, thus correcting the multivariate normality of the dataset (Gao et al., 2008). However, if slight normality exists, a transformation by itself is unlikely to rectify deviations from multivariate normality (Gao et al., 2008). Box-Cox transformations were attempted using the JMP software for the 12 JEPR Test Dataset attributes that were selected for the CFA effort. None of the transformations recommended by JMP resulted in a normally distributed univariate dataset for any of the 12 JEPR Test Dataset attributes. Additionally, both logarithmic and power transformations were attempted for transformation of the attributes to a normally distributed dataset. Again, none of the transformations were successful.

In an effort to alleviate the non-normality issue, a technique identified by West et al. and Yung & Bentler known as “the bootstrap”, as cited in (Byrne, 2001) was applied to the JEPR Test Dataset. Using a bootstrap technique enables better estimation of the sampling variance for a statistic, without incurring a normality assumption (Enders, 2005). With the majority of SEM and CFA analysis relying on ML or GLS, the ability to satisfy the assumption of normality is critical in deriving an accurate model that can relate causal relationships between observable variables (Hatcher, 1994, p. 250).

Bootstrapping involves the resampling process of data multiple times, where multiple samples are randomly drawn from the original sample with replacement (Byrne, 2001). The resampling process is replicated many times to in an effort to provide insight as to the variability of the SEM fit statistic and the fit indices (Byrne, 2001). Yung & Bentler noted, as cited by (Enders, 2005), that both the naïve bootstrap and the Bollen-Stine bootstrap (Bollen & Stine, 1992) have been presented in SEM research. Although a naïve bootstrap can generate accurate estimates, it is inappropriate for assessing model fit, as the fit statistics will misfit and fluctuates due to the original datasets covariance structure being inconsistent with the null hypothesis (Enders, 2005). Therefore, the Bollen-Stine bootstrap technique was selected to rectify the JEPR Test Datasets deviation from multivariate normality.

The Bollen-Stine bootstrap, which is used to estimate standard errors and to correct the inflation of the Chi-Square fit statistic due to the non-normality of input data (Enders, 2005). Bollen and Stine, as cited in (Enders, 2005), conveyed that before bootstrapping a dataset, the original data matrix must be transformed. Once transformed, the bootstrap will resample and replicate just as the naïve bootstrap does (Enders, 2005).

Therefore, for the JEPR Test Dataset, a Bollen-Stine bootstrap method was selected to address the non-normality of the data, as this technique can provide a more realistic estimators and standard errors where serious departures of multivariate normality are encountered (Stevanovic, 2009). For the CFA effort on the JEPR Test Dataset, 100 data samples will be subjected to bootstrapping, with 2000 replications generated during the each Bollen-Stine bootstrap application during the testing of the initial hypothesized model and all modified models.

Confirmatory Factor Analysis Evaluation of Fit Criteria (Test Dataset)

As was stated earlier, a CFA is a sub-model within a SEM construct (Byrne, 2009, p. 12-13). This CFA sub-model, which is also known as the measurement model, is focused on analyzing the relationships between the observed and latent variables (Byrne, 2009, p. 12-13). In AMOS, the CFA modeling effort is an iterative process, where modifications to the original model are recommended to improve the overall model fit (Arbuckle, 2009, p. 105). The initial model is constrained with no covariance terms allowed between observed variables. After the regression operation is applied, Modification Indices (MIs) are generated which provide recommendations to improve the fit of the model (Hox & Bechger, 1998). Jöreskog & Sörbom, as cited in (Byrne, 2001) noted that the concept of a miss fitting model can be captured by the Chi-Square statistic, with one degree of freedom. The MI value provided by AMOS for each recommended variable pair indicates the anticipated drop in the Chi-Square if the parameter was freely estimated and allowed to have covariance between error terms (Byrne, 2001; Hox & Bechger, 1998). Parameters are freed one at a time sequentially (Hox & Bechger, 1998).

with the model tested for fit between each modification (Hox & Bechger, 1998). Only parameter sets which are theoretically sound and do not deviate from the theoretical intent of the initial model should be modified (Schreiber et al., 2006). This process is repeated until model fit thresholds are achieved with no significant improvement through modification (Hox & Bechger, 1998). Modification of the original hypothesized model is possible since an observed covariance matrix cannot be perfectly replicated by SEM software.

To appraise the fit of the JEPR Test Dataset CFA model, both absolute and incremental fit indices were selected for reporting. McDonald and Ho noted, as cited in (Hooper, Coughlan, & Mullen, 2008), absolute fit indices indicate how well the appraised model structure fits the sampled dataset. For the JEPR Test Dataset, the Chi-Square statistic (with degrees of freedom and p-value), AGFI, RMSEA, and SRMR will be reported as absolute fit indices. Incremental indices are comparative indices that evaluate the fit of a model by comparing the models Chi-Square value versus a baseline models Chi-Square value (Hooper et al., 2008; Iacobucci, 2010). For the JEPR Test Dataset, the TLI and CFI will be reported as incremental indices. Table 68 provides a listing of the fit indices and a brief description of the indices used to determine the goodness of fit during the CFA effort on the JEPR Test Dataset.

Table 68. Absolute and Incremental Fit Indices Used for CFA (JEPR Test Dataset)

Fit Indices	Fit Type	Type	Range	Description
Chi-Square (χ^2) p-value	Absolute	Goodness of Fit	0 to 1	Used for Goodness of Fit determination
Bollen-Stine p-value	Absolute	Goodness of Fit	0 to 1	Uses adjusted Chi-Square after Bootstrapping for Normality. Used for Goodness of Fit determination
Adjusted Goodness of Fit Index (AGFI)	Absolute	Goodness of Fit	0 to 1	Compares relative amounts of variances and covariances accounted for with a penalty function for degrees of freedom
Root Mean Square Error of Approximation (RMSEA)	Absolute	Goodness of Fit	0 to 1	Measure how well the covariance matrix of the sample population fits due to approximation
Standardized Root Mean Square Residual (SRMR)	Absolute	Badness of Fit	0 to 1	Measures the difference in residuals between covariances of the data and covariances of the model
Tucker-Lewis Index (TLI)	Incremental	Goodness of Fit	0 to ∞	Forces a constrained model where the variables are uncorrelated, error variances are zero, all loadings are fixed to one, where only variables are estimated. Used to address underestimation of the model
Comparative Fit Index (CFI)	Incremental	Goodness of Fit	0 to 1	Compares covariances between the model under test and the null model which is completely uncorrelated

Jöreskog and Sörbom, as cited in (Marsh, Balla, & McDonald, 1988), described the AGFI index as the relative amount of variances and covariances accounted for jointly by the model with a penalty function that adjusts the GFI index based upon degrees of freedom (Hooper et al., 2008; Marsh et al., 1988). The AGFI index usually ranges between zero and one (Hooper et al., 2008; Schermelleh-Engel, Moosbrugger, & Müller, 2003). In SEM modeling, it is common place to classify an AGFI index value of 0.90 or

greater as a good fitting model (Hooper et al., 2008; Schermelleh-Engel et al., 2003). The AGFI absolute fit index is illustrated in Equation 30.

$$AGFI = 1 - \frac{\left(\frac{\chi_{target}^2}{df_{target}} \right)}{\left(\frac{\chi_{null}^2}{df_{null}} \right)} \quad (30)$$

Jöreskog & Sörbom, as cited by (Schermelleh-Engel et al., 2003), described the target model as the model under test, while the null model is more restrictive baseline model, with all parameters set to zero. The $\frac{\chi_{target}^2}{df_{target}}$ component represents the Chi-Square for the target model over the degrees of freedom for the target model, while $\frac{\chi_{null}^2}{df_{null}}$ portion represents the Chi-Square for the target model over the degrees of freedom for the target model (Schermelleh-Engel et al., 2003).

Byrne, as cited in (Hooper et al., 2008) described that the RMSEA index as a measure of how well the model fits the covariance matrix of the sampled population due to differences from approximation. The lower bound of the RMSEA index is zero (Schermelleh-Engel et al., 2003), with an upper bound of one. According to Browne and Cudeck, as cited in (Van Damme, Crombez, Bijttebier, Goubert & Van Houdenhove, 2002), an acceptable close fit for the RMSEA is 0.05, with an upper level of less than 0.08 representative reasonable approximation errors. The RMSEA absolute fit index is shown in Equation 31 where χ^2 and df are for the model under test, while N represents the number of samples (Iacobucci, 2010).

$$RMSEA = \sqrt{\frac{\left(\frac{\chi^2}{df} - df\right)}{df(N - 1)}} \quad (31)$$

The Standardized Root Mean Residual (SRMR) index is a measure of the difference in the residuals between the covariances of the dataset and the covariances of the model under test (Hooper et al., 2008; Iacobucci, 2010). The SRMR measures badness-of-fit index between the model and the data, with larger values indicating a worse model fit (Iacobucci, 2010). The SRMR index ranges from zero to one, with zero indicating a perfect fit between the hypothesized model and the data sample (Hooper et al., 2008). Byrne, along with Diamantopoulos and Siguaw, as cited in (Hooper et al., 2008), identified a SRMR value of 0.05 or lower as a threshold for a good fitting model, with a SRMR index of 0.08 or lower considered the threshold for an acceptable model. The SRMR absolute index value is shown in Equation 32.

$$SRMR = \sqrt{\frac{\left(\sum_{i=1}^p \sum_{j=1}^i \left[\frac{(s_{ij} - \hat{\sigma}_{ij})^2}{(s_{ii}s_{jj})}\right]\right)}{\left(\frac{k(k+1)}{2}\right)}} \quad (32)$$

For the SRMR absolute fit index, the s_{ij} term is an element of the data sample covariance matrix while the $\hat{\sigma}_{ij}$ term is an element of the model covariance matrix (Schermele-Engel et al., 2003). The SRMR is a non Chi-Square based absolute index. The k term in the SRMR absolute index is the number of observed variables in the data sample (Schermele-Engel et al., 2003). The s_{ii} term is a diagonal element of the sample data

covariance matrix, while the s_{jj} term is a diagonal element of the model covariance matrix (Schermelleh-Engel et al., 2003).

The TFI incremental index, also known as the Non-Normed Fit Index (NNFI), was created to address underestimation of model fit from small sample sizes by other incremental indices (Schermelleh-Engel et al., 2003). To correct for small sample size, the TLI considers both the Chi-Square degrees of freedom for the model being appraised, and the Chi-Square degrees of freedom for the independence model (Schermelleh-Engel et al., 2003). The TLI uses the independence model which postulates that the variables are uncorrelated, the error variances in the model are zero, and that all factor loadings are fixed to one (Schermelleh-Engel et al., 2003). This constrained model forces only the variables of the model to be estimated (Schermelleh-Engel et al., 2003). The TLI index is bounded at the lower limit by zero, however, the indices value can become unbounded, and exceed one (Hooper et al., 2008). Hu and Bentler, as cited in (Hooper et al., 2008), suggested that a TLI greater than or equal to 0.95 was indicative of a good model fit. The TLI incremental fit index is shown in Equation 33.

$$TLI = 1 - \left(\frac{\left[\left(\frac{\chi_{indep}^2}{df_{indep}} \right) - \left(\frac{\chi_{target}^2}{df_{target}} \right) \right]}{\left(\frac{\chi_{indep}^2}{df_{indep}} \right) - 1} \right) \quad (33)$$

The CFI incremental index is a variant of the Relative Noncentrality Index (NFI, Bentler and Bonnet, 1980) and is used to compare covariances between the model under analysis and a null model which is completely uncorrelated (Hooper et al., 2008; Schermelleh-Engel et al., 2003). The statistic compares the covariances between the

models using the Chi-Square statistics between the model under test and the independence model for goodness of fit (Hooper et al., 2008; Schermelleh-Engel et al., 2003). The CFI index is an evolution of a class of indices, resolving problems of underestimation, small sample size, and unbounded upper limit (Hooper et al., 2008; Iacobucci, 2010; Schermelleh-Engel et al., 2003). The CFI index ranges from zero to one, with higher values indicating a better fitting model (Schermelleh-Engel et al., 2003). Hu and Bentler, as cited in (Hooper et al., 2008), indicated that CFI index values greater than or equal to 0.95 indicate a good model fit. The CFI incremental index is shown in Equation 34.

$$CFI = 1 - \left(\frac{[\max(\chi_{target}^2 - df_{target}), 0]}{\max[(\chi_{target}^2 - df_{target})(\chi_{indep}^2 - df_{indep}), 0]} \right) \quad (34)$$

For each model modification, the maximum Likelihood (ML) method of estimation was used. In addition to the standard evaluation of fit criteria, the Bollen-Stine adjusted p-value, along with the bootstrapped distribution used for each of the model iterations will be reported. Following standard CFA reporting practices, the p-value for the traditional Chi-Square will be reported, even though it may be possibly inflated due to the non-normality of the raw data (Jackson et al., 2009). The Bollen-Stine p-value will be the statistic used for assessing fit, as the Bollen-Stine Chi-Square value more accurately reflects the true fit of the model since the data was multivariate non-normal. The standardized regression weights (predicted factor loadings); the covariances and correlations due to model modification, and the Squared Multiple Correlations (SMC) values, which as stated earlier during EFA, are the estimated variances, also known as the

communalities of the correlation matrix, will all be reported for each model modification. Additionally, the AMOS covariance modification indices will also be reported for each of the model iterations. These indices recommended additional minor changes, such as allowing covariance to exist between variables, to further improve the fit of model. The inclusion of covariance terms between variables in separate factors will not be considered. Finally, no modifications indices less than 6.00 will be considered for model inclusion. The full AMOS outputs for each model are located in Appendix VI.

Confirmatory Factor Analysis and Interpretation (Test Dataset)

Using the AMOS software, a representative CFA model of the JEPR Test Dataset was built based on the orthogonal loadings matrix generated during the EFA effort. The orthogonal matrix was used for simplicity as both the orthogonal and oblique rotations had provided almost identical solutions and had highlighted the same loadings and factor relationships as significant. Relationships between the observed attributes and the latent factors were built in AMOS based on all loadings values equal to or greater than 0.40. Each observed variable was connected to only one latent factor as is common practice in CFA analysis to control correlations (Beckstead, 2002). The crossloading Administration Actions independent correction factor was omitted from the CFA model (Farrell & Rudd, 2009), as this attribute was crossloading with loadings values in excess of 0.32 on both factors (Costello and Osborne, 2005). Error (residual) terms were also added to the model to capture the unexplained variance by the latent factors during the regression analysis that is performed during creation of the SEM structural sub-model (Beckstead, 2002).

As discussed earlier, the structural sub-model utilizes multivariate, multi-equation regression to create causal relationships among model variables (Fox, 2002). For the initial JEPR Test Dataset model, 12 equations generated to describe the regression paths of the model. The eight equations used to describe the Standards factor are reflected in Equation 35. Equation 36 reflects the regression equations used for the Professional Expectations factor, while Equation 37 reflects the covariance between factors.

$$\begin{aligned}
 \text{Duty Performance}_s &= \mu_1 + (\lambda_{11}) \text{Standards}_s + (0) \text{Prof Expectations}_s + e_{1s} \\
 \text{Duty Leadership}_s &= \mu_2 + (\lambda_{21}) \text{Standards}_s + (0) \text{Prof Expectations}_s + e_{2s} \\
 \text{Communication}_s &= \mu_3 + (\lambda_{31}) \text{Standards}_s + (0) \text{Prof Expectations}_s + e_{3s} \\
 \text{Respect for Serv and Stds}_s &= \mu_4 + (\lambda_{41}) \text{Standards}_s + (0) \text{Prof Expectations}_s + e_{4s} \\
 \text{Discipline and Self Control}_s &= \mu_5 + (\lambda_{51}) \text{Standards}_s + (0) \text{Prof Expectations}_s + e_{5s} \\
 \text{Honesty and Accountability}_s &= \mu_6 + (\lambda_{61}) \text{Standards}_s + (0) \text{Prof Expectations}_s + e_{6s} \\
 \text{Responsibility}_s &= \mu_7 + (\lambda_{71}) \text{Standards}_s + (0) \text{Prof Expectations}_s + e_{7s} \\
 \text{Teamwork and Followership}_s &= \mu_8 + (\lambda_{81}) \text{Standards}_s + (0) \text{Prof Expectations}_s + e_{8s} \quad (35)
 \end{aligned}$$

$$\begin{aligned}
 \text{Military Awards}_s &= \mu_9 + (0) \text{Standards}_s + (\lambda_{12}) \text{Prof Expectations}_s + e_{9s} \\
 \text{Education Level}_s &= \mu_{10} + (0) \text{Standards}_s + (\lambda_{22}) \text{Prof Expectations}_s + e_{10s} \\
 \text{Base \& Comm Involvement}_s &= \mu_{11} + (0) \text{Standards}_s + (\lambda_{32}) \text{Prof Expectations}_s + e_{11s} \\
 \text{Physical Fitness}_s &= \mu_{12} + (0) \text{Standards}_s + (\lambda_{42}) \text{Prof Expectations}_s + e_{12s} \quad (36)
 \end{aligned}$$

$$\text{Factor Covariance} = \text{Standards} * \text{Professional Expectations} \quad (37)$$

Looking at the hypothesized JEPR Test Dataset model in Figure 55, notice that 14 of the 24 paths have regression weights that are fixed at “1”. The AMOS software automatically forces these paths to have fixed regression weights of 1.00, as they are required in order to meet model identification issues and to establish a measurement scale for the unobserved

factors and error terms (Byrne, 2001). Having these paths set to “1” allows for the model to be overidentified, meaning the number of parameters that are estimated is less than the total number of parameters (Byrne, 2009, p. 34-35). An overidentified model results in positive degrees of freedom, which allows hypothesis testing of the model for statistical significance, and if unsatisfactory, the model can be rejected (Byrne, 2009, p. 34-35).

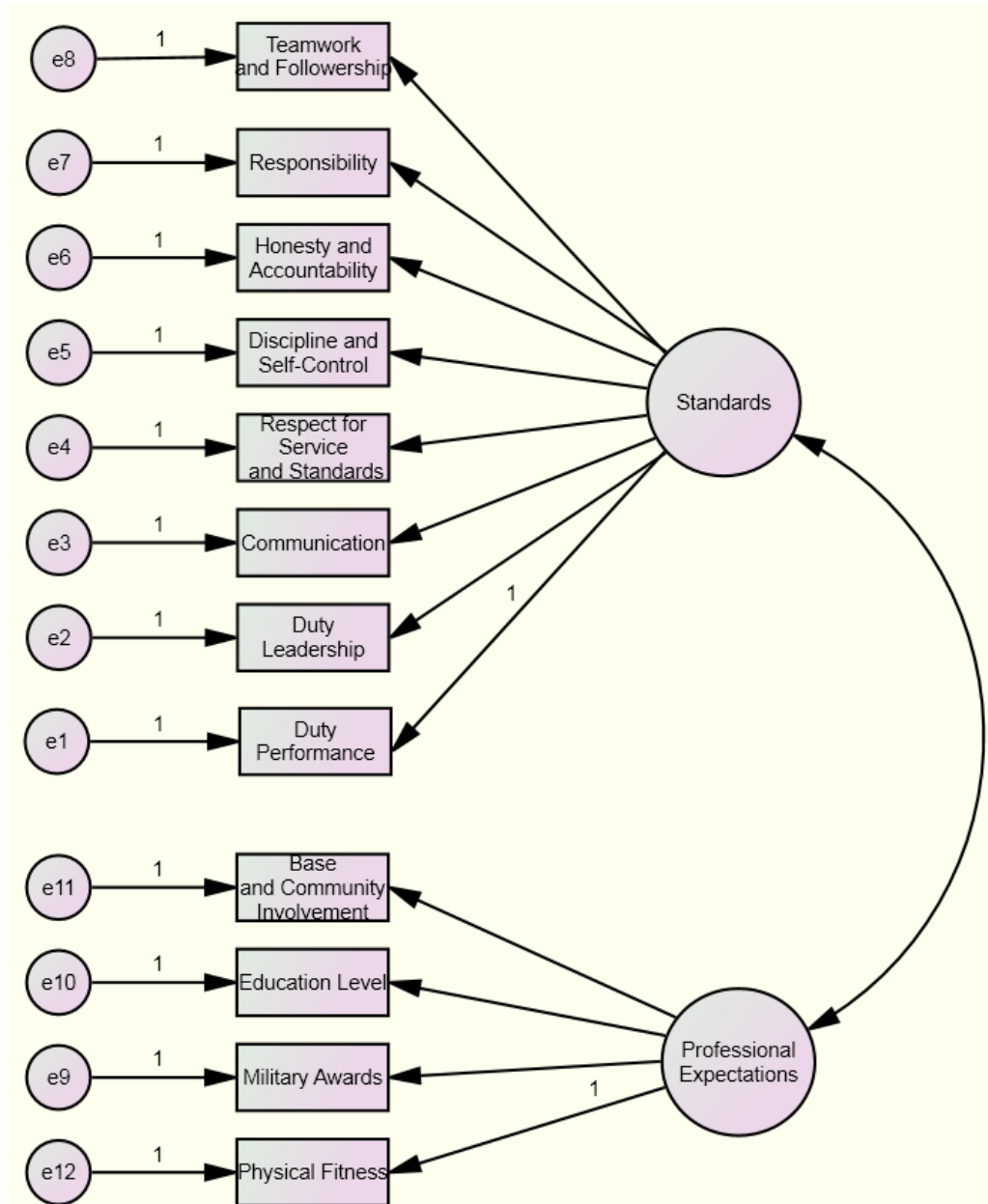


Figure 55. Overidentified SEM Model of JEPR Test Dataset

The Baseline model was run using the AMOS software to perform the CFA computations for the JEPR Test Dataset. The predicted factor loadings and SMCs are illustrated in Figure 56. The black numbers indicate the SMCs while the red numbers indicated the predicted loadings (regression weights) generated by the SEM regression.

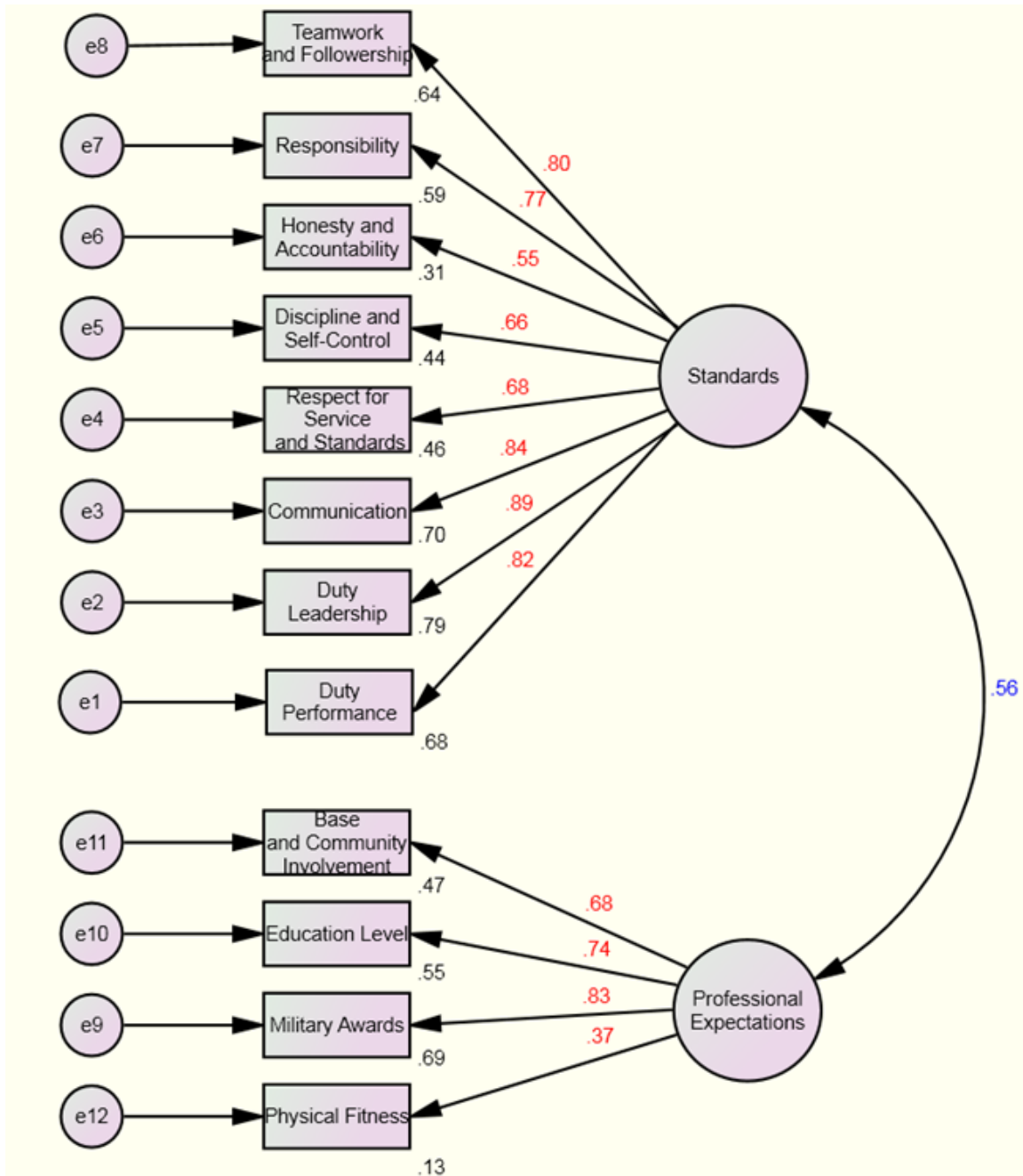


Figure 56. Hypothesized SEM of JEPR Test Dataset (Baseline Model)

The statistical test results of the Baseline Model are shown below in Table 69.

Table 69. Statistical Tests for CFA of JEPR Test Dataset (Baseline Model)

Confirmatory Factor Analysis Evaluation of Fit Criteria (Test Dataset)									
	χ^2	<i>Df</i>	χ^2 p value	Bollen Stine p value	AGFI	RMSEA	SRMR	TLI	CFI
Assumption (α)			0.05	0.05					
Range			0 to 1	0 to 1	0 to 1	0 to 1	0 to 1	0 to ∞	0 to 1
Decision Rule			\geq 0.05	\geq 0.05	\geq 0.90	\leq 0.05	\leq 0.05	\geq 0.95	\geq 0.95
Results of CFA Statistical Tests of JEPR Test Dataset (Baseline Model)									
Estimation Method (Maximum Likelihood)									
Baseline Model	122.615	53	0.000	0.030	0.826	0.091	0.047	0.914	0.931

Looking at the statistical results shown in Table 69, the only testing threshold that was met was for the SRMR statistic. A detailed listing of the baseline models outputs are captured in Appendix VI of this research. Looking at the MI values supplied by AMOS in Table 70, a very large improvement in the models Chi-Square value, 29.443, could be achieved in the model if a covariance could be added between e1, the error term for Duty Performance, and e2, the error term for the Duty Leadership attribute. Adding this relationship is intuitive as these components are measured separately, they may capture measurement features in the same domain, especially in the interpretation by the supervisor generating the JEPR appraisal report. Therefore, adding this covariance was logical.

Table 70. Recommended MIs for CFA of JEPR Test Dataset (Baseline Model)

Attribute Error Term One	Relationship (Covariance)	Attribute Error Term Two	Modification Indices (MI) Value (Hypothesized Improvement in χ^2 Value.)
e12	<-->	e9	5.142
e7	<-->	e8	6.325
e5	<-->	e6	6.137
e4	<-->	e5	6.015
e3	<-->	e8	7.482
e3	<-->	e7	8.986
e3	<-->	e6	5.193
e2	<-->	e6	11.288
e2	<-->	e3	6.262
e1	<-->	e11	8.003
e1	<-->	e9	4.765
e1	<-->	e6	6.788
e1	<-->	e3	4.331
e1	<-->	e2	29.443

This process required three iterations before all the designated goodness of fit criteria for the model were satisfied. The model iterations are included in detail in Appendix VI. The recommended modification indices and modification sequence for each iteration are reflected in Table 71. The overall goodness of fit evaluations for each model iteration are summarized while the Table 72. The highlights of the modification sequence and rationale for the modifications are discussed below.

The baseline model was modified with the AMOS recommended covariance added between e1, the error term for the Duty Performance, and e2, the error term for Duty Leadership, creating the modified model #1. The modified model #1 of the JEPR Test Dataset was again run using the AMOS software. The software generated a revised structural model for the SEM, revising the multivariate multi-equation regression

equations to include the added covariance. AMOS then ran the multivariate multi-equation regression, generating the new factor loadings (regression weights), SMCs (variance estimates), and the covariances for Modified Model #1. The predicted factor loadings, and the predicted SMCs, and predicted correlations for Modified Model #1 are illustrated in Figure 57.

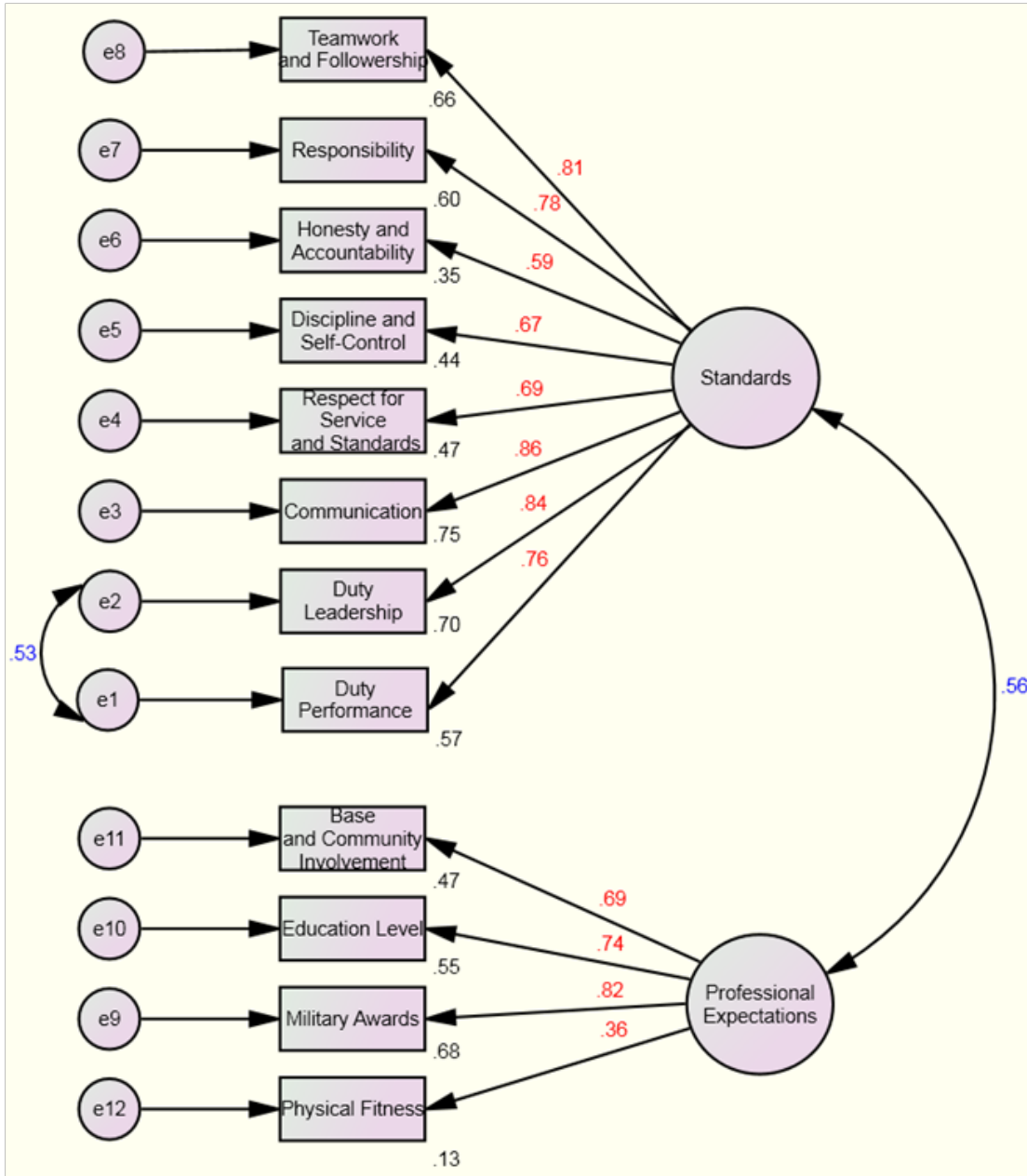


Figure 57. Hypothesized SEM of JEPR Test Dataset (Modified Model #1)

The blue numbers indicate the correlations between attributes, the black numbers indicate the SMCs, and the red numbers indicated the predicted attribute loadings (regression weights) of the model generated by the SEM regression.

Freeing of the model to allow covariance between Duty Performance and Duty Leadership improved the fit of Modified Model #1 substantially. The 0.53 correlation indicated a strong positive relationship existed between Duty Performance and Duty Leadership. The positive correlation indicated that if Duty Performance increased, so would Duty Leadership, and vice versa. Conversely, if Duty Performance decreased, Duty Leadership would be lower, and vice versa. This behavior was substantiated after discussion with the SNO SMEs supporting the JEPR analysis.

The results of Modified Model #1 showed that, in addition to the Bollen-Stine p-value meeting the testing threshold, the SRMR, TLI, and CFI goodness of fit indices also meet their respective thresholds for indicating a well fitting model. For a complete listing of all AMOS data generated for Modified Model #1, refer to Appendix VI section. A review of the AMOS MI values for Modified Model #1 showed that an improvement of at least 10.676 could be achieved in the Chi-Square value of the model if a covariance was added between e7, the error term for Responsibility, and e8, the error term for the Teamwork and Followership attribute. Again, adding this relationship seemed appropriate, as these separately measured attributes may share features in the same measurement domain during an appraisal. Therefore, the JEPR Test Dataset model was modified again from Modified Model #1 to include the covariance term between the Responsibility error term, e7, and the Teamwork and Followership error term, e8. This yielded Modified Model #2. Modified model #2 was run using the AMOS software.

Next, allowing the model to have covariance between the Responsibility and Teamwork and Followership attributes of Modified Model #2 further improved the model fit. This modification improved the fit of the four metrics previously satisfied in Modified Model #1. The -0.41 correlation indicated a strong negative relationship between Responsibility and the Teamwork and Followership attributes. This relation indicated that if Responsibility increased, Teamwork and Followership decreased, and vice versa. This relationship was intuitive, as when an employee's advance to leadership positions, then responsibility increases and they direct actions to subordinates. Conversely members with less responsibility are more reliant on teamwork. Again, this statistical behavior was validated by the SNO SMEs.

For this modification, the AGFI and RMSEA approached, but did not reach the established indices thresholds for Modified Model #2 to be deemed a good fitting model. The Chi-Square p-value also improved with this modification. For the indices already meeting the requirements for a good fitting model, the Bollen-Stine p-value, the SRMR, the TLI, and the CFI indices all substantially improved. Appendix VI details a complete list of data generated by this model. Inspection of the AMOS MI values for Modified Model #2 indicated that an improvement of at least 6.421 in the Chi-Square value could occur if a covariance term was added between the e4 error term for Respect for Service and Standards, and the e5 error term, representing the Discipline and Self-Control attribute. Including this relationship seemed appropriate, as Respect for Service and Standards and Discipline and Self-Control are measured independently, but likely share features in the same measurement domain.

The model was modified again to include a covariance between the Respect for Service and Standards error term and the Discipline and Self-Control the Responsibility error term. The model was run for a fourth time using the AMOS software. Adding the covariance between Respect for Service and Standards and Discipline and Self Control generated a 0.21 correlation, indicating a positive relationship between the attributes. The positive correlation indicated an increase in Respect for Service and Standards would also indicate an increase in Discipline and Self Control values, and vice versa. Conversely, a lower score for Respect for Service and Standards would also be indicative of a lower score in Discipline and Self Control, and vice versa. Again, these statistical observations were substantiated after discussion with the SNO SMEs.

This third model iteration met all the required criteria for a good fitting model as the Bollen-Stine p-value clearly exceeded the acceptance criteria, with all other fit indices exceeding standards for fit. A review of the AMOS MI values for Modified Model #3 showed that only a minor improvement could be gained by adding a covariance between e12 and e9. However, the MI value for this pair was below the 6.00 threshold stated at the beginning of the CFA, and since a statistically valid model had been obtained, this covariance was not included. Therefore, Modified Model #3 was determined to be the Final Model that represented the JEPR Test Dataset.

Table 71 represents the covariances added for each model modification and the associated MI value to achieve the final model.

Table 71. Covariances Added for JEPR Test Dataset CFA Modification

Model Iteration	Attribute Error Term One	Relationship (Covariance)	Attribute Error Term Two	Modification Indices (MI) Value (Hypothesized Improvement in χ^2 Value.)
Modified Model #1	e1	<-->	e2	29.443
Modified Model #2	e7	<-->	e8	10.676
Modified Model #3	e4	<-->	e5	6.421

Table 72 reflects a summary of the statistical tests used in deriving a each iteration of the CFA model enroute to generating the final CFA model.

Table 72. Statistical Tests Summary for CFA of JEPR Test Dataset

Confirmatory Factor Analysis Evaluation of Fit Criteria (Test Dataset)									
	χ^2	Df	χ^2 p value	Bollen Stine P value	AGFI	RMSEA	SRMR	TLI	CFI
Assumption (α)			0.05	0.05					
Range			0 to 1	0 to 1	0 to 1	0 to 1	0 to 1	0 to ∞	0 to 1
Decision Rule			\geq 0.05	\geq 0.05	\geq 0.90	\leq 0.05	\leq 0.05	\geq 0.95	\geq 0.95
Results of CFA Statistical Tests of JEPR Test Dataset									
Estimation Method (Maximum Likelihood)									
Baseline Model	122.615	53	0.00	0.030	0.826	0.091	0.0474	0.914	0.931
Modified Model #1	86.068	52	0.002	0.267	0.88	0.064	0.0432	0.957	0.966
Modified Model #2	71.580	51	0.03	0.485	0.895	0.051	0.0431	0.974	0.980
Modified Model #3 (Final Model)	64.935	50	0.076	0.604	0.905	0.043	0.0420	0.980	0.985

Figure 58 graphically illustrates the final model generated from the CFA effort. The predicted factor loadings (red numbers), and the predicted SMCs (black numbers), and predicted correlations (blue numbers) generated by the SEM regression.

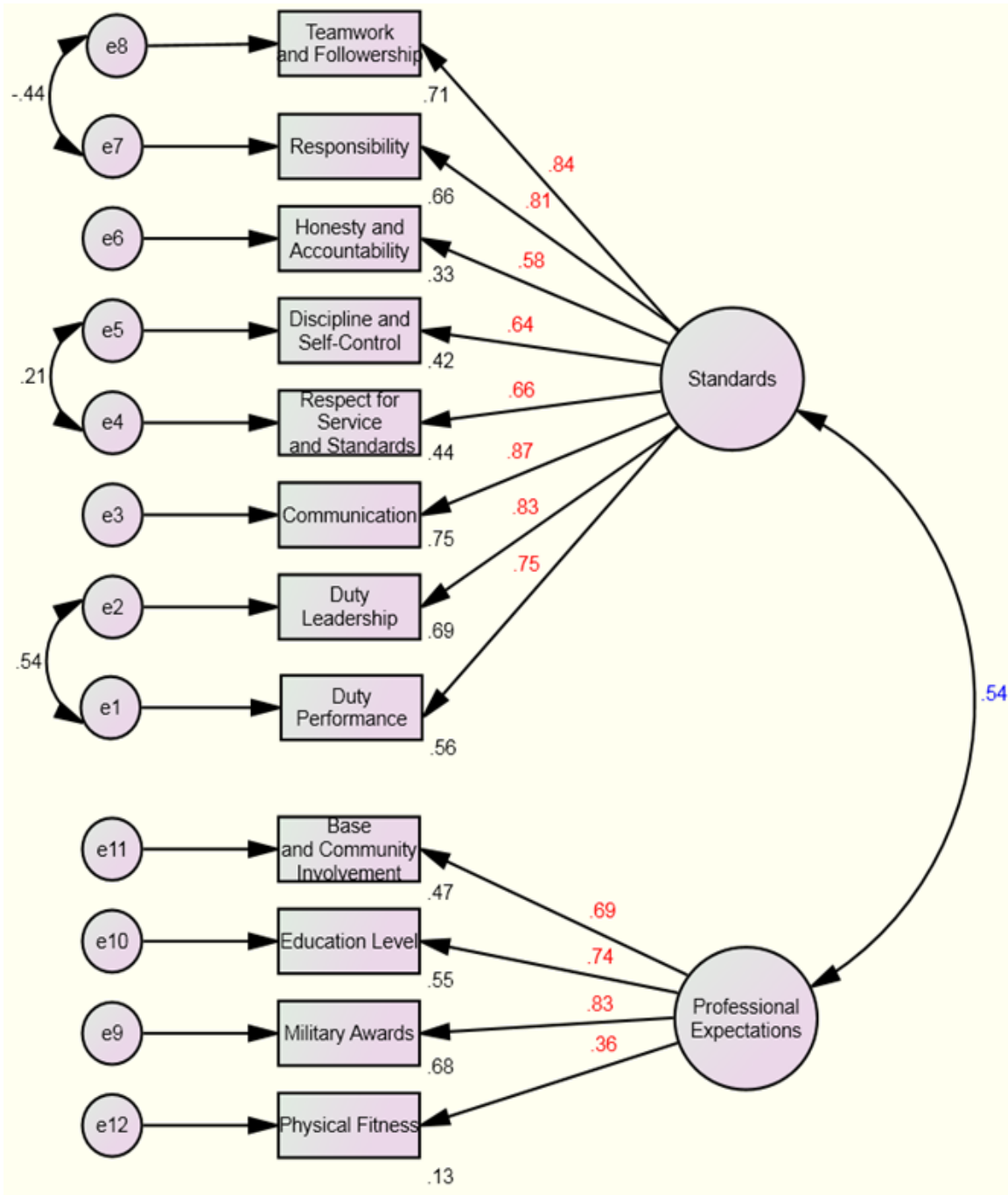


Figure 58. Hypothesized SEM of JEPR Test Dataset (Final Model)

The accuracy of the Final Model can be seen in Table 73 by comparing the rotated EFA loadings with the predicted CFA model regression weights. The maximum difference between the loadings and the weights was noted in the Responsibility attribute. The 10.7% difference was attributed to the sizable correlation (0.24629) which resided in the Professional Expectations factor. The differences between the loadings and weights for all other attributes can be traced to correlations with the less dominant factor.

Table 73. EFA Loadings vs. CFA Regression Weights (JEPR Test Dataset)

Attribute	EFA Orthogonal Loadings (Observed Data)	CFA Predicted Weights (SEM Model Regression)	Difference Between EFA Loadings and CFA Weights	% Diff	Factor
Duty Performance	0.77541	0.74900	0.02641	-3.5%	Standards
Duty Leadership	0.81120	0.83300	-0.02180	2.6%	Standards
Physical Fitness	0.39550	0.36300	0.03250	-9.0%	Professional Expectations
Communication	0.81171	0.86700	-0.05529	6.4%	Standards
Respect for Service and Standards	0.64336	0.66100	-0.01764	2.7%	Standards
Discipline and Self-Control	0.67708	0.64400	0.03308	-5.1%	Standards
Honesty and Accountability	0.55440	0.57800	-0.02360	4.1%	Standards
Responsibility	0.72615	0.81300	-0.08685	10.7%	Standards
Teamwork and Followership	0.77347	0.84200	-0.06853	8.1%	Standards
Military Awards	0.74733	0.82700	-0.07967	9.6%	Professional Expectations
Education Level	0.67986	0.74000	-0.06014	8.1%	Professional Expectations
Base and Community Involvement	0.65604	0.68700	-0.03096	4.5%	Professional Expectations

Qualitative Classification (Test Dataset)

Having statistically verified the JEPR construct with Confirmatory Factor Analysis, the JEPR research transitioned into studying the classification effectiveness of both the JEPR and current EPR systems. The JEPR system was designed with three classification classes in which to group appraisals in accordance with Air Force values and doctrine. These three classes were defined as Exceeds Standards, Meets Standards, or Below Standards. To allow for an analogous comparison between the JEPR system and the current EPR system, the same three class construct was applied to the current EPR system based on inputs from the SNCO SMEs supporting the JEPR research. The three class classification system devised for classifying EPR scores used the same classes, Exceeds Standards, Meets Standards, or Below Standards, as the JEPR model. The SMEs translated the overall EPR rating scheme, doctrine, and subject matter expert experience to devise the classification classes (Air Force Instruction 36-2406, 2013, p. 83). The classifications classes and the description of the classes are reflected in Table 74.

Table 74. JEPR and Translated EPR Classification Classes

Classification Category Class Descriptions (By Appraisal Method)		
Classification Class Name	Translated EPR (Current AF 910) Classification Class Description	JEPR Classification Class Description
Below Standards	Overall Rating ≤"2"	Overall Score ≤45.57 and/or Failure to Meet any Standard in the Standards group of attributes
Meets Standards	Overall Rating >"2 "and ≤"4"	Overall Score >47.57 and <85. Must meet Standards in all attributes in Standards group
Exceeds Standards	Overall Rating ="5"	Overall Score ≥85 Must meet Standards in all attributes in Standards group

With the classes defined, a qualitative analysis using a pivot table was performed to look at how each system classified individuals based on the known overall JEPR score and classification class, and the overall EPR rating and the translated EPR classification class. By contrasting the classification classes against each other using the 159 sample JEPR Test Dataset, insight was gained on inflation and the overall classification of airmen during appraisals.

Table 75. Pivot Table of Translated EPR Classes and JEPR Classification Classes

		Translated EPR Classification Classes			JEPR Totals (By Class)	% JEPR Totals (By Class)
		Below Standards	Meets Standards	Exceeds standards		
JEPR Classification Classes	Below Standards	4	16	1	21	13.2%
	Meets Standards	0	24	61	85	53.5%
	Exceeds Standards	0	0	53	53	33.3%
EPR Totals (By Class)		4	40	115		
% EPR Totals (By Class)		2.5%	25.2%	72.3%		

Table 75 illustrates that 13.2% (21 of 159) of the airmen appraisals sampled, were classified as “Below Standards” by the JEPR system. In contrast, only 2.5% (4 of 159) of the airmen appraisals sampled had were classified as “Below Standards” under the Translated EPR classification system. Table 76 details the classification discrepancies between the two systems and explains the rationale for the JEPR systems classification assignment.

Table 76. Classification Discrepancies (JEPR Below Standards Classification)

# Individuals With Classification Discrepancy	JEPR Classification	JEPR Overall Average Score	EPR Classification	EPR Overall Rating	JEPR Classification Rationale
13	Below Standards	27.70	Meets Standards	3	7 of 13 failed to meet a Standard outlined by doctrine. The average overall JEPR score was 17.87 points below the JEPR “Meets Standards” threshold
3	Below Standards	45.49	Meets Standards	4	2 of 3 failed to meet a Standard outlined by doctrine (Physical Fitness). The 3 rd test subject had a 32.7 overall JEPR score with low Duty Performance (12.3 of 40) and Duty Leadership scores (2.8 of 10), with documented Administrative Actions .
1	Below Standards	33.59	Exceeds Standards	5	Low Duty Performance score and documented Administrative Actions
17	Total				

Looking back at Table 75, the JEPR classification system classified 85 individuals as “Meets Standards”. Of these 85 individuals, the JEPR classification system and the Translated EPR classification system agreed on the classification of 24 individuals. However, 61 of the individuals classified as “Meets Standards” by the JEPR, were

classified as “Exceeds Standards” using the Translated EPR classification system. All 61 of these individuals were rated as “5” or “Truly Among the Best” on the EPR appraisals. Table 77 illustrates the classification discrepancies between the two systems and explains the rationale for the JEPR systems classification assignment.

Table 77. Classification Discrepancies (JEPR Meets Standards Classification)

# Individuals With Classification Discrepancy	JEPR Classification	JEPR Overall Average Score	EPR Classification	EPR Overall Rating	JEPR Classification Rationale
61	Meets Standards	76.40	Exceeds Standards	5	The average overall JEPR score was 8.60 points below the JEPR “Exceeds Standards” threshold
61	Total				

Reviewing the overall JEPR scores for these 61 individuals, the lowest JEPR overall score from these 61 airmen was a 48.3, only 0.72 points from the “Below Standards” classification by the JEPR. The highest JEPR overall score from this sub-population was 84.9, which was 0.1 points from being classified as “Exceeds Standards” by the JEPR system.

Finally, for the JEPR system, 33.2% (53 of the 159 airmen) were classified as “Clearly Exceeds Standards”. The Translated EPR classification system also classified these same 53 individuals as “Exceeds Standards”. However, as stated earlier, the Translated EPR system also classified an additional 62 airmen as “Exceeds Standards”, for a classification rate of 72.3% (115 of 159 airmen).

Inspecting at the JEPR overall scores plotted against the EPR ratings for the same 159 individuals, with the JEPR classification classes overlaid, the graph in Figure 59 clearly shows that the JEPR can delineate between “near peers” through the scoring construct. Additionally, if the current system is truly inflated, as senior Air Force leaders have stated (Losey, Sep 2013), and the JEPR VFT Framework is an accurate representation of Air Force doctrine and values, then JEPR system can substantially reduce inflation. The JEPR systems ability to control inflation is clearly seen in the blue points (between the green and red dotted lines) in Figure 59, where 61 airmen rated as “5” or Truly Among the Best” on their EPRs were classified by the JEPR as “Meets Standards” , with overall JEPR scores ranging from approximately 48.3 to 84.9. As for delineation, under the current EPR construct, all members in each of the ratings categories would receive the same number of promotion points for this rating period from this specific appraisal. However, under the JEPR construct, the individuals ability to test for promotion would be determined by their JEPR classification class, only individuals earning a “Meets Standards” or “Exceeds Standards” would be allowed to test, then their promotion points contributed by this appraisal would equate to their unique overall JEPR score. Table 75 through Table 77 along with Figure 59 clearly illustrated that there is a discrepancy between how airmen are currently evaluated and what the SNCO SMEs identified are important to the Air Force appraising the performance of junior enlisted airmen.

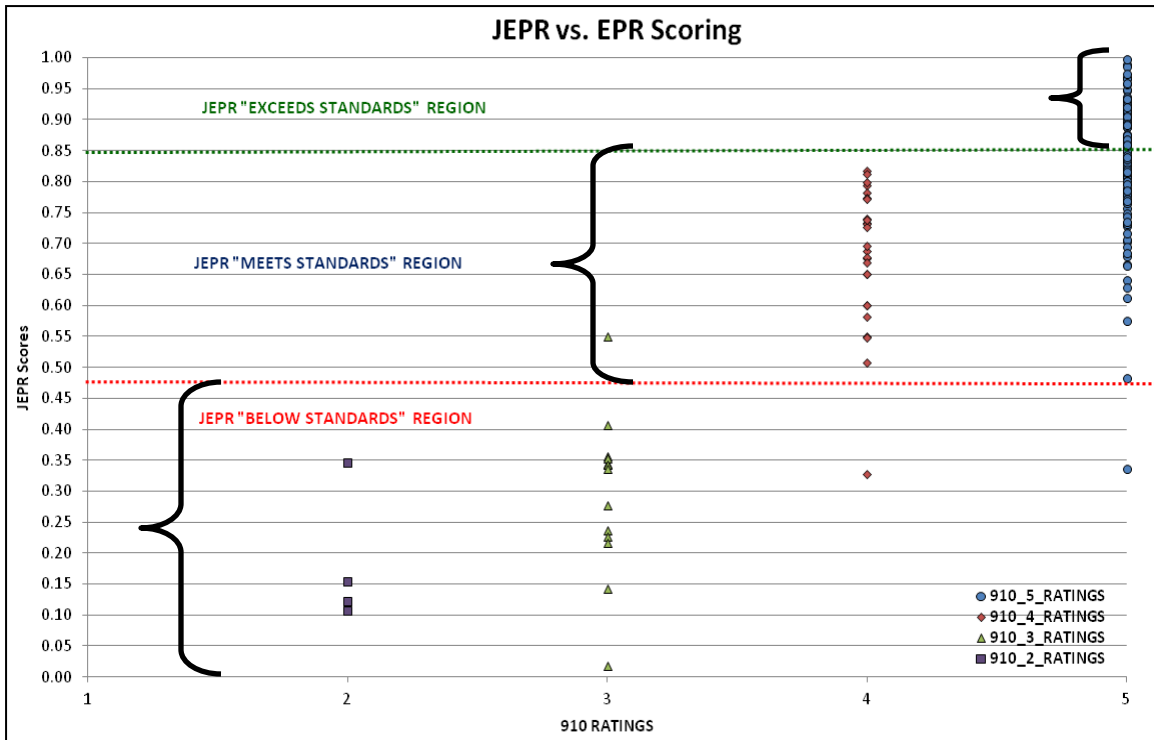


Figure 59. JEPR Versus EPR Scoring (JEPR Classification Classes Overlaid)

Artificial Neural Network Suitability (Test Dataset)

The ability to classify people, items, or ideas into predefined groups or classes based on observed attributes is one of the most essential decision tasks in human activity (Zhang, 2000). Civilian organizations use classification classes to appraise the employee's actual and potential contribution to the success of the organization (Berger & Berger, 2008, p. 7). In-turn, the classification determines the employee's promotion suitability, salary, and further retention. The Air Force is no different, as the Air Force uses its EPR ratings system to classify individuals for promotion, salary, and retention (Air Force Instruction 36-2406, 2013, p. 77).

To simplify classification, observed attributes are often used to assign objects or people into groups or classes that can be described by the attributes (Zhang, 2000). Based on the review of Air Force doctrine, input from the SNO SMEs, and the previous analysis completed in this research, the attributes which comprise the JEPR VFT Framework appear to be accurate observations of traits which the Air Force values as important in the appraisal and classification of airman. However, if these attributes are truly what the Air Force values in their junior enlisted force, then two additional research questions arise.

- 1. Are the values assigned as breakpoints for the JEPR classification system classes the correct points for accurately classifying airmen using the JEPR attributes?**
- 2. How effective is the EPR system at classifying airmen using the JEPR attributes?**

To answer these two research questions, two equivalent classification classes had to be first identified. Looking back at the Qualitative Classification (Test Dataset) section of this chapter, the classification structures used for the pivot table analysis met this criteria. The classification classes are shown in Table 78.

Table 78. JEPR and Translated EPR Classification Classes

Classification Category Class Descriptions (By Appraisal Method)		
Classification Class Name	Translated EPR (Current AF 910) Classification Class Description	JEPR Classification Class Description
Below Standards	Overall Rating ≤ "2"	Overall Score ≤ 45.57 and/or Failure to Meet any Standard in the Standards group of attributes
Meets Standards	Overall Rating > "2" and ≤ "4"	Overall Score > 47.57 and < 85. Must meet Standards in all attributes in Standards group
Exceeds Standards	Overall Rating = "5"	Overall Score ≥ 85 Must meet Standards in all attributes in Standards group

For the first question, a high successful classification rate would support the assumption that the JEPR classification system can accurately classify airman based on the attributes solicited to build the JEPR VFT Framework. On the otherhand, a high misclassification rate would support the belief that the JEPR classification system is ineffective at classifying airmen based on the attributes supplied by the JEPR VFT Framework. One way to test the classification effectiveness of both the Translated EPR and the JEPR classification methods is by using an Artificial Neural Network (ANN).

For the second question, a high classification rate would indicate that the Translated EPR classification scheme which is based on the current EPR rating system, can do an effective job of classifying airmen using the attributes solicited for the JEPR VFT Framework. Conversely, if there is considerable misclassification of airmen, then the Translated EPR classification scheme may not effective at classifying airmen based on the attributes from the JEPR VFT Framework.

Artificial Neural Network Background (Test Dataset)

ANNs have become a popular tool in the reseach community to assess classification accuracy and to determine the probability of correctly classifying future data based on input attributes also known as features (Zhang, 2000). There are several advantages to using ANNs. First, ANNs can model non-normal class distributions and provide better performance over other Bayesian methods (Hunter, Kennedy, Henry, & Ferguson, 2000). Second, traditional Bayesian methods are severely limited by the underlying assumption or conditions determined when they are studied (Zhang, 2000). ANNs on the other hand, are learning classifiers and are adaptive to the data (Zhang,

2000). ANN classifiers can adjust based on what being learned from the data by the ANN, without changing specific function or distributional changes (Zhang, 2000).

The foundation of the ANN architecture is the neuron (Shi, Liu, Kong, & Chen, 2013). The ANN neuron, inspired by the sensory processing abilities of the human brain, is a machine based processing element that can learn with experience (Shi et al., 2013; Krogh, 2008). In the human brain, tasks are accomplished by the transmission of electrical stimuli through a complex interwoven network of neurons (Krogh, 2008). In an ANN, input data is initially weighted randomly, and then the weights are replaced with minimized squared differences between the input and the known output (Krogh, 2008). This process is repeated for each data sample, which gradually reduces the error amount until the error value stabilizes (Krogh, 2008). This method is known as back-propagation (Krogh, 2008). Multiple sigmoid units, which are also known as threshold units as shown in Figure 60, receive weighted input data, and then partially classify the input data based on the known output in a network of hidden neurons (Krogh, 2008).

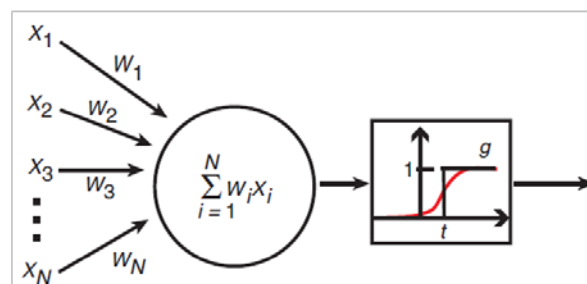


Figure 60. McCulloch-Pitts Model Neuron or Single Threshold Unit (Krogh, 2008)

The partially classified results are sent to an output layer of neurons, where they are reassembled, and receive a final classification determination (Krogh, 2008). This type of

ANN is known as a feed-forward multilayer network or Multi-Layer Perceptron (MLP) network, and is the most widely used ANN for classification of data (Krogh, 2008; Zhang, 2000).

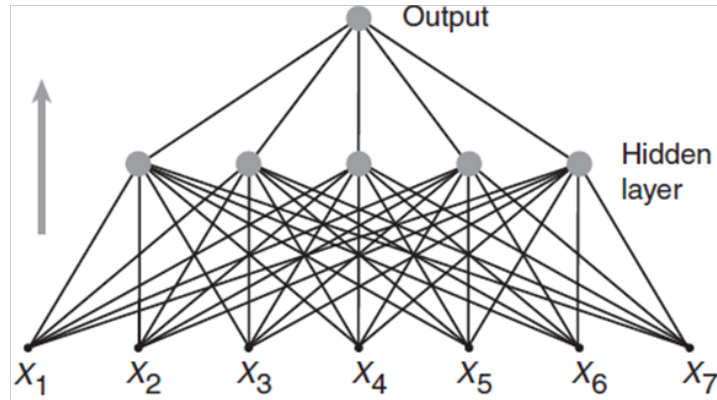


Figure 61. Feed-forward Two-Layer Network Example (Krogh, 2008)

Artificial Neural Network and Interpretation (Test Dataset)

To answer the first question posed in the

Artificial Neural Network Suitability (Test Dataset) section of this chapter, a neural network was constructed to it the right breakpoints or boundaries had been selected for the JEPR classes. If the ANN was able to effectively classify the appraisals into the classification classes that were selected based on the JEPR attributes, then the breakpoints of the JEPR classification classes could be deemed correct. If however, there was a high misclassification rate, then the breakpoints selected for the JEPR classes should be reanalyzed for accuracy.

For the ANN JEPR classifier, the 12 attributes from JEPR VFT Framework were supplied as inputs, in addition to the external Administrative Actions correction factor. The JEPR Test Dataset, with 159 observations, was used to supply these inputs, with the normalized referral markings vector and the random noise vector also included in the inputs to the ANN JEPR classifier. The referral markings vector was included as a quality indicator to identify whether or not a member had violated an Air Force standard, and to what extent. As was previously stated in chapter IV of this research, a referral appraisal occurs when the ratee fails to meet an established standard (Air Force Instruction 36-2406, 2013, p. 40). The ramifications of a referral report are severe, and could result in elimination for promotion consideration for the specific period, and possibly could impact continued service of the ratee, despite the ratees' overall appraisal rating of score. For example, under the current construct, an individual may receive a "4" or "Above Average" EPR rating, but may also receive a referral report due to failing their Physical Fitness test. A random noise vector was also included to randomize sample selection for the training, validation, and test populations for ANN operations. Therefore, there were 15 input vectors input into the ANN JEPR classifier: 12 JEPR VFT Framework attributes, the Administrative Actions correction factor, the normalized referral markings attribute, and the random noise vector.

The ANN JEPR classifier that was previously defined in Table 55, was used as the output classification classes. The classes were constructed based on Air Force Instruction 36-2406 guidance and inputs from the SMEs assisting with the JEPR research. Table 79 reflects the JEPR classification classes.

Table 79. JEPR Classification Classes

JEPR Classification Class Descriptions	
Classification Class Name	JEPR Classification Class Description
Below Standards	Overall Score ≤ 45.57 and/or Failure to Meet any Standard in the Standards group of attributes
Meets Standards	Overall Score >47.57 and <85 . Must meet Standards in all attributes in Standards group
Exceeds Standards	Overall Score ≥ 85 Must meet Standards in all attributes in Standards group

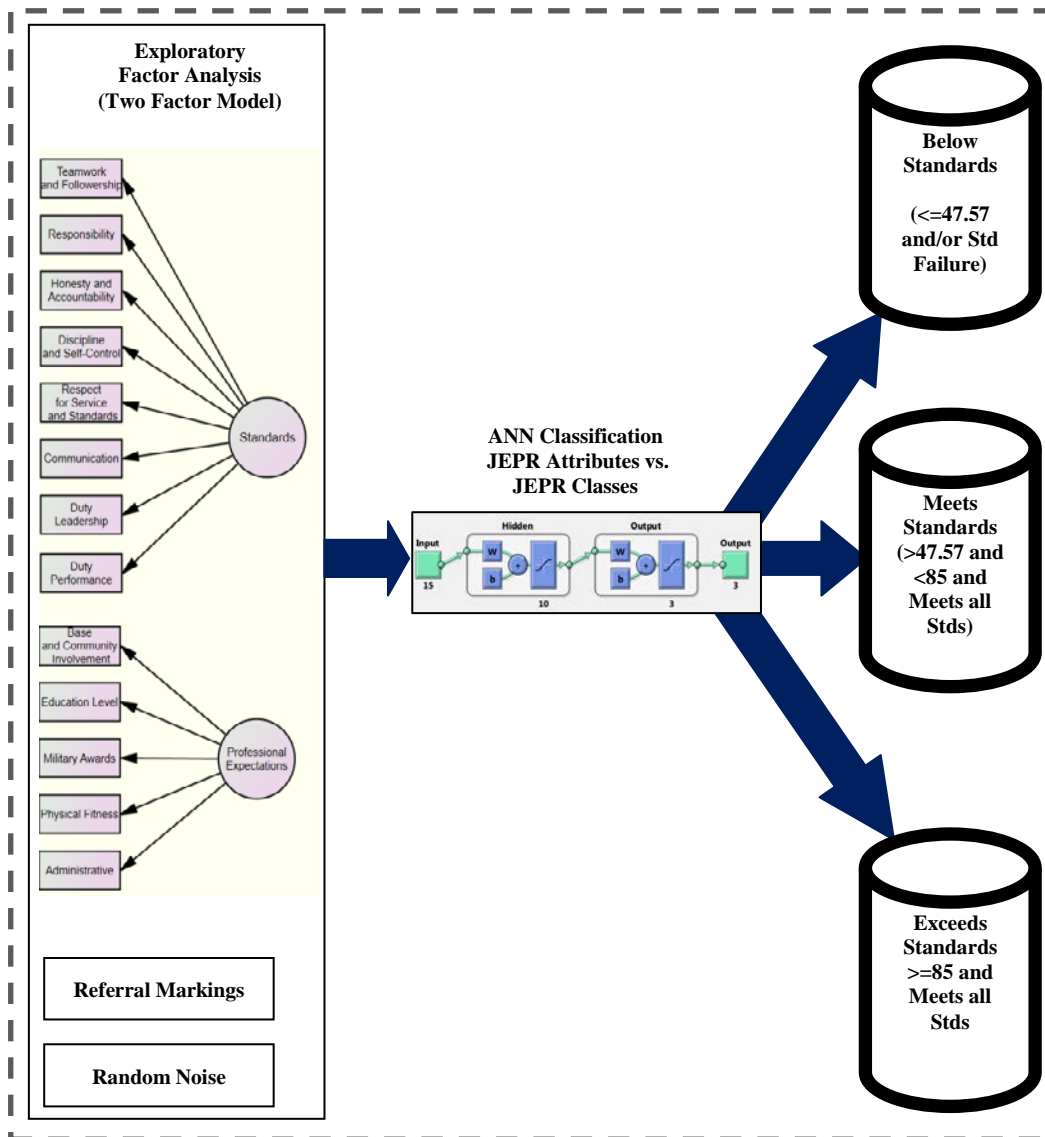


Figure 62. ANN JEPR Classifier (JEPR Classes Shown)

Using the MATLAB Software (MATLAB R2012b 12.0, 2012) environment, the Neural Network Pattern Recognition (NPR) tool was used to generate the ANN JEPR classifier for studying the classification effectiveness of the JEPR system. The NPR tool allows the user to solve two-layer (hidden and output neurons) feed-forward networks using back propagation through a series of Graphic User Interfaces (GUIs) in MATLAB (Shi et al., 2013). Ten neurons were selected for use in the ANN JEPR classifier based on the recommended MATLAB default, however several other configurations were tested with varying number of neurons between eight and 12 with similar results. A graphical representation of the ANN EPR classifier generated by MATLAB is shown in Figure 63.

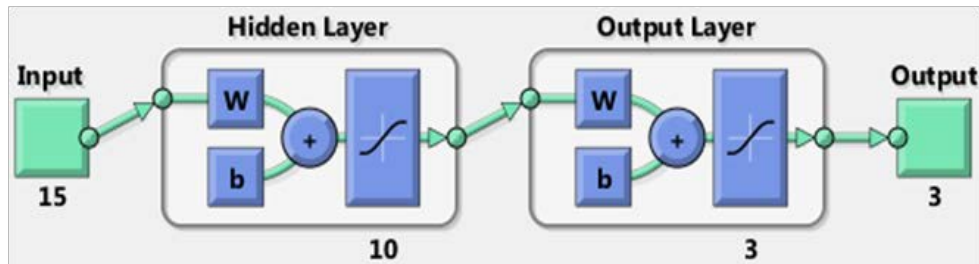


Figure 63. ANN JEPR Classifier (MATLAB NPR Tool, 2012)

The MATLAB NPR tool randomized the order of the 159 data samples, and then parsed the data into three distinct sub-datasets. The ANN JEPR Training Dataset consisted of 111 of the 159 samples, and was used to train the behavior of the ANN based on the known outcomes (Krogh, 2008) from the JEPR DSS interface. For each sample, the NPR tool iteratively reduced the Mean Squared Error (MSE) between the inputs and the known classification classes during training, until the MSE had stabilized, changing the network weights and biases. Equation 38 illustrates the function used by MATLAB

for computing the MSE, where f_k represents the known classification class for sample k , and t_k represents the predicted classification class by the network (Shi et al., 2013).

$$= \frac{1}{2} \sum_k (f_k - t_k)^2 \quad (38)$$

The MATLAB NPR tool algorithm for performing the iterations during training is shown in Equation 39, where x_k is a vector of the current weights for each input, a_k is the learning rate, while g_k is the current gradient for the current sample (Shi et al., 2013).

$$x_{k+1} = x_k - a_k g_k \quad (39)$$

In essence, the ANN EPR classifier was “learning” which characteristics in the JEPR input data yielded a known JEPR output, and then adjusted the classification thresholds of the ANN accordingly for the next data sample.

The ANN JEPR Validation Set consisted of 24 of the 159 samples. This dataset was used to ensure the network was generalizing and is used to prevent over-fitting (Shi et al., 2013). The NPR tool also created the ANN JEPR Test dataset, which was comprised of the remaining 24 samples. This dataset was used as an independent sample to test the classification effectiveness of the ANN after training and validation.

The ANN JEPR network was trained, and then was retrained six times to ensure consistency of output, preventing a local maximum or minimum. Training was ceased when the Signal to Noise Ratios (SNRs) for the network were all sizable positive values, indicating all the attribute of the JEPR VFT Framework were contributing to the

classification effort of the ANN. The larger the positive SNR value was, the more salient or relevant the input feature (attribute) was in determining the output classification for the data sample in the network (Bauer, Alsing, & Greene, 2000). The weights for the hidden neurons and the SNR values for the ANN JEPR are reflected in Appendix IX.

Looking at the confusion matrix for the ANN JEPR Training dataset, all 111 airmen accurately classified using the JEPR VFT Framework attributes into the JEPR classification classes. Additionally, the delineation capabilities of the JEPR classification system are clear to see with 34 of the 111 (31%) test subjects who were classified as “Exceeds Standards”, with 64 of 111 (57%) of the test subject classified as “Meets Standards” and 13 of 111 (12%) of the test subjects classified as “Below Standards”. The MATLAB confusion matrix for the ANN JEPR Training dataset is shown in Figure 64.

Training Confusion Matrix

Predicted Class From JEPR Attributes	BELOW	13 11.7%	0 0.0%	0 0.0%	100% 0.0%
	MEETS	0 0.0%	64 57.7%	0 0.0%	100% 0.0%
	EXCEEDS	0 0.0%	0 0.0%	34 30.6%	100% 0.0%
		100% 0.0%	100% 0.0%	100% 0.0%	100% 0.0%
		BELOW	MEETS	EXCEEDS	
		Known Translated JEPR Class			

Figure 64. ANN EPR Training Confusion Matrix (111 of 159 Randomly Sampled)

For the ANN JEPR Validation dataset, two of 24 sampled appraisals were misclassified. The ANN JEPR network misclassified two individuals as “Meets

Standards”, however, these two members were known to have been rated as “Exceeds Standards” by the JEPR model Decision Support System tool. Looking at the raw data, the two misclassifications were identified. The actual overall JEPR scores for these two appraisals were 85.04 and 85.09, which was very close to the lower limit “Meets Standards” threshold of 84.99, exceeding the threshold by only 0.04 and 0.09 of a point respectively. The ANN JEPR Validation confusion matrix is shown Figure 65.

Validation Confusion Matrix

Predicted Class From JEPR Attributes	BELOW	3 12.5%	0 0.0%	0 0.0%	100% 0.0%
	MEETS	0 0.0%	11 45.8%	2 8.3%	84.6% 15.4%
	EXCEEDS	0 0.0%	0 0.0%	8 33.3%	100% 0.0%
		100% 0.0%	100% 0.0%	80.0% 20.0%	91.7% 8.3%
		BELOW	MEETS	EXCEEDS	
		Known Translated JEPR Class			

Figure 65. ANN JEPR Validation Confusion Matrix (24 of 159 Randomly Sampled)

Looking at the ANN JEPR Test dataset, there were three individuals misclassified as “Meets Standards” which had been rated as “Exceeds Standards” by the JEPR model DSS. Inspection of the raw data revealed that the overall JEPR scores for these three misclassifications were 85.13, 85.38, and 85.74, very near the “Meets Standards” upper limit threshold of 84.99. There was also a misclassification where the ANN JEPR network predicted that a member should be classified as “Exceeds Standards”, and had actually been classified as “Meets Standards” by the JEPR model Decision Support

System tool. The raw data showed that this appraisal had an overall JEPR score of 84.93, which is approximately 0.07 away from the “Exceeds Standards” threshold, barely missing the 85.00 lower threshold requirement. The confusion matrix is shown in Figure 66. The ANN JEPR Combined dataset (all 159 samples) is also shown in Figure 67.

Test Confusion Matrix

Predicted Class From JEPR Attributes	BELOW	5 20.8%	0 0.0%	0 0.0%	100% 0.0%
	MEETS	0 0.0%	9 37.5%	3 12.5%	75.0% 25.0%
	EXCEEDS	0 0.0%	1 4.2%	6 25.0%	85.7% 14.3%
		100% 0.0%	90.0% 10.0%	66.7% 33.3%	83.3% 16.7%
		BELOW	MEETS	EXCEEDS	

Known Translated JEPR Class

Figure 66. ANN JEPR Test Confusion Matrix (24 of 159 Randomly Sampled)

All Confusion Matrix

Predicted Class From JEPR Attributes	BELOW	21 13.2%	0 0.0%	0 0.0%	100% 0.0%
	MEETS	0 0.0%	84 52.8%	5 3.1%	94.4% 5.6%
	EXCEEDS	0 0.0%	1 0.6%	48 30.2%	98.0% 2.0%
		100% 0.0%	98.8% 1.2%	90.6% 9.4%	96.2% 3.8%
		BELOW	MEETS	EXCEEDS	

Known Translated JEPR Class

Figure 67. ANN JEPR Combined Confusion Matrix (159 of 159 Randomly Sampled)

The analysis of the ANN JEPR network illustrated that if the VFT Framework attributes are what the Air Force values, then the breakpoints of the JEPR classification construct are accurate, with a 96.2% classification rate. Figure 68 graphically illustrates the clearly defined breakpoints, overlaying the JEPR and EPR scores for the 159 test subjects.

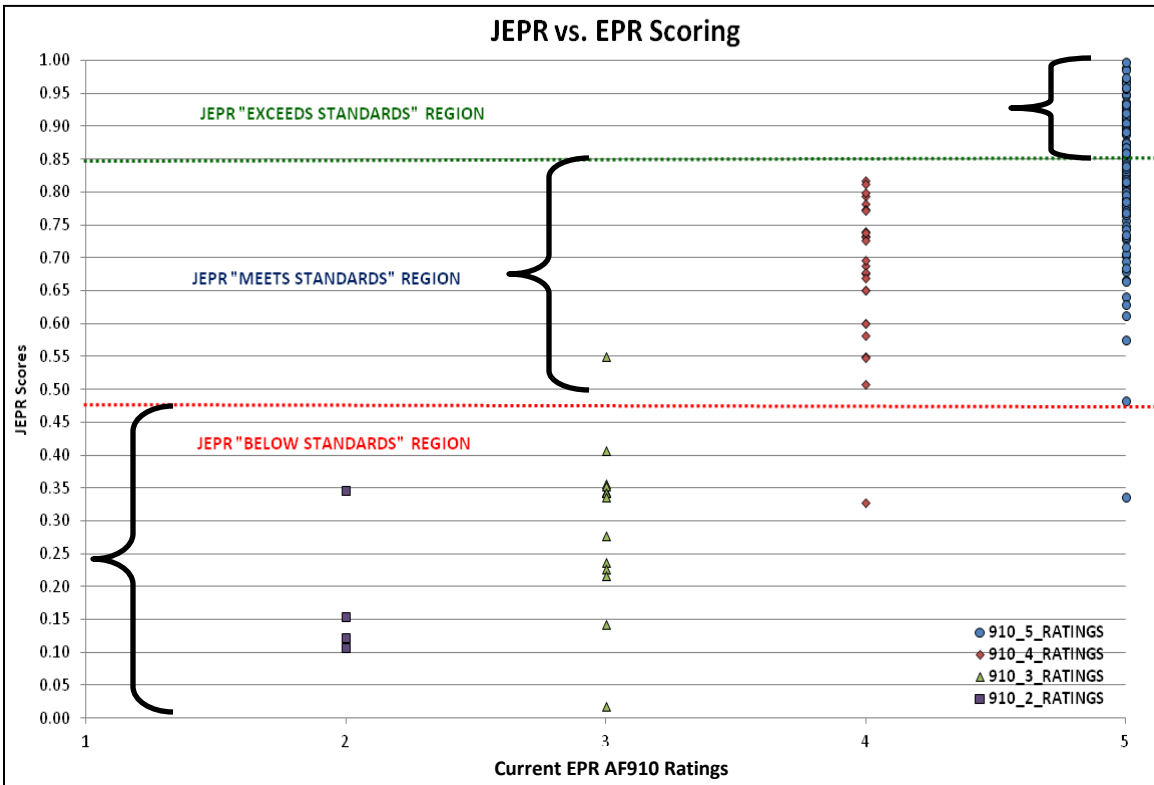


Figure 68. JEPR vs. EPR Scoring (JEPR Classification Classes Overlaid)

A second neural network was constructed to contrast how the current EPR system compared to the JEPR in classifying airman using the VFT Framework attributes. The ANN EPR classifier utilized the same 15 inputs as the ANN JEPR classifier with 12 of the inputs coming from the JEPR VFT Framework (159 observations), an external Administrative Actions correction factor vector, a normalized referral markings vector,

and a random noise vector. Table 80 reflects the Translated EPR classification classes while Figure 69 illustrates the ANN EPR classifier design and the classification classes.

Table 80. Translated EPR Classification Classes

Translated EPR Classification Class Descriptions	
Classification Class Name	Translated EPR (Current AF 910) Classification Class Description
Below Standards	Overall Rating \leq "2"
Meets Standards	Overall Rating $>$ "2 "and \leq "4"
Exceeds Standards	Overall Rating $=$ "5"

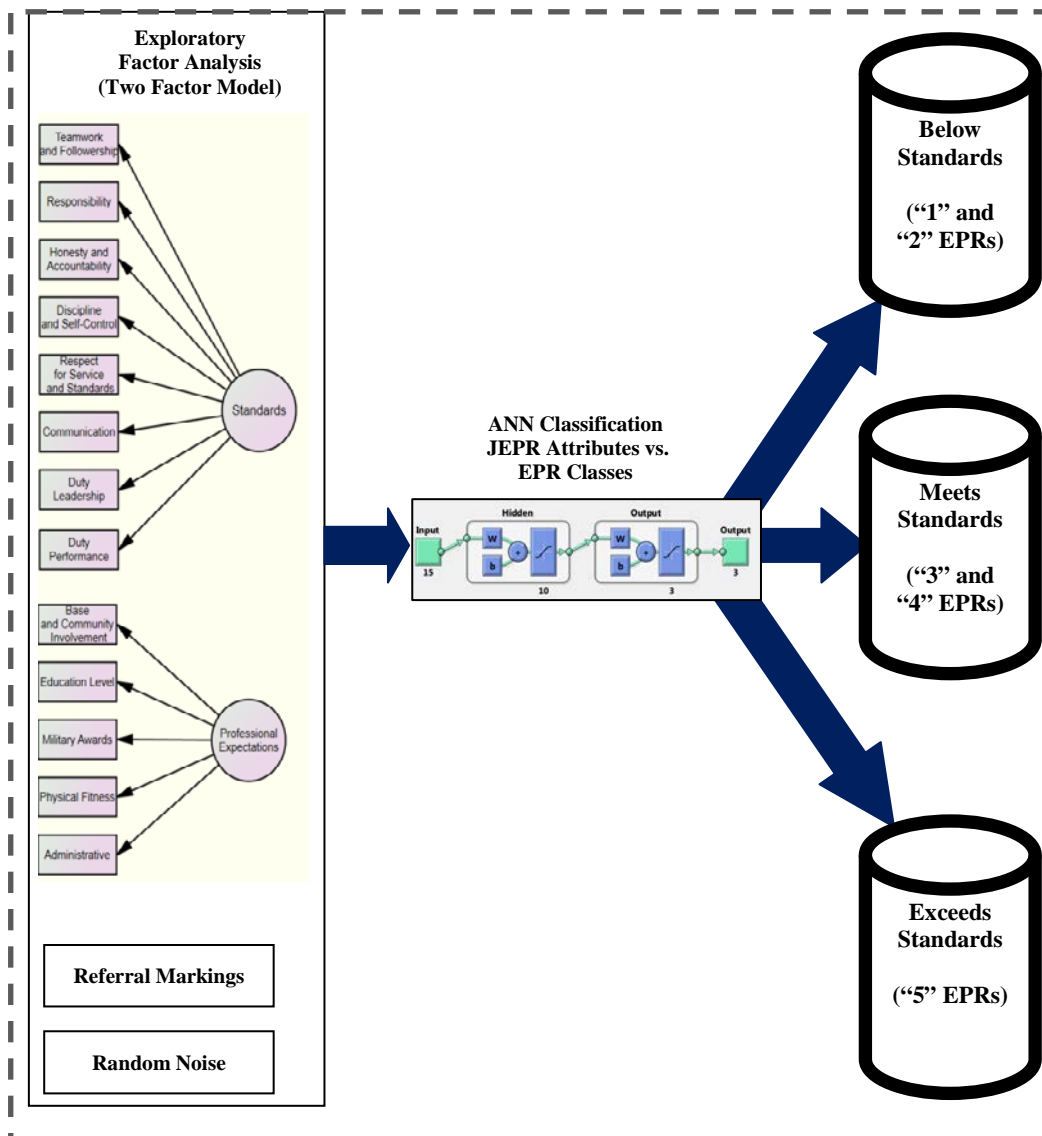


Figure 69. ANN EPR Classifier (EPR Classes Shown)

Using the MATLAB NPR tool, the ANN EPR classifier was generated to study the classification effectiveness of the current EPR system using the JEPR VFT Framework attributes as inputs and the known EPR outputs grouped as the Translated EPR classification classes. Ten neurons were again selected for use in the ANN EPR classifier based on the recommended MATLAB default, however several other configurations were tested with varying number of neurons between eight and 12 hidden neurons with similar results.

The MATLAB NPR tool randomized the order of the 159 data samples, and then parsed the data into three distinct sub-datasets. The ANN EPR Training dataset consisted of 111 of the 159 samples, and was used to train the behavior of the ANN based on the known outcomes (Krogh, 2008) from the current EPR system. For each sample, the NPR tool iteratively reduced the Mean Squared Error (MSE) between the inputs and the known Translated EPR classification classes during training, until the MSE had stabilized, thus changing the weights and biases for the network.

The ANN EPR Validation Set consisted of 24 of the 159 samples while the ANN EPR Test dataset was comprised of the remaining 24 samples. As was done with the ANN JEPR network, the ANN EPR network was trained, and then retrained to ensure output consistency and to prevent local maximums or minimums. Training was again ceased when the SNRs for the network were all sizable positive values, denoting all features or attributes from the VFT Framework were providing input in determining the overall classification. The weights for the hidden neurons and the SNR values for the ANN EPR are reflected in Appendix IX.

Looking at the confusion matrix for the ANN EPR Training dataset in Figure 70, 16 of the 111 airmen who had been given a “5” overall EPR rating in the under the current appraisal system, were predicted as “Meets Standards” using the JEPR inputs.

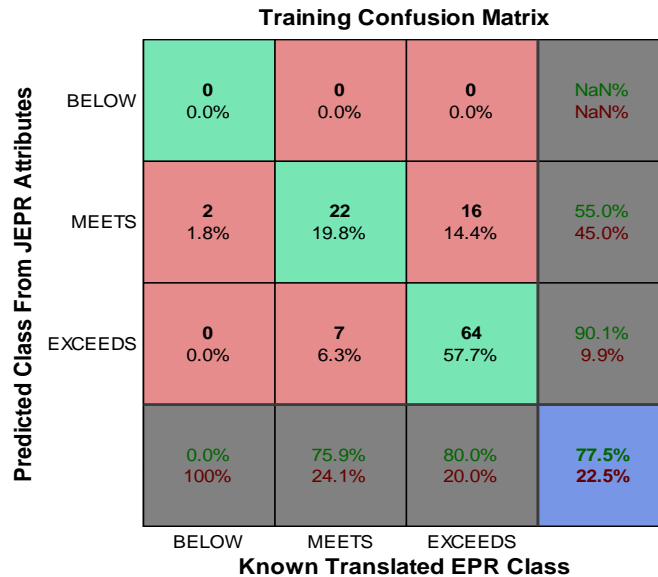


Figure 70. ANN EPR Training Confusion Matrix (111 of 159 Randomly Sampled)

Looking in detail at the misclassified data, these individuals had overall JEPR scores ranging from 81.93 to 84.93, which was below the 85.00 minimum thresholds for the JEPR classification class of “Exceeds Standards”. Seven of the 111 appraisals that were classified as “Meets Standards” under the Translated EPR classification system who had was rated as a “3” or “4” under the current EPR construct, were classified as “Exceeds Standards” based on the JEPR attribute inputs. Delving into the raw dataset, the overall JEPR scores for these seven appraisals ranged between 77.22 and 81.70. Finally, two appraisals were misclassified as “Meets Standards” based on the JEPR attribute inputs, yet had actually received “2” ratings on their EPRs, and were classified as “Below Standards” on the Translated EPR classification scheme.

From the 111 observations sampled, the Translated EPR system could only classify 77.5% of the known EPR appraisals using the JEPR attributes as inputs. There were two reasons identified that contributed to the high misclassification rate. First, there was a great deal of variance in the JEPR attribute input data in relationship to the Translated EPR classification classes. Since the “learning” design of the ANN EPR classifier attempts learn where to classify each subsequent data sample based on minimizing the MSE from previous samples iteratively, high variability in the randomly sampled observations can disrupt the learning process of the ANN, creating misclassifications. Second, the narrow range of EPR ratings (1 through 5) did not provide enough granularity in the design of the Translated EPR output classes for the ANN to effectively class the appraisals. This behavior continued to be noted during the analysis of ANN EPR Validation dataset and the ANN EPR Test dataset. The confusion matrices are for the ANN EPR Validation dataset, the ANN EPR Test dataset, and the ANN EPR Combined dataset (all 159 samples) are shown in Figure 71 through Figure 73.

Validation Confusion Matrix

Predicted Class From JEPR Attributes	BELOW	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	MEETS	0 0.0%	6 25.0%	2 8.3%	75.0% 25.0%
	EXCEEDS	0 0.0%	2 8.3%	14 58.3%	87.5% 12.5%
		NaN% NaN%	75.0% 25.0%	87.5% 12.5%	83.3% 16.7%
	BELOW	MEETS	EXCEEDS	Known Translated EPR Class	

Figure 71. ANN EPR Validation Confusion Matrix (24 of 159 Randomly Sampled)

Test Confusion Matrix

Predicted Class From JEP R Attributes	BELOW	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	MEETS	2 8.3%	2 8.3%	4 16.7%	25.0% 75.0%
	EXCEEDS	0 0.0%	1 4.2%	15 62.5%	93.8% 6.3%
		0.0% 100%	66.7% 33.3%	78.9% 21.1%	70.8% 29.2%
	BELOW	MEETS	EXCEEDS	Known Translated EPR Class	

Figure 72. ANN EPR Test Confusion Matrix (24 of 159 Randomly Sampled)

All Confusion Matrix

Predicted Class From JEP R Attributes	BELOW	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
	MEETS	4 2.5%	30 18.9%	22 13.8%	53.6% 46.4%
	EXCEEDS	0 0.0%	10 6.3%	93 58.5%	90.3% 9.7%
		0.0% 100%	75.0% 25.0%	80.9% 19.1%	77.4% 22.6%
	BELOW	MEETS	EXCEEDS	Known Translated EPR Class	

Figure 73. ANN EPR Combined Confusion Matrix (159 of 159 Randomly Sampled)

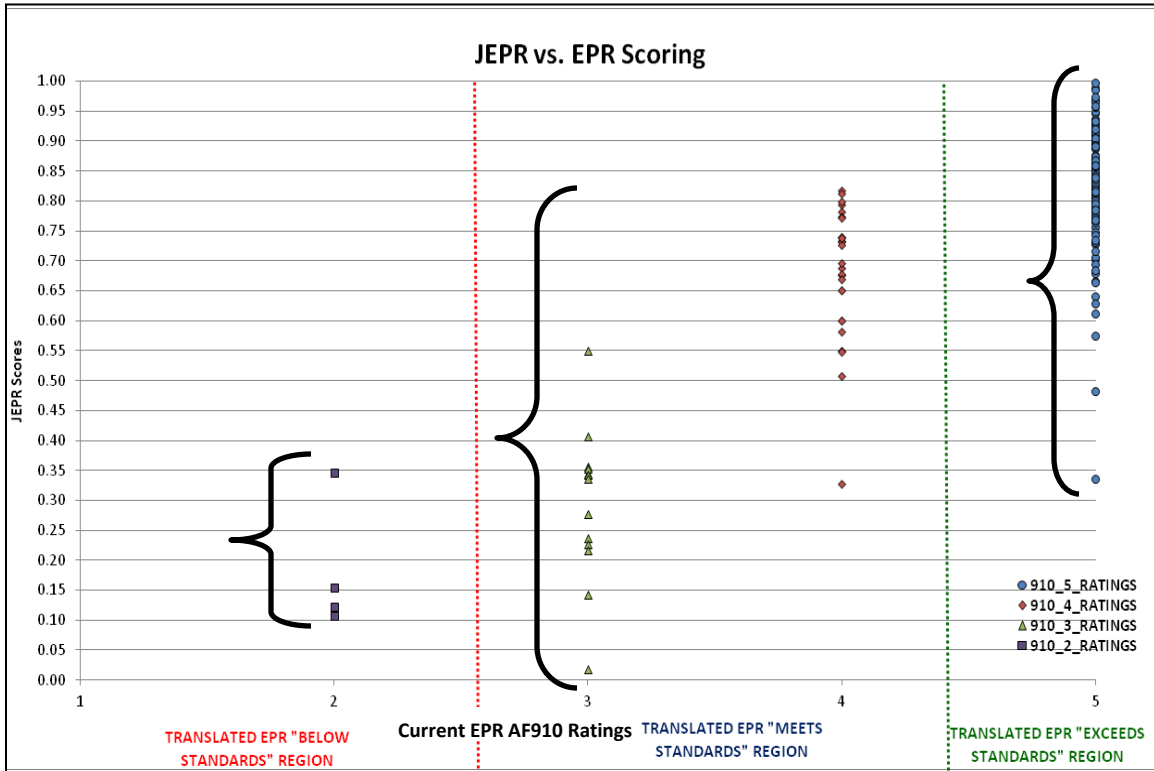


Figure 74. JEPR vs. EPR Scoring (Translated EPR Classification Classes Overlaid)

Looking at the “5” rated EPRs for the “Exceeds Standards” class in Figure 74, illustrates the variance between appraisals were classified as “Exceeds Standards” using Translated EPR classification system (based on EPR ratings) had overall JEPR overall that varied between 33.59 and 99.82. In comparison, the JEPR scores as shown in Figure 75, only varies from 85.04 to 99.82.

The JEPR classification system has demonstrated that it is better able to classify junior enlisted appraisals, if the JEPR VFT Framework is truly what the Air Force values, due to the more granular scoring design of the JEPR, which reduces in-class variability during classification. The 96.2% classification successful classification rate of the JEPR

was a considerable improvement over the 77.4% classification success rate of the current EPR system. There were two reasons noted for the variability differences.

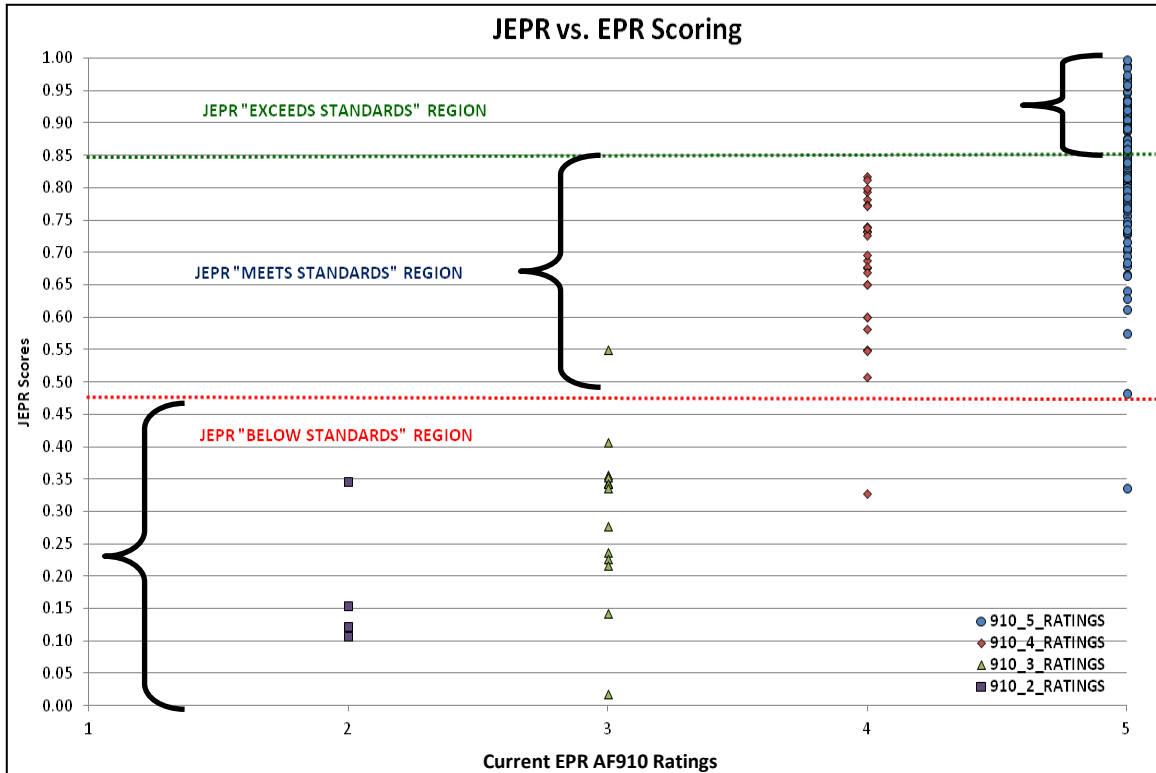


Figure 75. JEPR vs. EPR Scoring (JEPR Classification Classes Overlaid)

First, the consistency in evaluations from JEPR was provided by the scale design allowed the ANN JEPR network to better classify the ratee appraisals than the Translated EPR classification system. This was due to less variability in the known outcomes for the network to handle when trying to classify the appraisals versus the discrete 1 to 5 rating scheme. From detailed analysis of the data, the use of the larger 100 point scale parsed by the four explicitly defined distinct ratings categories in the JEPR helped the supervisors better appraise the airmen. The supervisor could effectively narrow down which category

best captured the observed behavior displayed by the airmen, then, using the range of that category, capture the strength or weakness of the observed behavior in this category with the rating value assigned. The use of the categories and ranges not only provided a scoring construct, within the category effectively provided a mechanism for feedback, highlighting measured performance and quantitative areas for improvement.

Second, and most important, variability is greatly reduced in the JEPR due to the fact that the JEPR overall scores are not independent of the attribute scores for the appraisal. This forces the overall score to be a relation of the attribute scorings entered by the supervisor. The overall rating (backside of the form) of the current EPR system is independent of the performance assessment appraisal ratings (front side of the form) creating an environment where overall ratings are not indicative of observed performance ratings documented by the supervisor, as illustrated by the large amount of variance for each of the Translated EPR classification classes shown in Figure 74. Review of the comments annotated in the JEPR appraisal comments supported the data findings as several individuals had been rated as “Exceeds Standards” in the overall rating of the EPR, yet had experienced Administrative Actions or had failed to meet an Air Force standard.

In this chapter, we conducted an EFA effort on a second, much larger dataset, the JEPR Test Dataset, to validate the loadings structure uncovered during the initial EFA effort with the JEPR Training Dataset. Not only did this validate that the initial EFA structure was correct, it also validated that the VFT Framework was an accurate representation of the doctrine and Air force values, which could be further explained by the two latent factors of Standards and Professional Expectations. Additionally, CFA was

used to validate that the EFA loadings construct was statistically and structurally accurate, which again, confirmed that the VFT Framework was accurately designed. Finally, this chapter confirmed that the breakpoints for the classifications classes of the JEPR are accurate for classification of airmen using the VFT Framework attributes, and that the current EPR system struggles to classify airmen using the VFT Framework attributes due to the variability encountered rating instruments design.

VI. Conclusions and Recommendations

Conclusion

The current Air Force junior enlisted appraisal system can be improved. Since 2009, 80% of the airmen within the Air Force have been rated as “Truly among the Best”. The JEPR appraisal process has clearly demonstrated the ability to accurately evaluate airmen based on doctrine and the criteria the Air Force values as most important. The JEPR design has shown that it directly aligns with Air Force doctrine and values and can generate more accurate and consistent appraisals. Using collected evaluation data, JEPR has also shown that it can reduce inflation through a rigorously validated framework design. Additionally, the JEPR system has demonstrated that it can delineate between “near peer” performers. The JEPR system has shown to be a flexible design that is capable of incorporating changes in leadership and mission priorities. The system can even be used to conduct defensible and value focused force management decisions. Finally, the efficient web-based system also enables unit leaders and supervisors to escape “management behind the desk” and be better utilized for direct leadership and mentorship of airmen, without being saturated with a labor intensive manual appraisal process. The JEPR system can appraise personnel in a fair and consistent manner based on the doctrine and values of the Air Force.

Significant Research Contributions

This research married multiple Operations Research and Management Science techniques to provide a solution to appraisal inflation and incongruency in ratings which have plagued the Air Force appraisal system since its inception. This research directly

mapped organizational values into the performance appraisal process. For the Air Force, this results in a stronger force more in-tune to doctrine, due to more accurate appraisals of performance and promotion of airmen whose performance reflects the values of the organization. For the ratee, this provides clear guidance on what is valued by the Air Force, providing direction for sustainment of performance expectations or a mechanism for behavioral modifications to occur.

This research introduced efficiencies in the appraisal process, while also providing a quantitative method to make efficient force cultivation and force management decisions. Leveraging informational technologies, career field managers and the Air Force Personnel Center (AFPC) have the ability to quickly query historical data enabling trend analysis and force management decisions to be studied quantitatively. The efficiencies attained through the use of informational technologies are not solely constrained to personnel decisions and trend analysis. Unit level leader and supervisors benefit from a web-based design, enabling supervisors to spend more time to providing “hands-on” leadership and mentorship to junior airmen instead of being saturated with the paperwork associated with a manual process.

Finally, this research has also provided a method for statistically validating Decision Analysis Value Hierarchies. Exploratory Factor analysis is used to validate assumptions pertaining to the alignment of Means Objectives under the Fundamental Objectives during construction of the VFT Framework. Studying the alignments and the strengths of factor correlation loadings between the attributes of the VFT Framework and the common factors can validate whether the assumed Value Hierarchy structure is correct, and if not, what the true underlying latent construct is.

This research also illustrated how Confirmatory Factor Analysis can be used to further statistically validate, the loadings structure revealed during the Exploratory Factor Analysis effort is statistically accurate and defensible. Confirmatory Factor Analysis is a tool that has been commonly used by psychologists and researchers to develop, refine, and assess the validity of measurement constructs (Jackson et al., 2009). Through use of multivariate multiple regression Structural Equation Modeling equations are applied along with multiple testing indices to test the hypothesized model design, Confirmatory Factor Analysis can validate the measurement constructs of the VFT Framework attributes and the validate the Framework design.

Finally, this research showcased how Artificial Neural Networks can be used to for classification of data derived from VFT Frameworks, and how an existing classification system can be studied for performance and anomalies using solicited VFT Framework attributes. The Artificial Neural Network provided a method for classifying Behavioral Science data, which often non-normal, without distributional assumptions or linearity (Krycha & Wagner, 1999). The Artificial Neural Networks enabled validation that the classification breakpoints selected for the VFT Framework were correctly selected during the VFT design. Finally, the Artificial Neural Network that studied the current EPR system revealed that the current classification system struggled to effectively classify test subjects using the VFT attributes. This was due primarily to the large amount of variance encountered in the current system stemming from the fact that the overall rating captured on the backside of the form is independent of the performance assessment appraisal ratings reflected on the front side of the current form.

Recommendations for Future Research

This type of technique should be considered for future research by both government and civilian organizations for conducting any type of personnel appraisal. In particular, this research could be the foundation for future research in the redesigning the military officer appraisal system to better capture the traits the Air Force values in its officer corps. Additionally, future research should be performed to study how a JEPR type system could control inflation in Senior Non-Commissioned Officer (SNCOs) appraisals ensuring only the highest performing SNCOs are selected as future leaders.

This technique could also be applied to facilitate force management decisions. With military force reductions on the horizon, this approach could be beneficial in quantitatively determining which members should be retained for continued service. The system could easily be adapted to changes in priorities of senior leaders, and can be modified to meet changing force retention requirements.

Civilian organizations could also benefit from the foundations provided by this research in appraising personnel or items. This approach can be utilized for acquisitions programs, ensuring that the acquisition aligns with the organizations portfolio. The technique could also be used for any type of corporate decision, such as determining which manufacturing project to undertake is more in-line with the company values. Finally, this method could be utilized as a framework for any type of evaluation or decision scenario.

Summary

It is possible to create a VFT model for performance appraisals consistent with leadership and organizational values. Exploratory and Confirmatory Factor Analysis, when used appropriately, can also be used to validate the framework of an evaluation model. Additionally, the use of Artificial Neural Networks validated the accuracy of the breakpoints selected for the JEPR classification classes which had originally been determined by the SMEs. The use of a web-based used interface for performing appraisals enables a data repository, which can be queried and studied by Air Force personnel managers and researchers for trends and force quantitative management decisions. A statistically validated evaluation model can aid in overcoming or mitigating common appraisal systems such as consistency, inflation, and the ability to delineate members. The end result of this research is that incorporation of the proposed system would result in better evaluations, better feedback, better promotion opportunities for better qualified members, and a more capable workforce.

Appendix I

Single Attribute Value Function for Duty Performance

Mid-Value Point	(Performance) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Performance Percentile of Employee versus Ideal Employee	Estimated Weighted Performance Category Score for EPR (Weighted at 40%)	Gamma
x(bottom)	0	0	0.000000	0.00%	0.00967939
x(.25)	0.25	15	0.217925	8.72%	Sum SQ Diff @ 0.25, 0.50, 0.75 0.001350206
x(.50)	0.5	40	0.517675	20.71%	
x(.75)	0.75	65	0.752999	30.12%	
x(top)	1	100	1.000000	40.00%	Function $\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^1 - x_1^0)}}, x_1 \in X_1$

NOTE

SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

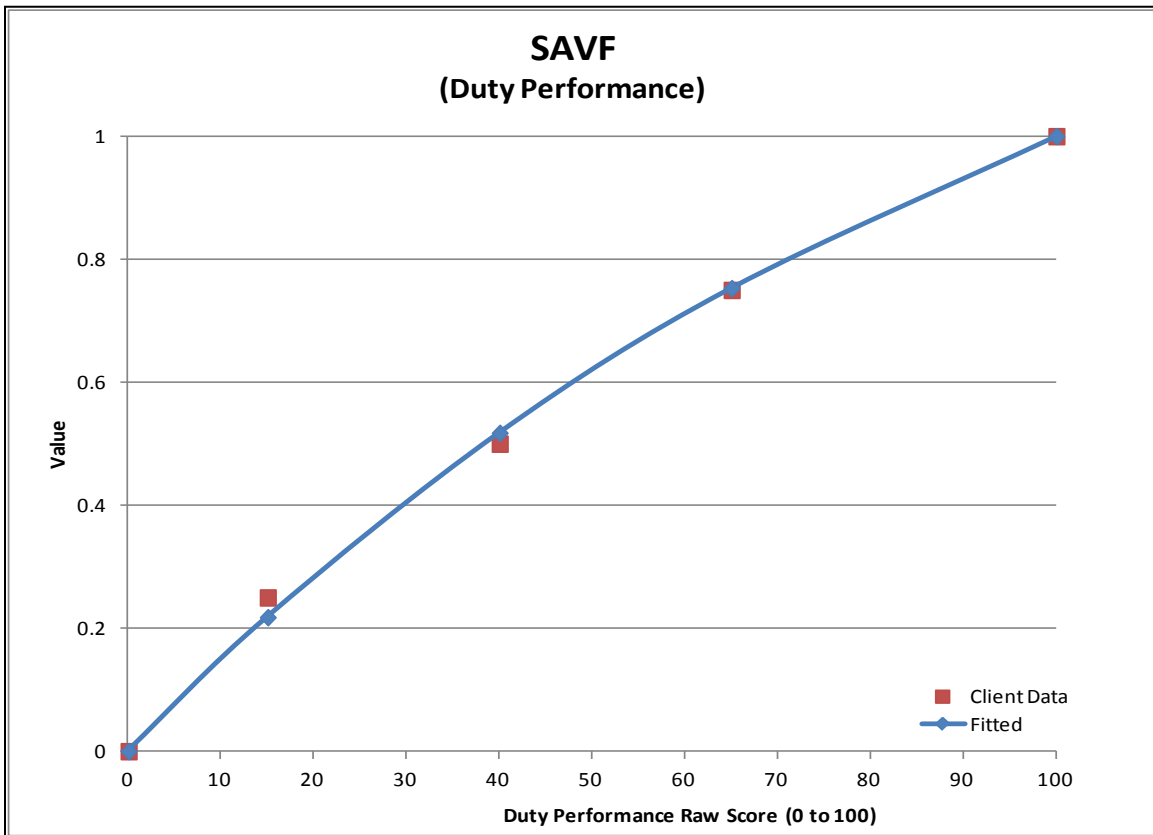


Figure 76. Duty Performance Single Attribute Value Function

Single Attribute Value Function for Duty Leadership

Mid-Value Point	(Leadership) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Leadership Percentile of Employee versus Ideal Employee	Estimated Weighted Leadership Category Score for EPR (Weighted at 10%)	Gamma
x(bottom)	0	0	0.000000	0.00%	0.00938621
x(.25)	0.25	20	0.281123	2.81%	Sum SQ Diff @ 0.25, 0.50, 0.75 0.002995379
x(.50)	0.5	40	0.514129	5.14%	
x(.75)	0.75	60	0.707255	7.07%	
x(top)	1	100	1.000000	10.00%	Function $\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^* - x_1^0)}}, x_1 \in X_1$

NOTE
SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

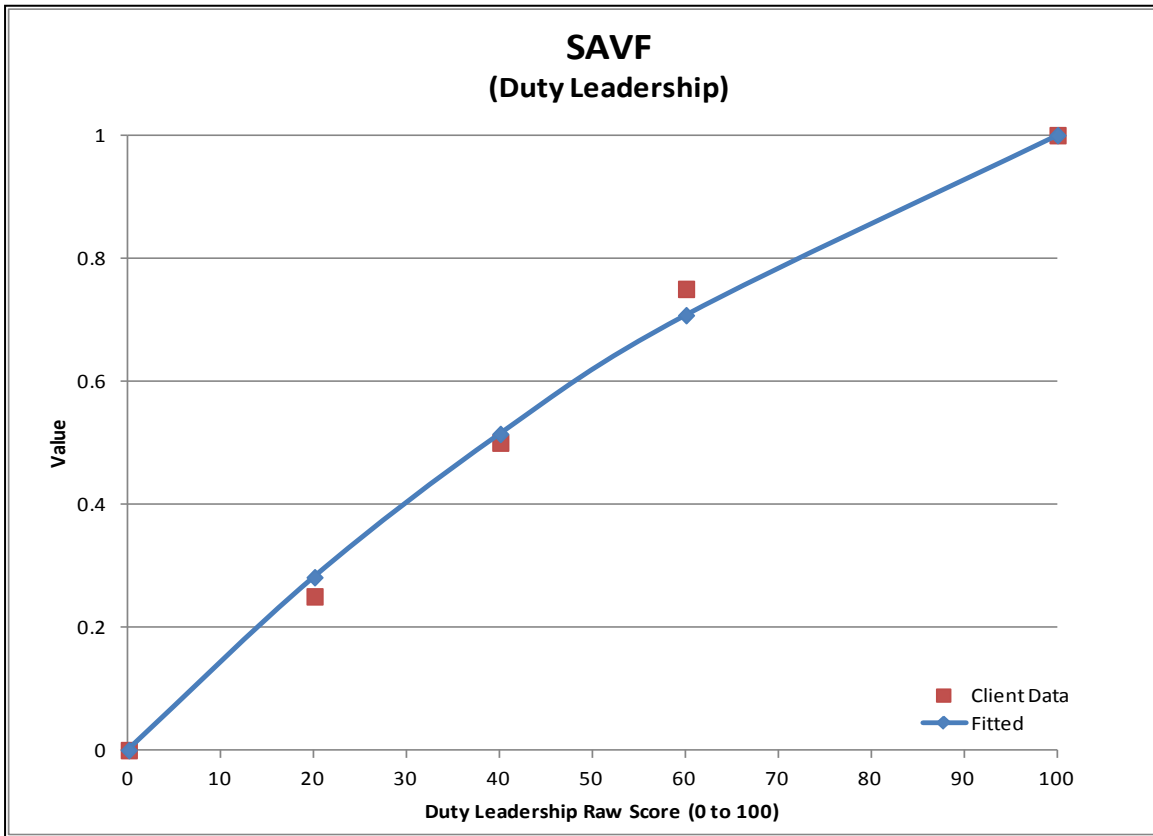


Figure 77. Duty Leadership Single Attribute Value Function

Single Attribute Value Function for Teamwork and Followership

Mid-Value Point	(Teamwork & Followership) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Section Slopes	Function Estimated Teamwork & Followership Percentile of Employee versus Ideal Employee	Estimated Weighted Teamwork & Followership Category Score for EPR (Weighted at 3%)
x(bottom)	0	0	0	0.000000	0.00%
x(.25)	0.25	30	120	0.250000	0.75%
x(.50)	0.5	45	60	0.500000	1.50%
x(.75)	0.75	65	80	0.750000	2.25%
x(top)	1	100	140	1.000000	3.00%

Function
Piecewise

NOTE

SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

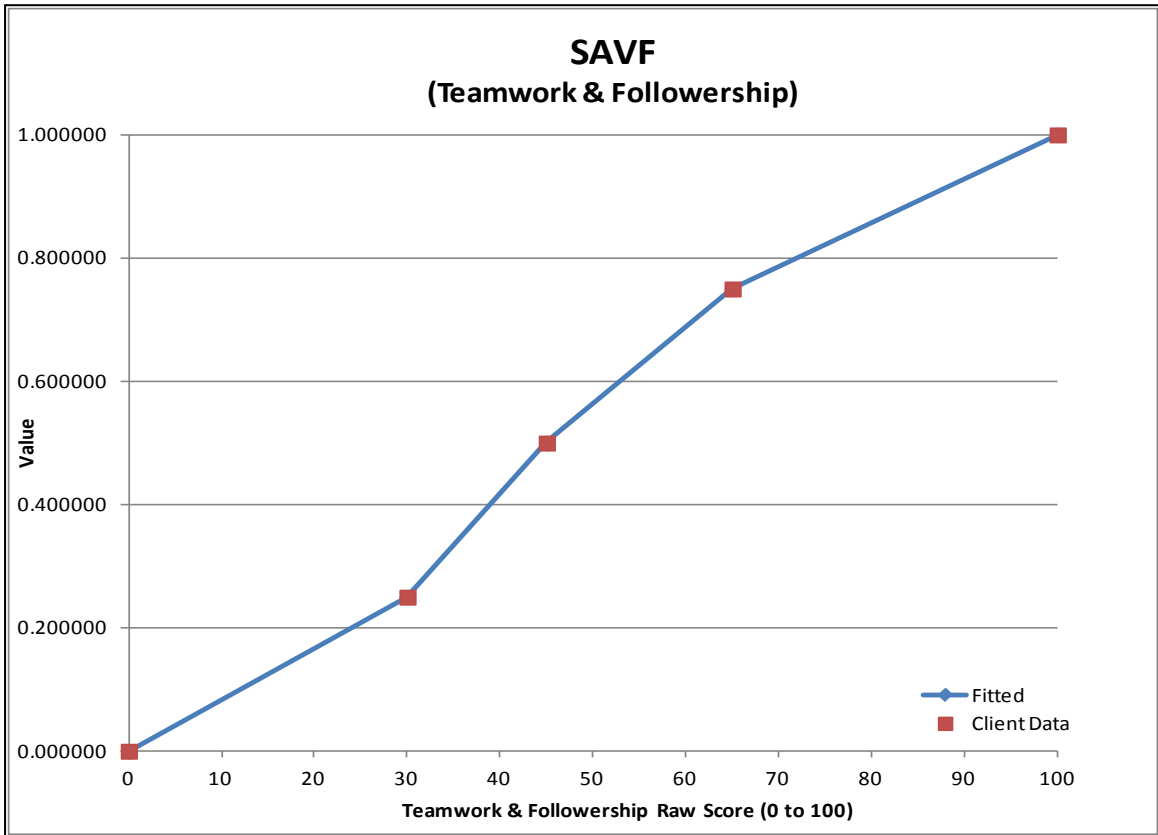


Figure 78. Teamwork and Followership Single Attribute Value Function

Single Attribute Value Function for Respect for Service and Standards

Mid-Value Point	(Respect for Standards) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Respect for Standards Percentile of Employee versus Ideal Employee	Estimated Weighted Respect for Standards Category Score for EPR (Weighted at 8%)	
x(bottom)	0	0	0.000000	0.00%	Gamma 0.0000000
x(.25)	0.25	25	0.250000	2.00%	
x(.50)	0.5	50	0.500000	4.00%	Sum SQ Diff @ 0.25, 0.50, 0.75 2.7111E-16
x(.75)	0.75	75	0.750000	6.00%	
x(top)	1	100	1.000000	8.00%	Function $\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^* - x_1^0)}}, x_1 \in X_1$

NOTE
SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

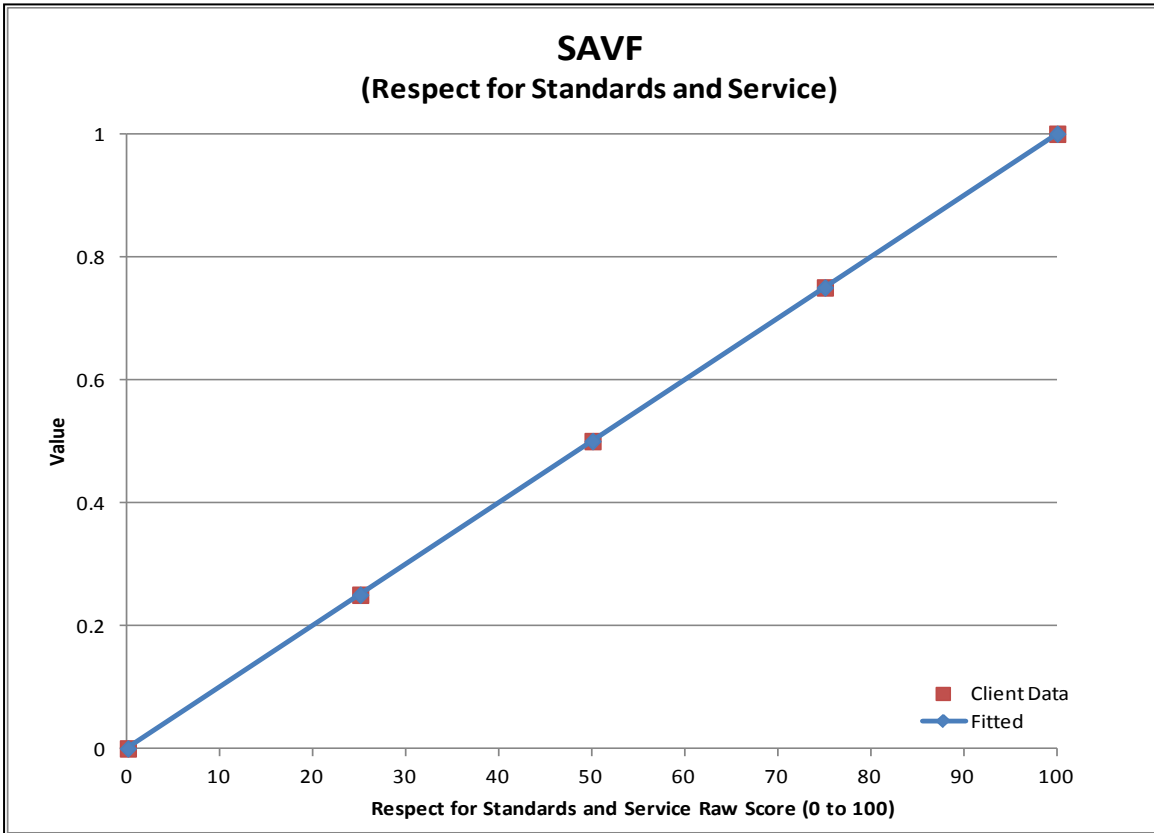


Figure 79. Respect for Service and Standards Single Attribute Value Function

Single Attribute Value Function for Discipline and Self-Control

Mid-Value Point	(Discipline & Self Control) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Discipline & Self Control Percentile of Employee versus Ideal Employee	Estimated Weighted Discipline & Self Control Category Score for EPR (Weighted at 5%)	
x(bottom)	0	0	0.000000	0.00%	Gamma 0.00938621
x(.25)	0.25	20	0.281123	1.41%	Sum SQ Diff @ 0.25, 0.50, 0.75 0.002995379
x(.50)	0.5	40	0.514129	2.57%	
x(.75)	0.75	60	0.707255	3.54%	
x(top)	1	100	1.000000	5.00%	Function $\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^* - x_1^0)}}, x_1 \in X_1$

NOTE

SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

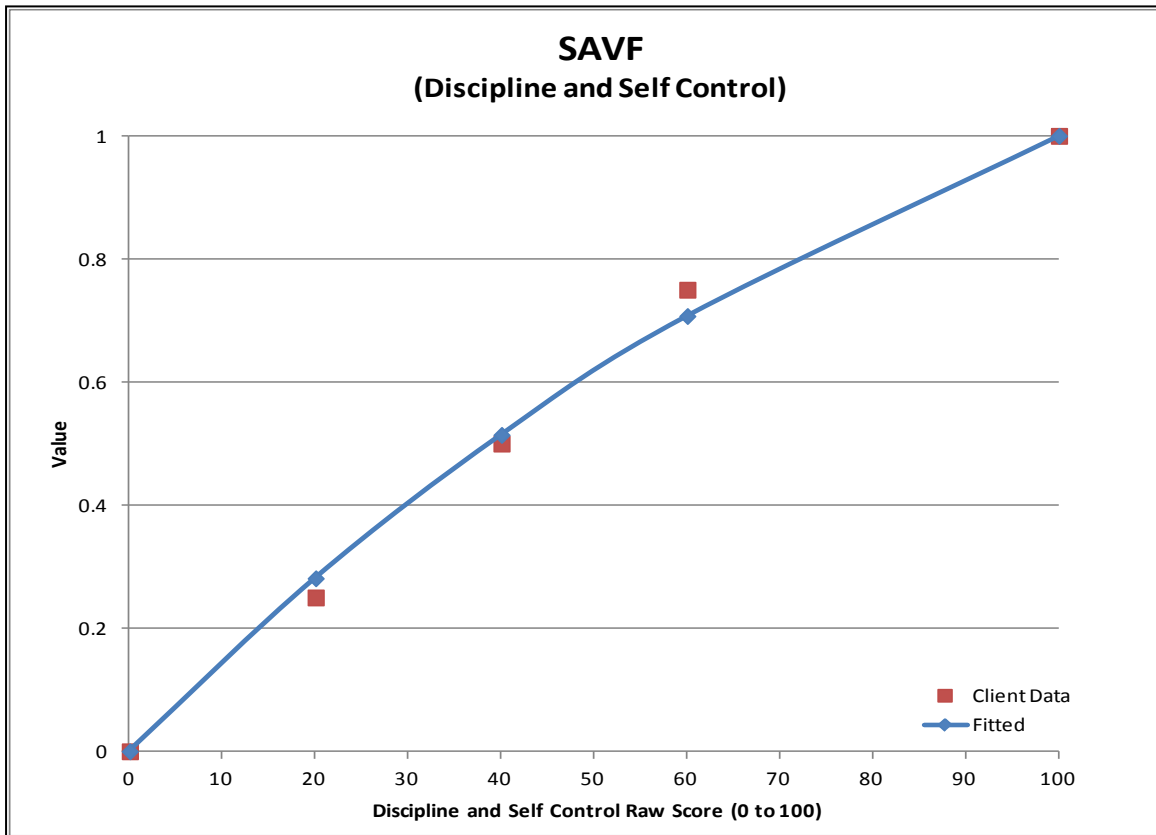


Figure 80. Discipline and Self Control Single Attribute Value Function

Single Attribute Value Function for Communication

Mid-Value Point	(Communication) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Communication Percentile of Employee versus Ideal Employee	Estimated Weighted Communication Category Score for EPR (Weighted at 5%)	
x(bottom)	0	0	0.000000	0.00%	Gamma 0.00938621
x(.25)	0.25	20	0.281123	1.41%	
x(.50)	0.5	40	0.514129	2.57%	Sum SQ Diff @ 0.25, 0.50, 0.75 0.002995379
x(.75)	0.75	60	0.707255	3.54%	
x(top)	1	100	1.000000	5.00%	Function $\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^+ - x_1^0)}}, x_1 \in X_1$

NOTE
SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

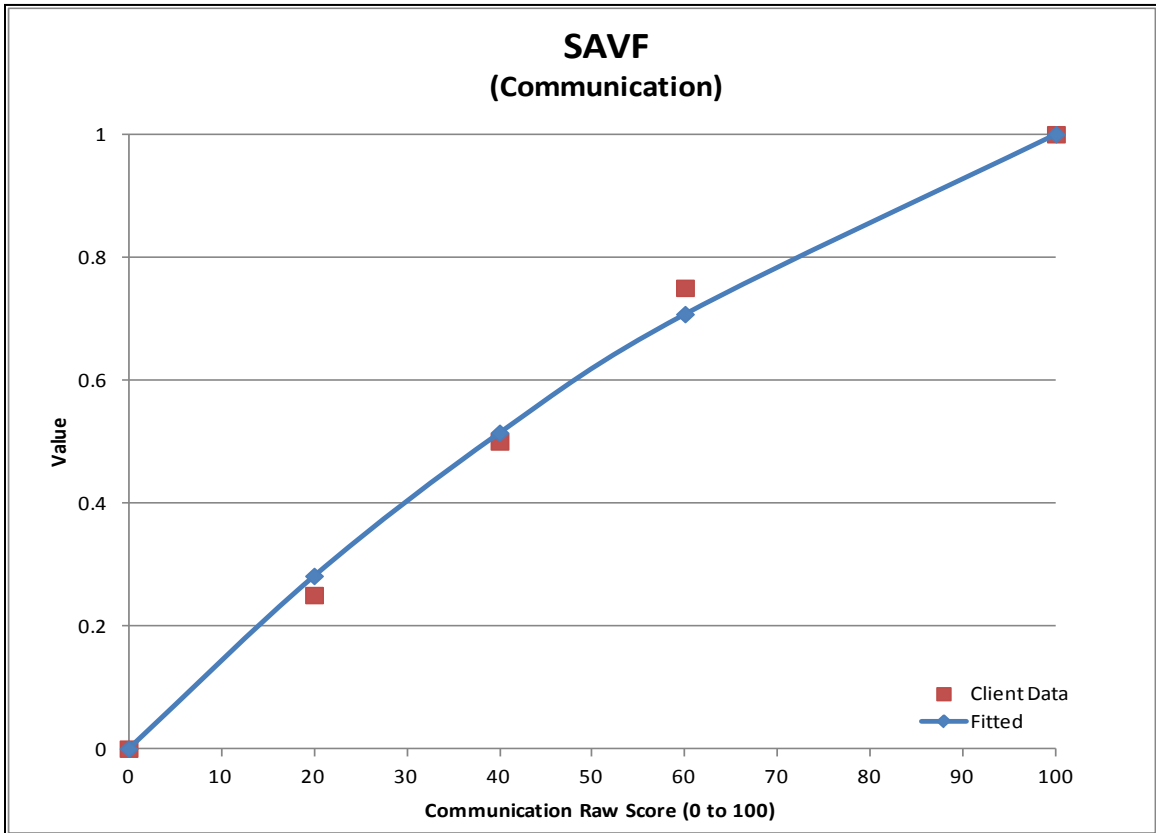


Figure 81. Communication Single Attribute Value Function

Single Attribute Value Function for Responsibility

Mid-Value Point	(Responsibility) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Responsibility Percentile of Employee versus Ideal Employee	Estimated Weighted Responsibility Category Score for EPR (Weighted at 4%)	
x(bottom)	0	0	0.000000	0.00%	Gamma
x(.25)	0.25	15	0.287015	1.15%	0.01843588
x(.50)	0.5	30	0.504689	2.02%	Sum SQ Diff @ 0.25, 0.50, 0.75
x(.75)	0.75	50	0.715408	2.86%	0.002588741
x(top)	1	100	1.000000	4.00%	Function

$$\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^* - x_1^0)}}, x_1 \in X_1$$

NOTE

SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

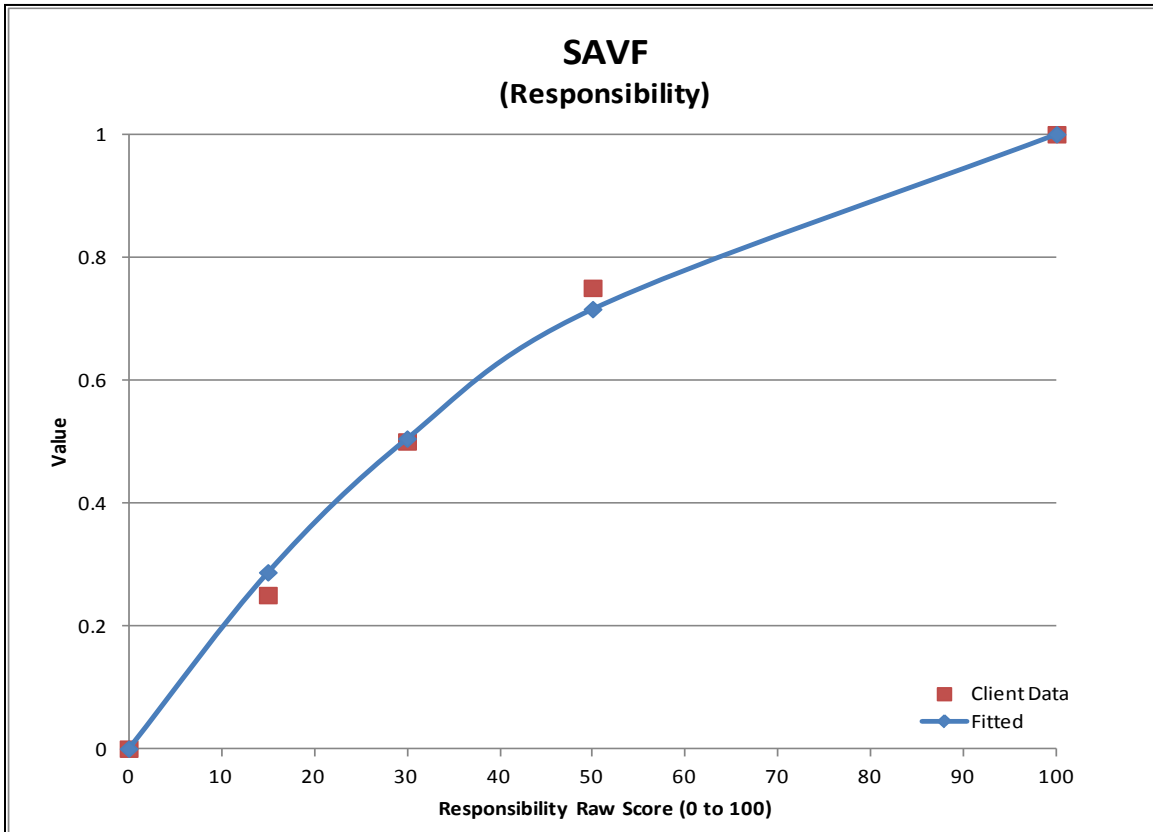


Figure 82. Responsibility Single Attribute Value Function

Single Attribute Value Function for Honesty and Accountability

Mid-Value Point	(Honesty & Accountability) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Honesty & Accountability Percentile of Employee versus Ideal Employee	Estimated Weighted Honesty & Accountability Category Score for EPR (Weighted at 5%)
x(bottom)	0	0	0.000000	0.00%
x(.25)	0.25	20	0.281123	1.41%
x(.50)	0.5	40	0.514129	2.57%
x(.75)	0.75	60	0.707255	3.54%
x(top)	1	100	1.000000	5.00%

Gamma
0.00938621
Sum SQ Diff @ 0.25, 0.50, 0.75
0.002995379
Function
$\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^* - x_1^0)}}, x_1 \in X_1$

NOTE

SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

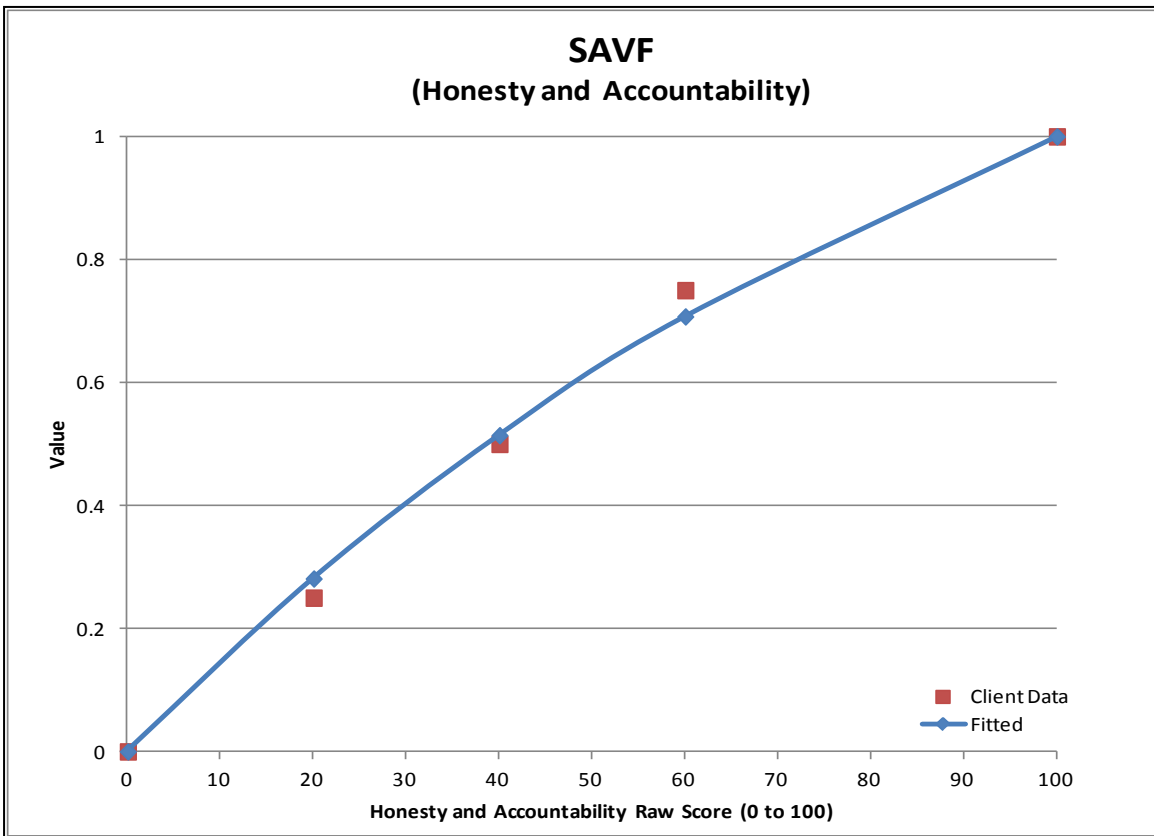


Figure 83. Honesty and Accountability Single Attribute Value Function

Single Attribute Value Function for Physical Fitness

Mid-Value Point	(Teamwork & Followership) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Section Slopes	Function Estimated Teamwork & Followership Percentile of Employee versus Ideal Employee	Estimated Weighted Teamwork & Followership Category Score for EPR (Weighted at 3%)
x(bottom)	0	0	0.00	0.000000	0.00%
x(.25)	0.25	74	296.00	0.250000	2.50%
x(.70)	0.65	75	2.50	0.650000	6.50%
x(.95)	0.95	90	50.00	0.950000	9.50%
x(top)	1	100	200.00	1.000000	10.00%

Function
Piecewise

NOTE

SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 65% of what an ideal employee would operate at, and at 95% of what an ideal employee would operate at

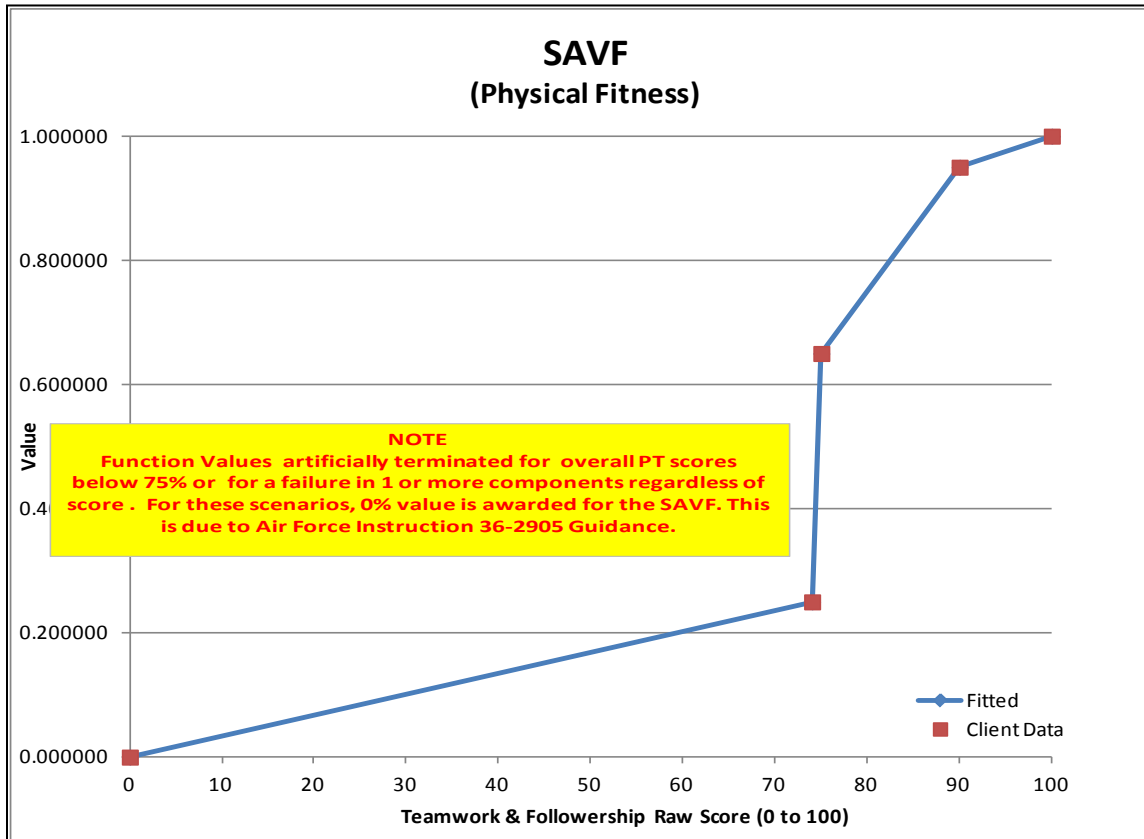


Figure 84. Physical Fitness Single Attribute Value Function

Single Attribute Value Function for Military Awards

Mid-Value Point	(Awards) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Awards Percentile of Employee versus Ideal Employee	Estimated Weighted Awards Category Score for EPR (Weighted at 4%)	Gamma
x(bottom)	0	0	0.000000	0.00%	0.01843588
x(.25)	0.25	15	0.287015	1.15%	Sum SQ Diff @ 0.25, 0.50, 0.75 0.002588741
x(.50)	0.5	30	0.504689	2.02%	
x(.75)	0.75	50	0.715408	2.86%	
x(top)	1	100	1.000000	4.00%	Function $\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^* - x_1^0)}}, x_1 \in X_1$

NOTE
SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

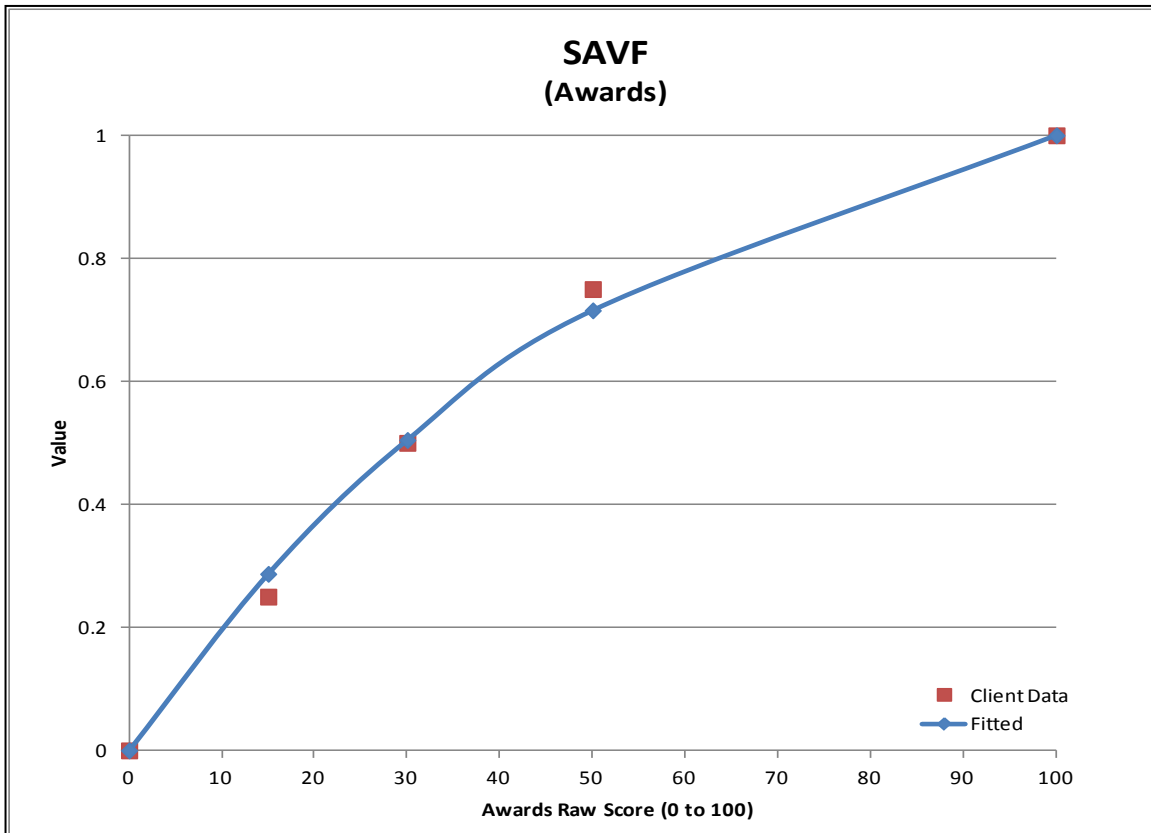


Figure 85. Military Awards Single Attribute Value Function

Single Attribute Value Function for Base and Community Involvement

Mid-Value Point	(Base & Community Involvement) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Function Estimated Base & Community Involvement Percentile of Employee versus Ideal Employee	Estimated Weighted Base & Community Involvement Category Score for EPR (Weighted at 3%)	
x(bottom)	0	0	0.000000	0.00%	Gamma -0.00281841
x(.25)	0.25	30	0.271003	0.81%	
x(.50)	0.5	50	0.464828	1.39%	Sum SQ Diff @ 0.25, 0.50, 0.75 0.002398685
x(.75)	0.75	80	0.776842	2.33%	
x(top)	1	100	1.000000	3.00%	Function $\frac{1 - e^{-\gamma_1(x_1 - x_1^0)}}{1 - e^{-\gamma_1(x_1^1 - x_1^0)}}, x_1 \in X_1$

NOTE
SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

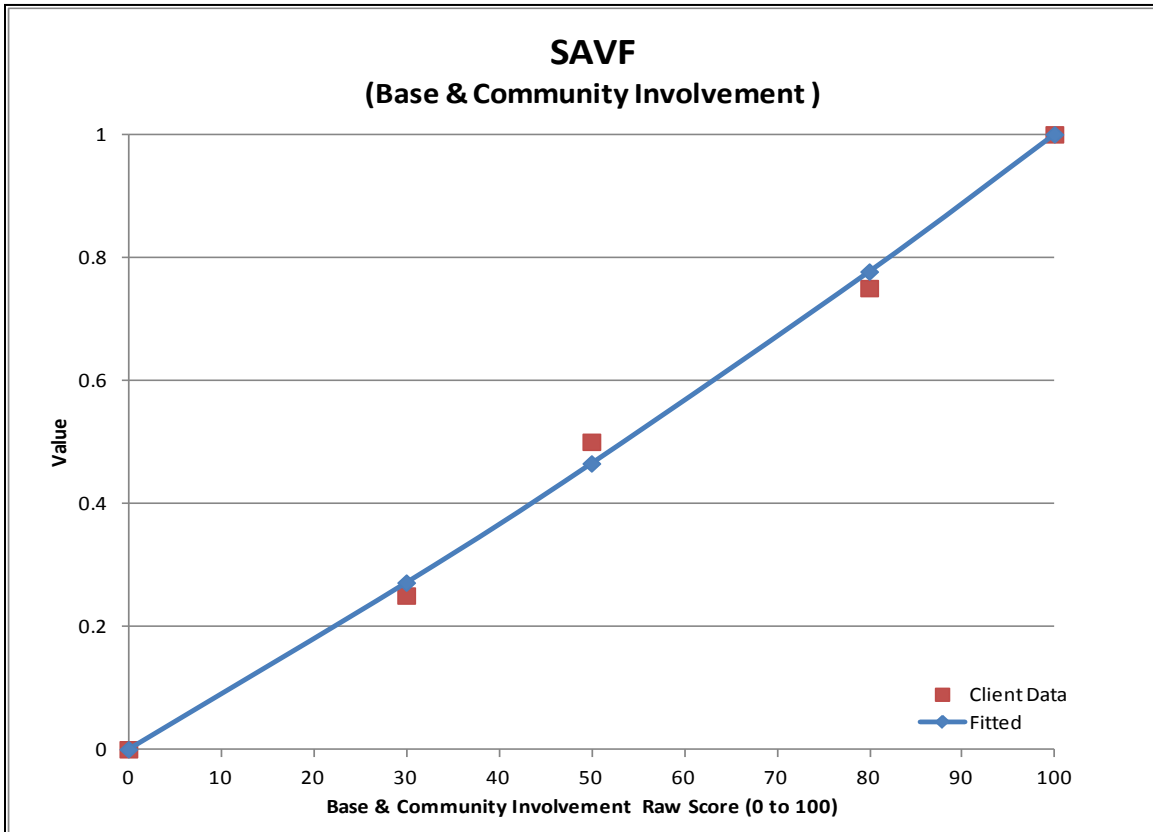


Figure 86. Base and Community Involvement Single Attribute Value Function

Single Attribute Value Function for Education

Mid-Value Point	(Education Level) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (0 to 100) Points	Section Slopes	Function Estimated Education Level Percentile of Employee versus Ideal Employee	Estimated Weighted Education Level Category Score for EPR (Weighted at 3%)
x(bottom)	0	0	0	0.000000	0.00%
x(.25)	0.25	1	4	0.250000	0.75%
x(.50)	0.5	50	196	0.500000	1.50%
x(.75)	0.75	70	80	0.750000	2.25%
x(top)	1	100	120	1.000000	3.00%

Function
Piecewise

NOTE
 SME provided score based on what he felt an employee operating at 25% of what a perfect employee would operate at, a score at 50% of what an ideal employee would operate at, and at 75% of what an ideal employee would operate at

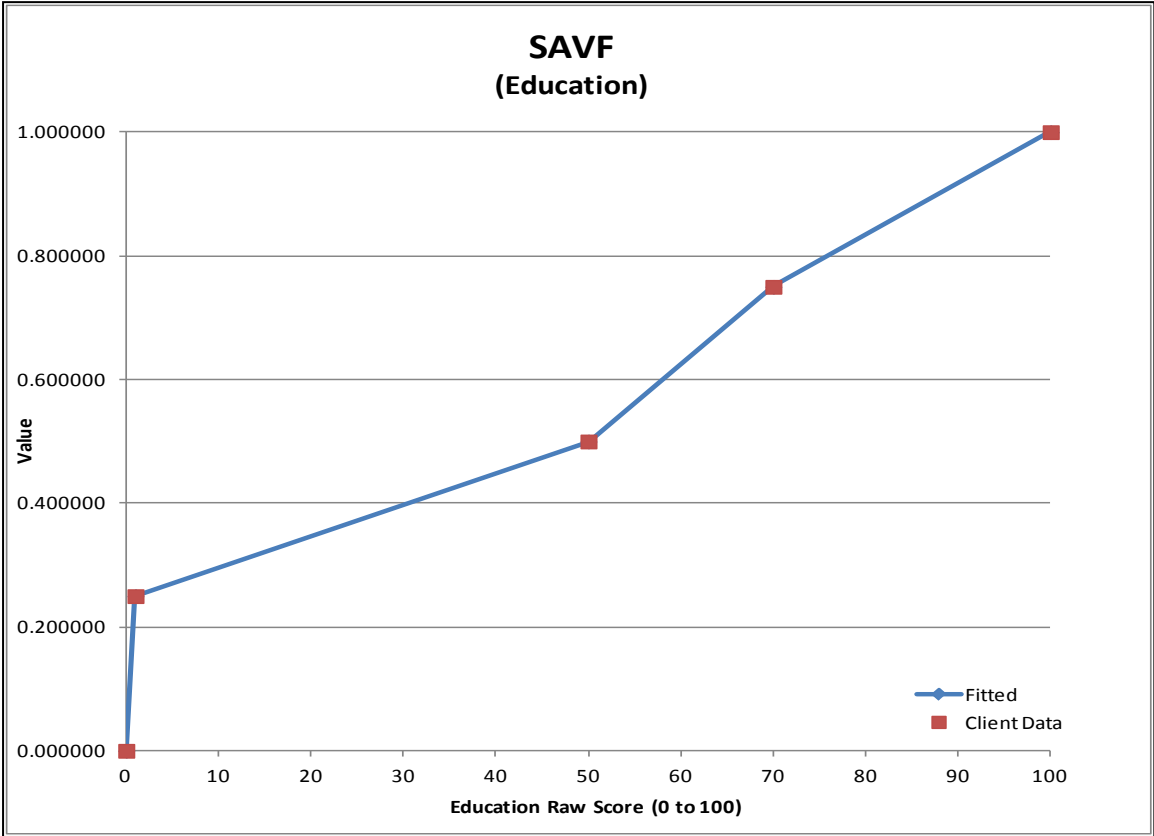


Figure 87. Education Single Attribute Value Function

Single Attribute Value Function for Administrative Actions Correction Factor

Mid-Value Point	(Administrative Actions) Percentile Employee Operates at vs. Ideal Employee	Raw SME Score for Employee (-100 to 0) Points	Section Slopes	Function Estimated Administrative Actions Percentile of Employee versus Ideal Employee	Estimated Weighted Administrative Actions Category Score for EPR (Weighted at 3%)
x(bottom)	-1.00	-100	0	-1.000000	-35.00%
x(.80)	-0.80	-80	100	-0.800000	-28.00%
x(.45)	-0.45	-60	57.14285714	-0.450000	-15.75%
x(.15)	-0.15	-30	100	-0.150000	-5.25%
x(top)	0.00	0	200	0.000000	0.00%

Function
Piecewise

NOTE

SME provided score based on what he felt an employee operating at 15% less of what a perfect employee would operate at, a score at 45% less of what an ideal employee would operate at, and at 80% less of what an ideal employee would operate at

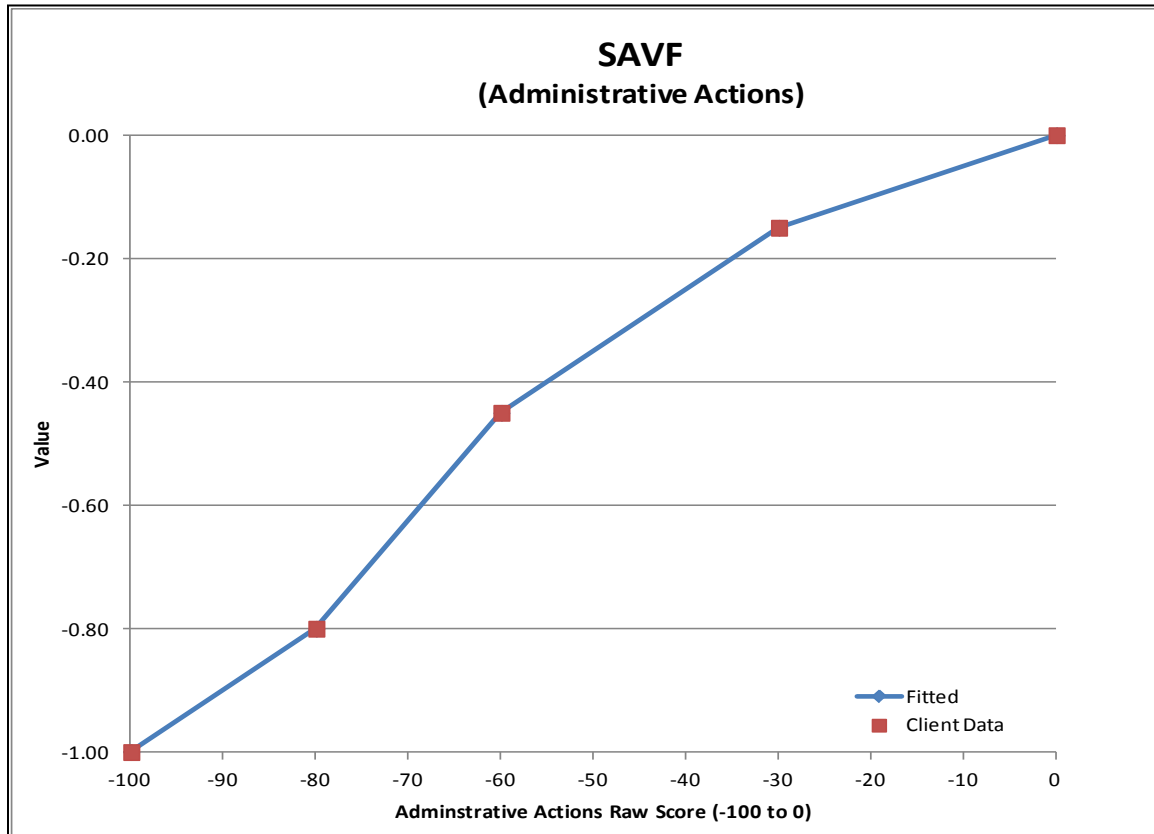


Figure 88. Administrative Actions Independent External Function

Appendix II

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 1

wi	Service Before Self					Integrity			Excellence				MAVF (Weighted)	Rank
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03	0.03		
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 89. Scores For Eight Notional Airmen Using Provided Weights

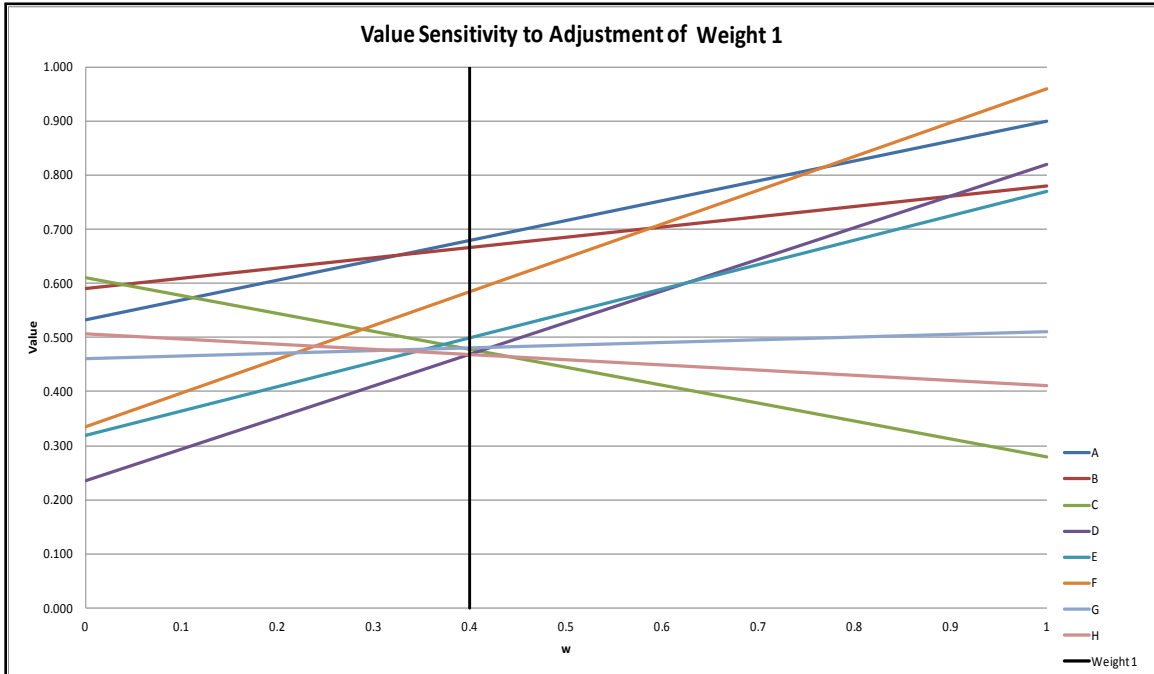


Figure 90. Score Changes From Performance Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 2

wi	Service Before Self				Integrity			Excellence				MAVF (Weighted)	Rank	
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03			0.03
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 91. Scores For Eight Notional Airmen Using Provided Weights

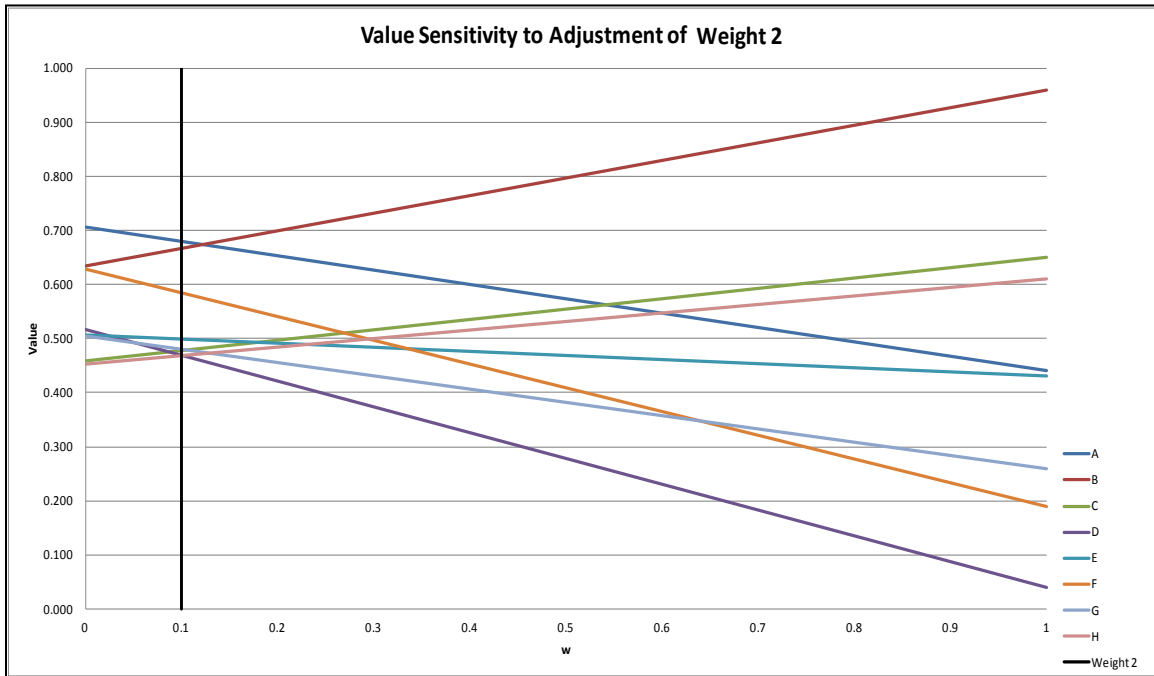


Figure 92. Score Changes From Duty Leadership Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 3

Alt	Service Before Self					Integrity			Excellence				MAVF (Weighted)	Rank
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03	0.03		
SAVF Scores After Function Applied Unweighted														
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 93. Scores For Eight Notional Airmen Using Provided Weights

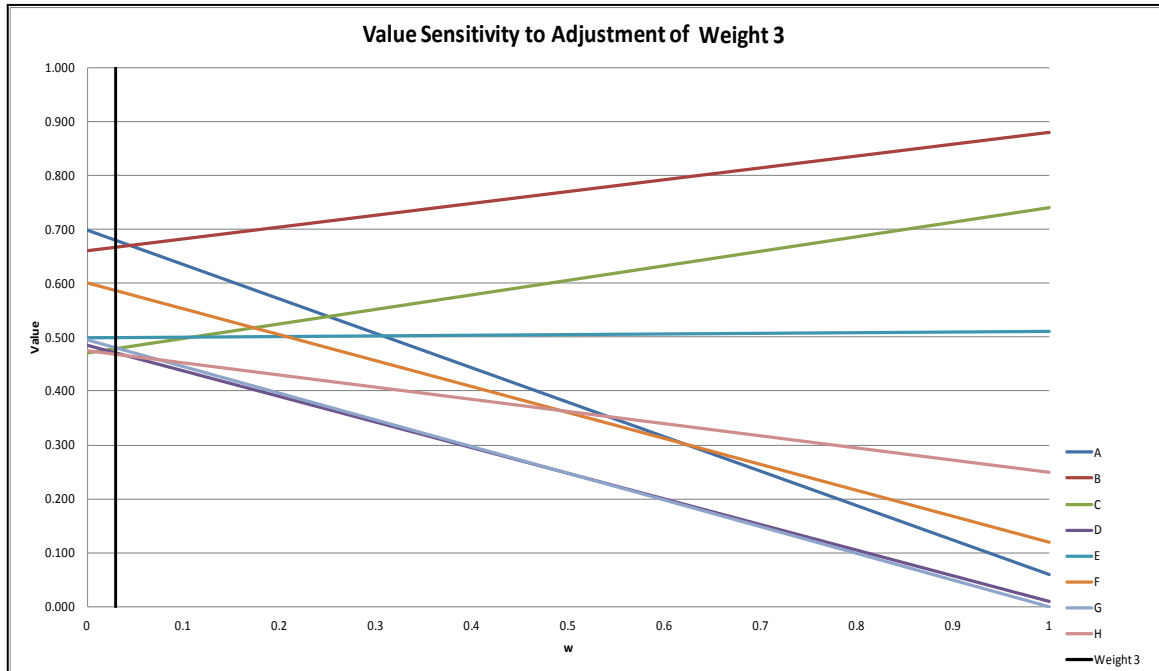


Figure 94. Score Changes From Teamwork and Followership Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 4

wi Alt	Service Before Self					Integrity			Excellence				MAVF (Weighted)	Rank
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03	0.03		
SAVF Scores After Function Applied Unweighted														
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 95. Scores For Eight Notional Airmen Using Provided Weights

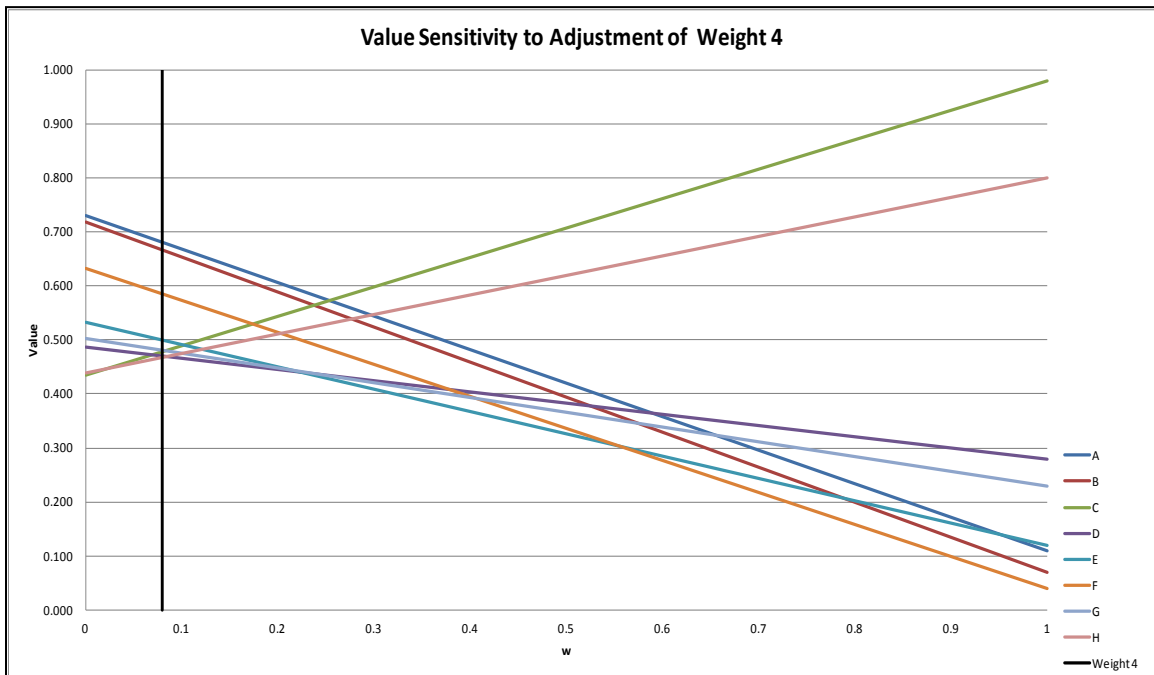


Figure 96. Score Changes From Respect for Service and Standards Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 5

wi	Service Before Self				Integrity			Excellence				MAVF (Weighted)	Rank	
	0.4	0.10	0.03	0.08	0.05	0.04	0.05	0.10	0.04	0.03	0.03			
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 97. Scores For Eight Notional Airmen Using Provided Weights

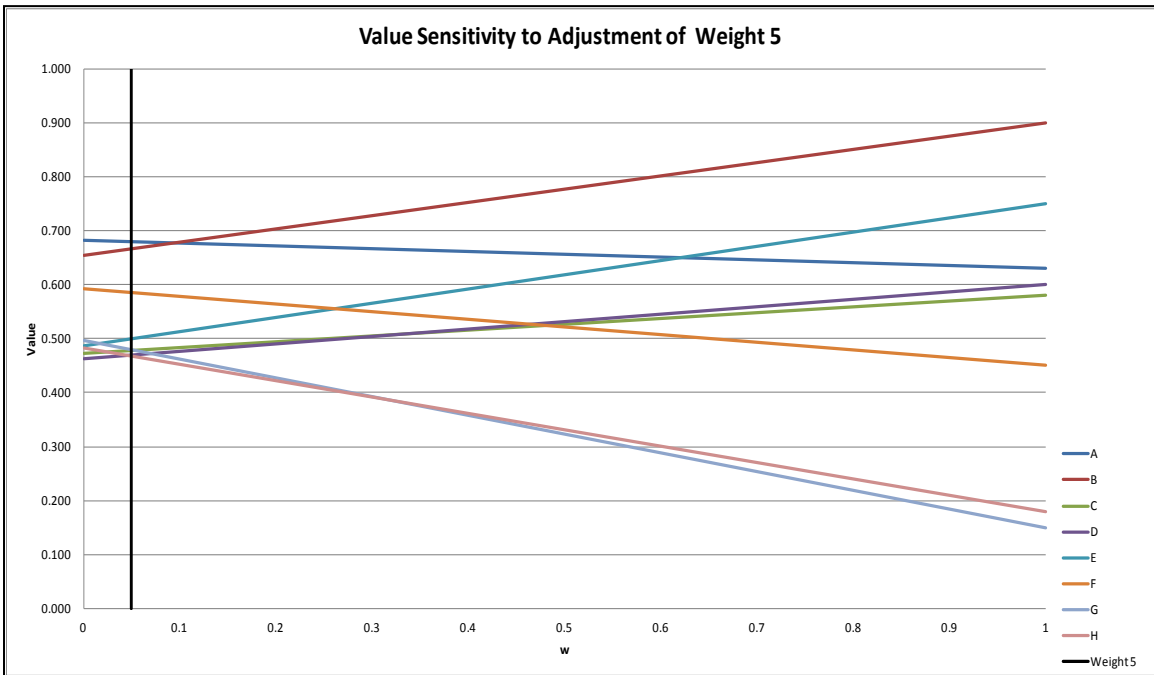


Figure 98. Score Changes From Discipline and Self-Control Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 6

wi	Service Before Self					Integrity			Excellence				MAVF (Weighted)	Rank
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03	0.03		
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl	MAVF (Weighted)	Rank
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 99. Scores For Eight Notional Airmen Using Provided Weights

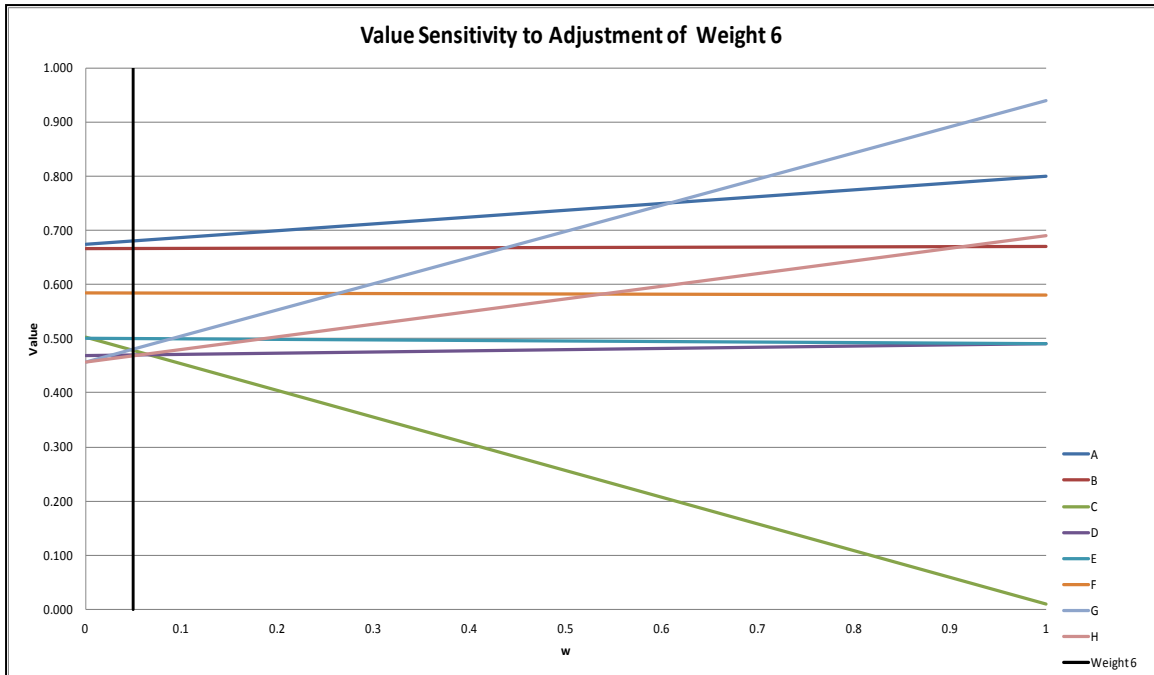


Figure 100. Score Changes From Communication Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 7

wi	Service Before Self				Integrity			Excellence				MAVF (Weighted)	Rank	
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03			0.03
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 101. Scores For Eight Notional Airmen Using Provided Weights

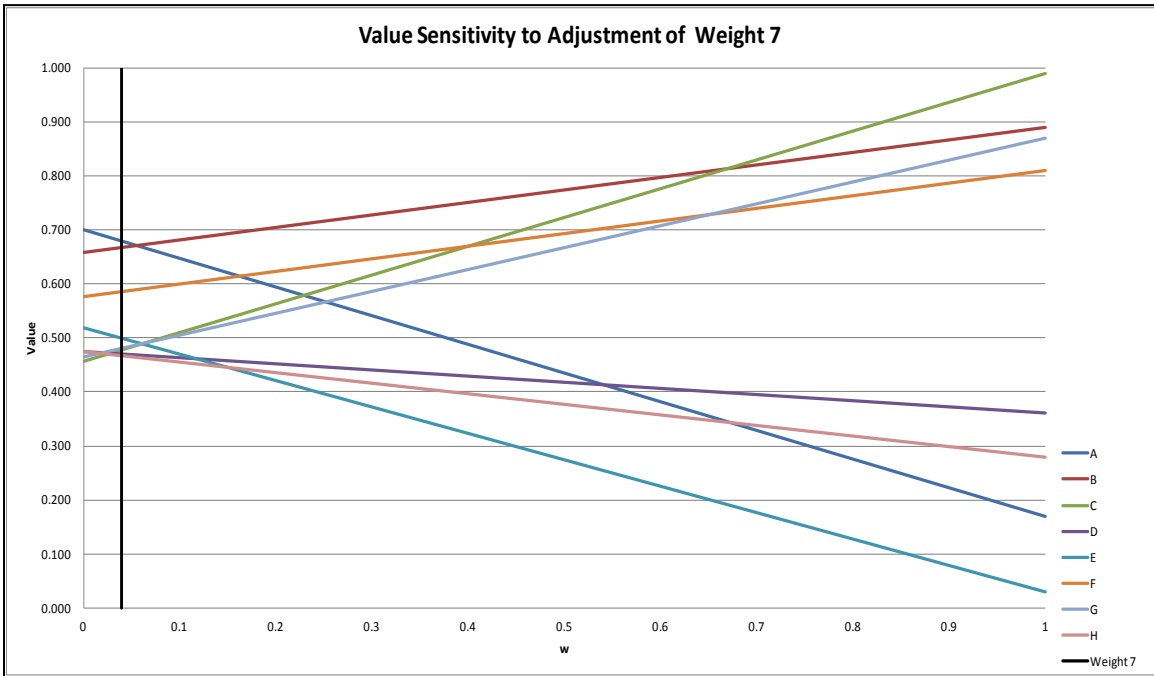


Figure 102. Score Changes From Responsibility Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 8

wi	Service Before Self				Integrity				Excellence				MAVF (Weighted)	Rank
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.20	0.03		
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 103. Scores For Eight Notional Airmen Using Provided Weights

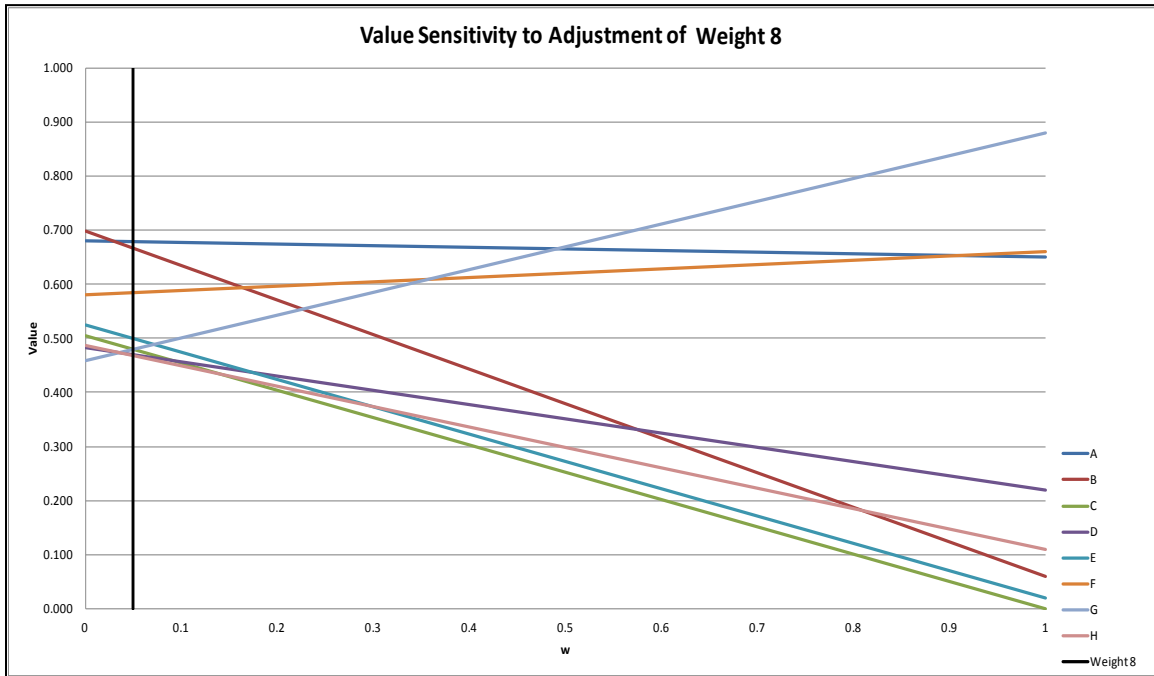


Figure 104. Score Changes From Honesty and Accountability Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 9

wi	Service Before Self				Integrity			Excellence				MAVF (Weighted)	Rank	
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03			0.03
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 105. Scores For Eight Notional Airmen Using Provided Weights

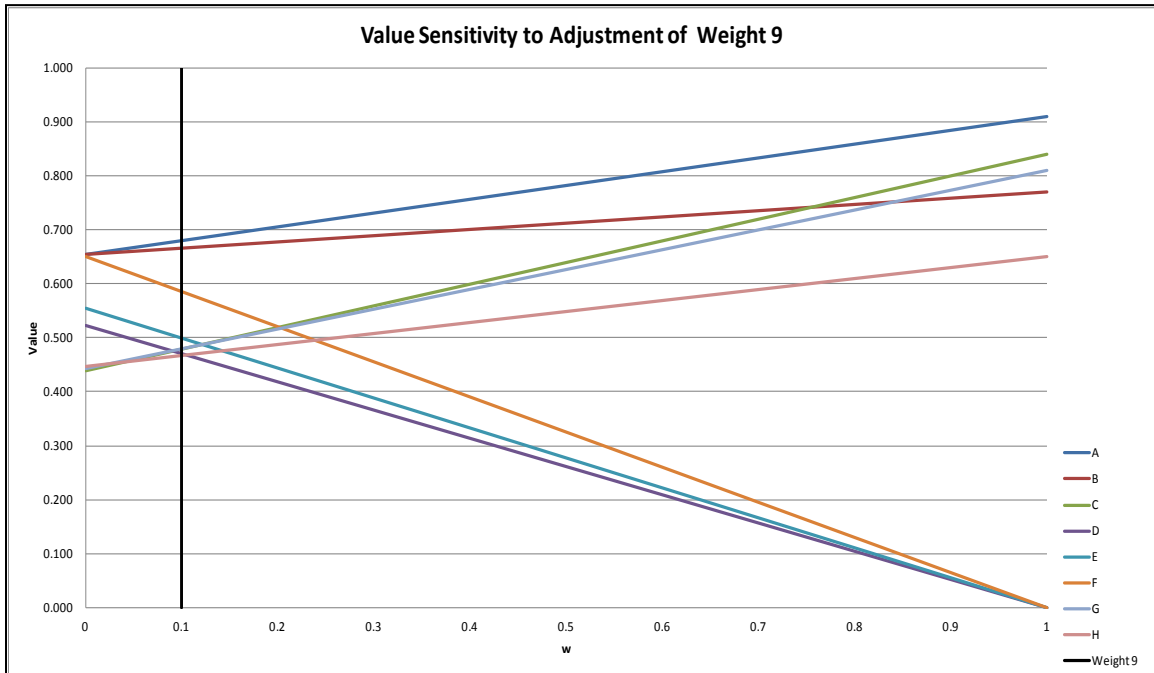


Figure 106. Score Changes From Physical Fitness Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 10

wi	Service Before Self				Integrity			Excellence				MAVF (Weighted)	Rank	
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03			0.03
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 107. Scores For Eight Notional Airmen Using Provided Weights

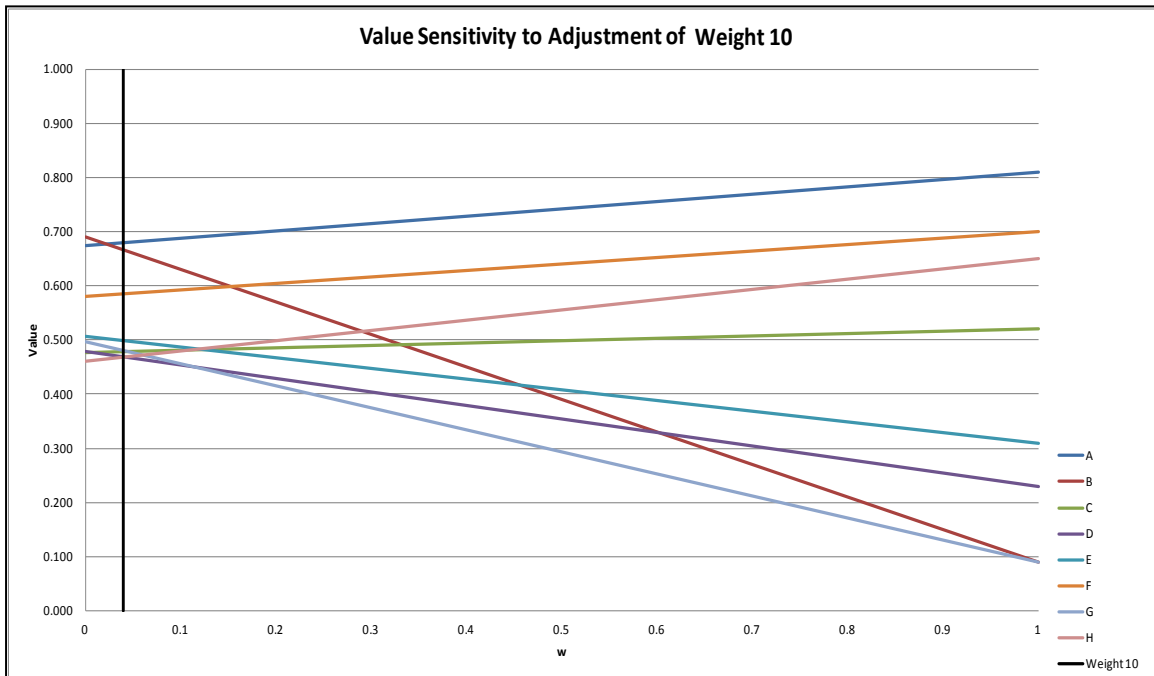


Figure 108. Score Changes From Military Awards Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 11

wi	Service Before Self				Integrity				Excellence				MAVF (Weighted)	Rank
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03	0.03		
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl		
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 109. Scores For Eight Notional Airmen Using Provided Weights

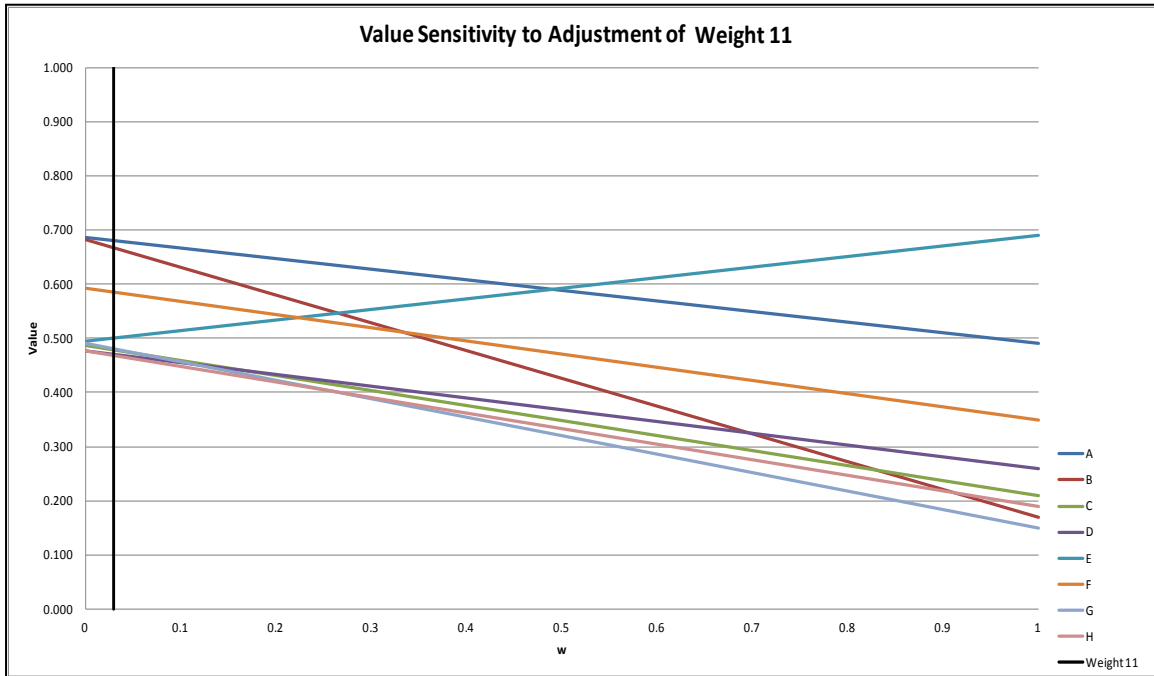


Figure 110. Score Changes From Base and Community Involvement Weight Change

Weight Sensitivity Analysis of Overall Scores For Eight Notional Airmen Sensitivity Analysis of Weight 12

wi	Service Before Self				Integrity				Excellence				MAVF (Weighted)	Rank
	0.4	0.10	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03	0.03		
Alt	SAVF Scores After Function Applied Unweighted													
	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl	MAVF (Weighted)	Rank
Utopia	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.000	
A	0.90	0.44	0.06	0.11	0.63	0.80	0.17	0.65	0.91	0.81	0.49	0.53	0.679	1
B	0.78	0.96	0.88	0.07	0.90	0.67	0.89	0.06	0.77	0.09	0.17	0.79	0.667	2
C	0.28	0.65	0.74	0.98	0.58	0.01	0.99	0.00	0.84	0.52	0.21	0.69	0.479	6
D	0.82	0.04	0.01	0.28	0.60	0.49	0.36	0.22	0.00	0.23	0.26	0.60	0.470	7
E	0.77	0.43	0.51	0.12	0.75	0.49	0.03	0.02	0.00	0.31	0.69	0.88	0.500	4
F	0.96	0.19	0.12	0.04	0.45	0.58	0.81	0.66	0.00	0.70	0.35	0.66	0.585	3
G	0.51	0.26	0.00	0.23	0.15	0.94	0.87	0.88	0.81	0.09	0.15	0.31	0.480	5
H	0.41	0.61	0.25	0.80	0.18	0.69	0.28	0.11	0.65	0.65	0.19	0.48	0.468	8

Min 0.65 Value
 for raw score of 75%
 per AFI 36-2905
 0.00 value < 75%
 raw score

Figure 111. Scores For Eight Notional Airmen Using Provided Weights

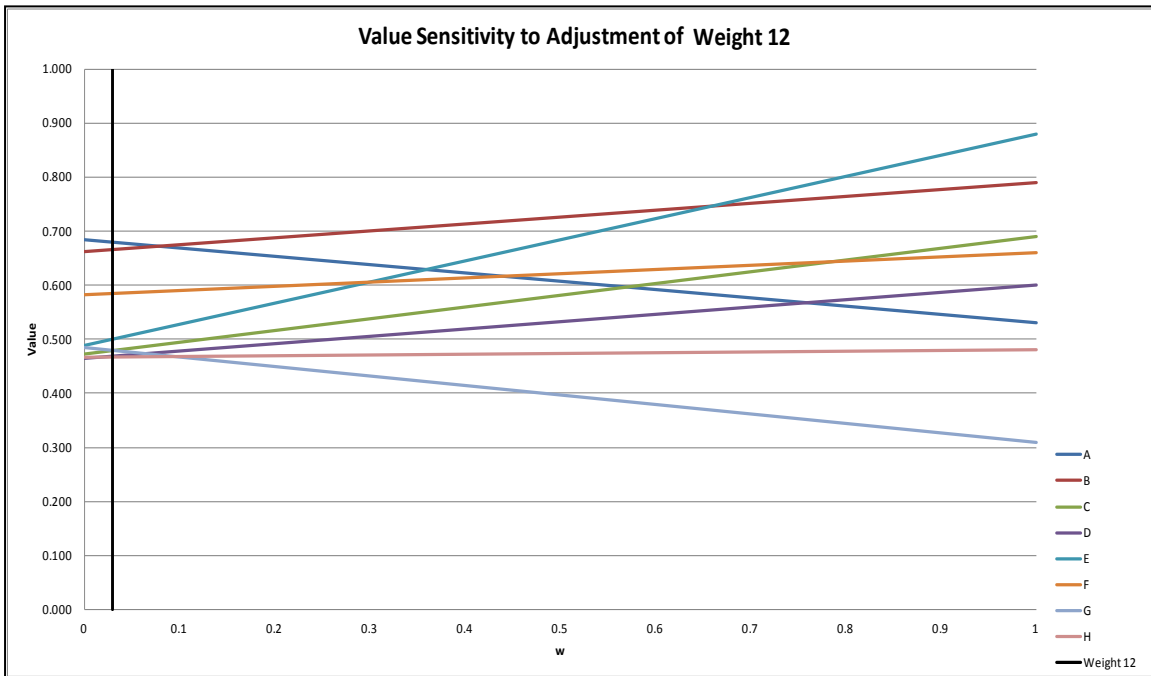


Figure 112. Score Changes From Education Level Weight Change

Appendix III

Value Breakout Attribute Contribution for Each JEPR Attribute For Scores of Eight Notional Airmen Using Provided Weights

Alt	1.0												MAVF (Weighted)	Rank
	0.66					0.14			0.20					
	0.4	0.1	0.03	0.08	0.05	0.05	0.04	0.05	0.10	0.04	0.03	0.03		
	1	2	3	4	5	6	7	8	9	10	11	12		
w1v1	w2v2	w3v3	w4v4	w5v5	w6v6	w7v7	w8v8	w9v9	w10v10	w11v11	w12v12			
Utopia	0.400	0.100	0.030	0.080	0.050	0.050	0.040	0.050	0.10	0.040	0.030	0.030	1.000	
A	0.360	0.044	0.002	0.009	0.032	0.040	0.007	0.033	0.09	0.032	0.015	0.016	0.679	1
B	0.312	0.096	0.026	0.006	0.045	0.034	0.036	0.003	0.08	0.004	0.005	0.024	0.667	2
C	0.112	0.065	0.022	0.078	0.029	0.001	0.040	0.000	0.08	0.021	0.006	0.021	0.479	6
D	0.328	0.004	0.000	0.022	0.030	0.025	0.014	0.011	0.00	0.009	0.008	0.018	0.470	7
E	0.308	0.043	0.015	0.010	0.038	0.025	0.001	0.001	0.00	0.012	0.021	0.026	0.500	4
F	0.384	0.019	0.004	0.003	0.023	0.029	0.032	0.033	0.00	0.028	0.011	0.020	0.585	3
G	0.204	0.026	0.000	0.018	0.008	0.047	0.035	0.044	0.08	0.004	0.005	0.009	0.480	5
H	0.164	0.061	0.008	0.064	0.009	0.035	0.011	0.006	0.07	0.026	0.006	0.014	0.468	8

Min 0.65 Value
for raw score of 75%
per AFI 36-2905
0.00 value < 75%
raw score

Figure 113. Scores for Eight Notional Airmen Using Provided Weights

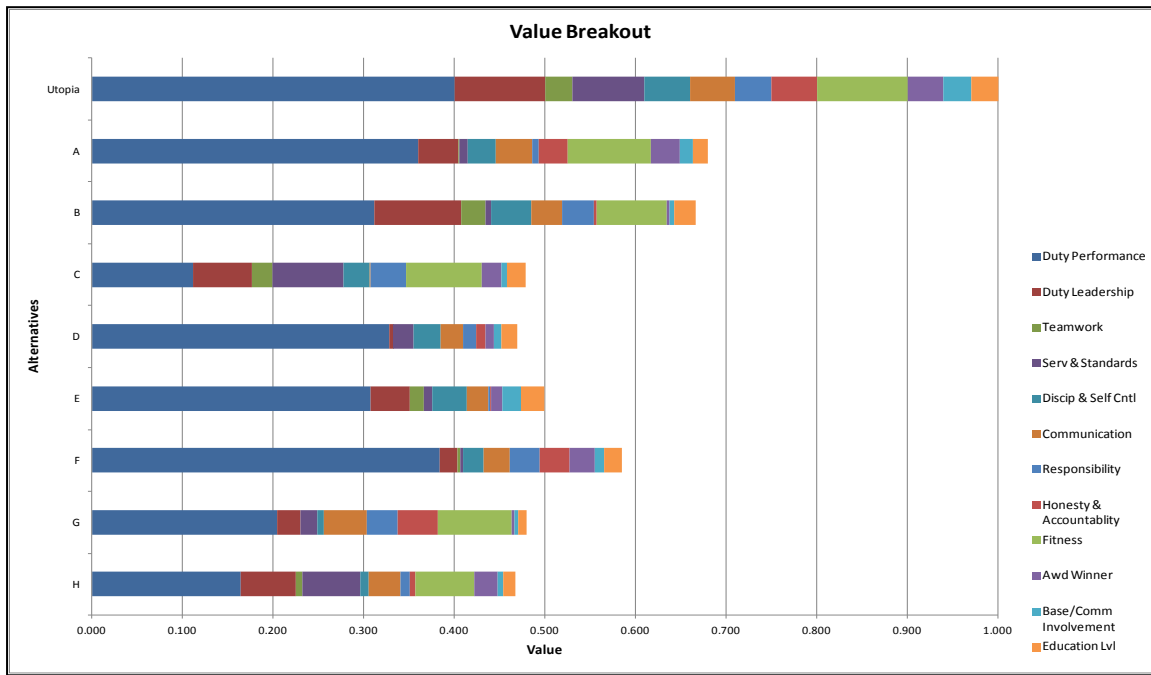


Figure 114. Contribution to Overall Score by Value Type for Eight Notional Airmen

Appendix IV

Value Breakout Contribution for Each JEPR Fundamental Objective For Scores of Eight Notional Airmen Using Provided Weights

	Service Before Self	Integrity	Excellence	
Alt	1-5	6-8	9-12	V
Utopia	0.660	0.140	0.200	1.000
A	0.446	0.079	0.154	0.679
B	0.485	0.072	0.109	0.667
C	0.307	0.040	0.132	0.479
D	0.385	0.050	0.035	0.470
E	0.413	0.027	0.060	0.500
F	0.432	0.094	0.058	0.585
G	0.256	0.126	0.098	0.480
H	0.306	0.051	0.111	0.468

Figure 115. Scores for Eight Notional Airmen Using Provided Weights

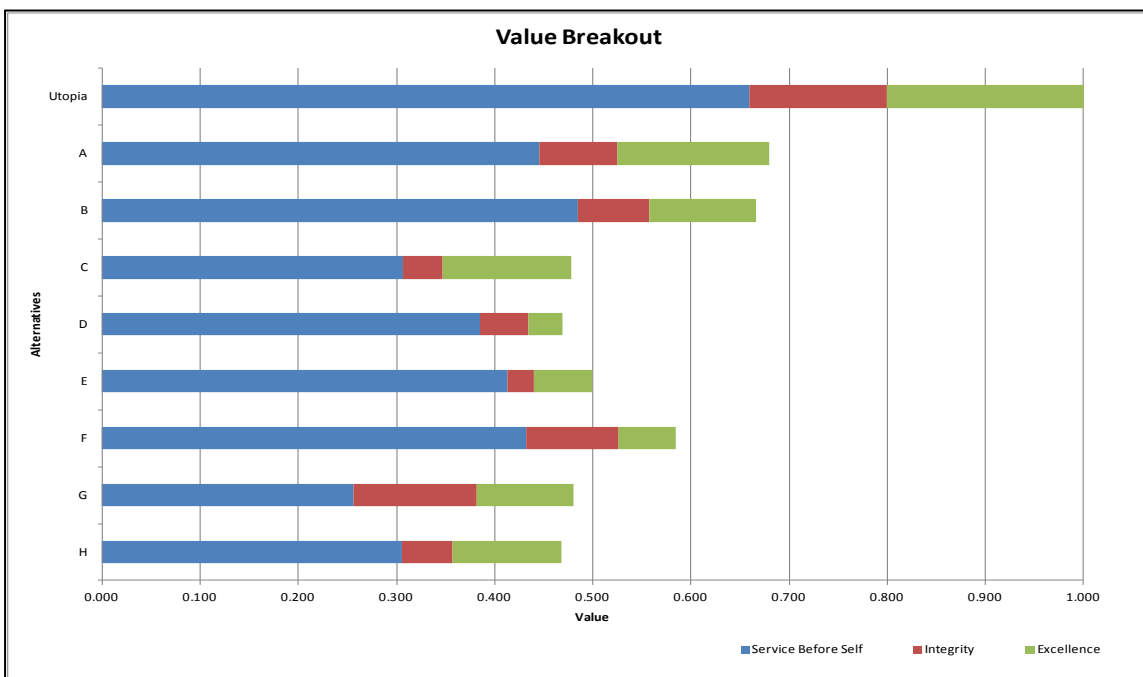


Figure 116. Contribution to Overall Score by Value Type for Eight Notional Airmen

Appendix V

JEPR Value Gap Feedback Scores Value-Gap Strengths and Shortfalls of Notional Airman A

Value Gap for Hypothetical Airmen A												
Hypothetical Airmen	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl
A	0.040	0.056	0.028	0.071	0.019	0.010	0.033	0.018	0.009	0.008	0.015	0.014
B	0.088	0.004	0.004	0.074	0.005	0.017	0.004	0.047	0.023	0.036	0.025	0.006
C	0.288	0.035	0.008	0.002	0.021	0.050	0.000	0.050	0.016	0.019	0.024	0.009
D	0.072	0.096	0.030	0.058	0.020	0.026	0.026	0.039	0.100	0.031	0.022	0.012
E	0.092	0.057	0.015	0.070	0.013	0.026	0.039	0.049	0.100	0.028	0.009	0.004
F	0.016	0.081	0.026	0.077	0.028	0.021	0.008	0.017	0.100	0.012	0.020	0.010
G	0.196	0.074	0.030	0.062	0.043	0.003	0.005	0.006	0.019	0.036	0.026	0.021
H	0.236	0.039	0.023	0.016	0.041	0.016	0.029	0.045	0.035	0.014	0.024	0.016
Attribute Score	0.3600	0.0440	0.0018	0.0088	0.0315	0.0400	0.0068	0.0325	0.0910	0.0324	0.0147	0.0159
Gap	0.0400	0.0560	0.0282	0.0712	0.0185	0.0100	0.0332	0.0175	0.0090	0.0076	0.0153	0.0141

Figure 117. Scores for Eight Notional Airmen (Airman A Highlighted)

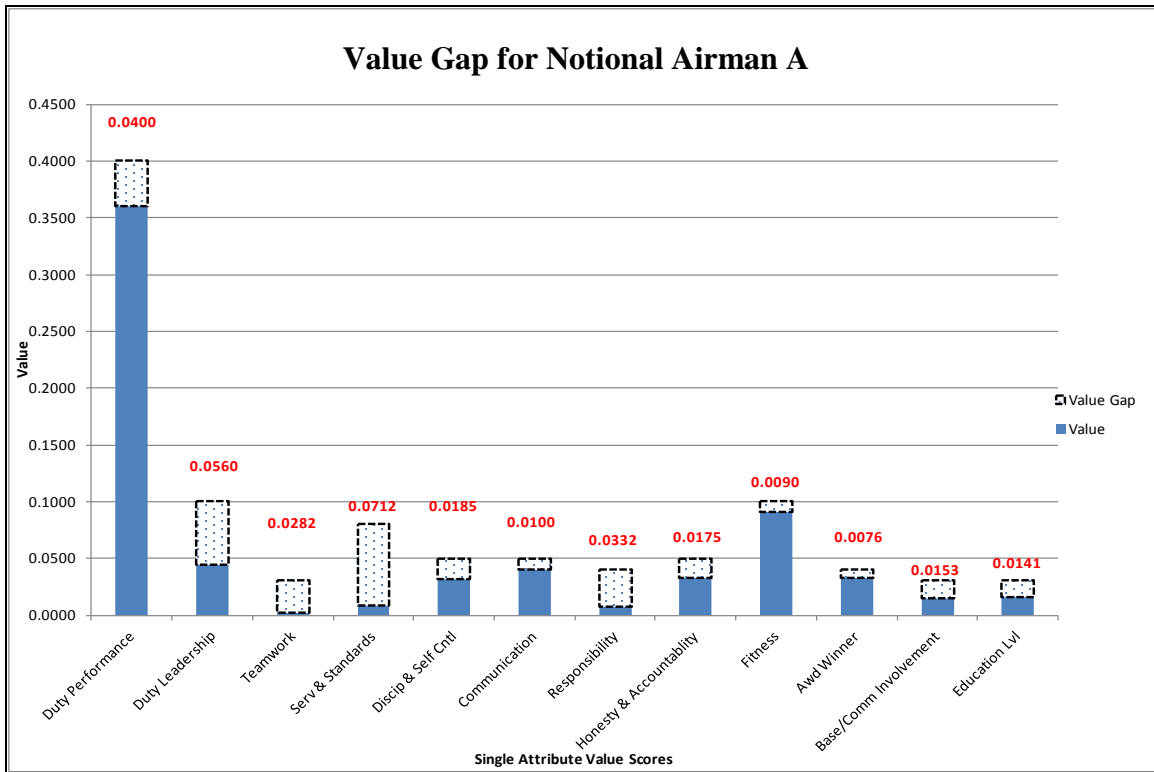


Figure 118. Value Gap Feedback for Notional Airman A

JEPR Value Gap Feedback Scores Value Gap Strengths and Shortfalls of Notional Airman B

Value Gap for Hypothetical Airmen B												
Hypothetical Airmen	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl
A	0.040	0.056	0.028	0.071	0.019	0.010	0.033	0.018	0.009	0.008	0.015	0.014
B	0.088	0.004	0.004	0.074	0.005	0.017	0.004	0.047	0.023	0.036	0.025	0.006
C	0.288	0.035	0.008	0.002	0.021	0.050	0.000	0.050	0.016	0.019	0.024	0.009
D	0.072	0.096	0.030	0.058	0.020	0.026	0.026	0.039	0.100	0.031	0.022	0.012
E	0.092	0.057	0.015	0.070	0.013	0.026	0.039	0.049	0.100	0.028	0.009	0.004
F	0.016	0.081	0.026	0.077	0.028	0.021	0.008	0.017	0.100	0.012	0.020	0.010
G	0.196	0.074	0.030	0.062	0.043	0.003	0.005	0.006	0.019	0.036	0.026	0.021
H	0.236	0.039	0.023	0.016	0.041	0.016	0.029	0.045	0.035	0.014	0.024	0.016
Attribute Score	0.3120	0.0960	0.0264	0.0056	0.0450	0.0335	0.0356	0.0030	0.0770	0.0036	0.0051	0.0237
Gap	0.0880	0.0040	0.0036	0.0744	0.0050	0.0165	0.0044	0.0470	0.0230	0.0364	0.0249	0.0063

Figure 119. Scores for Eight Notional Airmen (Airman B Highlighted)

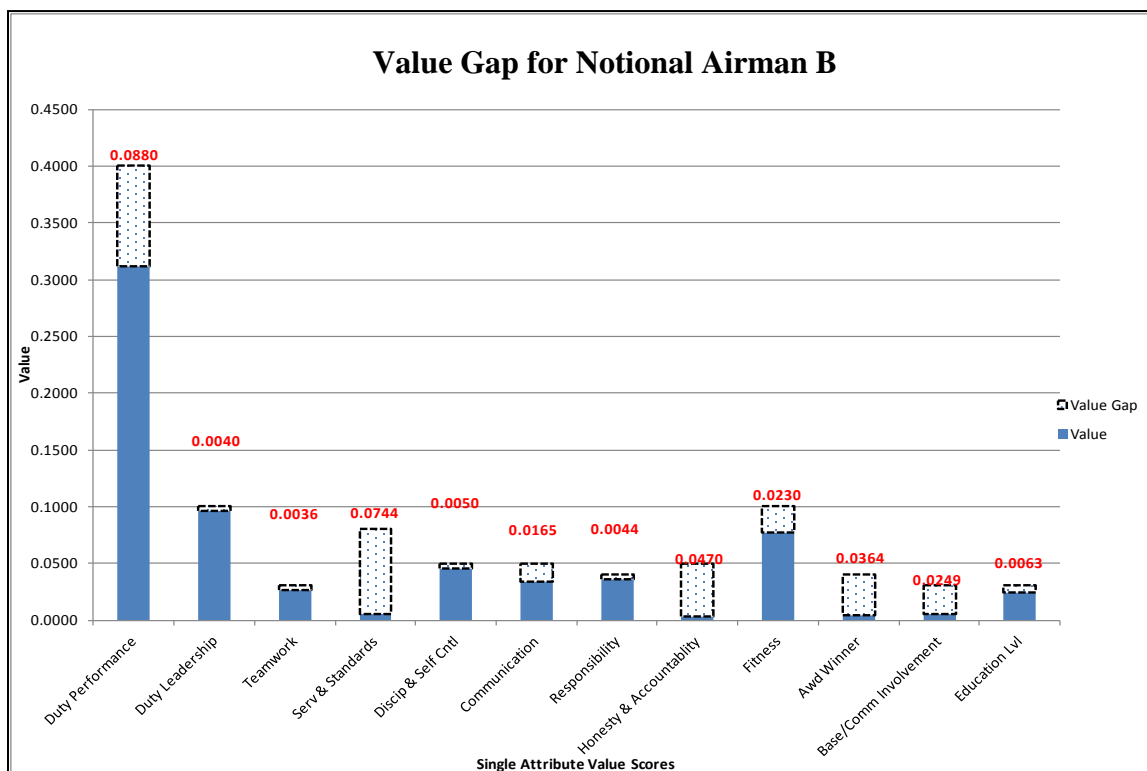


Figure 120. Value Gap Feedback for Notional Airman B

JEPR Value Gap Feedback Scores Value Gap Strengths and Shortfalls of Notional Airman C

<i>Value Gap for Hypothetical Airmen C</i>												
Hypothetical Airmen	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl
A	0.040	0.056	0.028	0.071	0.019	0.010	0.033	0.018	0.009	0.008	0.015	0.014
B	0.088	0.004	0.004	0.074	0.005	0.017	0.004	0.047	0.023	0.036	0.025	0.006
C	0.288	0.035	0.008	0.002	0.021	0.050	0.000	0.050	0.016	0.019	0.024	0.009
D	0.072	0.096	0.030	0.058	0.020	0.026	0.026	0.039	0.100	0.031	0.022	0.012
E	0.092	0.057	0.015	0.070	0.013	0.026	0.039	0.049	0.100	0.028	0.009	0.004
F	0.016	0.081	0.026	0.077	0.028	0.021	0.008	0.017	0.100	0.012	0.020	0.010
G	0.196	0.074	0.030	0.062	0.043	0.003	0.005	0.006	0.019	0.036	0.026	0.021
H	0.236	0.039	0.023	0.016	0.041	0.016	0.029	0.045	0.035	0.014	0.024	0.016
Attribute Score	0.1120	0.0650	0.0222	0.0784	0.0290	0.0005	0.0396	0.0000	0.0840	0.0208	0.0063	0.0207
Gap	0.2880	0.0350	0.0078	0.0016	0.0210	0.0495	0.0004	0.0500	0.0160	0.0192	0.0237	0.0093

Figure 121. Scores for Eight Notional Airmen (Airman C Highlighted)

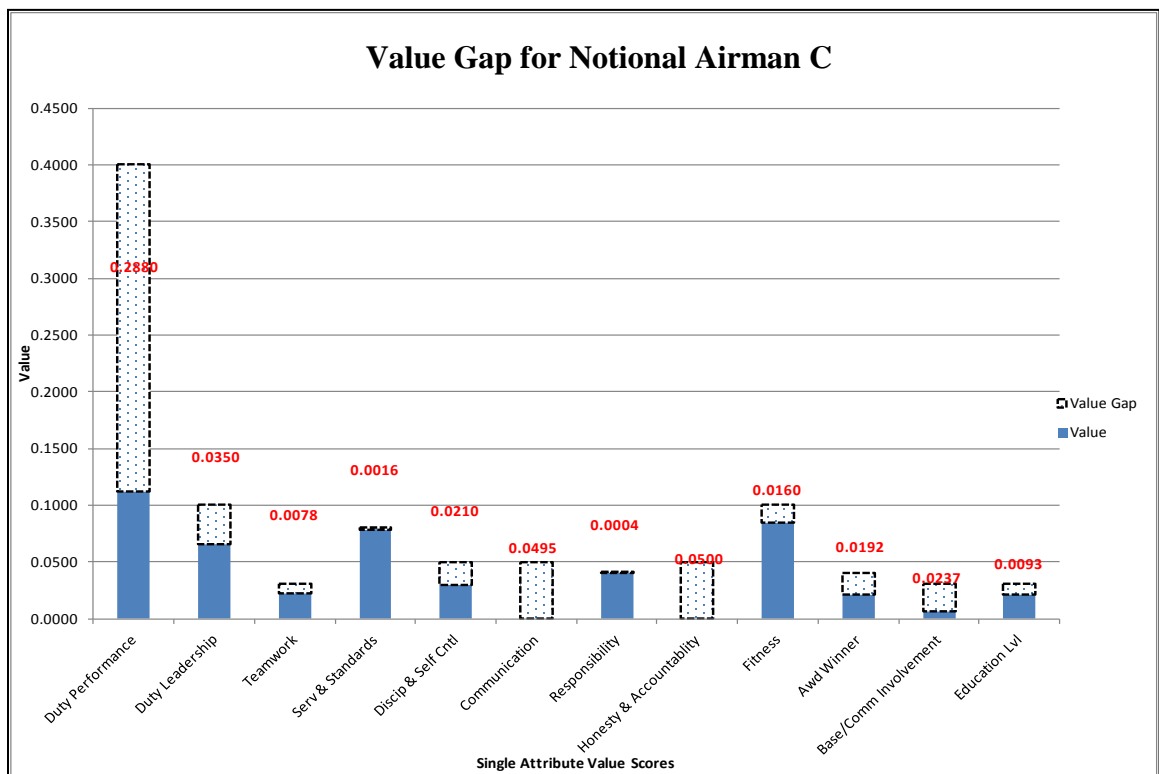


Figure 122. Value Gap Feedback for Notional Airman C

JEPR Value Gap Feedback Scores Value Gap Strengths and Shortfalls of Notional Airman D

Value Gap for Hypothetical Airmen D												
Hypothetical Airmen	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl
A	0.040	0.056	0.028	0.071	0.019	0.010	0.033	0.018	0.009	0.008	0.015	0.014
B	0.088	0.004	0.004	0.074	0.005	0.017	0.004	0.047	0.023	0.036	0.025	0.006
C	0.288	0.035	0.008	0.002	0.021	0.050	0.000	0.050	0.016	0.019	0.024	0.009
D	0.072	0.096	0.030	0.058	0.020	0.026	0.026	0.039	0.100	0.031	0.022	0.012
E	0.092	0.057	0.015	0.070	0.013	0.026	0.039	0.049	0.100	0.028	0.009	0.004
F	0.016	0.081	0.026	0.077	0.028	0.021	0.008	0.017	0.100	0.012	0.020	0.010
G	0.196	0.074	0.030	0.062	0.043	0.003	0.005	0.006	0.019	0.036	0.026	0.021
H	0.236	0.039	0.023	0.016	0.041	0.016	0.029	0.045	0.035	0.014	0.024	0.016
Attribute Score	0.3280	0.0040	0.0003	0.0224	0.0300	0.0245	0.0144	0.0110	0.0000	0.0092	0.0078	0.0180
Gap	0.0720	0.0960	0.0297	0.0576	0.0200	0.0255	0.0256	0.0390	0.1000	0.0308	0.0222	0.0120

Figure 123. Scores for Eight Notional Airmen (Airman D Highlighted)

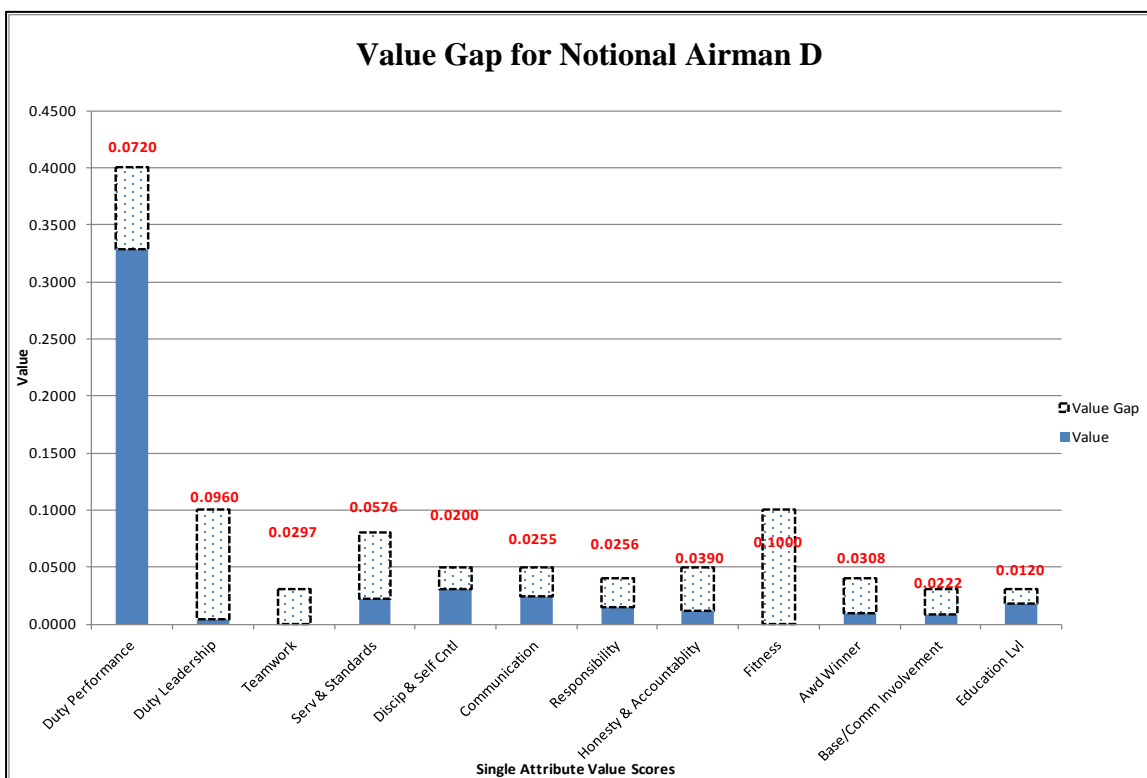


Figure 124. Value Gap Feedback for Notional Airman D

JEPR Value Gap Feedback Scores Value Gap Strengths and Shortfalls of Notional Airman E

Value Gap for Hypothetical Airmen E												
Hypothetical Airmen	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl
A	0.040	0.056	0.028	0.071	0.019	0.010	0.033	0.018	0.009	0.008	0.015	0.014
B	0.088	0.004	0.004	0.074	0.005	0.017	0.004	0.047	0.023	0.036	0.025	0.006
C	0.288	0.035	0.008	0.002	0.021	0.050	0.000	0.050	0.016	0.019	0.024	0.009
D	0.072	0.096	0.030	0.058	0.020	0.026	0.026	0.039	0.100	0.031	0.022	0.012
E	0.092	0.057	0.015	0.070	0.013	0.026	0.039	0.049	0.100	0.028	0.009	0.004
F	0.016	0.081	0.026	0.077	0.028	0.021	0.008	0.017	0.100	0.012	0.020	0.010
G	0.196	0.074	0.030	0.062	0.043	0.003	0.005	0.006	0.019	0.036	0.026	0.021
H	0.236	0.039	0.023	0.016	0.041	0.016	0.029	0.045	0.035	0.014	0.024	0.016
Attribute Score	0.3080	0.0430	0.0153	0.0096	0.0375	0.0245	0.0012	0.0010	0.0000	0.0124	0.0207	0.0264
Gap	0.0920	0.0570	0.0147	0.0704	0.0125	0.0255	0.0388	0.0490	0.1000	0.0276	0.0093	0.0036

Figure 125. Scores for Eight Notional Airmen (Airman E Highlighted)

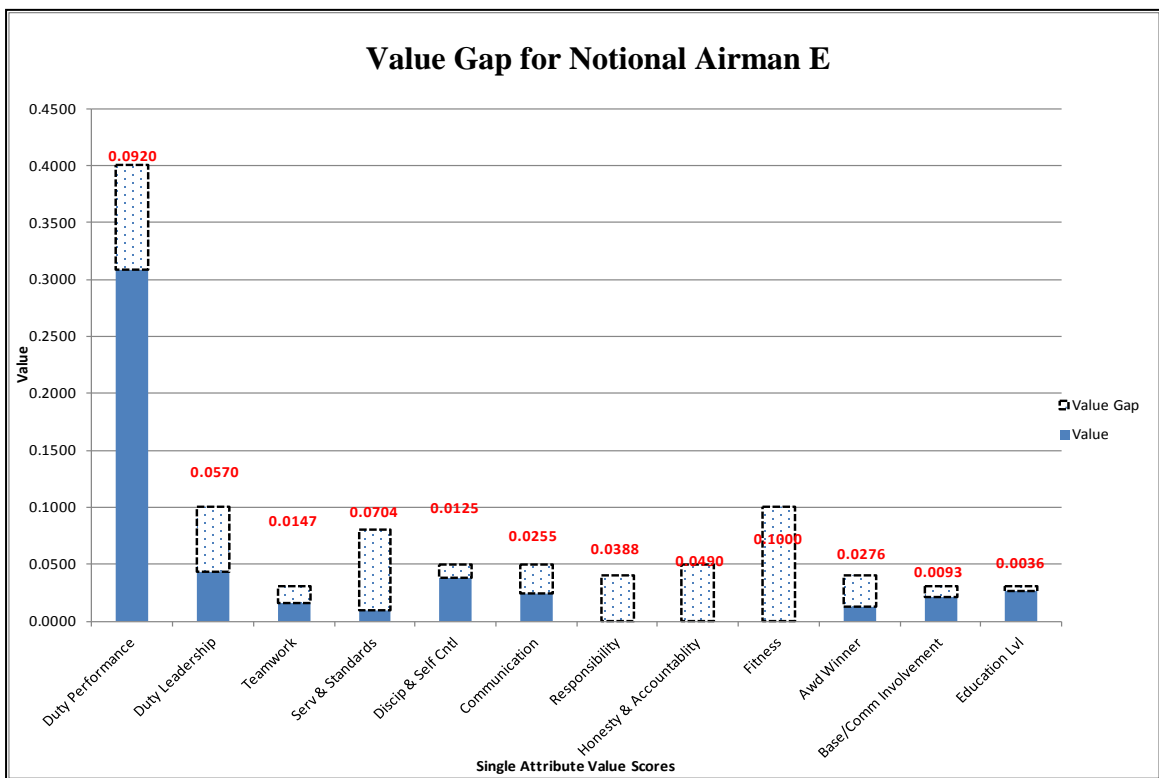


Figure 126. Value Gap Feedback for Notional Airman E

JEPR Value Gap Feedback Scores Value Gap Strengths and Shortfalls of Notional Airman F

Value Gap for Hypothetical Airmen F												
Hypothetical Airmen	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl
A	0.040	0.056	0.028	0.071	0.019	0.010	0.033	0.018	0.009	0.008	0.015	0.014
B	0.088	0.004	0.004	0.074	0.005	0.017	0.004	0.047	0.023	0.036	0.025	0.006
C	0.288	0.035	0.008	0.002	0.021	0.050	0.000	0.050	0.016	0.019	0.024	0.009
D	0.072	0.096	0.030	0.058	0.020	0.026	0.026	0.039	0.100	0.031	0.022	0.012
E	0.092	0.057	0.015	0.070	0.013	0.026	0.039	0.049	0.100	0.028	0.009	0.004
F	0.016	0.081	0.026	0.077	0.028	0.021	0.008	0.017	0.100	0.012	0.020	0.010
G	0.196	0.074	0.030	0.062	0.043	0.003	0.005	0.006	0.019	0.036	0.026	0.021
H	0.236	0.039	0.023	0.016	0.041	0.016	0.029	0.045	0.035	0.014	0.024	0.016
Attribute Score	0.3840	0.0190	0.0036	0.0032	0.0225	0.0290	0.0324	0.0330	0.0000	0.0280	0.0105	0.0198
Gap	0.0160	0.0810	0.0264	0.0768	0.0275	0.0210	0.0076	0.0170	0.1000	0.0120	0.0195	0.0102

Figure 127. Scores for Eight Notional Airmen (Airman F Highlighted)

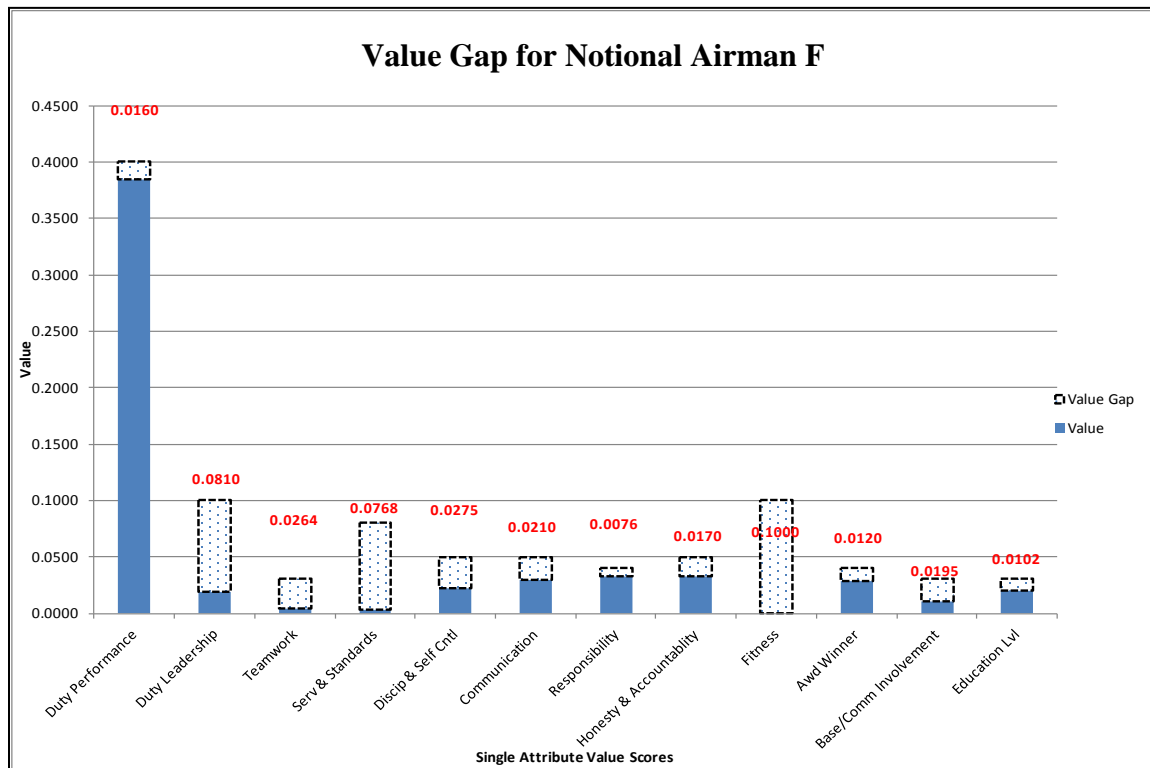


Figure 128. Value Gap Feedback for Notional Airman F

JEPR Value Gap Feedback Scores Value Gap Strengths and Shortfalls of Notional Airman G

Value Gap for Hypothetical Airmen G												
Hypothetical Airmen	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl
A	0.040	0.056	0.028	0.071	0.019	0.010	0.033	0.018	0.009	0.008	0.015	0.014
B	0.088	0.004	0.004	0.074	0.005	0.017	0.004	0.047	0.023	0.036	0.025	0.006
C	0.288	0.035	0.008	0.002	0.021	0.050	0.000	0.050	0.016	0.019	0.024	0.009
D	0.072	0.096	0.030	0.058	0.020	0.026	0.026	0.039	0.100	0.031	0.022	0.012
E	0.092	0.057	0.015	0.070	0.013	0.026	0.039	0.049	0.100	0.028	0.009	0.004
F	0.016	0.081	0.026	0.077	0.028	0.021	0.008	0.017	0.100	0.012	0.020	0.010
G	0.196	0.074	0.030	0.062	0.043	0.003	0.005	0.006	0.019	0.036	0.026	0.021
H	0.236	0.039	0.023	0.016	0.041	0.016	0.029	0.045	0.035	0.014	0.024	0.016
Attribute Score	0.2040	0.0260	0.0000	0.0184	0.0075	0.0470	0.0348	0.0440	0.0810	0.0036	0.0045	0.0093
Gap	0.1960	0.0740	0.0300	0.0616	0.0425	0.0030	0.0052	0.0060	0.0190	0.0364	0.0255	0.0207

Figure 129. Scores for Eight Notional Airmen (Airman G Highlighted)

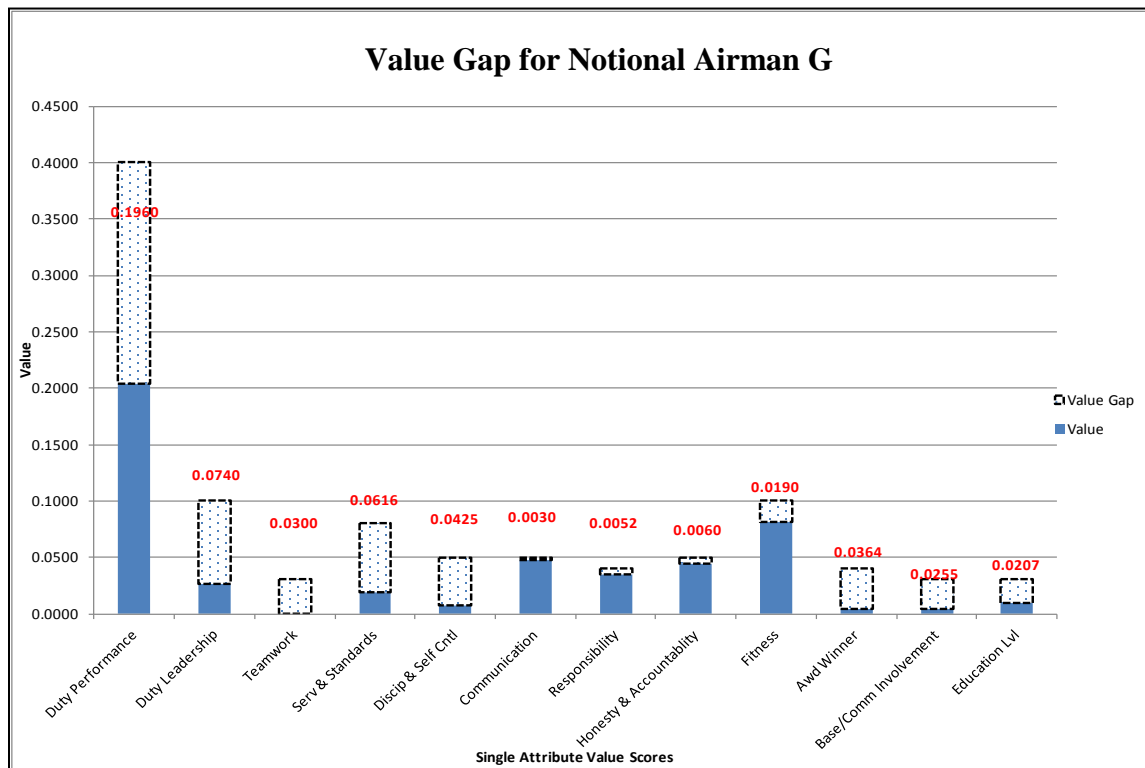


Figure 130. Value Gap Feedback for Notional Airman G

JEPR Value Gap Feedback Scores Value Gap Strengths and Shortfalls of Notional Airman H

Value Gap for Hypothetical Airmen H												
Hypothetical Airmen	Duty Performance	Duty Leadership	Teamwork	Serv & Standards	Discip & Self Cntl	Communication	Responsibility	Honesty & Accountability	Fitness	Awd Winner	Base/Comm Involvement	Education Lvl
A	0.040	0.056	0.028	0.071	0.019	0.010	0.033	0.018	0.009	0.008	0.015	0.014
B	0.088	0.004	0.004	0.074	0.005	0.017	0.004	0.047	0.023	0.036	0.025	0.006
C	0.288	0.035	0.008	0.002	0.021	0.050	0.000	0.050	0.016	0.019	0.024	0.009
D	0.072	0.096	0.030	0.058	0.020	0.026	0.026	0.039	0.100	0.031	0.022	0.012
E	0.092	0.057	0.015	0.070	0.013	0.026	0.039	0.049	0.100	0.028	0.009	0.004
F	0.016	0.081	0.026	0.077	0.028	0.021	0.008	0.017	0.100	0.012	0.020	0.010
G	0.196	0.074	0.030	0.062	0.043	0.003	0.005	0.006	0.019	0.036	0.026	0.021
H	0.236	0.039	0.023	0.016	0.041	0.016	0.029	0.045	0.035	0.014	0.024	0.016
Attribute Score	0.1640	0.0610	0.0075	0.0640	0.0090	0.0345	0.0112	0.0055	0.0650	0.0260	0.0057	0.0144
Gap	0.2360	0.0390	0.0225	0.0160	0.0410	0.0155	0.0288	0.0445	0.0350	0.0140	0.0243	0.0156

Figure 131. Scores for Eight Notional Airmen (Airman H Highlighted)

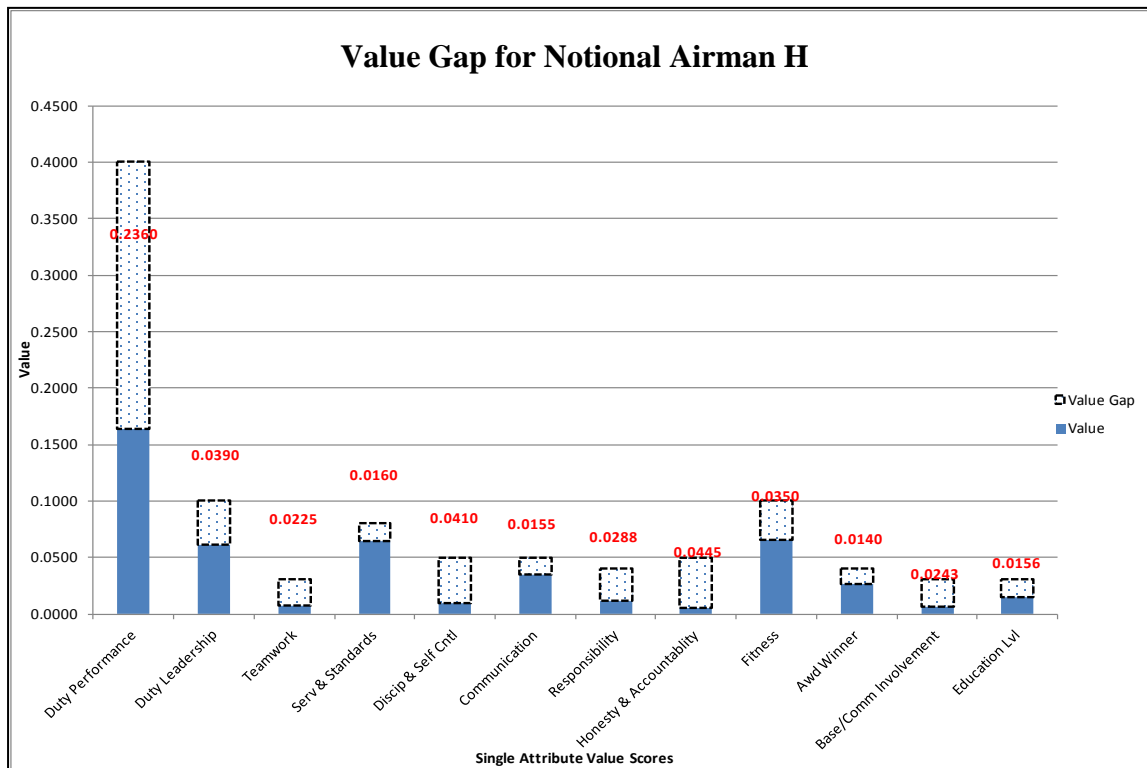


Figure 132. Value Gap Feedback for Notional Airman H

Appendix VI

Approved Exemption Request from Human Experimentation Requirements (32 CFR 219, DoDD 3216.2 and AFI 40-402)



DEPARTMENT OF THE AIR FORCE
AIR FORCE INSTITUTE OF TECHNOLOGY
WRIGHT-PATTERSON AIR FORCE BASE OHIO

1
18 Oct 2013

MEMORANDUM FOR Maj Jennifer Geffre

FROM: John Elshaw, Ph.D.
AFIT IRB Exempt Determination Official
2950 Hobson Way
Wright-Patterson AFB, OH 45433-7765

SUBJECT: Approval for exemption request from human experimentation requirements (32 CFR 219, DoDD 3216.2 and AFI 40-402) for "Research on Hybrid Workspace Implementation."

1. Your request was based on the Code of Federal Regulations, title 32, part 219, section 101, paragraph (b) (2) Research activities that involve the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior unless: (i) Information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) Any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.
2. Your study qualifies for this exemption because you are not collecting sensitive data, which could reasonably damage the subjects' financial standing, employability, or reputation. Further, the demographic data you are collecting cannot realistically be expected to map a given response to a specific subject.
3. This determination pertains only to the Federal, Department of Defense, and Air Force regulations that govern the use of human subjects in research. Further, if a subject's future response reasonably places them at risk of criminal or civil liability or is damaging to their financial standing, employability, or reputation, you are required to file an adverse event report with this office immediately.

X 

JOHN J. ELSHAW, Ph.D.
AFIT Exempt Determination Official

Appendix VII

Confirmatory Factor Analysis Model Outputs

JEPR Test Dataset CFA Model (Baseline)

Model Fit Summary
CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Baseline Model	25	122.615	53	.000	2.313
Saturated model	78	.000	0		
Independence model	12	1071.576	66	.000	16.236

RMR, RMR, GFI

Model	SRMR	RMR	GFI	AGFI	PGFI
Baseline Model	.0474	.000	.882	.826	.599
Saturated model		.000	1.000		
Independence model		.000	.312	.186	.264

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Baseline Model	.886	.858	.932	.914	.931
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Baseline Model	.803	.711	.747
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Baseline Model	69.615	41.151	105.797
Saturated model	.000	.000	.000
Independence model	1005.576	903.227	1115.340

FMIN

Model	FMIN	F0	LO 90	HI 90
Baseline Model	.776	.441	.260	.670
Saturated model	.000	.000	.000	.000
Independence model	6.782	6.364	5.717	7.059

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Baseline Model	.091	.070	.112	.001
Independence model	.311	.294	.327	.000

AIC

Model	AIC	BCC	BIC	CAIC
Baseline Model	172.615	177.098	249.338	274.338
Saturated model	156.000	169.986	395.375	473.375
Independence model	1095.576	1097.728	1132.403	1144.403

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Baseline Model	1.093	.912	1.321	1.121
Saturated model	.987	.987	.987	1.076
Independence model	6.934	6.286	7.629	6.948

HOELTER

Model	HOELTER	HOELTER
	.05	.01
Baseline Model	92	103
Independence model	13	15

Bollen-Stine Bootstrap (Baseline Model)

The model fit better in 98 bootstrap samples.
It fit about equally well in 0 bootstrap samples.
It fit worse or failed to fit in 2 bootstrap samples.
Testing the null hypothesis that the model is correct, Bollen-Stine bootstrap $p = .030$

Bootstrap Distributions (Default model)
ML discrepancy (implied vs. sample) (Baseline Model)

	33.334	*
	42.233	*****
	51.132	*****
	60.030	*****
	68.929	*****
	77.827	*****
	86.726	*****
N = 100	95.625	*****
Mean = 75.594	104.523	*****
S. e. = 2.197	113.422	*
	122.320	**
	131.219	
	140.118	
	149.016	*
	157.915	*

Scalar Estimates (JEPR Test Dataset - Baseline Model)
Maximum Likelihood Estimates
Regression Weights: (JEPR Test Dataset - Baseline Model)

			Estimate	S.E.	C.R.	P	Label
Duty Leadership	<---	Standards	.299	.022	13.758	***	
Communication	<---	Standards	.146	.012	12.580	***	
Respect for Service and Standards	<---	Standards	.226	.024	9.346	***	
Discipline and Self-Control	<---	Standards	.135	.015	9.057	***	
Honesty and Accountability	<---	Standards	.129	.018	7.316	***	
Responsibility	<---	Standards	.117	.011	11.124	***	
Physical Fitness	<---	Professional Expectations	1.000				
Military Awards	<---	Professional Expectations	1.527	.358	4.261	***	
Education Level	<---	Professional Expectations	.907	.216	4.192	***	
Base and Community Involvement	<---	Professional Expectations	.847	.206	4.112	***	
Teamwork and	<---	Standards	.088	.007	11.805	***	

			Estimate	S.E.	C.R.	P	Label
Followership							
Duty Performance	<---	Standards	1.000				

Standardized Regression Weights: (JEPR Test Dataset - Baseline Model)

			Estimate
Duty Leadership	<---	Standards	.888
Communication	<---	Standards	.837
Respect for Service and Standards	<---	Standards	.676
Discipline and Self Control	<---	Standards	.660
Honesty and Accountability	<---	Standards	.554
Responsibility	<---	Standards	.770
Physical Fitness	<---	Professional Expectations	.367
Military Awards	<---	Professional Expectations	.830
Education Level	<---	Professional Expectations	.740
Base and Community Involvement	<---	Professional Expectations	.682
Teamwork and Followership	<---	Standards	.802
Duty Performance	<---	Standards	.823

Covariances: (JEPR Test Dataset - Baseline Model)

			Estimate	S.E.	C.R.	P	Label
Standards	<-->	Professional Expectations	.000	.000	3.451	***	

Correlations: (JEPR Test Dataset - Baseline Model)

			Estimate
Standards	<-->	Professional Expectations	.560

Variances: (JEPR Test Dataset - Baseline Model)

			Estimate	S.E.	C.R.	P	Label
Standards			.004	.001	6.220	***	
Professional Expectations			.000	.000	2.178	.029	
e1			.002	.000	7.514	***	
e2			.000	.000	6.446	***	
e3			.000	.000	7.352	***	
e4			.000	.000	8.343	***	
e5			.000	.000	8.389	***	
e6			.000	.000	8.602	***	
e7			.000	.000	7.943	***	
e8			.000	.000	7.711	***	
e12			.000	.000	8.600	***	
e9			.000	.000	4.580	***	
e10			.000	.000	6.411	***	

			Estimate	S.E.	C.R.	P	Label
e11			.000	.000	7.171	***	

Squared Multiple Correlations: (JEPR Test Dataset - Baseline Model)

			Estimate
Base and Community Involvement			.466
Education Level			.548
Military Awards			.688
Physical Fitness			.135
Teamwork and Followership			.643
Responsibility			.592
Honesty and Accountability			.307
Discipline and Self Control			.436
Respect for Service and Standards			.457
Communication			.701
Duty Leadership			.788
Duty Performance			.678

Modification Indices (JEPR Test Dataset - Baseline Model)

Covariances: (JEPR Test Dataset - Baseline Model)

			M.I.	Par Change
e12	<-->	e9	5.142	.000
e7	<-->	e8	6.325	.000
e5	<-->	e6	6.137	.000
e4	<-->	e5	6.015	.000
e3	<-->	e8	7.482	.000
e3	<-->	e7	8.986	.000
e3	<-->	e6	5.193	.000
e2	<-->	e6	11.288	.000
e2	<-->	e3	6.262	.000
e1	<-->	e11	8.003	.000
e1	<-->	e9	4.765	.000
e1	<-->	e6	6.788	.000
e1	<-->	e3	4.331	.000
e1	<-->	e2	29.443	.000

Modification Indices (JEPR Test Dataset - Baseline Model)
 Variances: (JEPR Test Dataset - Baseline Model)
 Regression Weights: (JEPR Test Dataset - Baseline Model)

			M.I.	Par Change
Military Awards	<---	Physical Fitness	4.402	.080
Discipline and Self Control	<---	Honesty and Accountability	4.130	.110
Duty Leadership	<---	Honesty and Accountability	7.662	-.170
Duty Leadership	<---	Duty Performance	8.599	.034
Duty Performance	<---	Base and Community Involvement	4.580	-.869
Duty Performance	<---	Honesty and Accountability	4.584	-.547
Duty Performance	<---	Duty Leadership	4.926	.390

JEPR Test Dataset CFA Model (Modified Model #1)

Model Fit Summary CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Modified Model #1	26	86.068	52	.002	1.655
Saturated model	78	.000	0		
Independence model	12	1071.576	66	.000	16.236

SRMR, RMR, GFI

Model	SRMR	RMR	GFI	AGFI	PGFI
Baseline Model	.0432	.000	.920	.880	.613
Saturated model		.000	1.000		
Independence model		.000	.312	.186	.264

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Modified Model #1	.920	.898	.967	.957	.966
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Modified Model #1	.788	.725	.761
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Modified Model #1	34.068	12.437	63.591
Saturated model	.000	.000	.000
Independence model	1005.576	903.227	1115.340

FMIN

Model	FMIN	F0	LO 90	HI 90
Modified Model #1	.545	.216	.079	.402
Saturated model	.000	.000	.000	.000
Independence model	6.782	6.364	5.717	7.059

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Modified Model #1	.064	.039	.088	.159
Independence model	.311	.294	.327	.000

AIC

Model	AIC	BCC	BIC	CAIC
Modified Model #1	138.068	142.730	217.860	243.860
Saturated model	156.000	169.986	395.375	473.375
Independence model	1095.576	1097.728	1132.403	1144.403

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Modified Model #1	.874	.737	1.061	.903
Saturated model	.987	.987	.987	1.076
Independence model	6.934	6.286	7.629	6.948

HOELTER

Model	HOELTER .05	HOELTER .01
Modified Model #1	129	145
Independence model	13	15

Bollen-Stine Bootstrap (Modified Model #1)

The model fit better in 74 bootstrap samples.
It fit about equally well in 0 bootstrap samples.
It fit worse or failed to fit in 26 bootstrap samples.
Testing the null hypothesis that the model is correct, Bollen-Stine bootstrap $p = .267$

Bootstrap Distributions (Modified Model #1)
ML discrepancy (implied vs. sample) (Modified Model #1)

	33.201	*
	40.636	*****
	48.072	*****
	55.507	*****
	62.942	*****
	70.377	*****
	77.812	*****
N = 100	85.248	*****
Mean = 74.017	92.683	*****
S. e. = 2.078	100.118	****
	107.553	****
	114.988	*
	122.423	**
	129.859	*
	137.294	*

Scalar Estimates (JEPR Test Dataset - Modified Model #1)
Maximum Likelihood Estimates
Regression Weights: (JEPR Test Dataset - Modified Model #1)

			Estimate	S.E.	C.R.	P	Label
Duty Leadership	<---	Standards	.307	.019	15.779	***	
Communication	<---	Standards	.164	.015	11.292	***	
Respect for Service and Standards	<---	Standards	.249	.029	8.740	***	
Discipline and Self Control	<---	Standards	.148	.018	8.448	***	
Honesty and Accountability	<---	Standards	.149	.020	7.388	***	
Responsibility	<---	Standards	.129	.013	10.005	***	
Physical Fitness	<---	Professional Expectations	1.000				
Military Awards	<---	Professional Expectations	1.531	.364	4.211	***	
Education Level	<---	Professional Expectations	.919	.221	4.151	***	
Base and Community Involvement	<---	Professional Expectations	.861	.211	4.077	***	
Teamwork and	<---	Standards	.098	.009	10.553	***	

			Estimate	S.E.	C.R.	P	Label
Followership							
Duty Performance	<---	Standards	1.000				

Standardized Regression Weights: (JEPR Test Dataset - Modified Model #1)

			Estimate
Duty Leadership	<---	Standards	.839
Communication	<---	Standards	.865
Respect for Service and Standards	<---	Standards	.688
Discipline and Self Control	<---	Standards	.667
Honesty and Accountability	<---	Standards	.590
Responsibility	<---	Standards	.776
Physical Fitness	<---	Professional Expectations	.364
Military Awards	<---	Professional Expectations	.824
Education Level	<---	Professional Expectations	.743
Base and Community Involvement	<---	Professional Expectations	.687
Teamwork and Followership	<---	Standards	.814
Duty Performance	<---	Standards	.757

Covariances: (JEPR Test Dataset - Modified Model #1)

			Estimate	S.E.	C.R.	P	Label
Standards	<-->	Professional Expectations	.000	.000	3.373	***	
e1	<-->	e2	.000	.000	4.704	***	

Correlations: (JEPR Test Dataset - Modified Model #1)

			Estimate
Standards	<-->	Professional Expectations	.555
e1	<-->	e2	.530

Variances: (JEPR Test Dataset - Modified Model #1)

			Estimate	S.E.	C.R.	P	Label
Standards			.003	.001	5.443	***	
Professional Expectations			.000	.000	2.154	.031	
e1			.002	.000	7.714	***	
e2			.000	.000	6.991	***	
e3			.000	.000	6.547	***	
e4			.000	.000	8.190	***	
e5			.000	.000	8.266	***	
e6			.000	.000	8.476	***	
e7			.000	.000	7.702	***	
e8			.000	.000	7.347	***	
e12			.000	.000	8.602	***	

			Estimate	S.E.	C.R.	P	Label
e9			.000	.000	4.700	***	
e10			.000	.000	6.342	***	
e11			.000	.000	7.103	***	

Squared Multiple Correlations: (JEPR Test Dataset - Modified Model #1)

			Estimate
Base and Community Involvement			.472
Education Level			.552
Military Awards			.679
Physical Fitness			.132
Teamwork and Followership			.662
Responsibility			.603
Honesty and Accountability			.348
Discipline and Self Control			.445
Respect for Service and Standards			.473
Communication			.748
Duty Leadership			.703
Duty Performance			.574

Modification Indices (JEPR Test Dataset – Modified Model #1)

Covariances: (JEPR Test Dataset - Modified Model #1)

			M.I.	Par Change
e12	<-->	e9	5.559	.000
e7	<-->	e8	10.676	.000
e5	<-->	e6	4.001	.000
e4	<-->	e5	4.985	.000
e3	<-->	e7	4.815	.000
e3	<-->	e5	4.041	.000
e1	<-->	e11	7.809	.000
e1	<-->	e9	4.730	.000

Modification Indices (JEPR Test Dataset – Modified Model #1)

Variances: (JEPR Test Dataset - Modified Model #1)

Regression Weights: (JEPR Test Dataset - Modified Model #1)

			M.I.	Par Change
Military Awards	<---	Physical Fitness	4.770	.084
Duty Performance	<---	Base and Community Involvement	4.627	-.794

JEPR Test Dataset CFA Model (Modified Model #2)

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Modified model #2	27	71.580	51	.030	1.404
Saturated model	78	.000	0		
Independence model	12	1071.576	66	.000	16.236

SRMR, RMR, GFI

Model	SRMR	RMR	GFI	AGFI	PGFI
Baseline Model	.0431	.000	.931	.895	.609
Saturated model		.000	1.000		
Independence model		.000	.312	.186	.264

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Modified model #2	.933	.914	.980	.974	.980
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Modified model #2	.773	.721	.757
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Modified model #2	20.580	2.175	47.007
Saturated model	.000	.000	.000
Independence model	1005.576	903.227	1115.340

FMIN

Model	FMIN	F0	LO 90	HI 90
Modified model #2	.453	.130	.014	.298
Saturated model	.000	.000	.000	.000
Independence model	6.782	6.364	5.717	7.059

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Modified model #2	.051	.016	.076	.463
Independence model	.311	.294	.327	.000

AIC

Model	AIC	BCC	BIC	CAIC
Modified model #2	125.580	130.422	208.441	235.441
Saturated model	156.000	169.986	395.375	473.375
Independence model	1095.576	1097.728	1132.403	1144.403

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Modified model #2	.795	.678	.962	.825
Saturated model	.987	.987	.987	1.076
Independence model	6.934	6.286	7.629	6.948

HOELTER

Model	HOELTER .05	HOELTER .01
Modified model #2	152	171
Independence model	13	15

Bollen-Stine Bootstrap (Modified model #2)

The model fit better in 52 bootstrap samples.
It fit about equally well in 0 bootstrap samples.
It fit worse or failed to fit in 48 bootstrap samples.
Testing the null hypothesis that the model is correct, Bollen-Stine bootstrap $p = .485$

Bootstrap Distributions (Modified model #2)
ML discrepancy (implied vs. sample) (Modified model #2)

	33.547	*
	40.500	*****
	47.452	*****
	54.405	*****
	61.357	*****
	68.310	*****
	75.262	*****
N = 100	82.215	*****
Mean = 72.087	89.167	*****
S. e. = 1.992	96.120	*****
	103.072	****
	110.025	
	116.977	***
	123.930	*
	130.882	*

Scalar Estimates (JEPR Test Dataset - Modified model #2)
Maximum Likelihood Estimates

Regression Weights: (JEPR Test Dataset - Modified model #2)

			Estimate	S.E.	C.R.	P	Label
Duty Leadership	<---	Standards	.308	.019	15.885	***	
Communication	<---	Standards	.166	.015	11.406	***	
Respect for Service and Standards	<---	Standards	.246	.029	8.591	***	
Discipline and Self Control	<---	Standards	.147	.018	8.358	***	
Honesty and Accountability	<---	Standards	.148	.020	7.335	***	
Responsibility	<---	Standards	.135	.013	10.343	***	
Physical Fitness	<---	Professional Expectations	1.000				
Military Awards	<---	Professional Expectations	1.537	.365	4.207	***	
Education Level	<---	Professional Expectations	.917	.221	4.144	***	
Base and Community Involvement	<---	Professional Expectations	.862	.212	4.073	***	
Teamwork and	<---	Standards	.102	.009	10.835	***	

			Estimate	S.E.	C.R.	P	Label
Followership							
Duty Performance	<---	Standards	1.000				

Standardized Regression Weights: (JEPR Test Dataset - Modified model #2)

			Estimate
Duty Leadership	<---	Standards	.835
Communication	<---	Standards	.865
Respect for Service and Standards	<---	Standards	.673
Discipline and Self Control	<---	Standards	.656
Honesty and Accountability	<---	Standards	.582
Responsibility	<---	Standards	.808
Physical Fitness	<---	Professional Expectations	.363
Military Awards	<---	Professional Expectations	.826
Education Level	<---	Professional Expectations	.741
Base and Community Involvement	<---	Professional Expectations	.687
Teamwork and Followership	<---	Standards	.840
Duty Performance	<---	Standards	.751

Covariances: (JEPR Test Dataset - Modified model #2)

			Estimate	S.E.	C.R.	P	Label
Standards	<-->	Professional Expectations	.000	.000	3.355	***	
e1	<-->	e2	.000	.000	5.056	***	
e7	<-->	e8	.000	.000	-3.919	***	

Correlations: (JEPR Test Dataset - Modified model #2)

			Estimate
Standards	<-->	Professional Expectations	.544
e1	<-->	e2	.541
e7	<-->	e8	-.406

Variiances: (JEPR Test Dataset - Modified model #2)

			Estimate	S.E.	C.R.	P	Label
Standards			.003	.001	5.427	***	
Professional Expectations			.000	.000	2.152	.031	
e1			.002	.000	8.016	***	
e2			.000	.000	7.408	***	
e3			.000	.000	6.939	***	
e4			.000	.000	8.387	***	
e5			.000	.000	8.431	***	
e6			.000	.000	8.581	***	
e7			.000	.000	7.116	***	
e8			.000	.000	6.727	***	
e12			.000	.000	8.603	***	
e9			.000	.000	4.638	***	
e10			.000	.000	6.371	***	
e11			.000	.000	7.098	***	

Squared Multiple Correlations: (JEPR Test Dataset - Modified model #2)

			Estimate
Base and Community Involvement			.473
Education Level			.549
Military Awards			.682
Physical Fitness			.132
Teamwork and Followership			.706
Responsibility			.652
Honesty and Accountability			.338
Discipline and Self Control			.431
Respect for Service and Standards			.453
Communication			.749
Duty Leadership			.697
Duty Performance			.564

Modification Indices (JEPR Test Dataset - Modified model #2)

Covariances: (JEPR Test Dataset - Modified model #2)

			M.I.	Par Change
e12	<-->	e9	5.603	.000
e5	<-->	e6	4.767	.000
e4	<-->	e7	4.741	.000
e4	<-->	e5	6.421	.000
e1	<-->	e11	7.717	.000
e1	<-->	e9	4.759	.000

Modification Indices (JEPR Test Dataset - Modified model #2)

Variances: (JEPR Test Dataset - Modified model #2)

Regression Weights: (JEPR Test Dataset - Modified model #2)

			M.I.	Par Change
Military Awards	<---	Physical Fitness	4.811	.084
Duty Performance	<---	Base and Community Involvement	4.386	-.773

JEPR Test Dataset CFA Model (Modified Model #3)

Model Fit Summary

CMIN

Model	NPAR	CMIN	DF	P	CMIN/DF
Modified Model #3	28	64.935	50	.076	1.299
Saturated model	78	.000	0		
Independence model	12	1071.576	66	.000	16.236

SRMR, RMR, GFI

Model	SRMR	RMR	GFI	AGFI	PGFI
Baseline Model	.0420	.000	.939	.905	.602
Saturated model		.000	1.000		
Independence model		.000	.312	.186	.264

Baseline Comparisons

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Modified Model #3	.939	.920	.985	.980	.985
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

Parsimony-Adjusted Measures

Model	PRATIO	PNFI	PCFI
Modified Model #3	.758	.712	.746
Saturated model	.000	.000	.000
Independence model	1.000	.000	.000

NCP

Model	NCP	LO 90	HI 90
Modified Model #3	14.935	.000	39.860
Saturated model	.000	.000	.000
Independence model	1005.576	903.227	1115.340

FMIN

Model	FMIN	F0	LO 90	HI 90
Modified Model #3	.411	.095	.000	.252
Saturated model	.000	.000	.000	.000
Independence model	6.782	6.364	5.717	7.059

RMSEA

Model	RMSEA	LO 90	HI 90	PCLOSE
Modified Model #3	.043	.000	.071	.620
Independence model	.311	.294	.327	.000

AIC

Model	AIC	BCC	BIC	CAIC
Modified Model #3	120.935	125.955	206.864	234.864
Saturated model	156.000	169.986	395.375	473.375
Independence model	1095.576	1097.728	1132.403	1144.403

ECVI

Model	ECVI	LO 90	HI 90	MECVI
Modified Model #3	.765	.671	.923	.797
Saturated model	.987	.987	.987	1.076
Independence model	6.934	6.286	7.629	6.948

HOELTER

Model	HOELTER	HOELTER
	.05	.01
Modified Model #3	165	186
Independence model	13	15

Bollen-Stine Bootstrap (Modified Model #3)

The model fit better in 40 bootstrap samples.
It fit about equally well in 0 bootstrap samples.
It fit worse or failed to fit in 60 bootstrap samples.
Testing the null hypothesis that the model is correct, Bollen-Stine bootstrap $p = .604$

Bootstrap Distributions (Modified Model #3)
ML discrepancy (implied vs. sample) (Modified Model #3)

	33.559	**
	40.014	*****
	46.469	*****
	52.924	*****
	59.378	*****
	65.833	*****
	72.288	*****
N = 100	78.743	*****
Mean = 70.313	85.197	*****
S. e. = 1.942	91.652	****
	98.107	****
	104.561	***
	111.016	
	117.471	**
	123.926	**

Scalar Estimates (JEPR Test Dataset - Modified Model #3)
Maximum Likelihood Estimates

Regression Weights: (JEPR Test Dataset - Modified Model #3)

			Estimate	S.E.	C.R.	P	Label
Duty Leadership	<---	Standards	.309	.019	15.864	***	
Communication	<---	Standards	.166	.015	11.401	***	
Respect for Service and Standards	<---	Standards	.242	.029	8.410	***	
Discipline and Self Control	<---	Standards	.145	.018	8.177	***	
Honesty and Accountability	<---	Standards	.147	.020	7.281	***	
Responsibility	<---	Standards	.136	.013	10.373	***	
Physical Fitness	<---	Professional Expectations	1.000				
Military Awards	<---	Professional Expectations	1.538	.366	4.208	***	
Education Level	<---	Professional Expectations	.916	.221	4.144	***	
Base and Community Involvement	<---	Professional Expectations	.862	.212	4.074	***	
Teamwork and	<---	Standards	.102	.009	10.810	***	

			Estimate	S.E.	C.R.	P	Label
Followership							
Duty Performance	<---	Standards	1.000				

Standardized Regression Weights: (JEPR Test Dataset - Modified Model #3)

			Estimate
Duty Leadership	<---	Standards	.833
Communication	<---	Standards	.867
Respect for Service and Standards	<---	Standards	.661
Discipline and Self Control	<---	Standards	.644
Honesty and Accountability	<---	Standards	.578
Responsibility	<---	Standards	.813
Physical Fitness	<---	Professional Expectations	.363
Military Awards	<---	Professional Expectations	.827
Education Level	<---	Professional Expectations	.740
Base and Community Involvement	<---	Professional Expectations	.687
Teamwork and Followership	<---	Standards	.842
Duty Performance	<---	Standards	.749

Covariances: (JEPR Test Dataset - Modified Model #3)

			Estimate	S.E.	C.R.	P	Label
Standards	<-->	Professional Expectations	.000	.000	3.352	***	
e1	<-->	e2	.000	.000	5.086	***	
e7	<-->	e8	.000	.000	-4.109	***	
e4	<-->	e5	.000	.000	2.466	.014	

Correlations: (JEPR Test Dataset - Modified Model #3)

			Estimate
Standards	<-->	Professional Expectations	.542
e1	<-->	e2	.544
e7	<-->	e8	-.435
e4	<-->	e5	.214

Variiances: (JEPR Test Dataset - Modified Model #3)

			Estimate	S.E.	C.R.	P	Label
Standards			.003	.001	5.411	***	
Professional Expectations			.000	.000	2.152	.031	
e1			.002	.000	8.030	***	
e2			.000	.000	7.419	***	
e3			.000	.000	6.888	***	
e4			.000	.000	8.407	***	
e5			.000	.000	8.447	***	
e6			.000	.000	8.590	***	
e7			.000	.000	6.977	***	
e8			.000	.000	6.605	***	
e12			.000	.000	8.603	***	
e9			.000	.000	4.610	***	
e10			.000	.000	6.385	***	
e11			.000	.000	7.101	***	

Squared Multiple Correlations: (JEPR Test Dataset - Modified Model #3)

		Estimate
Base and Community Involvement		.472
Education Level		.548
Military Awards		.684
Physical Fitness		.132
Teamwork and Followership		.710
Responsibility		.662
Honesty and Accountability		.334
Discipline and Self Control		.415
Respect for Service and Standards		.437
Communication		.751
Duty Leadership		.695
Duty Performance		.562

Modification Indices (JEPR Test Dataset - Modified Model #3)

Covariances: (JEPR Test Dataset - Modified Model #3)

	M.I.	Par Change
e12 <--> e9	5.592	.000
e5 <--> e6	4.172	.000
e1 <--> e11	7.689	.000
e1 <--> e9	4.727	.000

Variances: (JEPR Test Dataset - Modified Model #3)
 Regression Weights: (JEPR Test Dataset - Modified Model #3)

		M.I.	Par Change
Military Awards	<--- Physical Fitness	4.802	.084
Duty Performance	<--- Base and Community Involvement	4.339	-.769

Appendix VIII

JEPR Test Dataset Artificial Neural Network (ANN) MATLAB Code

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
%
%                               ANN JEPR Test Dataset
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%

%%
% ***** Import Full Model Data
% *****

%% clear all variables
clc
clear all;

% Import JEPR Test Dataset data from spreadsheet (col B has random
uniform
% noise, col C to col O have JEPR attribute data, and col P has
normalized JEPR
% Standards violation discrepancy count).
[~, ~, raw] =
xlsread('I:\setup\Desktop\THESIS\MODEL_VERIFICATION\COMBINED\MASTER_JEP
R_SEQUENCE_SCORING.xlsx', 'ANN', 'B2:P160');

% Create output variable
THEISIS_ANN_IN = reshape([raw{:}], size(raw));

% Clear temporary variables
clearvars raw;

% Extract input matrix size
[m,n]=size(THEISIS_ANN_IN);

%%
% ***** Import Output Data matrix of known JEPR
results for JEPR Test Dataset to test ANN classification success based
on JEPR attributes *****
% Three categories - Below Standards, Meets Standards, Exceeds
Standards

% Import output data from spreadsheet
[~, ~, raw] =
xlsread('I:\setup\Desktop\THESIS\MODEL_VERIFICATION\COMBINED\MASTER_JEP
R_SEQUENCE_SCORING.xlsx', 'ANN', 'V2:X160');
```

```

% Create output variable
    JEPR_ANN_OUT = reshape([raw{:}],size(row));

% Clear temporary variables
    clearvars row;

%%
% ***** Import Output Data matrix of known EPR
results for JEPR Test Dataset to test ANN classification success based
on JEPR attributes *****
% Three categories - Below Standards, Meets Standards, Exceeds
Standards

% Import output data from spreadsheet
    [~, ~, raw] =
xlsread('I:\setup\Desktop\THESIS\MODEL_VERIFICATION\COMBINED\MASTER_JEP
R_SEQUENCE_SCORING.xlsx', 'ANN', 'Y2:AA160');

% Create output variable
    EPR_ANN_OUT = reshape([raw{:}],size(row));

% Clear temporary variables
    clearvars row;

%%
% ***** Implement MATLAB NNPR Tool
*****

% Call NPT tool from MATLAB
    nprtool

% Set breakpoint in code to pause before generating weights and
% signal to noise ratio values...verify well trained network

    dbstop in MODEL_VER_ANN at 68

% Generate weights
    Weights=results.net.IW{1}

% Create noise variable
    Noise=Weights(:,1)'*Weights(:,1)

% Generate SNR values of size n categories
    for i=1:n
        SNR(i)=10*log10((Weights(:,i)'*Weights(:,i))/Noise)
    end

```

Appendix IX

Artificial Neural Network (ANN) SNR Values and Feature Weights

SNR Values for ANN EPR Network (Retrained 8 Times)	
Input Feature	SNR Values
Noise	0.0000
Duty Performance	5.5408
Duty Leadership	4.9633
Physical Fitness	5.8402
Communication	4.1441
Respect for Service and Standards	4.3015
Discipline and Self-Control	4.5476
Honesty and Accountability	4.6557
Responsibility	5.5941
Teamwork and Followership	5.9109
Military Awards	4.7712
Education Level	6.4413
Base and Community Involvement	3.6163
Administrative(Correction Factor)	4.3730
Referral Markings	7.9517

Feature Weights for Hidden Neurons In ANN EPR Network										
Input Feature	Hidden Neuron #1	Hidden Neuron #2	Hidden Neuron #3	Hidden Neuron #4	Hidden Neuron #5	Hidden Neuron #6	Hidden Neuron #7	Hidden Neuron #8	Hidden Neuron #9	Hidden Neuron #10
Noise	-0.0892	0.2859	0.3338	-0.3714	-0.0701	0.3578	0.0290	0.2699	0.0283	-0.2258
Duty Performance	0.4308	0.3917	0.5248	-0.6664	-0.5742	0.0755	0.4893	-0.2529	0.6438	0.1679
Duty Leadership	0.2091	0.7441	0.4355	-0.4128	-0.6566	-0.3920	0.3562	-0.3896	-0.2167	-0.0752
Physical Fitness	-0.0882	-0.5517	0.1456	-0.6839	-0.3627	0.5396	0.0161	-0.0235	0.7393	0.7222
Communication	0.5346	0.1870	-0.3896	-0.2508	0.1683	0.0594	-0.4570	0.6497	-0.0592	-0.5913
Respect for Service and Standards	-0.3989	-0.4686	0.3582	0.5286	0.0732	-0.5528	-0.4616	0.3554	-0.0140	0.4144
Discipline and Self-Control	-0.4114	0.2573	-0.3696	-0.1019	-0.2197	0.6114	0.4999	0.2240	0.6645	-0.3954
Honesty and Accountability	-0.6445	0.2998	0.0404	0.0562	-0.4043	0.4207	0.4128	-0.6030	0.1575	0.5798
Responsibility	-0.1690	0.5974	0.0362	-0.1178	-0.8522	-0.0114	0.5084	0.6166	-0.0688	-0.6292
Teamwork and Followership	-0.7729	-0.2070	0.0866	0.5245	0.1764	0.4614	0.2743	-0.4046	0.9124	-0.3037
Military Awards	0.0005	0.7015	0.1267	-0.1607	-0.2995	-0.7306	-0.4015	-0.3350	-0.6012	-0.0136
Education Level	0.4695	0.0246	-0.8089	-0.1374	-0.6947	-0.6914	-0.7829	-0.1048	0.1172	-0.3749
Base and Community Involvement	0.2124	0.5411	0.2781	0.0573	-0.2649	0.0699	-0.2484	0.0727	0.6480	-0.6273
Admin(Correction Factor)	0.0291	0.1402	0.2551	-0.4738	0.2477	-0.4440	-0.2318	0.6791	0.7429	-0.0117
Referral Markings	-0.5450	-0.5345	-1.1527	0.9148	0.3456	-0.4589	0.2573	-0.0527	-0.4361	-0.6250

SNR Values for ANN JEPR Network (Retrained 6 Times)	
Input Feature	SNR Values
Noise	0.0000
Duty Performance	6.3512
Duty Leadership	4.9126
Physical Fitness	5.0064
Communication	1.4245
Respect for Service and Standards	1.0381
Discipline and Self-Control	0.3357
Honesty and Accountability	0.4631
Responsibility	1.5250
Teamwork and Followership	1.6543
Military Awards	3.0043
Education Level	2.8334
Base and Community Involvement	4.4766
Administrative(Correction Factor)	4.1708
Referral Markings	3.7117

Feature Weights for Hidden Neurons In ANN JEPR Network										
Input Feature	Hidden Neuron #1	Hidden Neuron #2	Hidden Neuron #3	Hidden Neuron #4	Hidden Neuron #5	Hidden Neuron #6	Hidden Neuron #7	Hidden Neuron #8	Hidden Neuron #9	Hidden Neuron #10
Noise	0.622	-0.5765	-0.5727	0.0957	0.1238	0.3361	0.2129	-0.157	0.3116	0.4808
Duty Performance	0.0342	1.2814	-0.8655	0.1391	-0.7946	0.3032	-1.2916	-0.9568	0.4101	-0.9724
Duty Leadership	0.1571	1.4523	-0.8508	-0.7565	-0.9335	-0.714	-0.1729	-0.2279	-0.0535	0.0998
Physical Fitness	-0.4042	0.096	0.5787	-0.8005	-0.5041	-0.3528	-1.3339	-0.1803	0.8361	-0.9876
Communication	0.1222	0.7279	-0.7359	-0.4758	0.1413	-0.4174	0.2952	0.4281	0.5731	0.3028
Respect for Service and Standards	-0.6529	0.3794	-0.3508	-0.1854	-0.8646	0.417	-0.3699	0.175	-0.1947	-0.3949
Discipline and Self-Control	-0.3408	-0.0247	-0.6859	0.2841	-0.5568	0.4833	0.2644	-0.3959	-0.345	-0.3911
Honesty and Accountability	0.5604	0.497	-0.0162	0.5101	-0.1296	0.0061	-0.3276	-0.5182	0.5807	-0.4579
Responsibility	-0.4382	0.0276	-0.0995	0.4117	0.1306	0.0187	-0.8525	0.9266	-0.0971	0.5143
Teamwork and Followership	-0.6537	0.5093	0.1089	-0.3978	1.1031	0.0425	-0.4013	0.1398	0.1671	-0.1813
Military Awards	0.082	0.8693	-0.2923	-0.4651	-0.2664	0.0959	-1.192	-0.0031	0.7721	0.0174
Education Level	-0.0108	0.2289	-0.4219	0.4353	-0.7374	-1.1294	0.3087	0.0631	-0.8134	0.1982
Base and Community Involvement	0.1863	1.1237	0.1391	1.552	0.5857	0.1715	0.3062	-0.0948	-0.4511	0.182
Admin(Correction Factor)	0.4915	-0.597	0.0085	0.1259	0.1614	-0.3569	-1.0682	1.207	0.8537	-0.2043
Referral Markings	0.2756	0.8372	-1.1217	-0.5813	-0.9363	-0.0873	0.0045	0.5961	-0.2704	-0.1871

Bibliography

Air Force Directive 070906-003, Air Force Core Values Manual (1997). Retrieved April 7, 2013, from <http://www.e-publishing.af.mil/shared/media/document/AFD-070906-003.pdf>

Air Force Instruction 36-2406, Officer and Enlisted Evaluation Systems (2013). Retrieved September 16, 2013, from http://static.e-publishing.af.mil/production/1/af_a1/publication/afi36-2406/afi36-2406.pdf

Air Force Instruction 36-2618, The Enlisted Force Structure (2009). Retrieved April 14, 2013, from http://static.e-publishing.af.mil/production/1/af_a1/publication/afi36-2618/afi36-2618.pdf

Air Force Pamphlet 36-2241, Professional Development Guide (2013). Retrieved October 10, 2013, from http://static.e-publishing.af.mil/production/1/af_a1/publication/afpam36-2241/afpam36-2241.pdf

Aguinis, H., Joo, H., & Gottfredson, R. K. (2012). Performance management universals: Think globally and act locally. *Business Horizons*, 55(4), 385-392.

Albright, J. J., & Park, H. M. (2009). Confirmatory factor analysis using Amos, LISREL, Mplus, and SAS/STAT CALIS. *The Trustees of Indiana University*, 1, 1-85.

Almond, P., Edwards, T., Colling, T., Gunnigle, P., Muller-Camen, M., Quintanilla, J., & Wachter, H. (2005). Unraveling home and host country effects: an investigation of the HR policies of an American multinational in four European countries. *Industrial Relations*, 44(2), 276-306.

Anusic, I., Schimmack, U., Pinkus, R. T., & Lockwood, P. (2009). The nature and structure of correlations among Big Five ratings: The halo-alpha-beta model. *Journal of Personality and Social Psychology*, 97(6), 1142.

Appelbaum, S. H., Nadeau, D., & Cyr, M. (2008). Performance evaluation in a matrix organization: a case study (Part two). *Industrial and Commercial Training*, 40(6), 295-299.

- Appelbaum, S. H., Roy, M., & Gilliland, T. (2011). Globalization of performance appraisals: theory and applications. *Management Decision*, 49(4), 570-585.
- Arbuckle, J. L. (2009). Amos 18.0 [computer software] SPSS. Inc, Chicago, IL.
- Bae, E.K. (2006). Major elements and issues in performance management system: a literature review. *Academy of Human Resource Development International Conference (AHRD)*, Columbus, OH, 22-26 February, pp. 1430-7.
- Baldwin, C. K., & Caldwell, L. L. (2003). Development of the free time motivation scale for adolescents. *Journal of Leisure Research*, 35(2), 129-151.
- Balzer, W. K. (1986). Biases in the recording of performance-related information: The effects of initial impression and centrality of the appraisal task. *Organizational Behavior and Human Decision Processes*, 37(3), 329-347.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44(1), 1-26.
- Bartlett, Maurice S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology*, 3.2, 77-85.
- Bauer Jr, K. W., Alsing, S. G., & Greene, K. A. (2000). Feature screening using signal-to-noise ratios. *Neurocomputing*, 31(1), 29-44.
- Beckstead, J. W. (2002). Confirmatory factor analysis of the Maslach Burnout Inventory among Florida nurses. *International Journal of Nursing Studies*, 39(8), 785-792.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238.
- Berger, J., Harbring, C., & Sliwka, D. (2013). Performance Appraisals and the Impact of Forced Distribution—An Experimental Investigation. *Management Science*, 59(1), 54-68.

- Berger, L. A., & Berger, D. R. (2008). *The compensation handbook: A state-of-the-art guide to compensation strategy and design*. New York, NY: McGraw-Hill.
- Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal*, 314(7080), 572.
- Blume, B. D., Baldwin, T. T., & Rubin, R. S. (2009). Reactions to different types of forced distribution performance evaluation systems. *Journal of Business and Psychology*, 24(1), 77-91.
- Boice, D. F., & Kleiner, B. H. (1997). Designing effective performance appraisal systems. *Work Study*, 46(6), 197-201.
- Bol, J. C. (2011). The determinants and performance effects of managers' performance evaluation biases. *The Accounting Review*, 86(5), 1549-1575.
- Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods and Research*, 21, 205-229.
- Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing*, 1(1), 55-86.
- Byrne, B. M. (2009). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. New York, NY: Routledge.
- Campbell, J. P. (1990). An overview of the army selection and classification project (Project A). *Personnel Psychology*, 43(2), 231-239.
- Cattell, R. B. (1966). The Scree test for the number of factors. *Multivariate behavioral research*, 1(2), 245-276.
- Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147-168.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98.

Costello, A. B., & Osborne, J. W. (2005). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment Research & Evaluation*, 10(7), 2.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2(3), 292.

DeCoster, J. (1998). Overview of Factor Analysis. Retrieved December 16, 2013, from <http://www.stat-help.com/factor.pdf>

Dees, R. A., Nestler, S. T., & Kewley, R. (2013). WholeSoldier performance appraisal to support mentoring and personnel decisions. *Decision Analysis*, 10(1), 82-97.

Dillon, W. R., & Goldstein, M. (1984). *Multivariate Analysis* (pp. 394-429). New York: Wiley.

Enders, C. K. (2005). An SAS macro for implementing the modified bollen-stine bootstrap for missing data: Implementing the bootstrap using existing structural equation modeling software. *Structural Equation Modeling*, 12(4), 620-641.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272.

Farrell, A.M., & Rudd, J.M. (2009). Factor Analysis and Discriminant Validity: A Brief Review of Some Practical Issues. *Australia-New Zealand Marketing Academy Conference (ANZMAC)*, December, Melbourne, Australia.

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39(2), 291-314.

Fox, J. (2002). *An R and S-PLUS Companion to Applied Regression*. Thousand Oaks, CA: Sage Publications.

Gagné, M. (2009). A model of knowledge-sharing motivation. *Human Resource Management*, 48(4), 571-589.

Gao, S., Mokhtarian, P. L., & Johnston, R. A. (2008). Nonnormality of data in structural equation models. *Transportation Research Record: Journal of the Transportation Research Board*, 2082(1), 116-124.

Ghorpade, J., Chen, M. M., & Caggiano, J. (1995). Creating quality-driven performance appraisal systems. *Academy of Management Executive*, 9(1), 32-39.

Giangreco, A., Carugati, A., Sebastiano, A., & Tamimi, H. A. (2012). War outside, ceasefire inside: An analysis of the performance appraisal system of a public hospital in a zone of conflict. *Evaluation and Program Planning*, 35(1), 161-170.

Gizaw, T. S. (2010). Performance Appraisal and Its Significance as a Human Resource Management Tool. *Proceedings of Conference of the Global Awareness Society International*, San Juan, PR (pp. 1-12).

Gliem, J. and Gliem, R. (2003). Calculating, Interpreting, and Reporting Cronbach's Alpha Reliability Coefficient for Likert-Type Scales. *2003 Midwest Research to Practice Conference in Adult, Continuing and Community Education*. Columbus, OH.

Greenberg, J. (1986). Determinants of perceived fairness of performance evaluations. *Journal of Applied Psychology*, 71(2), 340.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Erlbaum.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling*. Cary, NC: SAS Institute Inc.

Helfrich, C. D., Li, Y. F., Mohr, D. C., Meterko, M., & Sales, A. E. (2007). Assessing an organizational culture instrument based on the Competing Values Framework: Exploratory and confirmatory factor analyses. *Implementation Science*, 2(1), 13.

Henson, R. K., & Roberts, J. K. (2006). Use of Exploratory Factor Analysis in Published Research Common Errors and Some Comment on Improved Practice. *Educational and Psychological Measurement*, 66(3), 393-416.

Heslin, Peter A., and Don VandeWalle (2011). Performance appraisal procedural justice: The role of a manager's implicit person theory. *Journal of Management*, 37.6, 1694-1718.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *Journal of Business Research Methods*, 6, 53-60.

Hopkins, S. M., & Weathington, B. L. (2006). The relationships between justice perceptions, trust, and employee attitudes in a downsized organization. *Journal of Psychology*, 140(5), 477-498.

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review*, 11, 354-373.

Hubbell, A. P., & Chory-Assad, R. M. (2005). Motivating factors: Perceptions of justice and their relationship with managerial and organizational trust. *Communication Studies*, 56(1), 47-70.

Hunter, A., Kennedy, L., Henry, J., & Ferguson, I. (2000). Application of neural networks and sensitivity analysis to improved prediction of trauma survival. *Computer Methods and Programs in Biomedicine*, 62(1), 11-19.

Hutcheson, G.D., Sofroniou, N. (1999). *The multivariate social Scientist: Introductory statistics using generalized linear models*. London: Sage Publications Ltd.

Iacobucci, D. (2010). Structural equations modeling: Fit indices, sample size, and advanced topics. *Journal of Consumer Psychology*, 20(1), 90-98.

Jackson, D. J., & Ward, M. A. (1992). *An Analysis of the Air Force Enlisted Performance Feedback System* (Master's thesis, Air Force Institute of Technology, Wright Patterson AFB, OH School of Systems and Logistics). Retrieved from DTIC database. July 8, 2013, from <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA259732>

Jackson, D. L., Gillaspay Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: an overview and some recommendations. *Psychological Methods*, 14(1), 6.

Jafari, M., Bourouni, A., & Amiri, R. H. (2009). A new framework for selection of the best performance appraisal method. *European Journal of Social Sciences*, 7(3), 92-100.

JMP®, Version 11. SAS Institute Inc., Cary, NC, 1989-2013.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI, analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables, and Least Squares Methods*. Mooresville, IN: Scientific Software.

Kankainen, A., Taskinen, S., & Oja, H. (2007). Tests of multinormality based on location vectors and scatter matrices. *Statistical Methods and Applications*, 16(3), 357-379.

Kaplan, P. D., & Knowles, J. A. (2004). Kappa: A generalized downside risk-adjusted performance measure. *Journal of Performance Measurement*, 8, 42-54.

Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika*, 35(4), 401-415.

Keeney, R. L. (1982). Decision analysis: an overview. *Operations Research*, 30(5), 803-838.

Keeney, R. L. (1992). *Value-focused thinking: A path to creative decision-making*. Cambridge, MA: Harvard University Press.

Keeney, R. L. (1994). Creativity in decision making with value-focused thinking. *Sloan Management Review*, 35, 33-33.

Kirkwood, C. W. (1996). *Strategic Decision Making: Multiobjective Decision Analysis with Spreadsheets*. Belmont, CA: Duxbury Press

Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology*, 26(2), 195-197.

Krycha, K. A., & Wagner, U. (1999). Applications of artificial neural networks in management science: a survey. *Journal of Retailing and Consumer Services*, 6(4), 185-203.

Larter, D. (Dec 2011). NCOs offer ways to put meaning back in EPR. *Air Force Times*. Retrieved Sep 30, 2013, from <http://www.airforcetimes.com/apps/pbcs.dll/article?AID=2011112050329>

Lilley, D., & Hinduja, S. (2007). Police officer performance appraisal and overall satisfaction. *Journal of Criminal Justice*, 35(2), 137-150.

Longenecker, C. O., Sims, H. P., & Gioia, D. A. (1987). Behind the mask: The politics of employee appraisal. *The Academy of Management Executive*, 1(3), 183-193.

Losey, S. (Sep 16, 2013). EPR overhaul coming: 'Biggest single change' ever for enlisteds. *Air Force Times*. Retrieved Sep 30, 2013, from <http://www.airforcetimes.com/apps/pbcs.dll/article?AID=2013309160039>

Losey, S. (Sep 18, 2013). EPR overhaul coming: Cody: EPR overhaul will fix 'inflated system'. *Air Force Times*. Retrieved Sep 30, 2013, from <http://www.airforcetimes.com/apps/pbcs.dll/article?AID=2013309180031>

Leung, T. F., Wong, G. W., Ko, F. W., Lam, C. W., & Fok, T. F. (2005). Clinical and atopic parameters and airway inflammatory markers in childhood asthma: a factor analysis. *Thorax*, 60(10), 822-826.

Lubke, G., & Muthén, B. (2004). Factor-analyzing Likert scale data under the assumption of multivariate normality complicates a meaningful comparison of observed groups or latent classes. *Structural Equation Modeling*, 11, 514-534.

MATLAB version 8.0.0.783 (R2012b). Natick, Massachusetts: The MathWorks Inc., 2012.

Mayer, R. C., & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, 84(1), 123.

Mayer, R. C., & Gavin, M. B. (2005). Trust in management and performance: who minds the shop while the employees watch the boss? *Academy of Management Journal*, 48(5), 874-888.

Meyer, H., Kay, E. & French, J. R. (1965). Split roles in performance appraisal. *Harvard Business Review*, 43(January/February): 123-129.

Maciel, E. D. S., Savay-da-Silva, L. K., Vasconcelos, J. S., Galvão, J. A., Sonati, J. G., Silva, D. D., & Oetterer, M. (2013). Application of exploratory factor analysis to assess fish consumption in a university community. *Food Science and Technology (Campinas)*, 33(1), 99-106.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, 103(3), 391.

Merkle, L. A., Layne, C. S., Bloomberg, J. J., & Zhang, J. J. (1998). Using factor analysis to identify neuromuscular synergies during treadmill walking. *Journal of Neuroscience Methods*, 82(2), 207-214.

Merrick, J. R., Parnell, G. S., Barnett, J., & Garcia, M. (2005). A multiple-objective decision analysis of stakeholder values to identify watershed improvement needs. *Decision Analysis*, 2(1), 44-57.

Moers, F. (2005). Discretion and bias in performance evaluation: the impact of diversity and subjectivity. *Accounting, Organizations and Society*, 30(1), 67-80.

Monat, J. P. (2009). The benefits of global scaling in multi-criteria decision analysis. *Judgment and Decision Making*, 4(6), 492-508.

Mount, M., Ilies, R., & Johnson, E. (2006). Relationship of personality traits and counterproductive work behaviors: The mediating effects of job satisfaction. *Personnel Psychology*, 59(3), 591-622.

Murphy, K. R., Jako, R. A., & Anhalt, R. L. (1993). Nature and consequences of halo error: A critical analysis. *Journal of Applied Psychology*, 78(2), 218.

Murphy, K. R. (2008). Explaining the weak relationship between job performance and ratings of job performance. *Industrial and Organizational Psychology*, 1(2), 148-160.

Nankervis, A. R., & Compton, R. L. (2006). Performance management: theory in practice? *Asia Pacific Journal of Human Resources*, 44(1), 83-101.

Noe, R.A., Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (1997). *Human resource management: Gaining a competitive advantage*. Chicago: Irwin.

Orwat, M. E. (2008). *A Decision Framework for Enhancing Mobile Ad Hoc Network Stability and Security* (Doctoral dissertation, Naval Postgraduate School, Monterey, CA). Retrieved from DTIC database. Dec 21, 2013, from <http://www.dtic.mil/dtic/tr/fulltext/u2/a511848.pdf>

Parnell, G. S. (2007). Methods for Conducting Military Operational Analysis: Best Practices in Use Throughout the Department of Defense, chapter 19. In A. Loerch & L. Rainey (Eds.), *Value-Focused Thinking Using Multiple Objective Decision Analysis* (pp. 1-35). Arlington, VA: Military Operations Research Society.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology*, 88(5), 879-903.

Prendergast, C., & Topel, R. (1993). Discretion and bias in performance evaluation. *European Economic Review*, 37(2), 355-365.

Pöyhönen, M., & Hämäläinen, R. P. (2001). On the convergence of multiattribute weighting methods. *European Journal of Operational Research*, 129(3), 569-585.

Pruitt, J. L. (2011). *Decision Analysis and Validation of Value Focused Thinking Decision Models Using Multivariate Analysis Techniques* (Master's thesis, Air Force Institute of Technology, Wright Patterson AFB, OH School of Engineering and Management). Retrieved from DTIC database. Dec 14, 2013, from <http://www.dtic.mil/dtic/tr/fulltext/u2/a540067.pdf>

Quilty, L. C., Sellbom, M., Tackett, J. L., & Bagby, R. M. (2009). Personality trait predictors of bipolar disorder symptoms. *Psychiatry Research*, 169(2), 159-163.

Roberts, G. E. (2003). Employee performance appraisal system participation: A technique that works. *Public Personnel Management*, 32(1), 89-98.

Schogol, J. (Jan 2013). Airmen share ideas for CMSAF Cody. *Air Force Times*. Retrieved Sep 30, 2013, from <http://www.airforcetimes.com/apps/pbcs.dll/article?AID=2013301140325>

Schogol, J. (Feb 2013). New rules aimed at stopping EPR inflation. *Air Force Times*. Retrieved Sep 30, 2013, from <http://www.airforcetimes.com/apps/pbcs.dll/article?AID=2013302040326>

Schogol, J. (March 2013). Cody launches review of tape tests, EPRs. *Air Force Times*. Retrieved Sep 30, 2013, from <http://www.airforcetimes.com/apps/pbcs.dll/article?AID=2013303050356>

Schogol, J. (Sept 2013). Air Force may get even smaller if budget cuts continue, Cody says. *Air Force Times*. Retrieved Sep 30, 2013, from <http://www.airforcetimes.com/apps/pbcs.dll/article?AID=2013309170039>

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-338.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23-74.

Shi, F., Liu, Y. Y., Kong, X., & Chen, Y. (2013). Artificial Neural Network for search for metal poor galaxies. *Astronomy & Astrophysics*. Advance Online Publication. Retrieved Jan 23, 2014, from <http://arxiv.org/pdf/1312.1779>

Schiefer, M., Robbert, A. A., Crown, J. S., Manacapilli, T., & Wong, C. (2008). *The Weighted Airman Promotion System: Standardizing Test Scores*. Rand Corp Santa Monica, CA.

Spence, J. R., & Keeping, L. (2011). Conscious rating distortion in performance appraisal: A review, commentary, and proposed framework for research. *Human Resource Management Review*, 21(2), 85-95.

Spiliotopoulou, G. (2009). Reliability reconsidered: Cronbach's alpha and paediatric assessment in occupational therapy. *Australian Occupational Therapy Journal*, 56(3), 150-155.

SPSS Inc, Version 18.0, PASW Statistics for Windows, Chicago, IL, 2009.

Steiger, J. H., & Lind, J. C. (1980, May). Statistically based tests for the number of common factors. *Annual Meeting of the Psychometric Society*, Iowa City, IA (Vol. 758).

- Stevanovic, D. (2009). Serbian KINDL questionnaire for quality of life assessments in healthy children and adolescents: reproducibility and construct validity. *Health and Quality of Life Outcomes*, 7(1), 79.
- Stone, F. M. (1999). *Coaching, counseling & mentoring: How to choose & use the right technique to boost employee performance*. New York, NY: Amacom.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.
- Thompson, G. (2013). *Interview with CMSAF*. (2013). 386th Air Expeditionary Wing Public Webpage. Retrieved April 4, 2013, from <http://www.386aew.afcent.af.mil/news/story.asp?id=123334876>
- Tinsley, H. E., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology*, 34(4), 414.
- Trappey, A. J., Trappey, C. V., Wu, C. Y., & Lin, C. W. (2012). A patent quality analysis for innovative technology and product development. *Advanced Engineering Informatics*, 26(1), 26-34.
- Van Damme, S., Crombez, G., Bijttebier, P., Goubert, L., & Van Houdenhove, B. (2002). A confirmatory factor analysis of the Pain Catastrophizing Scale: invariant factor structure across clinical and non-clinical populations. *Pain*, 96(3), 319-324.
- Viswesvaran, C., & Ones, D. S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, 8(4), 216-226.
- von Winterfeldt, D., Edwards, W., 1986. *Decision Analysis and Behavioral Research*. Cambridge, UK: Cambridge University Press.
- Watson, S. R. (1987). *Decision Synthesis: The Principles and Practice of Decision Analysis*. Cambridge, UK: Cambridge University Press.
- Widaman, K. F., & Herrerger, L. G. (1985). Iterative least squares estimates of communality: Initial estimate need not affect stabilized value. *Psychometrika*, 50(4), 469-477.

Williams, B., Brown, T., & Onsmann, A. (2012). Exploratory factor analysis: A five-step guide for novices. *Journal of Emergency Primary Health Care*, 8(3), 1.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research a content analysis and recommendations for best practices. *Counseling Psychologist*, 34(6), 806-838.

Yang, J. (2005). *The Role of Trust in Organizations: Do Foci and Bases Matter?* (Doctoral dissertation, Louisiana State University in The Interdepartmental Program in Business Administration (Management), Nanjing University, China). Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.89.2172&rep=rep1&type=pdf>

Yates, Brian A. (2011). *Rethinking Military Personnel Evaluations* (Graduate research paper, Army War College Carlisle Barracks, PA). Retrieved from DTIC database. Jul 8, 2013, from <http://www.dtic.mil/dtic/tr/fulltext/u2/a560299.pdf>

Yee, C. C., & Chen, Y. Y. (2009). Performance appraisal system using multifactorial evaluation model. *World Academy of Science, Engineering and Technology*, 53, 231-235.

Zhang, G. P. (2000). Neural networks for classification: a survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions*, 30(4), 451-462.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological bulletin*, 99(3), 432.

Vita.

Captain Christopher Jones was born in Copperhill, Tennessee. He graduated from Fannin County High School in Blue Ridge, Georgia in 1990. In 1992, Captain Jones enlisted into the United States Air Force as an Aircraft Armament Systems Specialist. Captain Jones served in an aircraft armament positions over the next eight years with a myriad of duties ranging from a B-1B aircraft armament system specialist to a NATO weapons load monitor for Allied Command Europe. Captain Jones briefly left the Air Force in 2000, and working for Boeing Phantom Works supporting B-1B aircraft conventional weapons upgrades. In late 2000, Captain Jones reentered the United States Air Force ascending to the rank of Master Sergeant. In 2006, Captain Jones received a Bachelor of Science in Mathematics from Chapman University in Orange California. After graduation, Captain Jones was selected to attend Air Force Officer Training School and was commissioned in the United States Air Force. Captain Jones was assigned to Scott Air Force Base, Illinois where he served as a Joint Mobility analyst for United States Transportation Command and Air Mobility Command. During this period, Captain Jones deployed to Southwest Asia where he developed the USCENCOM, Theater-Express Value Focused Thinking algorithm and database which is used to move military cargo using commercial means. During this period, Captain Jones earned his Masters of Business Administration degree from McKendree University in Lebanon, Illinois. Upon return for deployment, Captain Jones was selected for career broadening returning to the munitions maintenance career field at Barksdale

Air Force Base, Louisiana. Captain Jones filled several positions at Barksdale, culminating as a flight commander for the 2d Munitions Squadron cruise missile flight. In August 2012, Captain Jones entered the Graduate School of Engineering and Management, Air Force Institute of Technology at Wright Patterson AFB, Ohio. Upon graduation, he will be assigned to the HQ USAFE/A9 at Ramstein AB, Germany.

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 074-0188</i>		
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 22-03-2014		2. REPORT TYPE Master's Thesis		3. DATES COVERED (From - To) September 2012 - March 2014	
TITLE AND SUBTITLE Value Focused Thinking Approach Using Multivariate Validation for Junior Enlisted Performance Reporting in the United States Air Force			5a. CONTRACT NUMBER		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
6. AUTHOR(S) Jones, Christopher M., Captain, USAF					
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/ENS) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865			8. PERFORMING ORGANIZATION REPORT NUMBER AFIT-ENS-14-M-13		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) :Intentionally left blank			10. SPONSOR/MONITOR'S ACRONYM(S)		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)		
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRUBTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.					
13. SUPPLEMENTARY NOTES This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.					
14. ABSTRACT From United States Air Force (USAF) doctrine, Air Force Instruction 1-1 lists three purposes for the USAF Enlisted Evaluation System. The first purpose is to provide feedback to individuals on how well they are meeting expectations. The second purpose is to provide a cumulative record of performance and potential based on observations. The third purpose is to identify the best qualified personnel. However, current Air Force leadership has expressed a need to revamp the enlisted appraisal process, requesting consistency in identifying the best performers, reduction in ratings inflation, and better delineation between "near peer" performers. This research proposes utilizing Value-Focused Thinking to perform junior enlisted performance reports, to better align with Air Force doctrine and values. Moreover, the multivariate Management Science techniques of Exploratory and Confirmatory Factor Analysis are applied to statistically validate the accuracy and defensibility of the design. Finally, Artificial Neural Networks are employed to showcase the classification accuracy of the proposed system. In addition to providing consistency, inflation reduction, and delineation during appraisals, this research advocates the use of a web-based design to reduce administrative demands and to provide query capability of appraisal data to the Air Force Personnel Center for trend and force management decisions.					
15. SUBJECT TERMS Value-Focused Thinking, Exploratory Factor Analysis, Confirmatory Factor Analysis, Artificial Neural Networks, EPRs, Appraisals					
16. SECURITY CLASSIFICATION OF:		17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 349	19a. NAME OF RESPONSIBLE PERSON Major Jennifer L. Geffre, AFIT/ENS	
a. REPORT U	b. ABSTRACT U			c. THIS PAGE U	19b. TELEPHONE NUMBER (Include area code) (937)255-3636, ext 4646 jennifer.geffre@afit.edu