6-1-2013

An Assessment of Character and Leadership Development Latent Factor Structures through Confirmatory Factor, Item Response Theory, and Latent Class Analyses

David L. Higginbotham
University of Denver

Follow this and additional works at: https://digitalcommons.du.edu/etd

Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation
https://digitalcommons.du.edu/etd/285

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu, dig-commons@du.edu.
AN ASSESSMENT OF CHARACTER AND LEADERSHIP DEVELOPMENT
LATENT FACTOR STRUCTURES THROUGH CONFIRMATORY FACTOR, ITEM
RESPONSE THEORY, AND LATENT CLASS ANALYSES

A Dissertation
Presented to
the Faculty of the Morgridge College of Education
University of Denver

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

by
David L. Higginbotham
June 2013
Advisor: Dr. Kathy E. Green
Abstract

This study leveraged the complementary nature of confirmatory factor (CFA), item response theory (IRT), and latent class (LCA) analyses to strengthen the rigor and sophistication of evaluation of two new measures of the Air Force Academy’s “leader of character” definition—the Character Mosaic Virtues (CMV) and the Leadership Mosaic Inventory (LMI). Special CFA methods involving robust weighted least squares estimation were implemented to analyze the rated responses through linear structural equation modeling. Most informative at the subscale level, the CFA techniques provided evidence of factorial validity and other desirable psychometric properties in support of previous exploratory results for a nine-factor CMV and a unidimensional LMI. As an alternative to CFA, IRT’s nonlinear approach and capability of estimating the probability of a response based on the amount of a latent trait was applied to the CMV and LMI data. Most informative at the item level, individual item difficulty parameters were estimated as the amount of the latent trait required for a cadet to give a particular response to an item and mapped against person ability estimates. The IRT analyses were extended to explore the underlying dimensions of the CMV and LMI through multidimensional IRT (MIRT)—results supporting the theoretical dimensional structure of the CMV were substantiated while new evidence of a multidimensional LMI structure was concluded. LCA techniques permitted the inference of classifying mixtures of unobserved cadet
subpopulations based on their responses to the CMV and LMI. The analyses uncovered the meaning and number of underlying subpopulations not evident through the other two traditional factor structure analysis techniques. The creation of the study’s latent class models complemented the more traditional dimensional approaches of structure assessment with an understanding of the unobserved cadet subpopulations which could lead to future targeted cadet developmental opportunities being applied at organizational levels or groups deficient in certain latent traits. By exposing more researchers, decision-makers, and other stakeholders to these three advanced psychometric evaluation methods, this study benefited the fields of moral development and leadership development, especially at the nation’s service academies.
Acknowledgements

For Dad—father, friend, encourager, exemplar
1935 - 2011

I give my sincerest gratitude to those who have guided and sustained me through this advanced education opportunity. I could not have completed this journey without the love, understanding, patience, and inspiration of my wonderful wife and my exceptional son. I am forever indebted to my parents who instilled in me the values of faith, family, and service.

I give my heartfelt thanks to Dr. Kathy Green, my advisor and dissertation director, for her wise counsel and overwhelming support of my pursuit of completing this degree in less than three years. I am also grateful for the expertise and generosity of time of Dr. Duan Zhang, Dr. Antonio Olmos-Gallo, Dr. Donald Bacon, and Dr. Walter LaMendola for guiding me through this degree program and research endeavor.

I am beholden to the following members of the faculty and staff from the Department of Mathematical Sciences and the Center for Character and Leadership Development of the United States Air Force Academy, whose encouragement, trust, and support made the selection and completion of this program a reality—Col John Andrew, PhD; Dr. Bradley Warner; Col Tony McKenzie; Col Joseph Sanders, PhD; Dr. Michael Rosebush; Col (Retired) John Norton; and Col (Retired) Tom Berry.
# Table of Contents

Chapter One: Introduction and Review of the Literature ........................................... 1
  Background .............................................................................................................. 1
  Statement of the Problem ..................................................................................... 17
  Purpose of the Study ............................................................................................ 20
  Research Questions ............................................................................................... 21
  Confirmatory Factor Analysis ............................................................................... 22
    Exploratory versus confirmatory ....................................................................... 23
    Dimensionality of measurement ....................................................................... 24
    Dimensionality of structure ............................................................................... 28
    Five-step CFA modeling approach ................................................................... 30
  CFA study to assess latent factor structure ......................................................... 43
  Item Response Theory ......................................................................................... 44
    The Rasch RSM. ................................................................................................. 45
    Dimensionality of measurement ....................................................................... 50
    Three approaches to assess dimensionality ....................................................... 54
    IRT study to assess dimensionality ................................................................... 59
  Latent Class Analysis ........................................................................................... 61
    Description of the model ................................................................................... 64
    Three-step modeling approach ....................................................................... 66
    Model estimation ............................................................................................... 69
    LCA study to explore population subtypes ....................................................... 70

Chapter Two: Method ............................................................................................... 72
  Participants ........................................................................................................... 72
    Character Mosaic Virtues (CMV): November 2011 ......................................... 72
    Leadership Mosaic Inventory (LMI): September 2012 ....................................... 73
    Leadership Mosaic Inventory (LMI): October 2012 .......................................... 73
  Instruments ............................................................................................................ 74
    CMV. ................................................................................................................... 74
    LMI. ..................................................................................................................... 75
  Procedure ............................................................................................................... 76
    CMV. .................................................................................................................... 76
    LMI. ...................................................................................................................... 77
    Human subjects protection ................................................................................. 79
  Analytical Strategy ................................................................................................. 79
    Research question one ....................................................................................... 79
    Research question two ....................................................................................... 81
    Research question three ..................................................................................... 82
    Research question four ....................................................................................... 84
    Research question five ....................................................................................... 85
    Research question six ......................................................................................... 87

Chapter Three: Results ............................................................................................ 89
List of Tables

Table 1  Hypothetical 3-Factor Model Convergent and Discriminant Validity Tests ............................................................................................................. 26
Table 2  CMV Eight-Factor Global Model-Fit Results ................................................. 92
Table 3  Standardized Parameter Estimates for CMV Eight-Factor Model ............... 93
Table 4  Reliability Estimates for CMV Eight-Factor Model .................................... 94
Table 5  CMV Nine-Factor Global Model-Fit Results ................................................. 98
Table 6  Standardized Parameter Estimates for CMV Nine-Factor Model ............... 98
Table 7  Reliability Estimates for CMV Nine-Factor Model .................................... 100
Table 8  CMV Nine-Factor Global Modified Model-Fit Results .............................. 104
Table 9  Standardized Parameter Estimates for CMV Nine-Factor Modified Model ............................................................................................................. 104
Table 10 Reliability Estimates for CMV Nine-Factor Modified Model.................... 105
Table 11 CMV Convergent and Discriminant Validity Tests .................................. 108
Table 12 CMV Construct-Level Correlation Matrix ................................................. 109
Table 13 LMI Self-Rating Construct-Level Correlation Matrix ............................... 114
Table 14 LMI Subordinate-Rating Construct-Level Correlation Matrix .................. 114
Table 15 LMI Self-Rating Unidimensional Global Model-Fit Results ..................... 117
Table 16 Standardized Parameter Estimates for LMI Self-Rating Unidimensional Model ............................................................................................................. 117
Table 17 Reliability Estimates for LMI Self-Rating Unidimensional Model ............ 118
Table 18 LMI Subordinate-Rating Unidimensional Global Model-Fit Results ........ 120
Table 19 Standardized Parameter Estimates for LMI Subordinate-Rating Unidimensional Model ............................................................................................................. 120
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Reliability Estimates for LMI Subordinate-Rating Unidimensional Model</td>
<td>121</td>
</tr>
<tr>
<td>21</td>
<td>LMI Self-Rating Item-Response Frequencies</td>
<td>123</td>
</tr>
<tr>
<td>22</td>
<td>LMI Self-Rating Unidimensional Global Model-Fit Results by Rating Scale</td>
<td>124</td>
</tr>
<tr>
<td>23</td>
<td>LMI Subordinate-Rating Unidimensional Global Modified Model-Fit Results</td>
<td>127</td>
</tr>
<tr>
<td>24</td>
<td>Standardized Parameter Estimates for LMI Subordinate-Rating Unidimensional Modified Model</td>
<td>127</td>
</tr>
<tr>
<td>25</td>
<td>Reliability Estimates for LMI Subordinate-Rating Unidimensional Modified Model</td>
<td>128</td>
</tr>
<tr>
<td>26</td>
<td>CMV Eight-Factor Model Consecutive Approach Fit Results</td>
<td>132</td>
</tr>
<tr>
<td>27</td>
<td>CMV Eight-Factor Model Multidimensional Reliabilities</td>
<td>134</td>
</tr>
<tr>
<td>28</td>
<td>CMV Eight-Factor Model Comparisons</td>
<td>135</td>
</tr>
<tr>
<td>29</td>
<td>CMV Eight-Factor Model Reliabilities</td>
<td>135</td>
</tr>
<tr>
<td>30</td>
<td>CMV Eight-Factor Consecutive and Multidimensional Correlation Matrix</td>
<td>137</td>
</tr>
<tr>
<td>31</td>
<td>CMV Nine-Factor Model Consecutive Approach Fit Results</td>
<td>140</td>
</tr>
<tr>
<td>32</td>
<td>CMV Nine-Factor Model Multidimensional Reliabilities</td>
<td>142</td>
</tr>
<tr>
<td>33</td>
<td>CMV Nine-Factor Model Comparisons</td>
<td>143</td>
</tr>
<tr>
<td>34</td>
<td>CMV Nine-Factor Model Reliabilities</td>
<td>143</td>
</tr>
<tr>
<td>35</td>
<td>CMV Nine-Factor Consecutive and Multidimensional Correlation Matrix</td>
<td>145</td>
</tr>
<tr>
<td>36</td>
<td>CMV Overall Best Model Fit Comparisons</td>
<td>146</td>
</tr>
<tr>
<td>37</td>
<td>Item Difficulty Estimates for CMV Nine-Factor Multidimensional Model</td>
<td>146</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>38</td>
<td>CMV Nine-Factor Post Hoc Model Multidimensional Reliabilities</td>
<td>150</td>
</tr>
<tr>
<td>39</td>
<td>Item Difficulty Estimates for CMV Nine-Factor Post Hoc Multidimensional Model</td>
<td>151</td>
</tr>
<tr>
<td>40</td>
<td>LMI Self-Rating Model Consecutive Approach Fit Results</td>
<td>157</td>
</tr>
<tr>
<td>41</td>
<td>LMI Self-Rating Model Multidimensional Reliabilities</td>
<td>159</td>
</tr>
<tr>
<td>42</td>
<td>LMI Self-Rating Model Comparisons</td>
<td>160</td>
</tr>
<tr>
<td>43</td>
<td>LMI Self-Rating Model Reliabilities</td>
<td>160</td>
</tr>
<tr>
<td>44</td>
<td>LMI Self-Rating Consecutive and Multidimensional Correlation Matrix</td>
<td>161</td>
</tr>
<tr>
<td>45</td>
<td>Item Difficulty Estimates for LMI Self-Rating Multidimensional Model</td>
<td>161</td>
</tr>
<tr>
<td>46</td>
<td>LMI Self-Rating Six-Factor Post Hoc Model Multidimensional Reliabilities</td>
<td>165</td>
</tr>
<tr>
<td>47</td>
<td>Item Difficulty Estimates for LMI Self-Rating Modified Multidimensional Model</td>
<td>166</td>
</tr>
<tr>
<td>48</td>
<td>LMI Subordinate-Rating Model Consecutive Approach Fit Results</td>
<td>173</td>
</tr>
<tr>
<td>49</td>
<td>LMI Subordinate-Rating Model Multidimensional Reliabilities</td>
<td>175</td>
</tr>
<tr>
<td>50</td>
<td>LMI Subordinate-Rating Model Comparisons</td>
<td>176</td>
</tr>
<tr>
<td>51</td>
<td>LMI Subordinate-Rating Model Reliabilities</td>
<td>176</td>
</tr>
<tr>
<td>52</td>
<td>LMI Subordinate-Rating Consecutive and Multidimensional Correlation Matrix</td>
<td>177</td>
</tr>
<tr>
<td>53</td>
<td>Item Difficulty Estimates for LMI Subordinate-Rating Multidimensional Model</td>
<td>177</td>
</tr>
<tr>
<td>54</td>
<td>LMI Subordinate-Rating Six-Factor Post Hoc Model Multidimensional Reliabilities</td>
<td>181</td>
</tr>
<tr>
<td>55</td>
<td>Item Difficulty Estimates for LMI Subordinate-Rating Modified Multidimensional Model</td>
<td>182</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Table 56</td>
<td>CMV LCA Model Comparisons</td>
<td>192</td>
</tr>
<tr>
<td>Table 57</td>
<td>CMV Final Latent Class Counts and Proportions</td>
<td>193</td>
</tr>
<tr>
<td>Table 58</td>
<td>CMV Average Latent Class Probabilities for Most Likely Latent Class Membership</td>
<td>193</td>
</tr>
<tr>
<td>Table 59</td>
<td>CMV 3-Class LCA Membership Probabilities</td>
<td>196</td>
</tr>
<tr>
<td>Table 60</td>
<td>LMI Self-Rating LCA Model Comparisons</td>
<td>202</td>
</tr>
<tr>
<td>Table 61</td>
<td>LMI Self-Rating Final Latent Class Counts and Proportions</td>
<td>203</td>
</tr>
<tr>
<td>Table 62</td>
<td>LMI Self-Rating Average Latent Class Probabilities for Most Likely Latent Class Membership</td>
<td>203</td>
</tr>
<tr>
<td>Table 63</td>
<td>LMI Self-Rating 3-Class LCA Membership Probabilities</td>
<td>205</td>
</tr>
<tr>
<td>Table 64</td>
<td>LMI Subordinate-Rating LCA Model Comparisons</td>
<td>211</td>
</tr>
<tr>
<td>Table 65</td>
<td>LMI Subordinate-Rating Final Latent Class Counts and Proportions</td>
<td>212</td>
</tr>
<tr>
<td>Table 66</td>
<td>LMI Subordinate-Rating Average Latent Class Probabilities for Most Likely Latent Class Membership</td>
<td>212</td>
</tr>
<tr>
<td>Table 67</td>
<td>LMI Subordinate-Rating 5-Class LCA Membership Probabilities</td>
<td>214</td>
</tr>
<tr>
<td>Table 68</td>
<td>LMI Self-Rating LCA Two Sample Comparisons</td>
<td>218</td>
</tr>
<tr>
<td>Table 69</td>
<td>LMI Subordinate-Rating LCA Two Sample Comparisons</td>
<td>221</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Figure 1</td>
<td>USAF Core Values and their Respective Virtues</td>
<td>4</td>
</tr>
<tr>
<td>Figure 2</td>
<td>USAF Institutional Leadership Competencies and Sub-Competencies by Leadership Level</td>
<td>5</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Officer Development System Foundation</td>
<td>7</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Character Mosaic Virtues’ Nine Factors, Definitions, and Item Allocations</td>
<td>13</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Leadership Mosaic Inventory Single Factor and Item Allocations Based on USAF Institutional Sub-Competencies</td>
<td>17</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Typical Standard Unidimensional Two-Factor CFA Model</td>
<td>28</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Typical Nonstandard Multidimensional Two-Factor CFA Model</td>
<td>28</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Unstandardized Two-Factor CFA Model with Ordered Categorical Items</td>
<td>33</td>
</tr>
<tr>
<td>Figure 9</td>
<td>Standardized Two-Factor CFA Model with Ordered Categorical Items</td>
<td>33</td>
</tr>
<tr>
<td>Figure 10</td>
<td>Illustration of Thresholds Underlying Ordered Categorical Items</td>
<td>36</td>
</tr>
<tr>
<td>Figure 11</td>
<td>RSM ORFs for Hypothetical Five-Point Likert Item</td>
<td>46</td>
</tr>
<tr>
<td>Figure 12</td>
<td>Composite IRT Modeling Approach</td>
<td>55</td>
</tr>
<tr>
<td>Figure 13</td>
<td>Consecutive IRT Modeling Approach</td>
<td>56</td>
</tr>
<tr>
<td>Figure 14</td>
<td>Multidimensional IRT Modeling Approach</td>
<td>58</td>
</tr>
<tr>
<td>Figure 15</td>
<td>Categorical Latent Class Variable Conceptualization</td>
<td>63</td>
</tr>
<tr>
<td>Figure 16</td>
<td>LCA Model with Two Observed Indicators</td>
<td>64</td>
</tr>
<tr>
<td>Figure 17</td>
<td>November 2011 CMV Administration Written Instructions</td>
<td>77</td>
</tr>
<tr>
<td>Figure 18</td>
<td>September/October 2012 LMI Element Leader Self-Rating Administration Written Instructions</td>
<td>78</td>
</tr>
</tbody>
</table>
Figure 39  LMI Self-Rating Consecutive Model Specification ........................................ 157
Figure 40  LMI Self-Rating Multidimensional Model Specification ........................................ 158
Figure 41  LMI Self-Rating Six-Factor Modified Multidimensional Model Specification ........................................ 163
Figure 42  LMI Self-Rating Six-Factor Post Hoc Multidimensional Model Specification ........................................ 164
Figure 43  LMI Self-Rating Six-Factor Post Hoc Multidimensional Model Latent Variable Item-Person Map ........................................ 168
Figure 44  LMI Self-Rating Six-Factor Post Hoc Multidimensional Model Latent Variable Item-Person Map ........................................ 170
Figure 45  LMI Subordinate-Rating Composite Model Specification ........................................ 171
Figure 46  LMI Subordinate-Rating Consecutive Model Specification ........................................ 172
Figure 47  LMI Subordinate-Rating Multidimensional Model Specification ........................................ 174
Figure 48  LMI Subordinate-Rating Six-Factor Modified Multidimensional Model Specification ........................................ 179
Figure 49  LMI Subordinate-Rating Six-Factor Post Hoc Multidimensional Model Specification ........................................ 180
Figure 50  LMI Subordinate-Rating Six-Factor Post Hoc Multidimensional Model Latent Variable Item-Person Map ........................................ 184
Figure 51  LMI Subordinate-Rating Six-Factor Post Hoc Multidimensional Model Latent Variable Item-Person Map ........................................ 186
Figure 52  CMV LCA Model ........................................ 188
Figure 53  CMV LCA Model Specification ........................................ 190
Figure 54  CMV LCA Replication Model Specification ........................................ 191
Figure 55  Conditional Probability Profiles of Endorsing “Like Me” for 3-Class CMV LCA Model ........................................ 197
Figure 56  LMI Self-Rating LCA Model ........................................ 199
Figure 57  LMI Self-Rating LCA Model Specification .................................................................200
Figure 58  LMI Self-Rating LCA Replication Model Specification.................................201
Figure 59  Conditional Probability Profiles of Endorsing “Like Me” for 3-Class LMI Self-Rating LCA Model ..............................................................................................................206
Figure 60  LMI Subordinate-Rating LCA Model .................................................................207
Figure 61  LMI Subordinate-Rating LCA Model Specification ..............................................209
Figure 62  LMI Subordinate-Rating LCA Replication Model Specification..............210
Figure 63  Conditional Probability Profiles of Endorsing “Like the Leader” for 5-Class LMI Subordinate-Rating LCA Model .................................................................215
Figure 64  LMI Self-Rating Confirmatory LCA 3-Class Model Specification ..........217
Figure 65  LMI Subordinate-Rating Confirmatory LCA 5-Class Model Specification .................................................................................................................................220
Chapter One: Introduction and Review of the Literature

“The vision of the United States Air Force Academy is to be the Air Force’s premier institution for developing leaders of character.” (U. S. Air Force Academy, 2010, p. 6)

This introduction discusses the theoretical underpinnings for the development of two new measures of the Air Force Academy’s “leader of character” definition and articulates the need for additional psychometric analyses to strengthen the rigor of the assessment of the latent factor structures. The review of the literature that follows explains the theory and application of confirmatory factor analysis, item response theory, and latent class analysis in assessing the latent factor structures of new measures in social science research. Current practices using each technique are highlighted with an example of an application.

Background

In the last several decades, the military service academies of the United States have created educational curricula, programs, and experiences to accelerate student character and leadership development, rather than relying on maturation alone, in order to better prepare their graduates to navigate the increasingly complex and uncertain global security environment (Adamshick, 2010; Bonadonna, 2010; Jackson, Lindsay, & Coyne, 2010; Shambach & Jackson, 2010; Sweeney & Fry, 2012; Turner, DeBos, & Licameli, 2010). These efforts to increase developmental opportunities are in direct response to the continued revelations of moral failures of leaders in business, athletics, politics, religion,
education, and the military. According to Klann (2007), lapses in character are nothing new for leaders, but “what is disturbing is the current frequency of failures, the range and depth of their impact, and their span across virtually every type of business and occupation” (p. vii). Having the ability to measure a student’s capacity to develop the character traits and leadership skills necessary to curb future moral failures through resolute moral actions will pay dividends toward the productivity, climate, and reputation of the graduate’s future professional organization (Klann, 2007).

Before examining two specific measures of undergraduate character and leadership development, an understanding of the integration of the two concepts is necessary for contextualization. Although the scholarship on moral development and leadership development is vast, much of the current literature addresses this integration with a call for the growth of not just effective leaders, but instead ones who lead with a moral authority based on specific character traits essential in navigating today’s complex society (Greenleaf, 2002; Hannah & Avolio, 2010; Jennings & Stahl-Wert, 2003; Klann, 2007; Peterson & Seligman, 2004).

Although much of the current literature addresses this relationship between character and leadership, the United States Air Force (USAF) has been integrating the two into the development of its personnel since its inception in 1947. From the very beginning, according to Brown (2002), leadership for Airmen was developed differently from their Soldier, Sailor, or Marine counterparts. In an effort to establish a unique identity distinctive from the Army Air Corps, the newly created USAF published Air Force Manual (AFM) 35-15 titled Leadership which Brown states “abandoned the
Army’s standard ‘traits and principles’ approach to the presentation of leadership doctrine in favor of a more nuanced interpretation of the leader’s role in a military unit” (p. 38). In doing so, AFM 35-15 published in 1948 described seven aspects of leadership—mission, integrity of character, responsibility, influencing men, knowing men, unity, and morale—while also listing six attributes of a leader—integrity of character, sense of responsibility, professional ability, energy, emotional stability, and humaneness (Department of the Air Force, 2011). According to Brown, these psychological aspects and attributes were paramount in fostering collaborative relationships between the leader and the follower in accomplishing the USAF’s highly sophisticated and technical missions, instead of trusting solely on the legal authority grounded in historical hierarchical command structures.

This collaborative approach, chronicled since the publishing of AFM 35-15, resulted in the requirements, theory, and lessons learned about leading others in the physical domains of air, space, and cyberspace of the 21st century. Air Force Doctrine Document (AFDD) 1-1 articulates the theoretical and intellectual underpinnings for today’s USAF declaration on Airmen leadership and leader development.

AFDD 1-1 explicates the Service’s three core values that fundamentally define the Airman identity—Integrity First, Service Before Self, and Excellence in All We Do—which trace back to AFM 35-15’s leadership aspects and attributes (Department of the Air Force, 2011). USAF core values are “a statement of those institutional values and principles of conduct that provide the moral framework for military activities” (Department of the Air Force, 2011, p. 9). Each USAF core value is a compilation of
virtues or moral attributes which assist in their universal and unchanging understanding and application (Department of the Air Force, 2006). Figure 1 lists each USAF core value and maps their respective virtues.

---

**Figure 1.** USAF Core Values and their Respective Virtues (Department of the Air Force, 2006, pp. 5-8).

Moreover, these core values are supported with enduring institutional leadership competencies and sub-competencies at the personal, team, and organizational level (Department of the Air Force, 2011). According to AFDD 1-1, an Airman is developed over time to lead with varying degrees of these competencies at three different levels—tactical expertise (i.e., personal leadership), operational competence (i.e., team leadership), and strategic vision (i.e., organizational leadership)—and it is the primary level at which the Airman functions which determines the necessary leadership competencies for mission success (Department of the Air Force, 2011). Figure 2 lists the
USAF enduring institutional leadership competencies and sub-competencies by leadership level.

![Figure 2](image_url)

While USAF leadership doctrine is focused toward officers, enlisted personnel, and civilian personnel, the United States Air Force Academy (USAFA) targets its mission “to educate, train, and inspire men and women to become officers of character motivated to lead the United States Air Force in service to our Nation” (U.S. Air Force Academy, 2010, p. 2) solely on future officers. USAFA’s Officer Development System (ODS) provides all Academy stakeholders with an integrated character-based framework consisting of three levels: 1) theoretical officer identity foundations, 2) institutional
outcomes, and 3) educational and training opportunities (U. S. Air Force Academy, 2008).

An examination into the components of the first ODS level (i.e., the theoretical officer identity foundations), consisting of the Constitution, the Oath of Office, Core Values, and Officership sheds further light on the USAF’s necessary and continued practice of the integration of the relationship between character and leadership (U. S. Air Force Academy, 2008). This foundation establishes an inspirational and enduring footing upon which each future officer develops a character-based identity—one that is willing to make the ultimate sacrifice in service to the nation (U. S. Air Force Academy, 2008). According to the ODS guide, “the Constitution provides the philosophical foundation; the Oath of Office affirms one’s commitment to this core set of ideals, while the Core Values guide all Airmen” (U. S. Air Force Academy, 2008, p. 4). Furthermore, the final component of the ODS foundation is Officership, whose sub-components contain the four AFDD 1-1 primary responsibilities of an Air Force officer—warfighter, servant of the Nation, member of the profession of arms, and a leader of character (Department of the Air Force, 2011; U. S. Air Force Academy, 2008). Figure 3 depicts the components and sub-components of the ODS foundational construct.
Whereas the Academy’s ODS was created and implemented by USAFA’s Directorate of Strategic Plans and Programs, Requirements, Assessments and Analyses (i.e., equivalent to a higher education institutional research office), the Center for Character and Leadership Development (CCLD) operates primarily across the academic/athletic departmental and military training (i.e., equivalent to a higher education school) levels. The vision of the CCLD is to become the “Air Force’s premier Center for integrating the development of character and leadership; the Academy’s catalyst for achieving USAFA’s highest purpose” (Center for Character and Leadership Development Online, 2012, “Vision,” para. 1). While pursuing this vision, the CCLD synthesized the constructs of AFDD 1-1 and ODS described above and expanded the notion of integrating character and leadership by incorporating evidence-based practices.

*Figure 3. Officer Development System Foundation (U. S. Air Force Academy, 2008).*
supported by theory from a wide variety of social science disciplines to create a conceptual framework for developing leaders of character (Sanders, 2011). In this five-part conceptual framework, the CCLD clearly articulates a comprehensive approach to developing leaders of character by 1) defining a leader of character, 2) inspiring cadets to take ownership of their own development, 3) offering a model of engagement in purposeful experiences, 4) providing a developmental model based on awareness, reasoning, and decision-making for cadets to practice exercising moral actions (i.e., the Awareness Reasoning Deciding Acting (ARDA) model), and 5) introducing four alignment mechanisms for institutional implementation of this conceptual framework (Sanders, 2011).

While AFDD 1-1, the ODS, and the CCLD’s conceptual framework each provide constructs involving the integration of character and leadership at various organizational levels, it is critical to understand exactly what a “leader of character” is in order to facilitate measurement of the trait. AFDD 1-1 states that a military officer has a responsibility as “a leader of character” but fails to define the term (Department of the Air Force, 2011, p. 4). The ODS guide broadly defines a “leader of character” as one who demonstrates “moral excellence reflected in their values and behavior… sets a personal example for all, whether in their units, organizations, or society…seeks to discover the truth, decides what is right, and then demonstrates the courage to act accordingly” (U. S. Air Force Academy, 2008, p. 18). Based on scholarship ranging from Greek moral philosophy to the current literature on character and leadership development, only the CCLD’s conceptual framework succinctly defines an Air Force
“leader of character” as someone who “lives honorably by consistently practicing the virtues embodied in the Air Force Core Values, lifts people to their best possible selves, and elevates performance toward a common and noble purpose” (Sanders, 2011, p. 9).

In an effort to clearly articulate the concepts in the definition, the CCLD’s conceptual framework provides a rationale for each phrase. The first phrase, “lives honorably,” is explained as:

The term “live honorably” has significant meaning and saliency to all Airmen—indeed to all men and women who serve in the military. We are bound by a code of behavior that defines our chosen profession. These standards bind and define us, and falling short of these high standards tarnishes the noble profession to which we have committed ourselves. In other words, from the moment of our oath of office, it becomes our responsibility to honor those who have come before us, especially those who have paid the ultimate price in service to our nation. In short, living honorably means committing ourselves to live by certain standards of behavior—standards that do not (necessarily) bind those outside the military. Notably, the concept is also an essential aspiration of the Cadet Wing Honor Oath: “I resolve to do my duty and to live honorably, so help me God.” (Sanders, 2011, p. 9)

The second phrase, “by consistently practicing the virtues embodied in the Air Force Core Values,” is expanded to mean:

Living honorably extends far beyond mere compliance with the technical and legal requirements of our commission. Instead, living honorably means understanding and consistently practicing the virtues essential to the core values of the military profession. Duty. Respect. Courage. These are just some of the virtues that define the military officer. There are also the virtues that enable us to practice the habits of integrity (honesty, fairness) and the virtues necessary to put “service before self” (self-sacrifice, humility). Being committed to a military career also demands that we strive to embody excellence in every facet of our character and conduct. Moreover, it becomes our responsibility to know what virtues are needed in a particular situation—and then exhibiting and modeling the competence and confidence to “do the right thing.” In short, living honorably means consistently “living the virtues” embodied in the Air Force Core Values. The Cadet Wing Honor Code also speaks to the essential role of habits (“consistently practicing the virtues”) affirming that “making the right decisions all the time no matter how seemingly insignificant the issue, will build a habit of
**honorable behavior** that will be with you when times are tough” (emphasis added). Significantly, there is growing research that suggests at the core of developing habits is keeping or honoring “one’s word.” (Sanders, 2011, pp. 9-10)

The third phrase, “lifts people to their best possible selves,” is described as:

There is growing recognition that the “best possible self” concept is integral to our development as leaders. The concept is steeped in the transformational leadership theory and has been developed over the past two decades by researchers interested in how the repertoire of our “possible selves” provide the meaning, organization and direction through which we set our goals and aspirations (as well as how we face our fears and threats). The concept of the ‘best possible self’ connotes that each one of us has the capacity to pursue the “best” of who we are (or want to become). One team of researchers summed up the potential and promise of the ‘best self’ concept when they wrote that the self-images, goals and aspirations of our ‘best selves’ serve as “both an anchor and a beacon, a personal touchstone of who we are and a guide for who we can become.” At USAFA, the challenge is how to provide ample opportunities for cadets to “lift others” in ways that optimize individual (and team) performance. For example, Sanders and his colleagues suggest that leaders “have the fundamental capacity to care about others, their feelings, and motives in such a way as to have a positive influence on followers. This concept is especially critical in the Cadet Wing where upperclassmen have a responsibility to develop themselves as role models as well as a responsibility to develop the cadets under their supervision. In sum, just as the Wright brothers were pioneers in trying to understand aeronautical lift—our 21st century vision at CCLD is that our cadets will begin to see themselves as pioneers in the discovery of human lift, the capacity to be laser-focused on mission and purpose while simultaneously having the ability to recognize, support and “lift” the strengths, passions and commitments of those around them. (Sanders, 2011, pp. 10-11)

The fourth phrase, “elevates performance,” is explicated as:

Historically, leaders “get things done” (e.g., accomplish objectives) by influencing others. Yet, at the Air Force Academy, a leader of character goes beyond simply “getting things done” to finding ways—large and small—to enhance and transform *how* things are done. In other words, leaders are always *striving*—they never simply rest on their laurels, rank, or current level of capability. The most outstanding leaders are always growing, developing, and searching for new ways to expand their capacities (and their mission) beyond the minimum standard of expected performance. (Sanders, 2011, p. 11)

The final phrase, “toward a common and noble purpose,” is clarified as:
Finally, our definition explicitly addresses why a leader engages in the exercise of leadership. We use the term “noble purpose” to denote that not all commitments are alike (indeed, some purposes and commitments are blatantly unethical). We are suggesting that a commitment is noble in the sense that it extends beyond one’s own narrow self-interest, and focuses instead on the common good (at the level of the squadron, Air Force, or world). These sorts of commitments enable us to experience (cognitively as well as emotionally) that there are important ideals and principles in the world that are right to care about. (Sanders, 2011, p. 11)

With this definition in place, the engagement model of development component of the CCLD’s conceptual framework contends that it is essential to assess, challenge, and support the development of a leader of character (Sanders, 2011). In this framework, three practices are necessary: 1) assessments are intentional in understanding a cadet’s strengths as well as future developmental opportunities, 2) cadets are motivated to change after being challenged (e.g., receiving measured feedback highlighting disparities between actual and desired ability levels), and 3) organizational support (e.g., coaching and mentoring) for each cadet is necessary in order to develop a leader of character (Sanders, 2011).

In an effort to place the engagement model of development component of the conceptual framework into practice (i.e., assess, challenge, and support), the CCLD created two new assessment instruments based on their “leader of character” definition: 1) the Character Mosaic, and 2) the Leadership Mosaic. Analyses of both instruments result in personalized reports for each cadet respondent to share with their respective developmental coaches in order to be challenged and supported beyond their self-perceived capabilities (Rosebush, 2011; Rosebush, 2012).
The first of these new assessments, the *Character Mosaic*, is a composite instrument designed to measure the first phrase of the “leader of character” definition and consists of two primary components: 1) a newly developed measure of a cadet’s self-identified strengths regarding nine virtues related to the Air Force Core Values, and 2) two previously developed and validated measures (e.g., the Defining Issues Test and the Character Behaviors and Acceptability Questionnaire) of a cadet’s ability to apply each of the ARDA model elements in various situations (Rosebush, 2011). The remainder of this study focused on the first component’s newly developed “virtues scale” and is referred to as the Character Mosaic Virtues (CMV).

The CMV was initially developed to measure a cadet’s self-identified strengths regarding the twenty virtue components of the USAF core values listed in AFDD 1-1 and illustrated in Figure 1. A majority of the items in the initial item pool were incorporated with permission from similar scales represented in the “Values in Action Survey” (VIA) (Peterson & Seligman, 2004) and from the public domain “International Personality Item Pool” (IPIP) (International Personality Item Pool Online, 2012; Rosebush, 2011). A validation study of the scale development process which included a developmental sample (*n* = 626) of first-year (i.e., basic) cadets, an exploratory factor analysis (EFA), and evidence for strong reliability and varying degrees of validity produced the CMV with nine final factors and forty-five items after a June 2011 administration (Rosebush, 2011). Figure 4 illustrates the nine independent CMV factors classified under each of the core values and matched with their definition and item allocations. Appendix A lists the rating scale and the nine CMV virtues with their corresponding items.
The second of these new assessments, the Leadership Mosaic, is an instrument designed to measure the relationship between a leader and his/her followers and consists of two primary components: 1) two criterion variables designed to directly measure the “lifts others” and “elevates performance” criteria of the “leader of character” definition, and 2) a leadership effectiveness scale that best explains the variance in the criterion variables (Rosebush, 2012). The remainder of this study focused on the second component’s newly developed “leadership effectiveness scale” and is referred to as the Leadership Mosaic Inventory (LMI).
The LMI, completed by the leader and his/her immediate subordinates, was developed to measure a cadet element leader’s effectiveness and provide a 180-degree feedback tool in order for the element leader to be challenged and supported by their developmental coach (Rosebush, 2012). At USAFA, an element is the lowest organized cadet grouping for command, usually consisting of about four to five first- or second-year cadets who report to a third-year (i.e., junior) cadet leader; with nine elements included in each of the 40 cadet squadrons, there are 360 element leaders in total. The single-factor LMI measures leadership effectiveness by incorporating six of AFDD 1-1’s institutional leadership sub-competencies listed in Figure 2 to include: 1) develops and inspires others, 2) takes care of people, 3) builds team and coalitions, 4) negotiating, 5) vision, and 6) adaptability.

While a majority of the items in the initial item pool were incorporated from twenty-two different leadership scales representing eight prevalent leadership theories (e.g., transformational, servant, toxic, authentic, virtuous, developmental, empowering, and organizational), the final six non-independent LMI areas represent theoretical leadership effectiveness constructs which exhibit face validity with AFDD 1-1’s institutional leadership sub-competencies (Rosebush, 2012). According to Rosebush (2012), the positively worded items were reallocated by face validation into conceptual sub-competencies by means of the following leadership effectiveness scales and supporting theory: 1) the Transformational Leadership Behavioral Scale (Podsakoff, MacKenzie, & Bommer, 1996) and the Developmental Leadership Questionnaire (Larsson, 2006) based on transformational leadership theory; 2) the Servant Leadership
Questionnaire (Barbuto & Wheeler, 2006) based on servant leadership theory; 3) the Character Mosaic Inventory (Rosebush, 2011) based on virtue-based leadership theory; 4) the Empowering Leadership Questionnaire (Arnold, Arad, Rhoades, & Drasgow, 2000) based on empowering leadership theory; and 5) pilot-test items on leadership adaptability (M. Rosebush, personal communication, September 21, 2012) based on several leadership models.

Each of these theoretical positions is supported in the current literature. For example, transformational leadership theory includes the fundamental practices of inspiring a shared vision, enabling others to act, and encouraging the heart (i.e., rewarding others) (Avolio, Walumbwa, & Weber, 2009; Northouse, 2013)—constructs measured in the LMI’s vision, develops and inspires others, and takes care of people theoretical underpinnings, respectively. Servant leadership includes functional (e.g., vision, demonstrating appreciation of others, empowerment, etc.) and accompany (e.g., good communicators and listeners, encouraging of others, teachers, delegators, etc.) attributes (Avolio, Walumbwa, & Weber, 2009) which are measured in the LMI’s vision, takes care of people, builds team and coalitions, and negotiating items. Virtue-based leadership theory originated in ancient Greek tradition and is associated with the Greek term aretaic meaning “excellence” (Northouse, 2013); this notion of ensuring excellent quality of work is included in the LMI’s develops and inspires others sub-competency. Empowering leadership theory’s “coaching” leader behavior classification includes those actions which educate teams and promote self-reliance (Arnold et al., 2000)—items from this subscale were matched to the LMI’s develops and inspires others, builds teams and
coalitions, and vision sub-competencies. Lastly, support for the LMI adaptability sub-competency may be found in the Skills Model of Leadership, in the Situational Leadership model, in team leadership theory, in complexity leadership theory, and as a positive psychological capacity which influences the Authentic Leadership model (Avolio, Walumbwa, & Weber, 2009; Northouse, 2013).

A validation study of the scale development process, which included a developmental sample \( n = 223 \) of first- and second-year cadets with their third-year element leaders, an EFA, a linear regression on the criterion variables in which the LMI explained 79% and 86% of the variance in “lift others” and “elevates performance” respectively, and evidence for strong reliability and varying degrees of validity, produced the LMI with a single factor solution and twenty-nine items after a March 2012 administration (Rosebush, 2012). Figure 5 illustrates the single LMI “leadership effectiveness” factor classified under the “lifts others, elevates performance” definition with item allocations based on the USAF institutional sub-competencies. Appendix B lists the rating scale and the LMI’s twenty-nine items with corresponding non-independent USAF institutional sub-competencies.
Figure 5. Leadership Mosaic Inventory Single Factor and Item Allocations Based on USAF Institutional Sub-Competencies (Rosebush, 2012, pp. 14-15).

Statement of the Problem

Data from the measurement of unobserved constructs (i.e., latent factors) such as character virtues and leadership effectiveness warrant rigorous analysis and substantive interpretation after newly developed scales have been administered. DeVellis (2003) cautions, “the validity of a scale is not firmly established during scale development...validation is a cumulative, ongoing process” (p. 159). Furthermore, DeVellis advises to analyze scales of ordered categorical response formats (e.g., Likert items) with interval-based methods.

These two issues apply to the psychometric evaluations of the CMV and LMI scales. While initial validation studies culminating in exploratory analyses of the factor structures on both measures have been completed, applying more sophisticated analytical
techniques provided stronger evidence for validity resulting in rigorously substantiated instruments.

Confirmatory factor analysis (CFA), originating from classical test theory (CTT), is one such psychometric evaluation technique capable of providing additional evidence to strengthen validity claims, especially regarding construct validity, even with ordinal outcomes. While parameter estimation of Likert scale items with most analytical software’s default maximum likelihood (ML) technique is not appropriate, special CFA methods involving robust weighted least squares (WLS) estimation are available in at least one software package (Kline, 2011). CFA is necessary to confirm previous EFA results or predicted factor structure based on theory, to provide statistical criteria regarding model fit to real data, to test and compare alternative models to the data, and to determine the dimensionality of measurement (DeVellis, 2003; Kline, 2011).

On the other hand, analysis of item-level data with CFA can be problematic. Unlike EFA where each item loads on every factor and some secondary loadings also account for a proportion of the variance, the more restrictive CFA model constrains secondary loadings to zero which may result in model misfit based on EFA factor structure (Kline, 2011). Furthermore, Kline states that a linear relationship between the items and their factors is assumed in CFA measurement models, which may not actually be the case. According to Kline, “in some situations, other statistical methods for item-level analyses are better alternatives than CFA” (p. 244).

One such alternative to CFA in the analysis of item-level data is the sophisticated nonlinear approach of creating item characteristic curves (ICC) through item response
theory (IRT). Regarded as “modern test theory,” IRT is a method for analyzing the underlying latent factor structure in new measures due to its ability to overcome certain limitations of CTT (Osteen, 2010). One of these CTT limitations evident in EFA, treating ordered categorical (i.e., Likert scale) data as continuous, is overcome with IRT’s logarithmic transformation of the data into a monotonically increasing ICC which estimates the probability of a response based on the amount of a latent trait (Osteen, 2010). By analyzing the ICC, individual item difficulty is computed as the amount of the latent trait required for a person to give a particular response to an item according to the rating scale (Osteen, 2010). Advances in software have extended IRT analyses to the exploration of the underlying dimensions of competing models through multidimensional IRT (MIRT).

Another interval-based technique, which is an analog to factor analysis except for its capacity to analyze models with both categorical indicators and latent variables, is latent class analysis (LCA). According to Kline (2011) mixtures of subpopulations, called classes of the categorical latent variable in which membership is unobserved, can be inferred from the data. For example, results of a LCA might reveal: 1) what makes one cadet more likely than another to have a certain mixture of the character virtues described above, or 2) in the absence of a “gold standard” scoring system, what is a cutoff point for belonging to a group having the latent trait of “leadership effectiveness.” Put simply, the goal of LCA is to uncover the meaning and number of underlying subpopulations or latent classes (Kline, 2011).
Wang and Hanges (2011) ascribe that dimensional approaches, such as factor and IRT analyses which focus on the interrelatedness of the observed variables under a latent factor to confirm underlying constructs, represent only one way of assessing latent factor structures. Wang and Hanges assert “an equally valid and complementary way is to consider the interrelatedness between or among different variables as a function of the unobserved heterogeneity of the population” (p. 24) through the application of LCA. While the stated goals of the CMV and LMI scales are to provide individual cadet assessment in order to be challenged and supported by their developmental coaches, creation of latent class models complements the more traditional dimensional approaches of structure assessment with an understanding of unobserved cadet subpopulations in which targeted developmental opportunities may be applied at organizational levels such as cadet squadrons or groups deficient in certain areas.

**Purpose of the Study**

The purpose of this research study was to evaluate the latent factor structures of the CMV and LMI scales through confirmatory factor, item response theory, and latent class analyses. The study aimed to leverage the complementary nature of the three analyses to strengthen the rigor and sophistication of evaluation of the newly developed scales. Additionally, this study purported to benefit the fields of moral development and leadership development, especially at the nation’s service academies, by exposing more researchers, decision-makers, and other stakeholders to these three advanced psychometric evaluation methods.
Research Questions

The following six questions were addressed by this research using CMV and LMI datasets gathered subsequent to the developmental samples:

1. Does analysis of CMV data using CFA techniques yield a dimensional structure consistent with Rosebush’s (2011) EFA results? Does analysis of CMV data using CFA techniques demonstrate desirable psychometric properties of acceptable model fit, construct reliability, and construct validity?

2. Does analysis of LMI data using CFA techniques yield a dimensional structure consistent with Rosebush’s (2012) EFA results? Does analysis of LMI data using CFA techniques demonstrate desirable psychometric properties of acceptable model fit, construct reliability, and construct validity?

3. Does analysis of CMV data using IRT techniques yield a dimensional structure consistent with the CFA results? Does analysis of CMV data using IRT techniques demonstrate desirable psychometric properties of acceptable model fit, item fit, and reliability?

4. Does analysis of LMI data using IRT techniques yield a dimensional structure consistent with the CFA results? Does analysis of LMI data using IRT techniques demonstrate desirable psychometric properties of acceptable model fit, item fit, and reliability?
5. Does analysis of CMV data using LCA techniques yield a cutoff point for classifying cadet subpopulations as either having a virtuous character or not? What combinations of Character Mosaic virtue endorsements in the CMV data distinguish cadets who have a virtuous character versus those who do not?

6. Does analysis of LMI data using LCA techniques yield a cutoff point for classifying cadet element leader subpopulations as either being an effective leader or not? Does analysis of LMI data using LCA techniques yield a cutoff point for classifying cadet subordinate subpopulations who view their element leaders as either being an effective leader or not? What combinations of USAF institutional leadership sub-competency endorsements in the LMI data distinguish cadets who are effective leaders versus those who are not?

**Confirmatory Factor Analysis**

Most constructs in the social sciences, such as character virtues and leadership effectiveness, cannot be measured directly since they are not observable. Instead, researchers create measures using items as proxies to reflect the underlying phenomenon called a latent variable (DeVellis, 2003). In order to represent the large number of relationships among the observed items in a more parsimonious way, researchers conduct factor analyses to identify or confirm a reduced set of latent variables which underlie the represented items (Gliner, Morgan, & Leech, 2009).
**Exploratory versus confirmatory.** Factor analytic techniques are divided between exploratory and confirmatory methods (DeVellis, 2003). EFA is normally conducted in the early stages of the scale development process as a method of data reduction by grouping correlated variables together to formulate hypotheses about the underlying constructs (Tabachnick & Fidell, 2007). On the other hand, CFA is routinely performed in the advanced stages of the scale development process through structural equation modeling (SEM) to statistically test the significance of a theoretical underlying latent variable construct (Schumacker & Lomax, 2010; Tabachnick & Fidell, 2007). Put simply, EFA is to theory development as CFA is to theory testing (Tabachnick & Fidell, 2007).

Another key difference between EFA and CFA which determines their appropriateness of use in the scale development process is CFA’s requirement for model identification. According to Byrne (2012), a CFA model is identified when a unique solution for every parameter value is computable—in this case there is no need to rotate the solution, as is usual in EFA, to clarify interpretation. Identification in CFA permits the parameters to be estimated and the model to be statistically tested to confirm the theoretical construct. Conversely, EFA models are generally not identified as they do not produce a unique set of parameter estimates since their solutions may be rotated infinitely in search for a solution to formulate a hypothesis about the underlying structure of the data (Kline, 2011).
Both EFA and CFA share a common origin from CTT in which three assumptions guide the relationship between items, latent variables, and sources of errors. DeVellis (2003) lists these three assumptions as:

1. The amount of error associated with individual items varies randomly. The error associated with individual items has a mean of zero when it is aggregated across a large number of people. Thus, items’ means tend to be unaffected by error when a large number of respondents complete the items.

2. One item’s error term is not correlated with another item’s error term; the only routes linking items pass through the latent variable, never through any error term.

3. Error terms are not correlated with the true score of the latent variable. Note that the paths emanating from the latent variable do not extend outward to the error terms. The arrow between an item and its error term aims the other way. (p. 20)

In addition to the three assumptions listed above, Kline (2011) describes two additional characteristics of a standard CFA model to include: 1) each indicator (i.e., item) is caused by either the single factor which it measures or by other sources represented by an error term, and 2) all between-factor associations are not analyzed and are thus allowed to covary. With each observed indicator in the standard CFA model having two causes—a score measuring a single factor and a random error component—consistency with the view of CTT is achieved such that observed score is the sum of true score and error (Gliner et al., 2009).

**Dimensionality of measurement.** The standard CFA model described above in which items load on a single latent factor—referred to as congeneric measurement (Jöreskog, 1971)—and with error terms independent of each other (i.e., their observed correlation is explained by their factor) is representative of **unidimensional measurement** (Kline, 2011). On the other hand, nonstandard **multidimensional measurement** is
modeled by allowing items to load on two or more factors or by permitting error terms to be correlated with other error terms (Kline, 2011).

An advantage of unidimensional measurement is greater precision regarding tests of convergent and divergent validity. For example, evidence for convergent validity is determined by the statistical significance of each item’s estimated loading on its hypothesized underlying factor (i.e., parameter estimates greater than two times its standard error) (Anderson & Gerbing, 1988). Furthermore, Kline (2011) describes construct measurement reliability ($\rho_\eta$) as the ratio of explained variance to total variance calculated amongst CFA parameters as another method to evaluate convergent validity. With $\lambda$ representing the subscale’s standardized factor loadings and $\epsilon$ representing the subscale’s item measurement error, the factor rho coefficient provides additional evidence for convergent validity with calculated values of 0.70 or greater through the application of Equation 1 (Fornell & Larcker, 1981):

$$\rho_\eta = \frac{(\Sigma \lambda)^2}{(\Sigma \lambda)^2 + \Sigma \epsilon}$$

(1)

Fornell and Larcker also recommend another more conservative test, average variance extracted (AVE), to capture the amount of variance in the construct related to the amount of variance due to measurement error. AVE ($\rho_{ve(\eta)}$) may be calculated as in Equation 2:

$$\rho_{ve(\eta)} = \frac{\Sigma (\lambda^2)}{\Sigma (\lambda^2) + \Sigma \epsilon}$$

(2)

While having an AVE in excess of 0.50 is ideal (i.e., the variance accounted for by the construct is greater than the variance due to measurement error), less stringent evidence
for convergent validity may be established on the basis of construct reliability alone (Fornell & Larcker, 1981).

Additionally, Fornell and Larcker (1981) suggest that factor AVE should exceed the shared variance between each pair of factors, and may be used to evaluate discriminant validity. For example, Table 1 illustrates the AVE in bold on the diagonal with the shared variances below the diagonal for a hypothetical three factor model. In cases such as Factor 2 in Table 1 where factor AVE does not exceed shared variance (e.g., Factor 2’s AVE is 0.48 yet its shared variance with Factor 1 is problematic at 0.59), evidence of discriminant validity between two constructs may be assessed by constraining the phi parameter (i.e., estimated correlation, $\hat{\phi}_{ij}$) to 1.0 then conducting a chi-square difference test (Anderson & Gerbing, 1988). Using the constrained phi approach, constructs which are not perfectly correlated provide evidence for the presence of discriminant validity (Anderson & Gerbing, 1988).

Table 1

Hypothetical 3-Factor Model Convergent and Discriminant Validity Tests

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Construct Reliability</th>
<th>Subscale</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>0.90</td>
<td><strong>0.68</strong></td>
<td></td>
</tr>
<tr>
<td>Factor 2</td>
<td>0.84</td>
<td>0.59</td>
<td><strong>0.48</strong></td>
</tr>
<tr>
<td>Factor 3</td>
<td>0.86</td>
<td>0.31</td>
<td>0.24</td>
</tr>
</tbody>
</table>

*Note. Average variance extracted is shown in **bold** along the diagonal. Shared variances are shown below the diagonal.*

According to Kline (2011) however, *multidimensional measurement* adds complexity to the CFA modeling process. Multidimensional measurement is typically specified to either achieve the best empirically derived model fit solution without conceptual justification to serve as a baseline for comparison with other theory-driven
models (Osteen, 2010), or in post hoc analyses to respecify models to detect and improve misfitting parameters based on substantively meaningful modifications (Byrne, 2012). Theoretical justification for some items measuring more than one construct (e.g., an aptitude test with text and figures may measure both verbal and spatial abilities) or two items having correlated error terms sharing sources of variability beyond the underlying factors (e.g., autocorrelated errors in repeated measures variables) is necessary to substantiate multidimensional measurement for the purpose of testing the factorial validity of theoretical constructs (Byrne, 2012; Kline, 2011). If multidimensional measurement is substantiated, tests of convergent and divergent validity as described above are much more complex (Kline, 2011).

Figure 6 illustrates a typical standard unidimensional two-factor CFA model with six indicators in which ovals represent continuous latent variables, rectangles represent observed continuous indicator variables, and circles represent the measurement error (i.e., indicator variance not explained). Here, lines with arrowheads depict presumed direct effect paths while the curved line with an arrowhead depicts the unanalyzed association between factors. Additionally, the number “1” that is illustrated along the paths are scaling constants which assign a metric to the factors and error terms in order for SEM software to make statistical estimates (Kline, 2011).

Figure 7 illustrates a typical nonstandard multidimensional two-factor CFA model. The model contains six indicators in which conceptual justification exists for Item 4 loading on both Factor 1 and Factor 2, and Error 4 and Error5 representing shared associations beyond the factors.
Dimensionality of structure. An important distinction must be made between
dimensionality of structure and the dimensionality of measurement. While the latter
refers to the SEM specification of every relationship between the observed and latent
variables such that “the true population model is deemed consistent with the implied
theoretical model being tested” (Schumacker & Lomax, 2010, p. 55), the former refers to
the composition of latent factors in the overall construct (Byrne, 2012; Osteen, 2010).
For example, a unidimensional structure refers to a single latent factor with multiple
indicators measuring a single theoretical construct, while a multidimensional structure
refers to two or more latent factors (i.e., subscales) with multiple indicators measuring a
single overall construct.

According to de Ayala (2009), it is implicitly assumed in CTT that constructs are
unidimensional. To further explain this notion, de Ayala states:

For observed scores to have any meaning they need to represent the sum of
responses to items that measure the same thing. For instance, assume that an
examination consists of five spelling questions and five single-digit addition
problems. Presumably our examination data would consist of two dimensions
representing spelling and addition proficiencies. If a person had an observed
score of 5, it would not be possible to determine whether he or she is perfect in
spelling, perfect in addition, or in some combination of spelling and addition
proficiencies. In this case, the observed score has no intrinsic meaning. In
contrast, if the examination consists of only spelling questions, then the score
would indicate how well a person could spell the questions on the test and would
have intrinsic meaning. (p. 10)

Nonetheless, Briggs and Wilson (2003) approach this implicit assumption as a balance
between the science of measurement (i.e., unidimensionality) and the art of measurement.

With regard to the art, Briggs and Wilson explain:

In general, a latent domain can be deconstructed into subcomponents, and these
subcomponents can in turn be deconstructed, and so on until the number of latent
domains requiring estimation may well equal the number of items being
administered! In such a scenario when items are allowed to contribute to more
than one domain, the number of dimensions are no longer identifiable parameters.
The art of assessing dimensionality is to find the smallest number of latent ability
domains such that they are both statistically well-defined and substantively
meaningful. (p. 88)
Two common cases exist in which the assumption of unidimensionality is problematic: 1) instruments initially designed unidimensionally where results are interpreted multidimensionally, and 2) instruments specifically designed to measure several ability domains (Briggs & Wilson, 2003). An example of the former is the Stanford 9 mathematics test (Briggs & Wilson, 2003), which produces an overall score yet also provides subscale scores; an example of the latter includes the SAT I which provides scores on independent math and verbal sections while a summed score is often used as a single indicator of performance (Briggs & Wilson, 2003). To resolve the conflict with the problematic CTT assumption Briggs and Wilson assert, “when performance on an instrument has a multidimensional interpretation, then the proper modeling of these as separate, though not necessarily unrelated, dimensions is a prerequisite before a measure can be properly constructed” (p. 89).

**Five-step CFA modeling approach.** The processes found in the literature to conduct a CFA to assess the dimensionality of structure and factorial validity of a theoretical construct vary only slightly by author (Byrne, 2012; Kline, 2011; Schumacker & Lomax, 2010). This study incorporated Schumacker and Lomax’s (2010) five-step modeling approach—*specification, identification, estimation, testing*, and *modification*.

Model *specification* is considered the most important (Kline, 2011) and the most difficult aspect of SEM (Schumacker & Lomax, 2010). Specification is representing every hypothesized relationship between a construct’s observed and latent variables so that SEM software will estimate model parameters using sample data later in the process; therefore, a properly specified model is important since results from later steps in the
CFA process assume that the model is correct (Kline, 2011). The difficulty for the applied researcher, according to Schumacker and Lomax (2010), is to specify the theoretical model which sufficiently reproduces the sample variance-covariance matrix (or correlation matrix for categorical data (Byrne, 2012)) such that it is consistent with the true population model. Prior research, theories, and exploratory analyses in the scale development process provide the researcher with a plausible rationale for model specification.

Model identification is the a priori determination of the ability of the SEM software to derive a unique set of parameter estimates based on the sample variance-covariance matrix (or correlation matrix for categorical data) to be estimated by the theoretical model (Schumacker & Lomax, 2010). Kline (2011) states three general requirements for model identification: 1) “every latent variable (including the residual terms) must be assigned a scale (metric)” (p. 124), 2) “the model degrees of freedom must be at least zero \( (df_M \geq 0) \)” (p. 124), and 3) standard CFA models that specify unidimensional measurement must meet or exceed the minimum number of indicators per factor.

Since latent variables are not observed, the first model identification requirement above is necessary in order to assign a scale to each factor related to the shared variance of a reference item in each congeneric set of loadings (Kline, 2011); however, contrary to observed continuous indicators, the residual terms of observed ordered categorical items in CFA models using the WLSMV estimator (discussed below under model estimation) do not have to be scaled since their variances are not identified or estimated (Byrne,
2012). Once all latent variables are scaled, the SEM software will only need to estimate the variance/covariances for each factor (Kline, 2011).

Figure 8 illustrates the latent variable scaling of unstandardized factors in a two-factor CFA model with ordered categorical items. Here, the first factor loading is fixed to “1” in each congeneric set of items, and is referred to as unit loading identification (ULI) (Kline, 2011).

While unstandardized estimates are not very useful in interpreting CFA models with ordered categorical items (Byrne, 2012), the scaling as standardized factors by fixing their variance to “1” through unit variance identification (UVI) is useful in simplifying interpretation (interpretation of standardized estimates is discussed below under model testing) (Kline, 2011). For example, with UVI constraints all factor loadings are free to be estimated.

Figure 9 illustrates the latent variable scaling of standardized factors in a two-factor CFA model with ordered categorical items, with each factor variance fixed to “1.” The constraints described above, ULI for unstandardized factors and UVI for standardized factors, are the default settings in the Mplus SEM software used in this study (Muthén & Muthén, 2012b).
Figure 8. Unstandardized Two-Factor CFA Model with Ordered Categorical Items (Amos Version 18)

Figure 9. Standardized Two-Factor CFA Model with Ordered Categorical Items (Amos Version 18)
The second requirement for model identification, \( df_M \geq 0 \), is typically calculated by subtracting the number of distinct values available in the sample variance-covariance matrix by the number of free parameters to be estimated (Schumacker & Lomax, 2010). When the number of distinct values in the matrix equals the number of free parameters to be estimated, then the model is said to be just-identified with \( df_M = 0 \); when the former is greater than the latter, the model is said to be over-identified with \( df_M > 0 \) (Schumacker & Lomax, 2010). It is important to note, however, that a priori determination of identification status using Mplus and the WLSMV estimator for ordered categorical observed variables is beyond the scope of this study since \( df_M \) is estimated according to highly technical formulas found in the software technical appendices (Byrne, 2010; Muthén, 2004). Moreover, according to Muthén (2007), “the chi-square and degrees of freedom are adjusted to obtain a correct \( p \)-value with WLSMV…only the \( p \)-value should be interpreted” (msg. 24).

The final requirement to identify standard CFA models with unidimensional measurement specification involves the minimum number of items per factor. According to Kline (2011), a standard single factor CFA model is identified with at least three items, while CFA models with at least two factors must have at least two items per factor.

Model estimation is the process of obtaining estimates for each of the specified parameters by choosing a fitting function to minimize the difference between the implied population variance-covariance matrix (or correlation matrix for categorical data) and the sample variance-covariance matrix (or correlation matrix for categorical variables) (Schumacker & Lomax, 2010). While several fitting functions (e.g., unweighted least
squares, ordinary least squares, generalized least squares, maximum likelihood, arbitrary
distribution function) are available to the researcher in various SEM software packages
(Kline, 2011; Schumacker & Lomax, 2010), the most theoretically appropriate method
for estimating CFA models with ordered categorical items involves robust weighted least
squares (WLSMV) currently available only in Mplus (Byrne, 2012; Flora & Curran,
assumes that a continuous, normal latent process determines each observed variable” (p.
466) whose bivariate associations are estimated with polychoric correlations. These
polychoric correlations are estimates for what the association would be between
categorical variables with three or more response categories if they were both continuous
and normally distributed; similarly, tetrachoric correlations would be estimated for
dichotomous variables (Kline, 2011). In support of this type of estimation, Flora and
Curran’s simulation study revealed the “estimation of polychoric correlations is robust to
modest violations of underlying normality” (p. 466) and WLSMV estimation performed
equally well amongst sample sizes of 100, 200, 500, and 1000.

The model estimation process implemented in Mplus, with its assumption of an
underlying normally distributed, continuous scale for each categorical variable is
explained by Kline (2011) such that:

Each observed ordinal indicator is associated with an underlying latent response
variable, which is the underlying amount of a continuous and normally distributed
trait or characteristic that is required to respond in a certain category of the
corresponding observed ordinal item. When the observed indicator is
dichotomous, such as for items with a true-false response format, this amount, or
threshold, is the point on the latent response variable where one answer is given
(e.g., true) when the threshold is exceeded. It is also the point where the other
response is given (e.g., false) when the threshold is not exceeded. Dichotomous items have a single threshold, but the number of thresholds for items with ≥ 3 response categories is the number of categories minus one. Each latent response variable is in turn represented as the continuous indicator of the underlying substantive factor that corresponds to a hypothetical construct. (p. 180)

Figure 10 illustrates a normally distributed, continuous latent response variable that underlies a hypothetical ordered categorical agreement variable. The categories of the hypothetical agreement variable are converted to thresholds of the underlying continuous latent response variable based on the proportional responses in each category; the thresholds are z-scores calculated from a standardized normal table (Ullman, 2007).

![Figure 10. Illustration of Thresholds Underlying Ordered Categorical Items](image)

Fit of CFA models is affected by the level of nonnormality, sample size, and model size (Hutchinson & Olmos, 1998). According to Chaney et al. (2007), the robust WLSMV assumes a distribution that is neither extremely skewed or leptokurtic (i.e.,
positive kurtosis). Violation of this assumption, according to simulation studies, would result in positively biased chi-square model fit statistics due to negatively biased standard errors (Flora & Curran, 2004). While screening data, using a value of ±3.0 for skewness and kurtosis as a cutoff for determining whether to discard any additional items is suggested (Chaney et al., 2007). Another simulation study found that increasing nonnormality led to poorer fit for all fit indices except $\chi^2$ and RMSEA when using WLS estimation (Hutchinson & Olmos, 1998). Moreover, Hutchinson and Olmos (1998) found that RMSEA remained unbiased with regards to sample size or model size. According to Byrne (2012), the WLSMV estimator was specifically designed for use with small and moderate sample sizes in comparison to those required for the WLS estimator. Furthermore, Byrne iterates:

Simulation research related to the WLSMV estimator has shown it to yield accurate test statistics, parameter estimates, and standard errors under both normal and nonnormal latent response distributions across sample sizes ranging from 100 to 1,000, as well as across four different CFA models (one-factor models with 5 and 10 indicators, and two-factor models with 5 and 10 indicators). (p. 132)

Model testing, to determine the extent a hypothesized model fits the sample data, involves a two-step process: 1) determining the adequacy of global model-fit criteria, and 2) examining the fit of the estimated free parameters (Schumacker & Lomax, 2010). To address step one, Mplus provides the following global model-fit criteria when estimating CFA models with WLSMV—the chi-square test of model fit, Comparative Fit Index (CFI), Tucker-Lewis Fit Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Weighted Root Mean Square Residual (WRMR) (Muthén & Muthén, 2012b). Regarding step two, a CFA with Mplus using WLSMV provides estimates for
standardized factor loadings, factor variances/covariances, item thresholds, and reliability with associated standard errors, test statistics, and $p$-values (Muthén & Muthén, 2012b).

The chi-square test of model fit incorporates a corrected chi-square statistic due to the categorical and non-normal aspects of the sample data (Byrne, 2012). In CFA, however, model fit assessment is not as straightforward as finding a non-significant chi-square. For example, three instances exist which make an overall assessment of fit using the chi-square statistic unreasonable: 1) trivial differences between the sample and estimated population matrices can often be significant in large samples, 2) the computed chi-square statistic may not be distributed as chi-square in small samples resulting in biased probability levels, and 3) violation of assumptions underlying the chi-square statistic also lead to biased probability levels (Ullman, 2007).

While several different types of global fit indices accommodate the shortcomings of the chi-square statistic (e.g., incremental fit indices, absolute fit indices, indices of proportion of variance accounted, degree of parsimony fit indices, and residual-based fit indices (Ullman, 2007)), Mplus outputs CFI and TLI as incremental indices using the WLSMV estimator. Both CFI and TLI, according to Byrne (2012), “measure the proportionate improvement in model fit by comparing the hypothesized model in which structure is imposed with the less restricted nested baseline model” (p. 70). However, values for CFI are normed ranging from zero to one while values for TLI are non-normed and can fall outside the zero to one range (Byrne, 2012). Furthermore, TLI includes a penalty function based on model complexity; for example, TLI values are reduced when parameters minimally improve model fit (Byrne, 2012). A cutoff value greater than 0.95
for both CFI and TLI is considered good model fit using categorical outcomes (Byrne, 2012; Muthén, 2004; Schreiber, Stage, King, Nora, & Barlow, 2006; Yu, 2002).

RMSEA is an absolute fit index provided by Mplus output using WLSMV estimation that evaluates how well a hypothesized model with unknown, yet optimally selected, parameter estimates reproduce the sample data by taking into account the population error of approximation (Byrne, 2012). Unlike incremental fit indices which increase with model fit improvement, RMSEA decreases with model fit improvement (Byrne, 2012). A cutoff value less than 0.06 for RMSEA is considered good model fit using categorical outcomes (Muthén, 2004; Schreiber et al., 2006; Yu, 2002).

Additionally, Mplus computes a 90% RMSEA confidence interval, highlighting the inaccuracy of the point estimate. Byrne (2012) conveys the importance of the confidence interval by stating:

Presented with a small RMSEA, albeit a wide confidence interval, a researcher would conclude that the estimated discrepancy value is quite imprecise, thereby negating any possibility to determine accurately the degree of fit in the population. In contrast, a very narrow confidence interval would argue for good precision of the RMSEA value in reflecting model fit in the population. (p. 75)

WRMR is a residual-based fit index computed by Mplus using WLSMV estimation that evaluates the minimum of the WLS fitting function to derive a weighted residual value such that small values represent good fit (Muthén, 2004). A cutoff value less than 0.90 for WRMR is considered good model fit using categorical outcomes (Muthén, 2004; Schreiber et al., 2006; Yu, 2002). However, according to Muthén (2010a), WRMR is considered an experimental fit statistic and may be disregarded if all of the remaining global model-fit criteria are met (msg. 2).
In addition to global model-fit criteria, evaluation of individual parameter estimates is also necessary since the assessment of global fit indices alone cannot determine the adequacy of model fit. Each standardized factor loading estimate should have significant $p$-values, indicating the estimates are statistically different from zero; for example, a non-significant loading indicates that the item is not important to the model or that the sample size is too small (Byrne, 2012). With UVI scaling (i.e., the factor variances are standardized to 1.0), each standardized factor loading estimate for the categorical items must be squared in order to be interpreted as the proportion of the variance in the underlying continuous item explained by the factor upon which it loads (Byrne, 2012). *Mplus* outputs these proportion of the variance explained results in a reliability table—also included are the residual variances which are the proportion of the variance not explained in the underlying continuous items by their respective latent factors (Byrne, 2012).

Item threshold estimates specific to each categorical item are also computed by *Mplus* using WLSMV estimation. These thresholds are $z$-scores resulting from the conversion of the categorical items to underlying continuous variables based on the proportional responses in each category (Ullman, 2007). The number of thresholds per item will be one less than the number of response categories (Byrne, 2012).

The final step of the modeling process, model *modification*, has traditionally involved three *post hoc* applications: 1) removing non-significant parameters from well-fitting models (Bandalos & Finney, 2010), 2) testing alternative hypotheses (Ullman, 2007), and 3) adding parameters to increase fit when the hypothesized model is not
adequate (Schumacker & Lomax, 2010). Regarding the first application, Bandalos and Finney (2010) recommend against removing paths from a priori specified models based on not obtaining a certain level of statistical significance for the following reasons:

(1) using the same sample to respecify and test a modified model capitalizes on sampling error and thus decreases the chance of obtaining replicable results; (2) the model no longer aligns with theory but instead is empirically based or data driven; and (3) respecified models are often presented as though they were a priori theoretically based models, thus misleading readers as to the initial models specified and tested. (p. 111)

The second application, testing of alternative hypotheses, is supported by Bandalos and Finney who view the utility of CFA as advantageous when comparing a set of a priori alternative models because the researcher is empowered to make a more informed decision about the adequacy of the hypothesized model based on fit and the rejection of competing models. The third application, adding parameters to increase model fit, is most often based on modification indices (MI) which predict the decrease in chi-square goodness-of-fit values if non-free parameters are allowed to be free (Schumacker & Lomax, 2010). Regarding this third application, Bandalos and Finney remark:

Unfortunately, many researchers lose sight of the purpose of CFA, which is to allow the testing of a priori models. If a model does not fit the data, that information, along with a diagnosis of the source of the misfit, is useful and should inform the domain. On the other hand, thoughtlessly modifying a model post hoc in an attempt to make it fit the data is not the purpose of CFA and may simply lead to models that do not replicate due to fitting the idiosyncrasies of the sample data. Researchers and reviewers must keep in mind that the purpose of conducting a CFA study is to gain a better understanding of the underlying structure of the variables, not to force models to fit. The former is a useful scientific endeavor; the latter is not. (p. 112)

In Mplus, there are two basic methods for conducting model modification: 1) chi-square difference tests, and 2) Lagrangian multiplier tests guided by MI (Muthén, 2004;
Muthén & Muthén, 2012b). In nested models (i.e., models that are hierarchically related to each other such that one model is a subset of the other (Byrne, 2012)) the chi-square difference statistic, $\chi^2_D$, tests the equal-fit hypothesis in which large values of $\chi^2_D$ result in the rejection of it (Kline, 2011). For example, as free parameters are removed (e.g., comparing an empirically derived best fit multidimensional measured model to a unidimensional measured model) a model could be interpreted as oversimplified with a rejection of the equal-fit hypothesis (Kline, 2011). According to Muthén and Muthén (2012b), it is not appropriate to conduct the chi-square difference test under WLSMV estimation in the conventional manner (e.g., subtracting the difference between the chi-square values and the degrees of freedom) since the chi-square difference is not distributed as a chi-square. However, Mplus incorporates a two-step “DIFFTEST” specification in which the derivatives of the less restrictive model are used in the estimation of the more restrictive model to compute the chi-square difference test (Muthén & Muthén, 2012b). With two models having approximate fit to the same observed data based on the chi-square difference test, the parsimony principle supports a preference toward the simpler model (Kline, 2011).

Misspecified parameters are determined in Mplus by examination of the MI to identify fixed parameters that if allowed to be free would result in a significant decrease in the value of the chi-square statistic in a modified model (Byrne, 2012). Expected parameter change (EPC) statistics are also computed and represent the predicted change in either direction for each fixed parameter if their respective constraints are freed (Byrne, 2012).
In situations where the hypothesized model does not adequately fit the data, Bandalos and Finney (2010) emphasize that model misfit should be determined using both standardized residual values as well as MI. Furthermore, Bandalos and Finney assert:

Given the sample-specific nature of model misfit, we encourage replication studies to evaluate the stability of the misfit. If the same area of misfit is found upon replication, it should be taken seriously and possible theoretical explanations of the misfit should be presented. Given plausible and thoughtful reasons for the misfit, the model may be modified and treated as an a priori specified model in future studies. (p. 112)

**CFA study to assess latent factor structure.** A study by Osteen (2010) compared CFA and MIRT techniques by assessing the multidimensional latent factor structure of an original measure of graduate students’ motivations toward a social work community of practice. The study examined the differences between the analyses to include the consistency of the results and how unique pieces of information from each technique informed the researcher. CFAs were performed on the hypothesized three-factor construct as well as on a four-factor competing model based on EFA results, treating the responses from a six-point rating scale as ordinal, and estimated with weighted least squares by the Lisrel 8.8 SEM software package. Three nested models were estimated and compared sequentially to include a four-factor empirically derived solution with MI guided cross-loadings as a baseline, the standard four-factor competing model, and the standard three-factor hypothesized model. The analysis of model fit in conjunction with conceptual support led to a conclusion to adopt the standard four-factor model, which also provided evidence for the hypothesized multidimensional structure due to non-significant correlations between the latent variables. Findings of this study
included the benefit of integrating CFA and MIRT results in the overall assessment of latent factor structures to minimize the impact of each technique’s limitations. CFA was found to be more beneficial regarding subscale construction and evaluating factor associations, while MIRT was demonstrated to be more effective at assessing individual item performance.

**Item Response Theory**

As an alternative to CTT, a measurement model may also be specified by IRT to address dimensional structure, model fit, item fit, reliability, and validity. Myers, Wolfe, and Feltz (2005) list three advantages of IRT which include:

(a) select diagnostic statistics that have proven useful in determining the optimal categorization of rating scale structures, (b) powerful diagnostic indexes that are available to assess both item- and model-level fit to the data, and (c) conditional standard errors that are routinely estimated and allow the precision of estimates to be explored at different levels of ability. (p. 141)

The guiding principle of IRT presumes the most parsimonious and effective predictor for any person’s response to any item involves the relationship between the item’s difficulty (i.e., characteristics of the item) and the person’s ability (i.e., amount of agreement or amount of latent trait) (Bond & Fox, 2007). Then, the probability of success (i.e., endorsement of the item) may be calculated as a function of the difference between the person’s ability and the item’s difficulty (Bond & Fox, 2007).

While IRT is considered “a general framework for specifying mathematical functions that describe the interactions of persons and test items” (Reckase, 2009, p. v), a family of Rasch IRT models exist based on the original model for dichotomous data pioneered by Georg Rasch (Bond & Fox, 2007). Common to all Rasch models is the
logarithmic transformation of ordinal level data into interval level data for persons and test items; this transformation converts “sample dependent data into inferential measures based on probabilistic functions” (Bond & Fox, 2007, p. 278). Of interest in this study is the Rasch Rating-Scale Model (RSM), which is an extension of the original dichotomous model to Likert-type scales (Andrich, 1978).

**The Rasch RSM.** The general form of the Rasch RSM states that the probability of any person $n$ selecting any given response $k$ (where $k$ is the threshold between response categories) on any item $i$ is a function of the agreeability ($\theta_n$) of person $n$ as well as the endorsability ($\delta_i$) of the entire item $i$ based on the difficulty ($\tau_k$) at the given threshold (Bond & Fox, 2007). The difficulty estimate ($\tau_k$), in which the number of thresholds is one less than the number of response categories, is based on the threshold where a person has a 50% probability of endorsing a certain category over another (Bond & Fox, 2007). Therefore, the probability of response is equal to the natural logarithmic transformation, based on the constant $e$ (i.e., 2.7183), raised to the difference between agreeability, endorsability, and threshold difficulty divided by the sum of one and that same difference. The mathematical expression for the Rasch RSM, where $P(\theta)$ is the probability of response given a person’s amount of trait along a latent continuum (de Ayala, 2009), is given in Equation 3 (Bond & Fox, 2007):

$$P_{niki} = P(\theta) = \frac{e^{(\theta_n - \delta_i - \tau_k)}}{1 + e^{(\theta_n - \delta_i - \tau_k)}}$$  

(3)

The basic unit of the Rasch RSM is the option response function (ORF), which is the graphical representation of the probability of endorsing any single response category as a function of $\theta$ (de Ayala, 2009). For any item, the ORFs will always be comprised of
at least a single monotonically nondecreasing function, at least a single monotonically nonincreasing function, and unimodal functions for all other response categories (de Ayala, 2009). Figure 11 illustrates the ORFs for a hypothetical five-point Likert item. For example, in the case where a person’s ability is equal to the difficulty of the item (i.e., logit of 0), the probability of endorsing a neutral response (i.e., the middle response category) is approximately 0.30.

![Figure 11. RSM ORFs for Hypothetical Five-Point Likert Item (Winsteps Version 1.0.0)](image)

The Rasch RSM produces estimates for item difficulty (i.e., endorsability), person ability (i.e., agreeability), threshold difficulty, overall model fit, item and person fit, item and person reliability, and step calibration as well as item-person maps which display the relative distributions of the item and person estimates (Bond & Fox, 2007). While a single set of threshold difficulty values are estimated for the entire model, these values
and individual item difficulty estimates are a function of the proportion of actual responses to overall possible responses, which are then transformed into log odd units (i.e., logits) (Bond & Fox, 2007). Likewise, individual person fit estimates are a function of the transformation of the log odds based on the ratio of the person’s total raw score to their maximum possible score (Bond & Fox, 2007). Individual item and person standard errors along with threshold standard errors are also reported as well as unstandardized and standardized fit estimates (Bond & Fox, 2007). It is important to note that the RSM estimated threshold structure is common to all items even though each item has a unique difficulty estimate; moreover, each item difficulty estimate is considered a balance point for the distribution of that item’s response categories (Bond & Fox, 2007). The variation of item difficulty estimates, as shown in the model’s item-person map, reveals the relative and unequal difficulty of the items to each other and to each person (Bond & Fox, 2007).

Rating scale diagnostics, to include category frequencies, average measures, threshold difficulty estimates, ORFs, and category fit are output to assist the researcher in determining the optimal number of response categories to fit the RSM (Bond & Fox, 2007). Category frequencies should be distributed as uniform, normal, bimodal, or slightly skewed, with a minimum number of responses per category limited to ten to accommodate stable threshold estimates (Linacre, 2002). Average measures, defined as the mean person ability (or logit score) estimate per response category, are expected to increase monotonically in size with each category, such that persons with greater abilities endorse higher categories and vice versa (Bond & Fox, 2007; Linacre, 2002). Threshold
difficulty estimates, referred to as step calibrations, should also advance monotonically across the scale by at least 1.4 logits but not more than 5 logits (Linacre, 2002). Visual inspection of the ORFs, which illustrate the probability of choosing a particular scale category based on every person-agreeability minus item-endorseability difference, should show a separation between the thresholds with each category curve having a distinct peak revealing the most probable response for some portion of the measured variable (Bond & Fox, 2007). Finally, the category outfit mean squares should be less than 2.0; when greater than 2.0, excessive randomness or noise is introduced into the system (Linacre, 2002). By using a combination of these diagnostics to eliminate noise and improve threshold clarity, the researcher may consider collapsing adjacent categories and reanalyzing the functioning of the category scale (Bond & Fox, 2007).

The Rasch RSM provides two fit statistics, inﬁt and outfit, at the model, item, and person levels. Both fit statistics provide information, based on the magnitude of discrepancy between the estimated parameter and the expected parameter, regarding inconsistencies in the item responses (de Ayala, 2009). Based on the squared standardized residual between observed and expected responses, the inﬁt mean-square is an information weighted fit statistic sensitive to unexpected responses near a person’s ability or item’s difficulty location, summed across either item or person observations, with values ranging from zero to infinity and an expectation of one (de Ayala, 2009). For example, an expected item response near a person’s ability location produces an inﬁt value close to one; conversely, an unexpected item response near a person’s ability location produces a large inﬁt value (de Ayala, 2009). In contrast, the outfit mean-square
statistic, while also based on the squared standardized residual between observed and expected responses with values ranging from zero to infinity and an expectation of one, is sensitive to unexpected responses away from a person’s ability or item’s difficulty location and is not weighted when summed across either item or person observations (de Ayala, 2009). For example, an expected item response away from a person’s ability location produces an outfit value close to one; conversely, an unexpected item response away from a person’s ability location produces a large outfit value (de Ayala, 2009). For both infit and outfit mean-squares, values substantially greater than one indicate noise, haphazard response patterns, and underfit; values substantially less than one indicate dependency, determined response patterns, and overfit (Bond & Fox, 2007; de Ayala, 2009). For adequate fit, a cutoff criteria between 0.5 and 1.5 for infit and outfit are considered acceptable (de Ayala, 2009). Infit and outfit mean-squares also have standardized forms based on the $t$ statistic; in these cases, the expected value of $t$ is zero with a cutoff criteria outside of the range $-2.0 \leq t \leq 2.0$ (Bond & Fox, 2007).

In the Rasch RSM, reliability indices—person reliability, item reliability, person separation, and item separation—assist the researcher in determining whether ample items are spread sufficiently along the latent continuum as well as whether ability levels are spread sufficiently amongst persons (Bond & Fox, 2007). Person reliability, bounded between zero and one and analogous to Cronbach’s alpha, estimates how well person placement may be replicated across parallel items designed to measure the same construct; similarly, item reliability, bounded between zero and one, estimates how well item placement may be replicated across persons of comparable ability (Bond & Fox,
2007). For instance, one can infer consistency in the development of a scale from high person reliability by accurately distinguishing some persons who score higher and some persons who score lower; likewise, one can infer consistency in the development of a scale from high item reliability by creating a well-defined variable in which some items are more difficult and some items are less difficult (Bond & Fox, 2007; de Ayala, 2009).

Person separation, expressed in units of standard error and bounded between zero and infinity, indicates how well the instrument distinguishes persons along the latent continuum; item separation, expressed in units of standard error and bounded between zero and infinity, indicates how well the instrument distinguishes items along the latent continuum (de Ayala, 2009). While adequate separation indices are difficult to determine since they do not have finite upper bounds, “large” separation values are desirable over smaller ones and are related to the bounded reliability indices (de Ayala, 2009).

**Dimensionality of measurement.** While the assumption of unidimensionality is implicit in CTT, it is explicit in IRT. Specifically, the unique continuous person location variable, \( \theta \), is conceptualized to reflect the single latent variable that accounts for the behavior in a person’s item responses (de Ayala, 2009). Based on model simulations, unidimensionality is indicated in the Winsteps Rasch-model software program (Linacre, 2006) by the magnitude of the eigenvalue of the raw unexplained variance in the first contrast (i.e., component or factor) being \( \leq 2.0 \) (Linacre, 2010; Linacre, 2012). Contrary to this measurement ideal, there are many situations in psychological testing where scores are disaggregated into subcomponents and reported as separate performance dimensions,
or where scores are aggregated from measure subcomponents and reported as a single performance dimension (Briggs & Wilson, 2003).

To address these situations in which IRT’s unidimensionality assumption would be violated, a multidimensional extension of many of the Rasch family measurement models was introduced by Adams, Wilson, and Wang (1997) as the Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM). Specified as a generalized Rasch MIRT model, the MRCMLM includes the flexibility to use just one model to fit a wide variety of multidimensional models to include dichotomous (e.g., simple logistic and others) and polytomous (e.g., RSM, partial credit, and others) (Wu, Adams, Wilson, & Haldane, 2007). Compared to IRT models, MIRT models provide greater clarity in understanding the dimensions that are being measured, how accurately the dimensions are being assessed, and how to best modify the multidimensional instrument (Ackerman, Gierl, & Walker, 2003).

The MRCMLM is described as being constructed from a basic conceptual building block approach (Briggs & Wilson, 2003). First, assume a unique dimension \( d \) (i.e., latent trait), among a larger set of dimensions \( d = 1, \ldots, D \), underlies an item \( i \) indexed by \( k \) ordered categorical responses. Second, the log odds of the probability of a response in category \( k \) versus category \( k-1 \) of item \( i \) \( (P_{ik} / P_{ik-1}) \) can be modeled as the difference between a person’s amount of latent trait on the dimension \( (\theta_d) \) and the relative difficulty \( (\delta_{ik}) \) (i.e., item difficulty) of category \( k \) (as opposed to category \( k-1 \)) to endorse item \( i \) with that level of latent ability as in Equation 4 (Allen & Wilson, 2006):

\[
\log \left( \frac{P_{ik}}{P_{ik-1}} \right) = \theta_d - \delta_{ik} \tag{4}
\]
Third, each person’s level of latent ability is measured on each dimension by the scale \( \theta = (\theta_1, \ldots, \theta_D) \), where the dimensions are free to be correlated (Allen & Wilson, 2006).

According to Briggs and Wilson (2003), the \( \theta_d \) in Equation 4:

\[
\theta_d
\]

represents the latent ability of the person as a function of the dimension of ability mapped onto item \( i \). Thus, for example, in an achievement testing context, the dimensions might be components of the curriculum. The mapping then would indicate that item \( i \) was related to only component \( d \), and the value of \( \delta_{ik} \) would indicate whether it was relatively easier or harder for a student to be classified as achieving category \( k-1 \) compared to \( k \). (p. 90)

In addition to modeling the dimensionality of a theoretical construct, an examination of the consistent use of the response categories across all items within a dimension is a feature of the MRCMLM as in the RSM (Allen & Wilson, 2006).

Two subclasses of the MRCMLM may be specified at the item level—multidimensional between-item models and multidimensional within-item models (Adams, Wilson, & Wang, 1997). The between-item subclass refers to tests containing several mutually exclusive subscales, which measure separate but related latent dimensions, where each item is associated with only one subscale (Wu et al., 2007).

Before the development of MIRT analyses, the application of unidimensional IRT modeling to each subscale separately resulted in less reliable estimation of item parameters and ability predictions due to the lack of use of all of the data; similarly, the practice of ignoring the multidimensionality altogether resulted in the lack of examination of the associations between the dimensions (Adams et al., 1997). The advantage of the MRCMLM with between-item data is:

(1) it explicitly recognizes the test developer’s intended structure, (2) it provides direct estimates of the relations between the latent dimensions, and (3) it draws on the (often strong) relationship between the latent dimensions to produce more
accurate parameter estimates and individual measurements. (Adams et al., 1997, p. 11)

The second subclass, within-item, refers to tests in which some of the items measure abilities from greater than one latent dimension (Adams et al., 1997). Three types of testing scenarios are present in which within-item specification is necessary: 1) when scale developers construct indicators requiring latent abilities amongst two or more subscales (e.g., an aptitude test containing verbal and quantitative components), 2) when a person’s overall ability is judged on two dimensions (e.g., an essay scored on thematic understanding and writing ability), and 3) when certain response patterns are not possible (e.g., three levels are identified with only two solving strategies) or when various ability levels require different latent abilities (e.g., a multi-step mathematics problem) (Adams et al., 1997).

Regardless of chosen subclass, the MRCMLM equations may be analyzed for a set of items with up to 30 latent dimensions with the ConQuest software package (Wu, Adams, Wilson, & Haldane, 2012). Marginal maximum likelihood estimation is implemented along with a Gauss-Hermite quadrature algorithm for models with less than three dimensions, while the Monte Carlo approach to the calculation of the integrals must be executed for models with greater than three dimensions (Wu et al., 2007). In addition to the estimated Rasch RSM parameters described above in which model fit is maximized, ConQuest outputs estimated populations means, variances, covariances, correlations, and multidimensional item-person maps of the $\theta_D$ parameters in relation to the logit scale (Allen & Wilson, 2006; Wu et al., 2007).
**Three approaches to assess dimensionality.** To formally examine dimensionality assumptions, Allen and Wilson (2006) present three approaches using IRT techniques—*composite, consecutive*, and *multidimensional*. With the *composite* approach, the total score based on responses to each item on an instrument is used to indicate a single estimate of an overall latent dimension (i.e., without regard to any of the subscales) such that the probability of a response to a single category relative to the previous one for each of the items is given by Equation 5 (Allen & Wilson, 2006):

\[
\log\left(\frac{p_{ik}}{p_{ik-1}}\right) = \theta - \delta_{ik}
\]

(5)

In ConQuest, the composite approach is modeled with the Random Coefficients Multinomial Logit Model (RCMLM), a unidimensional analogue to the MRCMLM which also integrates many of the Rasch family measurement models including the RSM (Adams et al., 1997). While an advantage of the composite approach is the parsimony in modeling achieved with a single estimate (and associated standard error) of latent ability based on an overall score, a disadvantage is the loss of differential information regarding the overall latent construct relative to each subscale (Briggs & Wilson, 2003). Figure 12 graphically illustrates the composite approach in which the sum of scores on items 1 through 12 is treated as a single estimate of the latent dimension \(\theta\). In this hypothetical example, ConQuest would provide estimates for 16 parameters including the mean and variance of \(\theta\), 11 item difficulty parameters (i.e., one parameter is constrained for model identification), and three step parameters (i.e., one parameter is constrained for model identification) (Wu et al., 2007).
As a means of comparison with the composite approach, the consecutive approach independently models each hypothesized subscale as unidimensional constructs to be analyzed separately (Briggs & Wilson, 2003). Since the consecutive approach is essentially the composite approach repeated for each dimension, the RCMLM model in ConQuest may be used again to estimate item and person parameters (Briggs & Wilson, 2003). While an advantage of the consecutive approach is producing $\theta_D$ estimates and standard errors for each hypothesized dimension, a disadvantage is the loss of understanding of the interrelatedness amongst the dimensions (Allen & Wilson, 2006). Since the ratio of items to dimension is less than that of the composite approach, the
standard errors for the consecutive estimates will be substantially larger and the resulting consecutive reliabilities will be lower compared to the composite approach (Allen & Wilson, 2006). Figure 13 graphically illustrates the consecutive approach in which the sum of scores associated with each hypothesized dimension is treated as a separate statistic. In this hypothetical example, ConQuest would provide estimates for 28 parameters including the means and variances of $\theta_{D1}$ through $\theta_{D4}$, eight item difficulty parameters (i.e., one parameter is constrained per dimension for model identification), and three step parameters for each dimension (i.e., one parameter is constrained for model identification) (Wu et al., 2007).

Figure 13. Consecutive IRT Modeling Approach (Amos Version 18)
According to Allen and Wilson (2006), “the multidimensional approach can be viewed as a compromise between the composite and consecutive approaches, one that incorporates the best of both approaches” (p. 179). By incorporating the correlations between the dimensions, the *multidimensional* approach’s reliability for each of the dimensions is closer in value (compared to the consecutive approach) to the composite approach since each item is influenced by its respective latent variable and by all other dimensions through latent associations (Allen & Wilson, 2006). This approach estimates abilities across each latent dimension simultaneously using the MRCMLM as formulated in Equation 4. Figure 14 graphically illustrates the multidimensional approach in which the sum of scores associated with each hypothesized dimension is treated as a separate statistic while incorporating the correlations between the dimensions. In this hypothetical example, ConQuest would provide estimates for 25 parameters including the means and variances of \( \theta_{D1} \) through \( \theta_{D4} \), eight item difficulty parameters (i.e., one parameter is constrained per dimension for model identification), three step parameters (i.e., one parameter is constrained for model identification), and the six unique elements of the variance-covariance matrix (Wu et al., 2007).
The likelihood ratio statistic, $G^2$ (also called the deviance), and Akaike’s Information Criterion (AIC) are useful in comparing the model fit regarding the three dimensionality evaluation approaches (Briggs & Wilson, 2003). Since the models are nested, model fit can be compared between the multidimensional approach and the composite approach by the difference in deviance, which approximates a $\chi^2$ distribution with degrees of freedom equal to the difference in estimated item parameters between the models (Briggs & Wilson, 2003). If the difference in deviance between the nested models is statistically significant, then evidence exists that the multidimensional model fits the data significantly better than the composite model; conversely, if the difference is not statistically significant, then the more parsimonious model (i.e., the composite one)

\textit{Figure 14. Multidimensional IRT Modeling Approach (Amos Version 18)}
should be chosen (de Ayala, 2010). AIC may be used to evaluate models that are estimated with maximum likelihood methods (Ullman, 2007) and to compare non-nested models that share the same data (Allen & Wilson, 2006). In the case of the latter, the lowest AIC value amongst the composite, consecutive, or multidimensional approaches would indicate the best model fit (Allen & Wilson, 2006). AIC is defined as follows in Equation 6 (Kang & Cohen, 2007):

$$AIC = d + 2p$$

(6)

where $d$ is the deviance and $p$ is the number of estimated parameters.

In addition to comparing the estimated model fit, reliabilities, and correlations, discrepant cases of the ability estimates can be computed across dimensions to highlight the ramifications of ignoring multidimensionality when assessing a latent trait (Briggs & Wilson, 2003). To compare standardized ability estimates across dimensions, $d$, the sum of squares indicator, $DI$, for each person, $p$ is calculated with Equation 7 (Briggs & Wilson, 2003):

$$DI_p = \sum_{d=1}^{d}(\bar{\theta} - \theta_d)^2$$

(7)

By setting an arbitrary threshold for a discrepant case at $DI_p = 0.5$ (Briggs & Wilson, 2003) or $DI_p = 1.0$ (Allen & Wilson, 2006), the resulting group of ability estimates will illustrate unidimensional measurement being either an underestimate or overestimate of latent ability on the excluded dimensions (Briggs & Wilson, 2003).

**IRT study to assess dimensionality.** A study by Allen and Wilson (2006) illustrated the composite, consecutive, and multidimensional approaches to item response theory modeling and interpretation of participant-reported attitudinal data regarding the
Treatment Self-Regulation Questionnaire. The data were analyzed to determine if self-regulation could be treated as a unidimensional construct, or if it contained multiple dimensions based on the type of regulation or motivation needed to cause one to seek improvements in healthy behavior. Using the RCMLM and MRCMLM with ConQuest software, model fit analyses determined that the multidimensional model fit significantly better than the composite model on the basis of the difference in deviance, and was also superior in fit to both the composite and consecutive models based on AIC comparisons. Allen and Wilson found the multidimensional approach achieved reliability enhancement with higher reliability values on each dimension, which were approximately equivalent to the single reliability value derived from the composite approach (compared to the consecutive approach). It was determined that the estimated multidimensional correlations, some negative while others relatively uncorrelated, may lead to a revision of the theoretical understanding of self-determination, but clearly indicated that modeling the construct with a unidimensional model does not explain the complexity of the data. Allen and Wilson noted also that the correlations from the consecutive approach were attenuated due to measurement error. To compare standardized ability estimates across dimensions, the sum of squares indicator was calculated for each person and revealed a reliance on a unidimensional estimate could falsely identify persons that may change their behavior. Allen and Wilson recommended that a comparison of multidimensional estimates with behavior change evidence is necessary to support theory and guide intervention.
Latent Class Analysis

LCA is a latent variable model used in the social, behavioral, and health sciences which is related to CFA and IRT when analyzing cross-sectional data (Collins & Lanza, 2010). LCA may be viewed as analogous to factor analysis in that both models utilize observed variables, assumed to be conditionally independent, which are a function of an underlying latent variable and error (Collins & Lanza, 2010; Samuelsen & Dayton, 2010). For example, it is assumed that observed variables in CFA are independent of each other once loaded on their respective factors; likewise, it is assumed that observed variables in LCA are mutually independent after the latent variable is conditioned out (Wang & Wang, 2012). However, in CFA the latent variable (i.e., factor) is continuous and has a normal distribution with indicators treated as continuous, while in LCA the latent variable (i.e., latent class variable) is categorical and has a multinomial distribution with indicators treated as categorical (Collins & Lanza, 2010). LCA is also related to Rasch IRT as a generalization of discrete response models (Samuelsen & Dayton, 2010); however, the latent variable is continuous in IRT while it is categorical in LCA (Collins & Lanza, 2010).

Another distinction between CFA, IRT, and LCA concerns the orientation of statistical analysis. Since CFA focuses on grouping items under a factor structure, it is considered a variable-oriented approach with an emphasis on the identification of relationships between variables applied across persons (Collins & Lanza, 2010; Wang & Wang, 2012). In contrast, LCA focuses on grouping unobserved subpopulations of individuals based on their patterns of responses, and therefore is considered a person-
oriented approach (Collins & Lanza, 2010). On the other hand, IRT analysis with both person ability and item difficulty estimates can be considered both a variable-oriented and a person-oriented (i.e., case-oriented) approach (Onwuegbuzie & Combs, 2010).

The person-oriented approach of LCA complements traditional variable-oriented approaches such as CFA and IRT for several reasons. According to Wang and Wang (2012), mixture modeling (i.e., the general framework for LCA), offers the opportunity for researchers to identify unknown a priori homogeneous groups/classes of individuals based on the measures of interest, examine the features of heterogeneity across the groups/classes, evaluate the effects of covariates on the group/class membership, assess the relationship between the group/class membership and other outcomes, and study transitions between the latent group/class memberships over time. (p. 289-290)

Formally, LCA is a model-based method for classifying cases (e.g., people or objects) into unobserved groups which are neither known or specified a priori based on similar response patterns identified with posterior membership probabilities (Samuelsen & Dayton, 2010; Wang & Wang, 2012). The overall objective in conducting an LCA is “to arrive at an array of latent classes that represents the response patterns in the data, and to provide a sense of the prevalence of each latent class and the amount of error associated with each variable in measuring these latent classes” (Collins & Lanza, 2010, p. 27).

To graphically illustrate the LCA modeling approach, consider three heterogeneous unobserved classes of people, each with a multinomial distribution, plotted along a line with positive slope with respect to two categorical outcomes, $x_1$ and $x_2$, as in Figure 15 (Muthén, 2001). While the line represents a strong association between the two observed indicators $x_1$ and $x_2$, the association is due to the mixture of the three
heterogeneous classes of people with unrelated outcomes (Muthén, 2001). Figure 16 illustrates the notion that the observed categorical indicators, $x_1$ and $x_2$, measure the latent variable, $c$; it is also important to recognize that the causal flow is from both the 3-class categorical latent variable and the error components, $e_1$ and $e_2$, associated with the indicators (Collins & Lanza, 2010).

Figure 15. Categorical Latent Class Variable Conceptualization (Boucher, 2012; Muthén, 2001)
Figure 16. LCA Model with Two Observed Indicators (Amos Version 18)

Description of the model. Samuelsen and Dayton (2010) list the primary assumptions of LCA as:

(1) the model correctly specifies the number of classes, (2) each respondent belongs to only one latent class, and (3) respondents within a class are homogeneous. Building on these, the fundamental concept of LCA is that of local (i.e., conditional) independence meaning that the observed manifest responses are independent given that latent class membership is known. (p. 175)

For the case of $r$ observed binary response variables, $x$, with the categorical latent variable $c$ having $K$ classes ($c = k$ and $k = 1, 2, \ldots, K$), assume conditional independence such that the joint probability of all of the response variables may be calculated as in Equation 8 (Muthén, 2001):

$$P(x_1, x_2, \ldots, x_r) = \sum_{k=1}^{K} P(c = k)P(x_1|c = k)P(x_2 = 1|c = k) \ldots P(x_r = 1|c = k) \quad (8)$$

The posterior probabilities, analogous to CFA factor scores, are then each individual’s most likely latent class membership given their observed pattern of responses (Collins & Lanza, 2010) and may be estimated as in Equation 9 (Muthén, 2001):
From the above equations, two types of measurement error adjusted model parameters are estimated—latent class probabilities (i.e., unconditional probabilities) and conditional item-response probabilities (Wang & Wang, 2012). The mean of the latent class (unconditional) probabilities indicate the population proportion most likely to belong to a latent class (Wang & Wang, 2012); these probabilities sum to one and are known as the latent class prevalence (Collins & Lanza, 2010). In contrast, the conditional item-response probabilities indicate the likelihood of selecting a specific category of the observed indicator, given a certain class membership; values close to 1.0 indicate that members of a latent class are very likely to endorse the respective item category, while values close to zero indicate that members of a latent class are not very likely to do the same (Wang & Wang, 2012).

With \( j \) representing the total number of item categories, the conditional item-response probability is expressed in logit form and defined as in Equation 10 and Equation 11 (Muthén, 2001; Wang & Wang, 2012):

\[
P(x_r = x_{rj} | c = k) = \frac{1}{1 + e^{-L_{jk}}} \quad \text{(10)}
\]

where

\[
L_{jk} = \ln \left( \frac{P_{jk}}{1-P_{jk}} \right) \quad \text{(11)}
\]

The conditional item-response probabilities are examined to assist the researcher in interpreting and labeling of the latent classes (Collins & Lanza, 2010).
**Three-step modeling approach.** The following steps are necessary to estimate a basic LCA model: 1) determine the optimal number of latent classes, 2) evaluate the quality of the classification of latent class membership, and 3) define the latent classes (Wang & Wang, 2012).

Despite the number of latent classes being unobserved and not estimated directly from the data, the optimal number is determined by analyzing the fit of a series of increasing class number models by comparing the \( k \)-class model with the \((k-1)\)-class model (Wang & Wang, 2012). While it is not appropriate to use likelihood ratio chi-square values to compare models which differ in the number of classes due to “inadmissible parameter values of zero class probabilities” (Muthén, 2004, p. 33), three other fit statistics are incorporated in the **Mplus** software: 1) the Lo-Mendell-Rubin likelihood ratio (LMR LR) test, 2) the adjusted LMR LR (ALMR LR) test, and 3) the bootstrap likelihood ratio test (BLRT) (Wang & Wang, 2012). First, the \( p \)-value from the LMR LR test should be evaluated which compares the \( k \)-class model to the \((k-1)\)-class model; a statistically significant \( p \)-value indicates the \( k \)-class model fits significantly better than the \((k-1)\)-class model (Wang & Wang, 2012). Then, these model comparisons should be continued iteratively until the LMR LR test is statistically non-significant between the \((k+1)\)-class model and the \( k \)-class model; this indicates that the optimal number of latent classes is \((k-1)\) since no additional statistically significant model fit improvement occurred by including the last analyzed class in the model (Wang & Wang, 2012). Furthermore, due to inflated Type I error when sample size is small, the ALMR LR test should be evaluated since it is adjusted for sample size and model degrees of
freedom (Wang & Wang, 2012). Finally the BLRT $p$-value, derived from the log-likelihood differences in bootstrap samples from both $k$-class and $(k-1)$-class models, should be evaluated in the same manner as the LRM LR test (Wang & Wang, 2012). Amongst all three LCA model fit statistics in Mplus, a Monte Carlo simulation study revealed that the BLRT was the most consistent in deciding on the number of classes in the study population (Nylund, Asparouhov, & Muthén, 2007).

In addition to the model fit statistics, the following information criterion indices based on model log-likelihood, parsimony, and/or sample size penalty terms are available in Mplus to assist in model comparison and selection of the number of latent classes: 1) the Akaike Information Criterion (AIC), 2) the Bayesian Information Criterion (BIC), and 3) the sample size Adjusted Bayesian Information Criterion (ABIC) (Wang & Wang, 2012). For each of these indices, the lowest value of the criterion amongst competing LCA models would be considered as justification for determining the best fit model (Samuelsen & Dayton, 2010). Amongst all three LCA information criterion indices in Mplus, a Monte Carlo simulation study revealed that the BIC has the best performance even when sample size is small (Nylund, Asparouhov, & Muthén, 2007). Nonetheless, a typical decision strategy regarding the preferred model should be based on multiple information criteria (Samuelsen & Dayton, 2010).

After the optimal number of classes is determined based on model fit, the quality of the classification of individuals is examined on the basis of the estimated posterior probabilities. While membership of individuals into a latent class is not definitely determined, individuals are assigned into a latent class based on their largest (i.e., most
likely) posterior probability; the probability of misclassification is low when an 
individual’s highest posterior probability is close to 1.0 (Wang & Wang, 2012). Nagin’s 
(2005) criterion for minimum acceptable class membership classification is when average 
posterior probability is at least 0.7 for all groups.

For example, consider a LCA model in which it has been determined to fit four 
optimal classes and the estimated posterior probabilities for an individual are 0.03, 0.04, 
0.88, and 0.05 for classes one through four, respectively. In this hypothetical case, the 
individual would be assigned to class three and the probability of correct class 
membership would be 0.88, while the probability of misclassification would be 0.12 (i.e., 
the sum of the remaining probabilities) (Wang & Wang, 2012).

Another criterion to summarize posterior misclassification is based on entropy, a 
single value summary of the degree of uncertainty or disorder in the model scaled such 
that large values indicate less classification error (Collins & Lanza, 2010). With \( \hat{p}_{ik} \) 
representing the estimated posterior probability for an individual \( i \) to be in class \( k \) of 
sample size \( n \), the entropy measure, \( E_K \), ranges from zero to 1.0 and is estimated in Mplus 
by Equation 12 (Muthén, 2004):

\[
E_K = 1 - \frac{\sum_i \sum_k (\hat{p}_{ik} \ln \hat{p}_{ik})}{n \ln K} 
\]  

(12)

In a Monte Carlo simulation study to examine selection of important covariates, Clark 
(2010) proposed high entropy values of 0.80, medium entropy values of 0.60, and low 
entropy values of 0.40. Regarding the use of the entropy measure as the sole source for 
evaluating the quality of latent class membership classification, Collins and Lanza (2010) 
assert:
Latent class assignment error can increase simply as a function of the number of latent classes, so indices like $E$ often decrease as the number of latent classes increases. In other words, class assignment can look better purely by chance in a two-latent-class model than in a comparable model with three or more latent classes. For this reason, entropy-based measures can be a poor tool for model selection. (p. 75)

After evaluating the quality of the classification of latent class membership, the size of each class should be considered since “the percentage of individuals in each class represents the prevalence of the corresponding subpopulation in the target population” (Wang & Wang, 2012, p. 295). According to Samuelsen and Dayton (2010), the researcher must determine if the class is substantively meaningful or an artifact of the sample data.

The final step in LCA modeling involves defining each latent class in a meaningful and interpretable manner such that the heterogeneity in the population is adequately described; this is analogous to defining the extracted factors in EFA (Wang & Wang, 2012). According to Samuelsen and Dayton (2010), post hoc analyses involving concomitant variables can enhance the naming of latent classes and should be considered a type of construct validation.

**Model estimation.** LCA models are estimated in Mplus by ML using an expectation-maximization (EM) algorithm (Muthén, 2004). Analogous to CFA modeling, the LCA model is identified when $df_M \geq 1$ and “the amount of the ‘known’ information exceeds the amount of ‘unknown’ information” (Collins & Lanza, 2010, p. 92). Even when models are properly identified, the model estimation may fail to converge on the global maximum of the likelihood, but rather provide incorrect parameter estimates based on local maxima (Wang & Wang, 2012). In order to report
evidence that the global maximum is reached, the model should be estimated with different sets of random starting values until the best log-likelihood value is the most frequent solution (Samuelsen & Dayton, 2010; Wang & Wang, 2012).

**LCA study to explore population subtypes.** A two-phase study by Gerber, Wittekind, Grote, and Staffelbach (2009) included an exploratory LCA in phase one to determine career orientation types of Swiss employees followed by a confirmatory LCA in phase two to validate the orientation types through work attitudes and sociodemographical relationships. A phase one random sample of 835 German-speaking Swiss employees was collected via telephone interviews regarding a career orientations survey which included nine dichotomous items of contrasting options based on three hypothesized dimensions—traditional, independent, and disengaged. Using two software programs (*Mplus* and Panmark) with 500 random sets of starting values, groups of BIC values were compared and global likelihood maxima solutions were obtained for the phase one data. The phase one analysis yielded plausible solutions for either a 3-class or a 4-class final model; the BLRT rejected the 3-class model in favor of the 4-class model and subsequently the classes were labeled independent, traditional/promotion, traditional/loyalty, and disengaged by the researchers based on the 4-class response probabilities. The phase one LCA confirmed the researcher’s hypothesis of three career orientations amongst Swiss employees; additionally, the analysis revealed the traditional career orientation is best described by two subtypes.

Phase two of the study included a random sample of 737 German-speaking Swiss employees collected via telephone interviews regarding the same career orientations
survey, as well as an additional measure of work attitudes involving employability, career success, intention to quit, and affective commitment. With a confirmatory LCA approach, the researchers attempted to replicate the phase one model by fixing the response probabilities in the phase two data set to the phase one values; thus, only the independent class size parameters were free to be estimated. The resulting LCA confirmed the adequacy of the 4-class model from phase one with the phase two data using the BLRT. To further validate the LCA results, the relationships between career orientations, the work attitudes measure, and demographics were assessed using analysis of variance (ANOVA) to investigate the statistical significance of the differences.
Chapter Two: Method

“I will listen to any suggested hypothesis, but on one condition—that you show me a method by which it can be tested.” - August Wilhelm von Hofmann, 1818-1892; (Gregory, 1916, p. 162)

This chapter explains in detail how this study was conducted. Included are sections identifying the study participants, the instruments used to measure the unobserved latent traits, the research procedure, and the analytical strategy to address the research questions.

Participants

Character Mosaic Virtues (CMV): November 2011. Following the June 2011 developmental field administration of the CMV, a revised measure was distributed by the Center for Character and Leadership Development (CCLD) to approximately 550 first-year (i.e., basic) cadets in the USAFA Class of 2015 on November 16, 2011 (M. Rosebush, personal communication, September 21, 2012). After discarding incomplete responses, the convenience sample collected by the CCLD returned 253 complete cases which were provided for analysis (M. Rosebush, personal communication, September 21, 2012). Since the measure was designed for character and leadership coaching, only the respondent’s named and cadet squadron were collected by the primary investigator in addition to the item responses. With personally identifiable information prohibited from being released, the only grouping variable provided was the cadet squadron identifier.
**Leadership Mosaic Inventory (LMI): September 2012.** Following the March 2012 developmental field administration of the LMI, a revised measure was distributed on September 10, 2012 by the CCLD to 360 third-year (i.e., junior) cadet element leaders from the USAFA Class of 2014 and to approximately 2,100 first-year and second-year (i.e., freshmen and sophomore) subordinate cadets in the USAFA Classes of 2015 and 2016, respectively (M. Rosebush, personal communication, October 9, 2012). After discarding incomplete responses, convenience samples of 357 cadet element leader self-ratings and 1,777 subordinate-ratings of their element leaders on the revised LMI scales were collected by the CCLD and provided for analysis (M. Rosebush, personal communication, November 9, 2012). Since the measures were designed for character and leadership coaching, only the respondent’s named and cadet squadron were collected by the primary investigator in addition to the item responses. With personally identifiable information prohibited from being released, the only grouping variable provided was the cadet squadron identifier.

**Leadership Mosaic Inventory (LMI): October 2012.** A second administration of the revised LMI was distributed on October 9, 2012 by the CCLD to the same 360 third-year (i.e., junior) cadet element leaders from the USAFA Class of 2014 and to approximately 2,100 first-year and second-year (i.e., freshmen and sophomore) cadets in the USAFA Classes of 2015 and 2016, respectively (M. Rosebush, personal communication, November 9, 2012). After discarding incomplete responses, convenience samples of 284 cadet element leader self-ratings and 1,535 subordinate-ratings of their element leaders on the revised LMI scales were collected by the CCLD
and provided for analysis (M. Rosebush, personal communication, November 9, 2012).
Since the measures were designed for character and leadership coaching, only the respondent’s named and cadet squadron were collected by the primary investigator in addition to the item responses. With personally identifiable information prohibited from being released, the only grouping variable provided was the cadet squadron identifier.

**Instruments**

**CMV.** Cadet character virtues were measured with 45 items developed by Rosebush (2011) reflecting nine theoretical dimensions including: 1) courage, 2) accountability, 3) humility, 4) duty, 5) care for others, 6) self-control, 7) respect for human dignity, 8) attention to detail, and 9) excellence. The 5-point rating scale included the following response options—*very much unlike me, unlike me, neutral, like me, and very much like me* (Rosebush, 2011). Appendix A lists the CMV rating scale and items grouped by theoretical dimension.

The CMV has demonstrated evidence of various forms of validity and reliability. The items initially considered to measure the 20 virtues that underlie the USAF core values were derived from previously validated instruments (i.e., IPIP and VIA) which exhibited face validity in representing the USAF virtues (Rosebush, 2011). Evidence for convergent and discriminant validity was demonstrated by conducting a series of pilot study EFAs in which the final solution produced nine independent factors and satisfied all tests of assumptions (Rosebush, 2011). Internal consistency reliability was determined by measuring Cronbach’s alpha for each scale. Since alpha values greater than or equal to 0.70 for each scale are a criteria for the consistency of scores from a set
of items, each of the nine CMV scales demonstrated reliability with a range of alpha from 0.72 to 0.89 (Rosebush, 2011). Further evidence of convergent and discriminant validity was demonstrated by moderately high and low correlations, respectively, of the CMV with the scales of the Moral Foundations Questionnaire (Haidt & Graham, 2007), a validated instrument which measures the degree a respondent makes decisions based upon avoidance of harm, fair treatment of others, endearment toward a group, authority, and acting properly (Rosebush, 2011).

LMI. The LMI has demonstrated evidence of various forms of validity and reliability. A content validation procedure resulted in the examination of eight different leadership theories and in a pilot study administration of 22 validated scales which claim to explain leadership effectiveness (Rosebush, 2012). Evidence for convergent and discriminant validity was demonstrated by conducting a series of EFAs in which the final solution produced two orthogonal factors based on positively and negatively worded leadership qualities (Rosebush, 2012). Using a face validation method, the items from the positively worded factor were reallocated into a conceptual unidimensional leadership effectiveness scale which included items representing the six non-independent USAF institutional sub-competencies (Rosebush, 2012). Internal consistency reliability was determined by measuring Cronbach’s alpha on each sub-competency item grouping, with alpha ranging from 0.75 to 0.92 (Rosebush, 2012). Strong evidence of criterion-related validity was demonstrated in which 79% of the variance in the “lifts others” criterion and 86% of the variance in the “elevates performance” criterion was explained in predicting the two operational definitions of being a “leader of character” (Rosebush, 2012).
**LMI element leader self-rating version.** Cadet element leader effectiveness was measured with 29 self-rating items developed by Rosebush (2012) on a unidimensional construct based on six USAF institutional sub-competencies to include: 1) develops and inspires others, 2) takes care of people, 3) builds teams and coalitions, 4) negotiating, 5) vision, and 6) adaptability. The 5-point rating scale included the following options for the responding element leaders—*very much unlike me, unlike me, neutral, like me, and very much like me* (Rosebush, 2012). Appendix B lists the LMI element leader self-rating scale and items; each item is annotated with the corresponding theoretically non-independent USAF institutional sub-competency.

**LMI element leader subordinate-rating version.** Cadet element leader effectiveness was also measured by their subordinates with the same 29 items as the element leaders rated themselves (Rosebush, 2012). This version was based on the same unidimensional construct as the self-rating version. The 5-point rating scale included the following options for the responding subordinates—*very much unlike the Leader, unlike the Leader, neutral, like the Leader, and very much like the Leader* (Rosebush, 2012). Appendix B lists the LMI element leader subordinate-rating scale and items; each item is annotated with the corresponding theoretically non-independent USAF institutional sub-competency.

**Procedure**

**CMV.** The CMV was administered in a classroom setting during designated military training time on November 16, 2011 to the first-year cadets from the 10 cadet squadrons who participated in the pilot study as well as first-year cadets from 10 other
randomly selected squadrons (M. Rosebush, personal communication, November 28, 2012). Both verbal and written instructions were provided, which included the purpose of the research, confidentiality, voluntary participation, and the anticipated time needed to complete the survey (M. Rosebush, personal communication, November 28, 2012). Cadets responded to the survey items using the USAFA Form 150 “General Answer Sheet Type A” multiple choice bubble sheet (M. Rosebush, personal communication, November 28, 2012). Figure 17 illustrates the CMV written instructions.

![CMV Written Instructions](image)

*Figure 17. November 2011 CMV Administration Written Instructions (M. Rosebush, personal communication, October 9, 2012)*

**LMI.** Both versions of the LMI were emailed to cadets using an Internet survey link as a pre-test on September 10, 2012 and again as a post-test on October 9, 2012, which included the purpose of the inventory and the instructions on how to participate (M. Rosebush, personal communication, November 28, 2012). Cadets were provided two days in which to complete the five-minute inventory; presumably, they utilized their own computers during their own free time to participate in the survey (M. Rosebush, personal communication, November 28, 2012). Figure 18 illustrates the LMI element leader self-
rating written instructions and Figure 19 illustrates the LMI element leader subordinate-rating written instructions.

Figure 18. September/October 2012 LMI Element Leader Self-Rating Administration Written Instructions (M. Rosebush, personal communication, October 9, 2012)

Figure 19. September/October 2012 LMI Element Leader Subordinate-Rating Administration Written Instructions (M. Rosebush, personal communication, October 9, 2012)
Human subjects protection. A research protocol exemption request was made to the USAFA Institutional Review Board (IRB) regarding the use of the following three de-identified datasets: 1) the June 2011 CMV developmental sample, 2) the November 2011 revised CMV sample, and 3) the March 2012 LMI developmental sample. The USAFA IRB exemption request is provided in Appendix C and the approval documentation is copied in Appendix D.

Since the September 2012 and October 2012 administrations of both revised LMI versions were conducted under the auspices of the CCLD and not under a USAFA IRB protocol, approval for use of the de-identified datasets was requested directly from the CCLD. The approval documentation for the use of these data is shown in Appendix E.

Additionally, exemption requests for the use of all of the above datasets were approved by the University of Denver IRB. The approval documentation is available in Appendix F and Appendix G.

Analytical Strategy

Research question one. The latent factor structure of the November 2011 CMV data was assessed using the CFA techniques described in the previous chapter with Mplus Version 7 (Muthén & Muthén, 2012a). The fit of the nine-factor theoretical model was compared with a competing hypothetical eight-factor model. This competing model was based on an a priori hypothesis that the attention to detail and excellence factors from the nine-factor model may be more accurately represented by a single factor labeled methodical—a construct in which two of Rosebush’s (2011) items (e.g., “I am exacting in
my work” and “I pay attention to details”) selected from the International Personality Item Pool were defined under a *methodical* factor in the Six Factor Personality Questionnaire (Jackson, Paunonen, & Tremblay, 2000).

The following criteria were examined to determine the adequacy of model fit:

1. A cutoff value less than 0.06 for RMSEA (Muthén, 2004; Schreiber et al., 2006; Yu, 2002)
2. A cutoff value greater than 0.95 for both CFI and TLI (Byrne, 2012; Muthén, 2004; Schreiber, Stage, King, Nora, & Barlow, 2006; Yu, 2002)
3. A cutoff value less than 0.90 for WRMR (Muthén, 2004; Schreiber et al., 2006; Yu, 2002)
4. Each standardized factor loading estimate should have significant $p$-values, indicating the estimates were statistically different from zero (Byrne, 2012)
5. Each standardized factor loading estimate should be greater than 0.70 such that the factor explained the majority of the variance (e.g., $R^2 > 0.50$) of each measured item (Fornell & Larcker, 1981; Kline, 2011)
6. Each factor’s construct measurement reliability should be greater than or equal to 0.70 (Fornell & Larcker, 1981)
7. AVE greater than 0.50 (i.e., the variance accounted for by the construct was greater than the variance due to measurement error) (Fornell & Larcker, 1981)
8. Factor AVE should exceed the shared variance between each pair of factors, and may be used to evaluate discriminant validity (Fornell & Larcker, 1981)

9. In nested models (i.e., models that are hierarchically related to each other such that one model is a subset of the other (Byrne, 2012)) the chi-square difference statistic, $\chi^2_D$, was used to test the equal-fit hypothesis in which large values of $\chi^2_D$ and $p < 0.05$ result in the rejection of it (Kline, 2011) (e.g., the two-step “DIFFTEST” specification in Mplus was used to compute the chi-square difference test (Muthén & Muthén, 2012b))

In the event the two competing models had approximate fit to the same observed data based on the chi-square difference test, the parsimony principle supported a preference toward the simpler model (Kline, 2011). If an acceptable model was found, *post hoc* modification to eliminate redundant items while maintaining congeneric measurement was performed based on retaining at least three items with the highest standardized factor loadings in each dimension (Bandalos & Finney, 2010). If an acceptable model was not found, recommendations for model fit improvement were to be discussed.

**Research question two.** The latent factor structures of the September 2012 LMI data, to include separate models for the self-rating version and the subordinate-rating version, were assessed using the CFA techniques described in the previous chapter with Mplus Version 7 (Muthén & Muthén, 2012a). The fit of the unidimensional theoretical models was compared with a competing hypothetical multidimensional six-factor model. This competing model was based on an *a priori* hypothesis that the six USAF
institutional sub-competencies including *develops and inspires others, takes care of people, builds teams and coalitions, negotiating, vision,* and *adaptability* were better modeled as interrelated subscales. The same criteria as in research question one were examined to determine the adequacy of model fit, model selection, and *post hoc* modification. If an acceptable model was not found, recommendations for model fit improvement were planned to be discussed.

**Research question three.** The latent factor structure of the November 2011 CMV data was assessed using the IRT approaches (composite, consecutive, and multidimensional) described in the previous chapter with ConQuest 3.0 (Wu et al., 2012). The fit of the nine-factor theoretical model was compared with a competing hypothetical eight-factor model. This competing model was based on an *a priori* hypothesis that the *attention to detail* and *excellence* factors from the nine-factor model may be more accurately represented by a single factor described as *methodical*—a construct in which two of Rosebush’s (2011) items (e.g., “I am exacting in my work” and “I pay attention to details”) selected from the International Personality Item Pool were defined under a *methodical* factor in the Six Factor Personality Questionnaire (Jackson, Paunonen, & Tremblay, 2000).

First, to determine the initial best fitting models (i.e., one for the hypothetical and one for the theoretical) after application of the three IRT approaches (composite, consecutive, and multidimensional), the following criteria were examined:

1. Nested models (i.e., multidimensional compared with composite) were compared by the likelihood ratio test ($G^2$) with $p < 0.05$ indicating a
statistically significant better fit of the model with more $df$ (Allen & Wilson, 2006)

2. Non-nested models (i.e., multidimensional compared with consecutive) were compared on the basis of the lowest AIC (Allen & Wilson, 2006)

3. Reliabilities were compared between multidimensional and consecutive to determine which approach was closer in value to the composite approach (i.e., evaluation of multidimensional reliability enhancement) (Allen & Wilson, 2006)

4. Estimated correlations between the consecutive and multidimensional approaches were evaluated to determine if multidimensional modeling was justified (Allen & Wilson, 2006)

Second, to select the best overall model after comparing the one’s selected using the criteria above, the following tests were conducted:

1. Nested models between the theoretical and hypothetical were compared (e.g., nine-factor multidimensional versus eight-factor multidimensional) by the likelihood ratio test ($G^2$) with $p < 0.05$ indicating a statistically significant better fit of the model with more $df$ (Allen & Wilson, 2006)

2. Non-nested models (e.g., nine-factor multidimensional compared with eight-factor consecutive) were compared on the basis of the lowest AIC (Allen & Wilson, 2006)

Finally, to arrive at the best fitting modified model, the following criteria were evaluated:
1. For adequate fit, a cutoff criteria between 0.5 and 1.5 for infit and outfit was considered acceptable (de Ayala, 2009)

2. Infit and outfit mean-squares also have standardized forms based on the $t$ statistic; in these cases, the expected value of $t$ was zero with a cutoff criteria outside of the range $-2.0 \leq t \leq 2.0$ (Bond & Fox, 2007)

3. The item-person map for the best fitting model was evaluated regarding the latent ability distribution and the distribution of items relative to their endorsability

In the event the two competing models had approximate fit to the same observed data based on the likelihood ratio test, the parsimony principle supported a preference toward the simpler model (Kline, 2011). If an acceptable model was found, post hoc modification to eliminate redundant or non-fitting items while maintaining congeneric measurement was performed based on retaining at least three items in each dimension (Bandalos & Finney, 2010). If an acceptable model was not found, recommendations for model fit improvement would be discussed.

**Research question four.** The latent factor structures of the September 2012 LMI data, to include separate models for the self-rating version and the subordinate-rating version, were assessed using the IRT approaches (composite, consecutive, and multidimensional) described in the previous chapter with ConQuest 3.0 (Wu et al., 2012). The fit of the unidimensional theoretical models was compared with competing hypothetical multidimensional six-factor models. These competing models were based on an *a priori* hypothesis that the six USAF institutional sub-competencies including
develops and inspires others, takes care of people, builds teams and coalitions, negotiating, vision, and adaptability were better modeled as interrelated subscales.

The same criteria as in research question three were examined to determine the adequacy of model fit, model selection, and post hoc modification. If an acceptable model was not found, recommendations for model fit improvement would be discussed.

**Research question five.** The typological latent class structure of the best fitting post hoc modified November 2011 CMV model was assessed using the LCA techniques described in the previous chapter with Mplus Version 7 (Muthén & Muthén, 2012a). Due to highly skewed responses and to ensure sufficient values in each cell of the contingency table, the rating scales were recoded to dichotomous responses (Collins & Lanza, 2010); for example, the item responses “very much like me” and “like me” were recoded as “like me” while the item responses “neutral,” “unlike me,” and “very much unlike me” were recoded as “unlike me.”

The following criteria were examined to determine the best model fit:

1. The optimal number of classes was determined by analyzing the fit of a series of increasing class number models by comparing the $k$-class model with the $(k-1)$-class model (Wang & Wang, 2012)

2. The $p$-value from the LMR LR test was evaluated which compares the $k$-class model to the $(k-1)$-class model; a statistically significant $p$-value indicated the $k$-class model fits significantly better than the $(k-1)$-class model (Wang & Wang, 2012)
3. These model comparisons were continued iteratively until the LMR LR test was statistically non-significant between the \((k+1)\)-class model and the \(k\)-class model; this indicated that the optimal number of latent classes was \((k-1)\) since no additional statistically significant model fit improvement occurred by including the last analyzed class in the model (Wang & Wang, 2012)

4. The ALMR LR test was also evaluated since it is adjusted for sample size and model degrees of freedom (Wang & Wang, 2012)

5. The BLRT \(p\)-value, derived from the log-likelihood differences in bootstrap samples from both \(k\)-class and \((k-1)\)-class models, was evaluated in the same manner as the LMR LR test (Wang & Wang, 2012)

6. AIC, BIC, and ABIC were evaluated; the lowest value of these criterion amongst competing LCA models was considered as justification for determining the best fit model (Samuelsen & Dayton, 2010).

7. The quality of the classification of individuals was examined on the basis of the estimated posterior probabilities; individuals were assigned into a latent class based on their largest (i.e., most likely) posterior probability (e.g., Nagin’s (2005) criterion for minimum acceptable class membership classification was exceeded when average posterior probability was at least 0.7 for all groups)

8. Entropy was evaluated as a criterion to summarize posterior misclassification (e.g., Clark (2010) proposed high entropy values of 0.80, medium entropy values of 0.60, and low entropy values of 0.40)
9. Each latent class was defined in a meaningful and interpretable manner such that the heterogeneity in the population was adequately described (Wang & Wang, 2012)

10. The best fitting model was estimated with different sets of random starting values until the best log-likelihood value was the most frequent solution to provide evidence that the global maximum was reached (Samuelsen & Dayton, 2010; Wang & Wang, 2012)

If an acceptable model was not found, recommendations for model fit improvement would be discussed.

**Research question six.** The typological latent class structures of the September 2012 and the October 2012 LMI data, to include separate models for the self-rating version and the subordinate-rating version, were assessed using the LCA techniques described in the previous chapter with Mplus Version 7 (Muthén & Muthén, 2012a). As in Gerber et al. (2009), analyses of the September data were conducted in an exploratory manner, while analyses of the October data were conducted in a confirmatory manner. The October data were analyzed in an attempt to replicate the final September LCA models by fixing all response probabilities in the October models with those estimated in the September models (Gerber et al., 2009). Due to highly skewed responses and to ensure sufficient values in each cell of the contingency table, the rating scales were recoded to dichotomous responses (Collins & Lanza, 2010); for example, the item responses “very much like me/the Leader” and “like me/the Leader” were recoded as
“like me/the Leader” while the item responses “neutral,” “unlike me/the Leader,” and “very much unlike me/the Leader” were recoded as “unlike me/the Leader.”

The same criteria as in research question five were examined to determine the adequacy of model fit and model selection. Additionally, all response probabilities in the October model were constrained to the September values in order to confirm the time one results with time two data. If an acceptable model was not found, recommendations for model fit improvement would be discussed.
Chapter Three: Results

“An approximate answer to the right problem is worth a good deal more than an exact answer to an approximate problem.” – John Tukey, 1915-2000; (Li & Klette, 2011, p. 213)

This chapter reports the results of the assessment of two latent structures based on the “leader of character” definition, the Character Mosaic Virtues (CMV) and the Leadership Mosaic Inventory (LMI), through confirmatory factor (CFA), item response theory (IRT), and latent class (LCA) analyses. Results regarding the research questions posed in Chapter One are addressed in each section that follows.

Research Question One

The latent factor structure of the November 2011 CMV was assessed by CFA techniques on the nine-factor theoretical model (Rosebush, 2011) and on the competing hypothetical eight-factor model by incorporating the five-step modeling approach (Schumacker & Lomax, 2010) with Mplus Version 7 structural equation modeling software (Muthén & Muthén, 2012a).

Eight-factor hypothetical model. The eight-factor competing model, based on an *a priori* hypothesis that the *attention to detail* and *excellence* factors from the nine-factor theoretical model (Rosebush, 2011) may be more accurately represented by a single factor labeled *methodical* (Jackson, Paunonen, & Tremblay, 2000), contained 45 items forming a multidimensional construct consisting of the following factors: *courage,*
accountability, humility, duty, care for others, self-control, respect for human dignity, and methodical.

**Model specification.** The Mplus input file specification for testing the factorial validity of the eight-factor hypothetical model is displayed in Figure 20. The results of the specification illustrate that the dependent variables *item1* through *item45*, representing the polytomously scored Likert-type items, were treated as ordered categorical variables in the model and estimation process through the Mplus CATEGORICAL option (Muthén & Muthén, 2012b). Additionally, the Mplus input file specification included the MODEL command in which the courage factor was measured by *item1* through *item6*, the accountability factor was measured by *item7* through *item10*, the humility factor was measured by *item11* through *item16*, the duty factor was measured by *item17* through *item20*, the care for others factor was measured by *item21* through *item26*, the self-control factor was measured by *item27* through *item31*, the respect for human dignity factor was measured by *item32* through *item38*, and the methodical factor was measured by *item39* through *item45*. Finally, the specification also included the hypothesis that the eight factors were correlated, a default setting in Mplus (Muthén & Muthén, 2012b). A graphical representation of the specified measurement model is provided in Appendix H.
**Figure 20.** CMV Eight-Factor Hypothetical Model Specification (*Mplus* Version 7)

**Model identification.** The *Mplus* STANDARDIZED option following the OUTPUT command as depicted in Figure 20 provided UVI scaling and standardized factors by fixing their variance to “1” such that all factor loadings were free to be estimated (Byrne, 2012; Muthén & Muthén, 2012b). The model was over-identified with $d_{M} = 917$ (Schumacker & Lomax, 2010) along with each factor consisting of at least two items per factor (Kline, 2011).

**Model estimation.** The *Mplus* ESTIMATOR = WLSMV option following the ANALYSIS command as depicted in Figure 20 selected the robust weighted least squares (WLSMV) fitting function for the analysis (Byrne, 2012; Muthén & Muthén, 2012b). Prior to analysis, the input data file was screened using a cutoff value of $\pm 3.0$ for skewness and kurtosis—all items in the CMV sample met this criterion (Chaney et al., 2007).
Model testing. The eight-factor model was analyzed, serving as a means of comparison with the nine-factor model, and produced the following global model-fit results: \( \chi^2 = 1472.98, df = 917, p < 0.001; \) CFI = 0.94; TLI = 0.93; RMSEA = 0.05; RMSEA 90% CI = [0.04, 0.05]; WRMR = 1.15. A summary of the global model-fit results and their respective cutoff criteria is provided in Table 2. Only RMSEA and the RMSEA 90% CI exceeded the global cutoff criteria for good CFA model-fit based on categorical outcomes.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi^2 ) (df)</td>
<td>1472.98 (917)</td>
<td>N/A</td>
</tr>
<tr>
<td>p-value</td>
<td>&lt; 0.001</td>
<td>N/A</td>
</tr>
<tr>
<td>CFI</td>
<td>0.94</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>TLI</td>
<td>0.93</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.05</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>RMSEA 90% CI</td>
<td>[0.04, 0.05]</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>WRMR</td>
<td>1.15</td>
<td>&lt; 0.90</td>
</tr>
</tbody>
</table>

Individual standardized parameter estimates, whose significant \( p \)-values indicated the items were important to model fit (Byrne, 2012), are given in Table 3. These standardized loadings are estimated correlations between the item and its factor and indicate the reliability of the measure; the squared standardized loadings may be interpreted as proportions of variance explained (i.e., squared multiple correlation, \( R^2_{\text{smc}} \)) (Kline, 2011). For example, \textit{item1} had a standardized loading of 0.79; therefore, the courage factor explained \( 0.79^2 = 0.62 \) or 62\% of the variance of the item. According to Kline (2011), the CFA model should ideally explain the majority of the variance (e.g., \( R^2_{\text{smc}} > 0.50 \)) of each measured item. The proportion of the variance explained in the
items by their latent factors (i.e., reliability estimates) as well as the residual variances (i.e., proportion of the variance not explained) are provided in Table 4.

Table 3

*Standardized Parameter Estimates for CMV Eight-Factor Model*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1</td>
<td>0.79</td>
<td>0.04</td>
<td>18.81</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.79</td>
<td>0.04</td>
<td>19.04</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.38</td>
<td>0.06</td>
<td>6.09</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.67</td>
<td>0.04</td>
<td>15.14</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.78</td>
<td>0.05</td>
<td>14.60</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.45</td>
<td>0.06</td>
<td>7.20</td>
</tr>
<tr>
<td>Accountability</td>
<td>7</td>
<td>0.75</td>
<td>0.03</td>
<td>25.11</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.90</td>
<td>0.03</td>
<td>32.18</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.73</td>
<td>0.05</td>
<td>16.08</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.76</td>
<td>0.04</td>
<td>18.19</td>
</tr>
<tr>
<td>Humility</td>
<td>11</td>
<td>0.78</td>
<td>0.03</td>
<td>26.76</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.81</td>
<td>0.03</td>
<td>24.49</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.75</td>
<td>0.04</td>
<td>20.10</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.80</td>
<td>0.03</td>
<td>25.05</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.78</td>
<td>0.03</td>
<td>24.02</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.72</td>
<td>0.04</td>
<td>18.55</td>
</tr>
<tr>
<td>Duty</td>
<td>17</td>
<td>0.94</td>
<td>0.02</td>
<td>52.34</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.88</td>
<td>0.02</td>
<td>38.83</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.86</td>
<td>0.03</td>
<td>32.12</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.76</td>
<td>0.03</td>
<td>25.49</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21</td>
<td>0.74</td>
<td>0.04</td>
<td>20.63</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.88</td>
<td>0.03</td>
<td>27.90</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>0.71</td>
<td>0.04</td>
<td>19.40</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.69</td>
<td>0.04</td>
<td>16.73</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.69</td>
<td>0.05</td>
<td>15.11</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.81</td>
<td>0.04</td>
<td>22.44</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27</td>
<td>0.89</td>
<td>0.02</td>
<td>46.78</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.92</td>
<td>0.02</td>
<td>49.92</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>0.87</td>
<td>0.02</td>
<td>38.92</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.87</td>
<td>0.02</td>
<td>42.10</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>0.73</td>
<td>0.03</td>
<td>21.65</td>
</tr>
<tr>
<td>Respect for Human</td>
<td>32</td>
<td>0.72</td>
<td>0.03</td>
<td>21.23</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>0.74</td>
<td>0.04</td>
<td>19.68</td>
</tr>
<tr>
<td>Dignity</td>
<td>34</td>
<td>0.80</td>
<td>0.03</td>
<td>29.97</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.67</td>
<td>0.04</td>
<td>17.70</td>
</tr>
<tr>
<td>Factor</td>
<td>Item</td>
<td>Estimate</td>
<td>Std. Error</td>
<td>Estimate/S.E.</td>
</tr>
<tr>
<td>------------</td>
<td>------</td>
<td>----------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Courage</td>
<td>1</td>
<td>0.62</td>
<td>0.07</td>
<td>9.40</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.62</td>
<td>0.07</td>
<td>9.52</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.15*</td>
<td>0.05</td>
<td>3.04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.45</td>
<td>0.06</td>
<td>7.57</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.60</td>
<td>0.08</td>
<td>7.30</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.20</td>
<td>0.06</td>
<td>3.60</td>
</tr>
<tr>
<td>Accountability</td>
<td>7</td>
<td>0.56</td>
<td>0.05</td>
<td>12.56</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.82</td>
<td>0.05</td>
<td>16.09</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.54</td>
<td>0.07</td>
<td>8.04</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.58</td>
<td>0.06</td>
<td>9.10</td>
</tr>
<tr>
<td>Humility</td>
<td>11</td>
<td>0.61</td>
<td>0.05</td>
<td>13.38</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.66</td>
<td>0.05</td>
<td>12.24</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.56</td>
<td>0.06</td>
<td>10.05</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.64</td>
<td>0.05</td>
<td>12.52</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.61</td>
<td>0.05</td>
<td>12.01</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.52</td>
<td>0.06</td>
<td>9.28</td>
</tr>
<tr>
<td>Duty</td>
<td>17</td>
<td>0.87</td>
<td>0.03</td>
<td>26.17</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.77</td>
<td>0.04</td>
<td>19.42</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.74</td>
<td>0.05</td>
<td>16.06</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.57</td>
<td>0.05</td>
<td>12.74</td>
</tr>
<tr>
<td>Care for</td>
<td>21</td>
<td>0.55</td>
<td>0.05</td>
<td>10.32</td>
</tr>
<tr>
<td>Others</td>
<td>22</td>
<td>0.78</td>
<td>0.06</td>
<td>13.95</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>0.50</td>
<td>0.05</td>
<td>9.70</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.47</td>
<td>0.06</td>
<td>8.37</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.47</td>
<td>0.06</td>
<td>7.56</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.66</td>
<td>0.06</td>
<td>11.22</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27</td>
<td>0.80</td>
<td>0.03</td>
<td>23.39</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.85</td>
<td>0.03</td>
<td>24.96</td>
</tr>
</tbody>
</table>

*Note.* All estimates significant at the 0.001 level (2-tailed).

Table 4

*Reliability Estimates for CMV Eight-Factor Model*
<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>0.75</td>
<td>0.04</td>
<td>19.46</td>
<td>0.25</td>
</tr>
<tr>
<td>30</td>
<td>0.75</td>
<td>0.04</td>
<td>21.05</td>
<td>0.25</td>
</tr>
<tr>
<td>31</td>
<td>0.53</td>
<td>0.05</td>
<td>10.83</td>
<td>0.47</td>
</tr>
<tr>
<td>32</td>
<td>0.51</td>
<td>0.05</td>
<td>10.61</td>
<td>0.49</td>
</tr>
<tr>
<td>33</td>
<td>0.55</td>
<td>0.06</td>
<td>9.84</td>
<td>0.45</td>
</tr>
<tr>
<td>34</td>
<td>0.64</td>
<td>0.04</td>
<td>14.99</td>
<td>0.36</td>
</tr>
<tr>
<td>35</td>
<td>0.45</td>
<td>0.05</td>
<td>8.85</td>
<td>0.55</td>
</tr>
<tr>
<td>36</td>
<td>0.62</td>
<td>0.06</td>
<td>10.80</td>
<td>0.38</td>
</tr>
<tr>
<td>37</td>
<td>0.65</td>
<td>0.05</td>
<td>12.81</td>
<td>0.35</td>
</tr>
<tr>
<td>38</td>
<td>0.52</td>
<td>0.06</td>
<td>9.06</td>
<td>0.48</td>
</tr>
<tr>
<td>39</td>
<td>0.60</td>
<td>0.04</td>
<td>13.55</td>
<td>0.40</td>
</tr>
<tr>
<td>40</td>
<td>0.79</td>
<td>0.04</td>
<td>19.68</td>
<td>0.21</td>
</tr>
<tr>
<td>41</td>
<td>0.52</td>
<td>0.05</td>
<td>10.84</td>
<td>0.48</td>
</tr>
<tr>
<td>42</td>
<td>0.39</td>
<td>0.05</td>
<td>7.15</td>
<td>0.61</td>
</tr>
<tr>
<td>43</td>
<td>0.63</td>
<td>0.05</td>
<td>12.55</td>
<td>0.37</td>
</tr>
<tr>
<td>44</td>
<td>0.63</td>
<td>0.05</td>
<td>12.67</td>
<td>0.37</td>
</tr>
<tr>
<td>45</td>
<td>0.61</td>
<td>0.06</td>
<td>10.99</td>
<td>0.39</td>
</tr>
</tbody>
</table>

*Note.* * Item 3 significant at $p = 0.002$ (2-tailed). All other estimates significant at the 0.001 level (2-tailed).

Based on the global model-fit results and evaluation of the individual parameter estimates, the overall fit of the eight-factor model was deemed only marginally adequate. This subjective assessment is justified 1) since only the RMSEA and the RMSEA 90% CI exceeded the global model-fit cutoff criteria for categorical outcomes, and 2) while each of the standardized parameter estimates were statistically significant (see Table 3), eight item’s (e.g., item3, item4, item6, item23, item24, item25, item35, and item42) proportion of the variance *not explained* by their latent factors exceeded their proportion of the variance *explained* (see Table 4).

**Model modification.** Results from chi-square difference testing with the nine-factor theoretical model and any *post hoc* model modifications are discussed under the nine-factor model modification section.
Nine-factor theoretical model. The nine-factor theoretical model, based on Rosebush’s (2011) EFA results, contained 45 items forming a multidimensional construct consisting of the following factors: courage, accountability, humility, duty, care for others, self-control, respect for human dignity, attention to detail, and excellence.

Model specification. The Mplus input file specification for testing the factorial validity of the nine-factor theoretical model is displayed in Figure 21. The results of the specification illustrate that the dependent variables item1 through item45, representing the polytomously scored Likert-type items, were treated as ordered categorical variables in the model and estimation process through the Mplus CATEGORICAL option (Muthén & Muthén, 2012b). Additionally, the Mplus input file specification included the MODEL command in which the courage factor was measured by item1 through item6, the accountability factor was measured by item7 through item10, the humility factor was measured by item11 through item16, the duty factor was measured by item17 through item20, the care for others factor was measured by item21 through item26, the self-control factor was measured by item27 through item31, the respect for human dignity factor was measured by item32 through item38, the attention to detail factor was measured by item39 through item41, and the excellence factor was measured by item42 through item45. Finally, the specification also included the hypothesis that the nine factors were correlated, a default setting in Mplus (Muthén & Muthén, 2012b). A graphical representation of the specified measurement model is provided in Appendix I.
**Figure 21.** CMV Nine-Factor Theoretical Model Specification (Mplus Version 7)

**Model identification.** The Mplus STANDARDIZED option following the OUTPUT command as depicted in Figure 21 provided UVI scaling and standardized factors by fixing their variance to “1” such that all factor loadings were free to be estimated (Byrne, 2012; Muthén & Muthén, 2012b). The model was over-identified with $df_M = 909$ (Schumacker & Lomax, 2010) along with each factor consisting of at least two items per factor (Kline, 2011).

**Model estimation.** The Mplus ESTIMATOR = WLSMV option following the ANALYSIS command as depicted in Figure 21 selected the robust weighted least squares (WLSMV) fitting function for the analysis (Byrne, 2012; Muthén & Muthén, 2012b). Prior to analysis, the input data file was screened using a cutoff value of ±3.0 for skewness and kurtosis—all items in the CMV sample met this criterion (Chaney et al., 2007).
**Model testing.** The nine-factor model was analyzed, serving as a comparison with the eight-factor competing model, and produced the following global model-fit results: \( \chi^2 = 1384.72, df = 909, p < 0.001; \) CFI = 0.95; TLI = 0.94; RMSEA = 0.05; RMSEA 90% CI = [0.04, 0.05]; WRMR = 1.07. A summary of the global model-fit results and their respective cutoff criteria is provided in Table 5. Only RMSEA and the RMSEA 90% CI exceeded the global cutoff criteria for good CFA model-fit based on categorical outcomes.

Table 5

*CMV Nine-Factor Global Model-Fit Results*

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \chi^2 ) (df)</td>
<td>1384.72(909)</td>
<td>N/A</td>
</tr>
<tr>
<td>( p )-value</td>
<td>&lt; 0.001</td>
<td>N/A</td>
</tr>
<tr>
<td>CFI</td>
<td>0.95</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>TLI</td>
<td>0.94</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.05</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>RMSEA 90% CI</td>
<td>[0.04, 0.05]</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>WRMR</td>
<td>1.07</td>
<td>&lt; 0.90</td>
</tr>
</tbody>
</table>

Individual standardized parameter estimates, whose significant \( p \)-values indicated the items were important to model fit (Byrne, 2012), are given in Table 6. The proportion of the variance explained in the items by their latent factors (i.e., reliability estimates) as well as the residual variances (i.e., proportion of the variance not explained) are provided in Table 7.

Table 6

*Standardized Parameter Estimates for CMV Nine-Factor Model*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1</td>
<td>0.79</td>
<td>0.04</td>
<td>18.81</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.79</td>
<td>0.04</td>
<td>19.11</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.38</td>
<td>0.06</td>
<td>6.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Value</td>
<td>Value</td>
<td>Value</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.67</td>
<td>0.04</td>
<td>15.16</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.78</td>
<td>0.05</td>
<td>14.63</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.44</td>
<td>0.06</td>
<td>7.19</td>
<td></td>
</tr>
<tr>
<td>Accountability</td>
<td>7</td>
<td>0.75</td>
<td>0.03</td>
<td>25.11</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.90</td>
<td>0.03</td>
<td>32.17</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.73</td>
<td>0.05</td>
<td>16.06</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.76</td>
<td>0.04</td>
<td>18.22</td>
</tr>
<tr>
<td>Humility</td>
<td>11</td>
<td>0.78</td>
<td>0.03</td>
<td>26.78</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.81</td>
<td>0.03</td>
<td>24.49</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.75</td>
<td>0.04</td>
<td>20.11</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.80</td>
<td>0.03</td>
<td>25.06</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.78</td>
<td>0.03</td>
<td>24.00</td>
</tr>
<tr>
<td>Duty</td>
<td>16</td>
<td>0.72</td>
<td>0.04</td>
<td>18.51</td>
</tr>
<tr>
<td>Care for</td>
<td>17</td>
<td>0.94</td>
<td>0.02</td>
<td>52.13</td>
</tr>
<tr>
<td>Others</td>
<td>18</td>
<td>0.88</td>
<td>0.02</td>
<td>38.68</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.86</td>
<td>0.03</td>
<td>31.99</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.76</td>
<td>0.03</td>
<td>25.48</td>
</tr>
<tr>
<td>Self-Control</td>
<td>21</td>
<td>0.74</td>
<td>0.04</td>
<td>20.62</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.88</td>
<td>0.03</td>
<td>27.94</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>0.71</td>
<td>0.04</td>
<td>19.40</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.69</td>
<td>0.04</td>
<td>16.76</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.69</td>
<td>0.05</td>
<td>15.16</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.81</td>
<td>0.04</td>
<td>22.45</td>
</tr>
<tr>
<td>Respect for</td>
<td>27</td>
<td>0.89</td>
<td>0.02</td>
<td>46.85</td>
</tr>
<tr>
<td>Human</td>
<td>28</td>
<td>0.92</td>
<td>0.02</td>
<td>49.92</td>
</tr>
<tr>
<td>Dignity</td>
<td>29</td>
<td>0.87</td>
<td>0.02</td>
<td>39.00</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.87</td>
<td>0.02</td>
<td>42.15</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>0.73</td>
<td>0.03</td>
<td>21.55</td>
</tr>
<tr>
<td>Attention</td>
<td>32</td>
<td>0.72</td>
<td>0.03</td>
<td>21.23</td>
</tr>
<tr>
<td>to Detail</td>
<td>33</td>
<td>0.74</td>
<td>0.04</td>
<td>19.68</td>
</tr>
<tr>
<td>Excellence</td>
<td>34</td>
<td>0.80</td>
<td>0.03</td>
<td>29.98</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.67</td>
<td>0.04</td>
<td>17.71</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>0.79</td>
<td>0.04</td>
<td>21.61</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>0.81</td>
<td>0.03</td>
<td>25.62</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>0.72</td>
<td>0.04</td>
<td>18.13</td>
</tr>
<tr>
<td>Attention</td>
<td>39</td>
<td>0.81</td>
<td>0.03</td>
<td>31.31</td>
</tr>
<tr>
<td>to Detail</td>
<td>40</td>
<td>0.97</td>
<td>0.02</td>
<td>46.56</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>0.78</td>
<td>0.03</td>
<td>24.36</td>
</tr>
<tr>
<td>Excellence</td>
<td>42</td>
<td>0.64</td>
<td>0.05</td>
<td>14.04</td>
</tr>
<tr>
<td></td>
<td>43</td>
<td>0.83</td>
<td>0.03</td>
<td>27.48</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>0.82</td>
<td>0.03</td>
<td>25.83</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>0.82</td>
<td>0.04</td>
<td>23.11</td>
</tr>
</tbody>
</table>

*Note.* All estimates significant at the 0.001 level (2-tailed).
Table 7
*Reliability Estimates for CMV Nine-Factor Model*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
<th>Residual Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1</td>
<td>0.62</td>
<td>0.07</td>
<td>9.41</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.62</td>
<td>0.07</td>
<td>9.56</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.15*</td>
<td>0.05</td>
<td>3.04</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.45</td>
<td>0.06</td>
<td>7.58</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.60</td>
<td>0.08</td>
<td>7.32</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.20</td>
<td>0.06</td>
<td>3.60</td>
<td>0.80</td>
</tr>
<tr>
<td>Accountability</td>
<td>7</td>
<td>0.56</td>
<td>0.05</td>
<td>12.55</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.82</td>
<td>0.05</td>
<td>16.09</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.54</td>
<td>0.07</td>
<td>8.03</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.58</td>
<td>0.06</td>
<td>9.11</td>
<td>0.42</td>
</tr>
<tr>
<td>Humility</td>
<td>11</td>
<td>0.61</td>
<td>0.05</td>
<td>13.39</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.66</td>
<td>0.05</td>
<td>12.25</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.56</td>
<td>0.06</td>
<td>10.05</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.64</td>
<td>0.05</td>
<td>12.53</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.61</td>
<td>0.05</td>
<td>12.00</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.52</td>
<td>0.06</td>
<td>9.25</td>
<td>0.48</td>
</tr>
<tr>
<td>Duty</td>
<td>17</td>
<td>0.88</td>
<td>0.03</td>
<td>26.06</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.77</td>
<td>0.04</td>
<td>19.34</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.74</td>
<td>0.05</td>
<td>15.99</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.57</td>
<td>0.05</td>
<td>12.74</td>
<td>0.43</td>
</tr>
<tr>
<td>Care for</td>
<td>21</td>
<td>0.55</td>
<td>0.05</td>
<td>10.31</td>
<td>0.45</td>
</tr>
<tr>
<td>Others</td>
<td>22</td>
<td>0.78</td>
<td>0.06</td>
<td>13.97</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>0.50</td>
<td>0.05</td>
<td>9.70</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.47</td>
<td>0.06</td>
<td>8.38</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.47</td>
<td>0.06</td>
<td>7.58</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.65</td>
<td>0.06</td>
<td>11.22</td>
<td>0.35</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27</td>
<td>0.80</td>
<td>0.03</td>
<td>23.43</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.85</td>
<td>0.03</td>
<td>24.96</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>0.76</td>
<td>0.04</td>
<td>19.50</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.75</td>
<td>0.04</td>
<td>21.07</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>0.53</td>
<td>0.05</td>
<td>10.78</td>
<td>0.47</td>
</tr>
<tr>
<td>Respect for</td>
<td>32</td>
<td>0.51</td>
<td>0.05</td>
<td>10.61</td>
<td>0.49</td>
</tr>
<tr>
<td>Human</td>
<td>33</td>
<td>0.55</td>
<td>0.06</td>
<td>9.84</td>
<td>0.45</td>
</tr>
<tr>
<td>Dignity</td>
<td>34</td>
<td>0.64</td>
<td>0.04</td>
<td>14.99</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>0.45</td>
<td>0.05</td>
<td>8.85</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>36</td>
<td>0.62</td>
<td>0.06</td>
<td>10.80</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>0.65</td>
<td>0.05</td>
<td>12.81</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>38</td>
<td>0.52</td>
<td>0.06</td>
<td>9.07</td>
<td>0.48</td>
</tr>
<tr>
<td>Attention</td>
<td>39</td>
<td>0.66</td>
<td>0.04</td>
<td>15.65</td>
<td>0.34</td>
</tr>
</tbody>
</table>
Based on the global model-fit results and evaluation of the individual parameter estimates, the overall fit of the nine-factor model was also deemed only marginally adequate. This subjective assessment is based on the result that 1) only the RMSEA and the RMSEA 90% CI exceeded the global model-fit cutoff criteria for categorical outcomes, and 2) while each of the standardized parameter estimates were statistically significant (see Table 6), eight item’s (e.g., item3, item4, item6, item23, item24, item25, item35, and item42) proportion of the variance not explained by their latent factors exceeded their proportion of the variance explained (see Table 7).

**Model modification.** A two-step “DIFFTEST” specification is incorporated in Mplus Version 7 in which the derivatives of the less restrictive model are used in the estimation of the more restrictive model allowing a chi-square difference test using the WLSMV estimators (Muthén & Muthén, 2012b). According to Muthén and Muthén (2012b), it is not appropriate to conduct the test in the conventional manner (e.g., subtracting the difference between the chi-square values and the degrees of freedom) since the chi-square difference under WLSMV estimation is not distributed as a chi-square. Using this specification, the chi-square difference statistic, $\chi^2_D (8) = 56.56, p < 0.001$, indicated that the overall fit of the nine-factor theoretical model was a statistical improvement over the eight-factor hypothetical model. Therefore, with regard to

|    | 40 | 0.93 | 0.04 | 23.28 | 0.07 |
|    | 41 | 0.61 | 0.05 | 12.18 | 0.39 |
| Excellence | 42 | 0.41 | 0.06 | 7.02  | 0.59 |
|    | 43 | 0.69 | 0.05 | 13.74 | 0.31 |
|    | 44 | 0.68 | 0.05 | 12.92 | 0.32 |
|    | 45 | 0.66 | 0.06 | 11.55 | 0.34 |

*Item 3 significant at $p = 0.002$ (2-tailed). All other estimates significant at the 0.001 level (2-tailed).*
research question one, analysis of the CMV data using CFA techniques did yield a dimensional structure consistent with Rosebush’s (2011) EFA results.

While this analysis supported a nine-factor dimensional structure, additional post hoc model modification was necessary in order to yield the desirable psychometric properties of acceptable model fit, construct reliability, and construct validity. In order to eliminate redundant and poor performing items while maintaining congeneric measurement, another CFA was specified based on retaining the three items in each dimension with the highest standardized factor loadings. The nine-factor modified model contained 27 items and was specified for testing according to the Mplus input file displayed in Figure 22. A graphical representation of the CMV nine-factor modified model is illustrated in Figure 23.

```
Mplus VERSION 7
MUTHEN & MUTHEN
12/07/2012  3:12 PM

INPUT INSTRUCTIONS
TITLE: CFA--USAFA CMV--Nov 11--27 items--n=253--9 Factors--3 Items Each Factor:
DATA:
FILE IS C:\Users\David\Desktop\Dissertation\CFA\CMV\SFactorx3Items\CMV_45_Items\n=253_AppA.dat;
VARIABLE:
NAMES ARE item1-item45;
CATEGORICAL ARE item1-item45;
USEVARIABLES ARE item1-item2 item5 item7-item8 item10-item12 item14 item17-item20 item21-item22 item26-item29 item34 item36-item37 item39-item41 item43-item45;
ANALYSIS:
TYPE = GENERAL;
ESTIMATOR = MLXV;
MODEL:
Courage BY item1-item2 item3;
ACCOUNT BY item7-item8 item10;
HUMILITY BY item11-item12 item14;
DUTY BY item17-item19;
SC BY item21-item22 item26;
RFHD BY item34 item36-item37;
ADP BY item39-item41;
EXCEL BY item43-item45;

OUTPUT: STANDARDIZED MOD;
```

*Figure 22. CMV Nine-Factor Modified Model Specification (Mplus Version 7)*
The nine-factor modified model produced the following global model-fit results:

\[ \chi^2 = 540.23, \text{df} = 288, p < 0.001; \ CFI = 0.96; \ TLI = 0.95; \ RMSEA = 0.06; \ RMSEA 90\% \ CI = [0.05, 0.07]; \ WRMR = 0.94. \]

A summary of the global model-fit results and their respective cutoff criteria is provided in Table 8. Every criterion except WRMR and the RMSEA upper bound exceeded the global cutoff criteria for good CFA model-fit based
on categorical outcomes. However, according to Muthén (2010a), WRMR is considered an experimental fit statistic and may be disregarded if all of the remaining global model-fit criteria are met (msg. 2).

Table 8
CMV Nine-Factor Global Modified Model-Fit Results

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2_{(df)}$</td>
<td>540.23 (288)</td>
<td>N/A</td>
</tr>
<tr>
<td>$p$-value</td>
<td>&lt; 0.001</td>
<td>N/A</td>
</tr>
<tr>
<td>CFI</td>
<td>0.96</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>TLI</td>
<td>0.95*</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.06*</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>RMSEA 90% CI</td>
<td>[0.05, 0.07]</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>WRMR</td>
<td>0.94</td>
<td>&lt; 0.90</td>
</tr>
</tbody>
</table>

*Note. TLI = 0.951 and RMSEA = 0.059.

Individual standardized parameter estimates, whose significant $p$-values indicated the items were important to model fit (Byrne, 2012), are given in Table 9. The proportion of the variance explained in the items by their latent factors (i.e., reliability estimates) as well as the residual variances (i.e., proportion of the variance not explained) are provided in Table 10.

Table 9
Standardized Parameter Estimates for CMV Nine-Factor Modified Model

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1</td>
<td>0.77</td>
<td>0.05</td>
<td>17.24</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.80</td>
<td>0.05</td>
<td>17.80</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.76</td>
<td>0.06</td>
<td>13.50</td>
</tr>
<tr>
<td>Accountability</td>
<td>7</td>
<td>0.73</td>
<td>0.03</td>
<td>21.34</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.93</td>
<td>0.04</td>
<td>26.32</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.78</td>
<td>0.04</td>
<td>18.72</td>
</tr>
<tr>
<td>Humility</td>
<td>11</td>
<td>0.77</td>
<td>0.04</td>
<td>19.70</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.81</td>
<td>0.04</td>
<td>18.41</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.81</td>
<td>0.04</td>
<td>22.13</td>
</tr>
<tr>
<td>Factor</td>
<td>Item</td>
<td>Estimate</td>
<td>Std. Error</td>
<td>Estimate/S.E.</td>
</tr>
<tr>
<td>-------------</td>
<td>------</td>
<td>----------</td>
<td>------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Courage</td>
<td>1</td>
<td>0.59</td>
<td>0.07</td>
<td>8.62</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.63</td>
<td>0.07</td>
<td>8.90</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.57</td>
<td>0.08</td>
<td>6.75</td>
</tr>
<tr>
<td>Accountability</td>
<td>7</td>
<td>0.53</td>
<td>0.05</td>
<td>10.67</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.87</td>
<td>0.07</td>
<td>13.16</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.61</td>
<td>0.07</td>
<td>9.36</td>
</tr>
<tr>
<td>Humility</td>
<td>11</td>
<td>0.59</td>
<td>0.06</td>
<td>9.85</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.65</td>
<td>0.07</td>
<td>9.20</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.66</td>
<td>0.06</td>
<td>11.06</td>
</tr>
<tr>
<td>Duty</td>
<td>17</td>
<td>0.83</td>
<td>0.04</td>
<td>23.11</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.78</td>
<td>0.04</td>
<td>19.98</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.76</td>
<td>0.05</td>
<td>16.73</td>
</tr>
<tr>
<td>Care for</td>
<td>21</td>
<td>0.44</td>
<td>0.06</td>
<td>7.41</td>
</tr>
<tr>
<td>Others</td>
<td>22</td>
<td>0.88</td>
<td>0.07</td>
<td>13.41</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.60</td>
<td>0.07</td>
<td>8.84</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27</td>
<td>0.83</td>
<td>0.04</td>
<td>22.64</td>
</tr>
<tr>
<td></td>
<td>28</td>
<td>0.84</td>
<td>0.04</td>
<td>21.71</td>
</tr>
</tbody>
</table>

Note. All estimates significant at the 0.001 level (2-tailed).

Table 10
Reliability Estimates for CMV Nine-Factor Modified Model
<table>
<thead>
<tr>
<th></th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect for Human Dignity</td>
<td>0.76</td>
<td>0.44</td>
<td>0.63</td>
<td>0.73</td>
<td>0.67</td>
</tr>
<tr>
<td>Attention to Detail Excellence</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>18.11</td>
<td>8.41</td>
<td>9.85</td>
<td>12.03</td>
<td>16.99</td>
</tr>
<tr>
<td></td>
<td>0.24</td>
<td>0.56</td>
<td>0.37</td>
<td>0.27</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>0.93</td>
<td>0.82</td>
<td>0.93</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>11.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. All estimates significant at the 0.001 level (2-tailed).

In addition to the evidence for construct validity provided by the significant factor loadings on each subscale (i.e., dimension) (Anderson & Gerbing, 1988), the following equation repeated from Chapter One was used to calculate the construct reliability (i.e., convergent validity), $\rho_\eta$, where $\lambda$ represents item standardized loadings and $\epsilon$ represents item measurement error (Fornell & Larcker, 1981):

$$
\rho_\eta = \frac{(\Sigma \lambda)^2}{(\Sigma \lambda)^2 + \Sigma \epsilon}
$$

Table 1 lists the construct reliabilities which ranged from 0.82 to 0.93 for each of the CMV factors; values greater than 0.70 demonstrated evidence for adequate convergent validity.

Fornell and Larcker (1981) also recommended another more conservative test, average variance extracted (AVE), to capture the amount of variance in the construct related to the amount of variance due to measurement error. AVE was calculated with the following equation repeated from Chapter One:

$$
\rho_{\text{AVE}}(\eta) = \frac{\Sigma(\lambda^2)}{\Sigma(\lambda^2) + \Sigma \epsilon}
$$
While having an AVE in excess of 0.50 is ideal (i.e., the variance accounted for by the construct is greater than the variance due to measurement error), less stringent evidence for convergent validity may be established on the basis of construct reliability alone (Fornell & Larcker, 1981).

Additionally, Fornell and Larcker (1981) suggested factor AVE should exceed the shared variance between each pair of factors, and may be used to evaluate discriminant validity. Table 11 lists the AVE in bold on the diagonal with the shared variances below the diagonal; in all cases, the factor AVE of the CMV modified model exceeded the shared variance between each pair of factors and therefore demonstrated adequate discriminant validity. Table 12 provides the construct-level correlation matrix.
Table 11

**CMV Convergent and Discriminant Validity Tests**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Reliability</th>
<th>Courage</th>
<th>Account</th>
<th>Humility</th>
<th>Duty</th>
<th>CFO</th>
<th>SC</th>
<th>RfHD</th>
<th>AtD</th>
<th>Excellence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.60</td>
</tr>
<tr>
<td>Account</td>
<td>0.86</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>Humility</td>
<td>0.84</td>
<td>0.06</td>
<td>0.18</td>
<td>0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duty</td>
<td>0.92</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFO</td>
<td>0.84</td>
<td>0.24</td>
<td>0.18</td>
<td>0.21</td>
<td>0.04</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>0.93</td>
<td>0.24</td>
<td>0.17</td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
<td>0.81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RfHD</td>
<td>0.82</td>
<td>0.11</td>
<td>0.23</td>
<td>0.31</td>
<td>0.05</td>
<td>0.49</td>
<td>0.15</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AtD</td>
<td>0.89</td>
<td>0.07</td>
<td>&lt;0.01</td>
<td>&lt;0.01</td>
<td>0.17</td>
<td>0.02</td>
<td>0.05</td>
<td>0.01</td>
<td>0.73</td>
<td></td>
</tr>
<tr>
<td>Excellence</td>
<td>0.86</td>
<td>0.16</td>
<td>0.02</td>
<td>0.01</td>
<td>0.26</td>
<td>0.06</td>
<td>0.21</td>
<td>0.01</td>
<td>0.48</td>
<td>0.67</td>
</tr>
</tbody>
</table>

*Note.* Average variance extracted is shown in **bold** along the diagonal. Shared variances are shown below the diagonal. Account = Accountability; CFO = Care for Others; SC = Self-Control; RfHD = Respect for Human Dignity; AtD = Attention to Detail.
Table 12

**CMV Construct-Level Correlation Matrix**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Courage</th>
<th>Account</th>
<th>Humility</th>
<th>Duty</th>
<th>CFO</th>
<th>SC</th>
<th>RfHD</th>
<th>AtD</th>
<th>Excellence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Account</td>
<td>0.41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humility</td>
<td>0.24</td>
<td>0.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duty</td>
<td>0.26</td>
<td>0.24</td>
<td>0.22</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CFO</td>
<td>0.49</td>
<td>0.43</td>
<td>0.46</td>
<td>0.20</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC</td>
<td>0.49</td>
<td>0.41</td>
<td>0.38</td>
<td>0.39</td>
<td>0.38</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RfHD</td>
<td>0.33</td>
<td>0.48</td>
<td>0.55</td>
<td>0.21</td>
<td>0.70</td>
<td>0.39</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AtD</td>
<td>0.25</td>
<td>0.04</td>
<td>0.02</td>
<td>0.42</td>
<td>0.12</td>
<td>0.23</td>
<td>0.07</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Excellence</td>
<td>0.40</td>
<td>0.15</td>
<td>0.12</td>
<td>0.51</td>
<td>0.25</td>
<td>0.46</td>
<td>0.09</td>
<td>0.69</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note. Account = Accountability; CFO = Care for Others; SC = Self-Control; RfHD = Respect for Human Dignity; AtD = Attention to Detail.*
Based on the global model-fit results, evaluation of the individual parameter estimates, and the convergent and discriminant validity tests, the overall fit of the nine-factor modified model was determined to be adequate. This subjective assessment was justified 1) since every fit statistic except WRMR and the RMSEA upper bound exceeded the global cutoff criteria for good CFA model-fit based on categorical outcomes, and 2) while each of the standardized parameter estimates were statistically significant (see Table 9), only two item’s (e.g., item21 and item34) proportion of the variance not explained by their latent factors exceeded their proportion of the variance explained (see Table 10). The inclusion of item21 and item34 in the modified model with resulting residual variances of 0.56 and 0.56 respectively, was necessary to comply with the study design of eliminating redundant items while maintaining congeneric measurement with at least three items per dimension.

**Research Question Two**

The latent factor structures of the September 2012 LMI self-rating and subordinate-rating versions were assessed by CFA techniques on the unidimensional theoretical model (Rosebush, 2012) and on the competing six-factor hypothetical model by incorporating the five-step modeling approach (Schumacker & Lomax, 2010) with Mplus Version 7 structural equation modeling software (Muthén & Muthén, 2012a).

**Six-factor hypothetical model.** The six-factor competing model, based on an a priori hypothesis that the six USAF institutional sub-competencies including develops and inspires others, takes care of people, builds teams and coalitions, negotiating, vision, and adaptability were better modeled as interrelated subscales, contained 29 items
forming a multidimensional construct consisting of these six respective factors for both the *self-rating* and *subordinate-rating* versions.

**Model specification.** The *Mplus* input file specification for testing the factorial validity of the six-factor hypothetical model for the *self-rating* version is displayed in Figure 24. The only difference in the input file specification for the *subordinate-rating* version was the data file. The results of the specification illustrate that the dependent variables *item1* through *item29*, representing the polytomously scored Likert-type items, were treated as ordered categorical variables in the model and estimation process through the *Mplus* CATEGORICAL option (Muthén & Muthén, 2012b). Additionally, the *Mplus* input file specification included the MODEL command in which the *develops and inspires others* factor was measured by *item1* through *item6*, the *takes care of people* factor was measured by *item7* through *item9*, the *builds teams and coalitions* factor was measured by *item10* through *item15*, the *negotiating* factor was measured by *item16* through *item18*, the *vision* factor was measured by *item19* through *item24*, and the *adaptability* factor was measured by *item25* through *item29*. Finally, these specifications also included the hypothesis that the six factors were correlated, a default setting in *Mplus* (Muthén & Muthén, 2012b). A graphical representation of the specified measurement model is provided in Appendix J.
Model identification. The Mplus STANDARDIZED option following the OUTPUT command as depicted in Figure 24 provided UVI scaling and standardized factors by fixing their variance to “1” such that all factor loadings were free to be estimated (Byrne, 2012; Muthén & Muthén, 2012b). Both analyses were over-identified with $df_M = 362$ (Schumacker & Lomax, 2010) along with each factor consisting of at least two items per factor (Kline, 2011).

Model estimation. The Mplus ESTIMATOR = WLSMV option following the ANALYSIS command as depicted in Figure 24 selected the robust weighted least squares (WLSMV) fitting function for the analyses (Byrne, 2012; Muthén & Muthén, 2012b). Prior to the analyses, the input data files were screened using a cutoff value of ±3.0 for skewness and kurtosis—all items in the self-rating sample met this criterion; all items in the subordinate-rating sample met this criterion except for item6 (kurtosis = 3.49), item8 (kurtosis = 3.94), item11 (kurtosis = 3.10), and item21 (kurtosis = 3.61) (Chaney et al.,
Violation of this assumption, according to simulation studies, result in positively biased chi-square model fit statistics due to negatively biased standard errors (Flora & Curran, 2004) which should be taken into account when interpreting the results.

**Model testing.** Both the self-rating and the subordinate-rating six-factor models were analyzed, but their solutions were found inadmissible since the latent variable covariance matrix was not positive definite (Muthén & Muthén, 2012a). The inadmissibility of the solutions was due to several of the model estimated correlations being greater than or equal to one between the following subscales—for the self-rating analysis, correlations between builds teams and coalitions and develops and inspires others, between builds teams and coalitions and takes care of people, and between vision and develops and inspires others were all greater than one; for the subordinate-rating analysis, correlations between builds teams and coalitions and develops and inspires others, between vision and develops and inspires others, between vision and takes care of people, and between vision and builds teams and coalitions were greater than or equal to one. According to Muthén (2006), when the estimated correlations between two latent variables are greater than or equal to one, the solutions are not admissible since the respective factors are not statistically distinguishable (msg. 15). Table 13 and Table 14 provide the self-rating and subordinate-rating construct-level correlation matrices, respectively; problematic correlations are highlighted in bold.
Table 13
*LMI Self-Rating Construct-Level Correlation Matrix*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>DaIO</th>
<th>TCoP</th>
<th>BTaC</th>
<th>Neg</th>
<th>Vision</th>
<th>Adapt</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCoP</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTaC</td>
<td>1.02</td>
<td>1.01</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg</td>
<td>0.83</td>
<td>0.87</td>
<td>0.84</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vision</td>
<td>1.06</td>
<td>0.99</td>
<td>0.99</td>
<td>0.79</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Adapt</td>
<td>0.64</td>
<td>0.83</td>
<td>0.69</td>
<td>0.69</td>
<td>0.64</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* Problematic estimated correlations in **bold.** DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions; Neg = Negotiating; Adapt = Adaptability.

Table 14
*LMI Subordinate-Rating Construct-Level Correlation Matrix*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>DaIO</th>
<th>TCoP</th>
<th>BTaC</th>
<th>Neg</th>
<th>Vision</th>
<th>Adapt</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCoP</td>
<td>0.98</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BTaC</td>
<td>1.00</td>
<td>0.99</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neg</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vision</td>
<td>1.02</td>
<td>1.00</td>
<td>1.00</td>
<td>0.97</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Adapt</td>
<td>0.92</td>
<td>0.95</td>
<td>0.92</td>
<td>0.92</td>
<td>0.93</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* Problematic estimated correlations in **bold.** DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions; Neg = Negotiating; Adapt = Adaptability.

**Model modification.** CFA *post hoc* model modifications were not appropriate for these data. According to Muthén (2010b), when correlations between factors are high, dimensionality should be reexamined through Exploratory Factor Analysis (EFA) (msg. 36).

**Unidimensional theoretical model.** The unidimensional theoretical model, based on Rosebush’s (2012) EFA results, contained 29 items forming a single construct to measure *leadership effectiveness* for both the *self-rating* and *subordinate-rating* versions.
**Model specification.** The Mplus input file specification for testing the factorial validity of the unidimensional theoretical model for the self-rating version is displayed in Figure 25. The only difference in the input file specification for the subordinate-rating version was the data file. The results of the specification illustrate that the dependent variables item1 through item29, representing the polytomously scored Likert-type items, were treated as ordered categorical variables in the model and estimation process through the Mplus CATEGORICAL option (Muthén & Muthén, 2012b). Additionally, the Mplus input file specification included the MODEL command in which the leadership effectiveness factor was measured by item1 through item29. A graphical representation of the specified measurement model is provided in Appendix K.

```
Mplus VERSION 7
MUTHEN & MUTHEN
02/11/2013  2:58 PM

INPUT INSTRUCTIONS
TITLE: CFA--USAFA LMI Self--Sep 12--29 items--n=357--1 Factor;
DATA:
FILE IS C:\Users\David\Desktop\Dissertation\CFA\LMI\Selffactor\LMI_self_t1_29_items_n=357_App4.dat;
VARIABLE:
NAMES ARE item1-item29;
CATEGORICAL ARE item1-item29;
ANALYSIS:
TYPE = GENERAL;
ESTIMATOR = WLSMV;
MODEL:
Ldrship by item1-item29;
OUTPUT: STANDARDIZED MOD;
```

*Figure 25. LMI Self-Rating Unidimensional Theoretical Model Specification (Mplus Version 7)*

**Model identification.** The Mplus STANDARDIZED option following the OUTPUT command as depicted in Figure 25 provided UVI scaling and a standardized factor by fixing the variance to “1” such that the factor loadings were free to be estimated (Byrne, 2012; Muthén & Muthén, 2012b). Both analyses were over-identified with $df_M =$
377 (Schumacker & Lomax, 2010) along with the single factor consisting of at least three items (Kline, 2011).

**Model estimation.** The *Mplus* ESTIMATOR = WLSMV option following the ANALYSIS command as depicted in Figure 25 selected the robust weighted least squares (WLSMV) fitting function for the analyses (Byrne, 2012; Muthén & Muthén, 2012b). Prior to the analyses, the input data files were screened using a cutoff value of ±3.0 for skewness and kurtosis—all items in the *self-rating* sample met this criterion; all items in the *subordinate-rating* sample met this criterion except for *item 6* (kurtosis = 3.49), *item 8* (kurtosis = 3.94), *item 11* (kurtosis = 3.10), and *item 21* (kurtosis = 3.61) (Chaney et al., 2007). Violation of this assumption, according to simulation studies, result in positively biased chi-square model fit statistics due to negatively biased standard errors (Flora & Curran, 2004) which should be taken into account when interpreting the results.

**Model testing.** CFA results from the *self-rating* data are provided followed by CFA results from the *subordinate-rating* data.

**Self-rating data.** The unidimensional modeling of the *self-rating* data produced the following global model-fit results: \( \chi^2 = 1566.39, df = 377, p < 0.001; \) CFI = 0.88; TLI = 0.87; RMSEA = 0.09; RMSEA 90% CI = [0.09, 0.10]; WRMR = 1.79. A summary of the global model-fit results and their respective cutoff criteria is provided in Table 15. None of the fit statistics exceeded the global cutoff criteria for good CFA model-fit based on categorical outcomes.
Table 15
LMI Self-Rating Unidimensional Global Model-Fit Results

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2_{(df)}$</td>
<td>1566.39</td>
<td>N/A</td>
</tr>
<tr>
<td>$p$-value</td>
<td>&lt; 0.001</td>
<td>N/A</td>
</tr>
<tr>
<td>CFI</td>
<td>0.88</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>TLI</td>
<td>0.87</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.09</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>RMSEA 90% CI</td>
<td>[0.09, 0.10]</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>WRMR</td>
<td>1.79</td>
<td>&lt; 0.90</td>
</tr>
</tbody>
</table>

Individual standardized parameter estimates, whose significant $p$-values indicated the items were important to model fit (Byrne, 2012), are given in Table 16. The proportion of the variance explained in the items by the latent factor (i.e., reliability estimates) as well as the residual variances (i.e., proportion of the variance not explained) are provided in Table 17.

Table 16
Standardized Parameter Estimates for LMI Self-Rating Unidimensional Model

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership</td>
<td>1</td>
<td>0.70</td>
<td>0.03</td>
<td>23.56</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.70</td>
<td>0.03</td>
<td>27.24</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.54</td>
<td>0.04</td>
<td>14.07</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.70</td>
<td>0.03</td>
<td>25.77</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.69</td>
<td>0.03</td>
<td>22.51</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.59</td>
<td>0.04</td>
<td>16.27</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.71</td>
<td>0.03</td>
<td>25.22</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.45</td>
<td>0.05</td>
<td>9.18</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.61</td>
<td>0.04</td>
<td>16.45</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.65</td>
<td>0.03</td>
<td>22.34</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.59</td>
<td>0.04</td>
<td>16.38</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.60</td>
<td>0.04</td>
<td>16.89</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.59</td>
<td>0.04</td>
<td>16.17</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.69</td>
<td>0.03</td>
<td>23.91</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.70</td>
<td>0.03</td>
<td>24.78</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.63</td>
<td>0.03</td>
<td>18.84</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.68</td>
<td>0.03</td>
<td>22.26</td>
</tr>
</tbody>
</table>
Table 17

*Reliability Estimates for LMI Self-Rating Unidimensional Model*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
<th>Residual Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership</td>
<td>1</td>
<td>0.49</td>
<td>0.04</td>
<td>11.78</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.49</td>
<td>0.04</td>
<td>13.62</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.30</td>
<td>0.04</td>
<td>7.04</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.49</td>
<td>0.04</td>
<td>12.88</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.47</td>
<td>0.04</td>
<td>11.25</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.35</td>
<td>0.04</td>
<td>8.13</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.51</td>
<td>0.04</td>
<td>12.61</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.20</td>
<td>0.04</td>
<td>4.59</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.38</td>
<td>0.05</td>
<td>8.22</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.42</td>
<td>0.04</td>
<td>11.17</td>
<td>0.58</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.34</td>
<td>0.04</td>
<td>8.19</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.36</td>
<td>0.04</td>
<td>8.44</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.35</td>
<td>0.04</td>
<td>8.09</td>
<td>0.65</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.48</td>
<td>0.04</td>
<td>11.96</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.49</td>
<td>0.04</td>
<td>12.39</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.40</td>
<td>0.04</td>
<td>9.42</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.46</td>
<td>0.04</td>
<td>11.13</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.23</td>
<td>0.04</td>
<td>6.29</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.30</td>
<td>0.04</td>
<td>7.05</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.49</td>
<td>0.04</td>
<td>11.44</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>0.47</td>
<td>0.05</td>
<td>9.16</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.49</td>
<td>0.04</td>
<td>12.05</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>0.48</td>
<td>0.04</td>
<td>11.55</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.44</td>
<td>0.04</td>
<td>11.84</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.46</td>
<td>0.04</td>
<td>12.29</td>
<td>0.54</td>
</tr>
</tbody>
</table>

*Note.* All estimates significant at the 0.001 level (2-tailed).
Based on the global model-fit results and evaluation of the individual parameter estimates, the overall fit of the unidimensional model was deemed inadequate. This subjective assessment is based on 1) none of the fit statistics exceeding the global model-fit cutoff criteria for categorical outcomes, and 2) while each of the standardized parameter estimates were statistically significant (see Table 16), 26 item’s (e.g., item1-item6, item8-item25, item27, and item29) proportion of the variance not explained by the latent factor exceeded their proportion of the variance explained (see Table 17).

**Subordinate-rating data.** The unidimensional modeling of the subordinate-rating data produced the following global model-fit results: $\chi^2 = 4259.92, df = 377, p < 0.001$; CFI = 0.98; TLI = 0.98; RMSEA = 0.08; RMSEA 90% CI = [0.07, 0.08]; WRMR = 2.15. A summary of the global model-fit results and their respective cutoff criteria is provided in Table 18. Only CFI and TLI exceeded the global cutoff criteria for good CFA model-fit based on categorical outcomes.
Table 18
*LMI Subordinate-Rating Unidimensional Global Model-Fit Results*

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ (df)</td>
<td>4259.92 (377)</td>
<td>N/A</td>
</tr>
<tr>
<td>$p$-value</td>
<td>&lt; 0.001</td>
<td>N/A</td>
</tr>
<tr>
<td>CFI</td>
<td>0.98</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>TLI</td>
<td>0.98</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.08</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>RMSEA 90% CI</td>
<td>[0.07, 0.08]</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>WRMR</td>
<td>2.15</td>
<td>&lt; 0.90</td>
</tr>
</tbody>
</table>

Individual standardized parameter estimates, whose significant $p$-values indicated the items were important to model fit (Byrne, 2012), are given in Table 19. The proportion of the variance explained in the items by the latent factor (i.e., reliability estimates) as well as the residual variances (i.e., proportion of the variance not explained) are provided in Table 20.

Table 19
*Standardized Parameter Estimates for LMI Subordinate-Rating Unidimensional Model*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership</td>
<td>1</td>
<td>0.85</td>
<td>0.01</td>
<td>107.12</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.88</td>
<td>0.01</td>
<td>137.56</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.83</td>
<td>0.01</td>
<td>94.64</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.88</td>
<td>0.01</td>
<td>137.61</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.87</td>
<td>0.01</td>
<td>123.41</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.82</td>
<td>0.01</td>
<td>79.68</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.91</td>
<td>0.01</td>
<td>169.52</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.81</td>
<td>0.01</td>
<td>79.58</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.85</td>
<td>0.01</td>
<td>107.45</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.85</td>
<td>0.01</td>
<td>107.78</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.89</td>
<td>0.01</td>
<td>132.28</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.82</td>
<td>0.01</td>
<td>87.43</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.86</td>
<td>0.01</td>
<td>103.78</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.86</td>
<td>0.01</td>
<td>107.63</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.84</td>
<td>0.01</td>
<td>95.78</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.84</td>
<td>0.01</td>
<td>104.58</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.88</td>
<td>0.01</td>
<td>142.55</td>
</tr>
</tbody>
</table>
Table 20
*Reliability Estimates for LMI Subordinate-Rating Unidimensional Model*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
<th>Residual Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership</td>
<td>1</td>
<td>0.72</td>
<td>0.01</td>
<td>53.56</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>0.78</td>
<td>0.01</td>
<td>68.78</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>0.68</td>
<td>0.01</td>
<td>47.32</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.78</td>
<td>0.01</td>
<td>68.80</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.76</td>
<td>0.01</td>
<td>61.71</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>0.67</td>
<td>0.02</td>
<td>39.84</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.82</td>
<td>0.01</td>
<td>84.76</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.66</td>
<td>0.02</td>
<td>39.79</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.73</td>
<td>0.01</td>
<td>53.72</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.72</td>
<td>0.01</td>
<td>53.89</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.78</td>
<td>0.01</td>
<td>66.14</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>0.67</td>
<td>0.02</td>
<td>43.72</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.73</td>
<td>0.01</td>
<td>51.89</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.74</td>
<td>0.01</td>
<td>53.82</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>0.71</td>
<td>0.02</td>
<td>47.89</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.71</td>
<td>0.01</td>
<td>52.29</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.78</td>
<td>0.01</td>
<td>71.27</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.60</td>
<td>0.02</td>
<td>36.04</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.74</td>
<td>0.01</td>
<td>56.46</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.78</td>
<td>0.01</td>
<td>67.07</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>0.74</td>
<td>0.01</td>
<td>54.01</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.84</td>
<td>0.01</td>
<td>93.60</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>0.73</td>
<td>0.01</td>
<td>54.27</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>0.70</td>
<td>0.01</td>
<td>50.49</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.78</td>
<td>0.01</td>
<td>68.57</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.81</td>
<td>0.01</td>
<td>80.61</td>
<td>0.19</td>
</tr>
</tbody>
</table>

*Note.* All estimates significant at the 0.001 level (2-tailed).
Based on the global model-fit results and evaluation of the individual parameter estimates, the overall fit of the unidimensional model was deemed adequate. This subjective assessment was supported 1) since CFI and TLI exceeded the global cutoff criteria for good CFA model-fit based on categorical outcomes, and 2) each of the standardized parameter estimates were statistically significant (see Table 16) and every item’s proportion of the variance explained by the latent factor exceeded their proportion of the variance not explained (see Table 17).

**Model modification.** Modification results from the *self-rating* data are provided followed by modification results from the *subordinate-rating* data.

*Self-rating data. Post hoc* modification of the unidimensional model, based on the *self-rating* data, was not appropriate since the model fit was determined to be inadequate (Bandalos & Finney, 2010). Regarding this notion, Bandalos and Finney remark:

Unfortunately, many researchers lose sight of the purpose of CFA, which is to allow the testing of *a priori* models. If a model does not fit the data, that information, along with a diagnosis of the source of the misfit, is useful and should inform the domain. On the other hand, thoughtlessly modifying a model *post hoc* in an attempt to make it fit the data is not the purpose of CFA and may simply lead to models that do not replicate due to fitting the idiosyncrasies of the sample data. Researchers and reviewers must keep in mind that the purpose of conducting a CFA study is to gain a better understanding of the underlying structure of the variables, not to force models to fit. The former is a useful scientific endeavor; the latter is not. (p. 112)
In addition to the evidence resulting from the 26 items whose proportion of the variance *not explained* by the latent factor exceeded their proportion of the variance *explained*, further diagnosis of the source of the misfit was gleaned by reanalyzing the unidimensional model by consecutively collapsing the rating scale from 5-points to 2-points and observing the global fit statistics. According to Byrne (2012):

In working with categorical variables, analyses must proceed *from* a frequency table comprising the number of thresholds, multiplied by the number of observed variables, *to* estimation of the correlation matrix. The problem here lies with the occurrence of cells having zero or near-zero cases, which can subsequently lead to estimation difficulties. This problem can arise because (a) sample size is small relative to the number of response categories (i.e., specific category scores across all categorical variables), (b) the number of variables is excessively large, and/or (c) the number of thresholds is large. Taken in combination, then, the larger the number of observed variables and/or number of thresholds for these variables, and the smaller the sample size, the greater the chance of having cells comprising zero to near-zero cases. (p. 131)

The item-response frequencies regarding the *self-rating* data’s 5-point rating scale, in which all 29 items had zero or near-zero frequencies in at least one rating category, are provided in Table 21. Tabulated in Table 22 are the global fit statistics for the *self-rating* unidimensional model with 5-point through 2-point rating scales, analyzed consecutively. Only minor model fit improvement was gained by collapsing to a 2-point rating scale.

**Table 21**

*LMI Self-Rating Item-Response Frequencies*

<table>
<thead>
<tr>
<th>Item</th>
<th>Very much unlike me</th>
<th>Unlike me</th>
<th>Neutral</th>
<th>Like me</th>
<th>Very much like me</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>5</td>
<td>60</td>
<td>192</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
<td>50</td>
<td>219</td>
<td>86</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>37</td>
<td>232</td>
<td>87</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>5</td>
<td>70</td>
<td>196</td>
<td>86</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>8</td>
<td>54</td>
<td>213</td>
<td>82</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>154</td>
<td>186</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>4</td>
<td>73</td>
<td>210</td>
<td>70</td>
</tr>
</tbody>
</table>
Table 22
LMI Self-Rating Unidimensional Global Model-Fit Results by Rating Scale

<table>
<thead>
<tr>
<th>Scale</th>
<th>$\chi^2_{(df)}$</th>
<th>p-value</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA</th>
<th>RMSEA 90% CI</th>
<th>WRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-point</td>
<td>1566.39(377)</td>
<td>&lt; 0.001</td>
<td>0.88</td>
<td>0.87</td>
<td>0.09</td>
<td>[0.09, 0.10]</td>
<td>1.79</td>
</tr>
<tr>
<td>4-point</td>
<td>1506.31(377)</td>
<td>&lt; 0.001</td>
<td>0.88</td>
<td>0.88</td>
<td>0.09</td>
<td>[0.09, 0.10]</td>
<td>1.78</td>
</tr>
<tr>
<td>3-point</td>
<td>1362.50(377)</td>
<td>&lt; 0.001</td>
<td>0.88</td>
<td>0.87</td>
<td>0.09</td>
<td>[0.08, 0.09]</td>
<td>1.68</td>
</tr>
<tr>
<td>2-point</td>
<td>655.07(377)</td>
<td>&lt; 0.001</td>
<td>0.90</td>
<td>0.89</td>
<td>0.05</td>
<td>[0.04, 0.05]</td>
<td>1.23</td>
</tr>
</tbody>
</table>

Subordinate-rating data. While the analysis of the subordinate-rating data confirmed the adequacy of the unidimensional structure, additional post hoc model modification was necessary in order to demonstrate the desirable psychometric properties of acceptable model fit, construct reliability, and construct validity. In order to eliminate redundant and less performing items while maintaining representation for each of the USAF institutional sub-competencies, another CFA was specified based on retaining...
three items per sub-competency with the highest standardized factor loadings. The unidimensional modified model contained 18 items and was specified for testing according to the *Mplus* input file displayed in Figure 26. A graphical representation of the LMI subordinate-rating unidimensional modified model is illustrated in Figure 27.

```
TITLE: CFA--USAFA LMI Subs--Sep 12--18 items--r=1777--1 Factor improvement;
DATA:
  FILE IS C:\Users\david\Desktop\Dissertation\CFA\LMI\SubsFactorX3\LMI_subb_t1_29_items_r=1777_AppB.dat;
VARIABLE:
  NAMES ARE Item1-Item29;
  USEARIABLES ARE Item2 Item4-Item5 Item7-Item9 Item11 Item13-Item14 Item16-Item20 Item22 Item26-Item27 Item29;
  CATEGORICAL ARE Item2 Item4-Item5 Item7-Item9 Item11 Item13-Item14 Item16-Item20 Item22 Item26-Item27 Item29;
ANALYSIS:
  TYPE = GENERAL;
  ESTIMATOR = WLSMV;
MODEL:
  Ldrshp BY Item2 Item4-Item5 Item7-Item9 Item11 Item13-Item14 Item16-Item20 Item22 Item26-Item27 Item29;
OUTPUT: STANDARDIZED MOD;
```

*Figure 26. LMI Subordinate-Rating Unidimensional Modified Model Specification (Mplus Version 7)*
The unidimensional modified model produced the following global model-fit results: $\chi^2 = 1079.93$, $df = 135$, $p < 0.001$; CFI = 0.99; TLI = 0.99; RMSEA = 0.06; RMSEA 90% CI = [0.06, 0.07]; WRMR = 1.45. A summary of the global model-fit results and their respective cutoff criteria is provided in Table 23. The CFI, TLI, and the RMSEA lower bound exceeded the global cutoff criteria for good CFA model-fit based on categorical outcomes.
Table 23
*LMI Subordinate-Rating Unidimensional Global Modified Model-Fit Results*

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Value</th>
<th>Cutoff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$ (df)</td>
<td>1079.93$^{(135)}$</td>
<td>N/A</td>
</tr>
<tr>
<td>$p$-value</td>
<td>$&lt; 0.001$</td>
<td>N/A</td>
</tr>
<tr>
<td>CFI</td>
<td>0.99</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>TLI</td>
<td>0.99</td>
<td>&gt; 0.95</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.06</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>RMSEA 90% CI</td>
<td>[0.06*, 0.07]</td>
<td>&lt; 0.06</td>
</tr>
<tr>
<td>WRMR</td>
<td>1.45</td>
<td>&lt; 0.90</td>
</tr>
</tbody>
</table>

*Note.* RMSEA lower bound = 0.059.

Individual standardized parameter estimates, whose significant $p$-values indicated the items were important to model fit (Byrne, 2012), are given in Table 24. The proportion of the variance explained in the items by the latent factor (i.e., reliability estimates) as well as the residual variances (i.e., proportion of the variance not explained) are provided in Table 25.

Table 24
*Standardized Parameter Estimates for LMI Subordinate-Rating Unidimensional Modified Model*

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership</td>
<td>2</td>
<td>0.88</td>
<td>0.01</td>
<td>132.84</td>
</tr>
<tr>
<td>Effectiveness</td>
<td>4</td>
<td>0.88</td>
<td>0.01</td>
<td>135.08</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.87</td>
<td>0.01</td>
<td>124.22</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.91</td>
<td>0.01</td>
<td>166.53</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.80</td>
<td>0.01</td>
<td>75.77</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.86</td>
<td>0.01</td>
<td>110.02</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.88</td>
<td>0.01</td>
<td>122.71</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.85</td>
<td>0.01</td>
<td>104.08</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.86</td>
<td>0.01</td>
<td>113.40</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.84</td>
<td>0.01</td>
<td>104.88</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.89</td>
<td>0.01</td>
<td>146.30</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.78</td>
<td>0.01</td>
<td>74.44</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.86</td>
<td>0.01</td>
<td>110.81</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.88</td>
<td>0.01</td>
<td>126.19</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.92</td>
<td>0.01</td>
<td>197.10</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.89</td>
<td>0.01</td>
<td>148.03</td>
</tr>
</tbody>
</table>
Note. All estimates significant at the 0.001 level (2-tailed).

Table 25
Reliability Estimates for LMI Subordinate-Rating Unidimensional Modified Model

<table>
<thead>
<tr>
<th>Factor</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Estimate/S.E.</th>
<th>Residual Var.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leadership</td>
<td>2</td>
<td>0.78</td>
<td>0.01</td>
<td>66.42</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.78</td>
<td>0.01</td>
<td>67.54</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.76</td>
<td>0.01</td>
<td>62.11</td>
<td>0.24</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>0.82</td>
<td>0.01</td>
<td>83.27</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>0.65</td>
<td>0.02</td>
<td>37.89</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>0.73</td>
<td>0.01</td>
<td>55.01</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.77</td>
<td>0.01</td>
<td>61.36</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.73</td>
<td>0.01</td>
<td>52.04</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>0.75</td>
<td>0.01</td>
<td>56.70</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>0.71</td>
<td>0.01</td>
<td>52.44</td>
<td>0.29</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>0.79</td>
<td>0.01</td>
<td>73.15</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>0.61</td>
<td>0.02</td>
<td>37.22</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.74</td>
<td>0.01</td>
<td>55.41</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.77</td>
<td>0.01</td>
<td>63.09</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>0.85</td>
<td>0.01</td>
<td>98.55</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>0.80</td>
<td>0.01</td>
<td>74.01</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>0.80</td>
<td>0.01</td>
<td>73.99</td>
<td>0.20</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>0.83</td>
<td>0.01</td>
<td>87.55</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Note. All estimates significant at the 0.001 level (2-tailed).

In addition to the evidence for construct validity provided by the significant factor loadings on the dimension (Anderson & Gerbing, 1988), the following equation repeated from Chapter One was used to calculate the construct reliability (i.e., convergent validity), $\rho_\eta$, where $\lambda$ represents item standardized loadings and $\epsilon$ represents item measurement error (Fornell & Larcker, 1981):

$$
\rho_\eta = \frac{(\Sigma \lambda)^2}{(\Sigma \lambda)^2 + \Sigma \epsilon}
$$

(15)

The leadership effectiveness factor’s construct reliability was 0.98; a value greater than 0.70 demonstrates evidence for adequate convergent validity.
Fornell and Larcker (1981) also recommended another more conservative test, average variance extracted (AVE), to capture the amount of variance in the construct related to the amount of variance due to measurement error. An AVE of 0.76 was calculated (e.g., having an AVE in excess of 0.50 is ideal since the variance accounted for by the construct is greater than the variance due to measurement error) with the following equation repeated from Chapter One (Fornell & Larcker, 1981):

\[ \rho_{vc(n)} = \frac{\sum(\lambda^2)}{\sum(\lambda^2) + \sum \epsilon} \]  

(16)

Based on the global model-fit results, evaluation of the individual parameter estimates, and the reliability and validity tests, the overall fit of the unidimensional modified model based on subordinate-rating data was determined to be adequate. This subjective assessment was justified 1) since the CFI, TLI, and the RMSEA lower bound exceeded the global cutoff criteria for good CFA model-fit based on categorical outcomes, 2) each of the standardized parameter estimates were statistically significant (see Table 24), and 3) each of the item’s proportion of the variance explained by the latent factor exceeded their proportion of the variance not explained (see Table 25).

**Research Question Three**

The latent factor structure of the November 2011 CMV was assessed by IRT techniques on the nine-factor theoretical model (Rosebush, 2011) and on the competing eight-factor hypothetical model by incorporating Allen and Wilson’s (2006) three phased approach—composite, consecutive, and multidimensional—with ConQuest 3.0 modeling software (Wu et al., 2012).
Eight-factor hypothetical model. The eight-factor competing model, based on an *a priori* hypothesis that the *attention to detail* and *excellence* factors from the nine-factor theoretical model (Rosebush, 2011) may be more accurately represented by a single factor labeled *methodical* (Jackson, Paunonen, & Tremblay, 2000), contained 45 items forming a multidimensional construct consisting of the following dimensions: *courage, accountability, humility, duty, care for others, self-control, respect for human dignity,* and *methodical*.

Composite approach. The composite approach was applied as a means of comparison with the consecutive and multidimensional approaches. The total score, based on responses to each item on the CMV, was treated as the indicator of a single estimate (i.e., unidimensional) of a cadet’s perception of their overall *virtue* (i.e., θ). The ConQuest command file executed for fitting the Rasch rating scale model using Gauss-Hermite Quadrature estimation to the CMV data to implement the composite approach is provided in Figure 28. A graphical representation of the composite approach is illustrated in Appendix L.

```
title rating scale Analysis;
datafile C:\Users\David\Desktop\Dissertation\IRT\CMV9\Composite\CMV_45_items_n=253_Appendix_A_order.dat;
format responses 1-45;
codes 0,1,2,3,4;
recode (1,2,3,4,5) (0,2,2,3,4);
model item + step:
/*rating scale*/
estimate;
show estimates=latent >> c:\users\David\Desktop\Dissertation\IRT\CMV9\composite\composite.shw;
tana1 >> c:\users\David\Desktop\Dissertation\IRT\CMV9\composite\composite.itn;
```

*Figure 28. CMV Eight-Factor Composite Model Specification (ConQuest 3.0)*

The analysis produced estimates for 49 parameters—including the mean and variance of θ, 44 item difficulty parameters (e.g., one parameter was constrained for model identification), and three step parameters (e.g., one parameter was constrained for
model identification)—with a model fit statistic, $G^2 = 26943.35$, and a degree of parsimony fit index, AIC = 27041.35. The reliability of the perceived cadet virtue estimates was 0.86.

**Consecutive approach.** The consecutive approach was applied as a means of comparison with the composite and multidimensional approaches. This approach modeled each hypothesized CMV subscale separately as unidimensional constructs which produced independent $\theta_D$ estimates and standard errors for each dimension. The ConQuest command file executed for fitting the Rasch rating scale model using Gauss-Hermite Quadrature estimation to the CMV data to implement the consecutive approach for the methodical subscale is provided in Figure 29 (command files for the other subscales were similar). A graphical representation of the consecutive approach is illustrated in Appendix M.

---

**Figure 29.** CMV Eight-Factor Consecutive Model Specification (Methodical Subscale Only—Other Subscales Similar) (ConQuest 3.0)

Results of these independent analyses to include the model fit statistics, the number of parameters, the degree of parsimony fit indices, and the reliability of the $\theta_D$ estimates are displayed in Table 26. The analyses produced estimates for 77 parameters—including the means and variances of each $\theta_D$, 37 item difficulty parameters (e.g., one parameter was constrained per model for identification), and 24 step parameters (e.g., one parameter was constrained per model for identification).
Table 26
CMV Eight-Factor Model Consecutive Approach Fit Results

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>$G^2$</th>
<th>Parameters</th>
<th>AIC</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1-6</td>
<td>3545.78</td>
<td>10</td>
<td>3565.78</td>
<td>0.76</td>
</tr>
<tr>
<td>Accountability</td>
<td>7-10</td>
<td>2074.48</td>
<td>8</td>
<td>2090.48</td>
<td>0.81</td>
</tr>
<tr>
<td>Humility</td>
<td>11-16</td>
<td>3484.15</td>
<td>10</td>
<td>3504.15</td>
<td>0.86</td>
</tr>
<tr>
<td>Duty</td>
<td>17-20</td>
<td>1804.59</td>
<td>8</td>
<td>1820.59</td>
<td>0.87</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21-26</td>
<td>3078.21</td>
<td>10</td>
<td>3098.21</td>
<td>0.83</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27-31</td>
<td>2697.16</td>
<td>9</td>
<td>2715.16</td>
<td>0.89</td>
</tr>
<tr>
<td>RfHD</td>
<td>32-38</td>
<td>3888.29</td>
<td>11</td>
<td>3910.29</td>
<td>0.86</td>
</tr>
<tr>
<td>Methodical</td>
<td>39-45</td>
<td>3866.77</td>
<td>11</td>
<td>3888.77</td>
<td>0.84</td>
</tr>
</tbody>
</table>

*Note.* RfHD = Respect for Human Dignity.

**Multidimensional approach.** The multidimensional approach was applied as a means of comparison with the composite and consecutive approaches. By incorporating the correlations between the dimensions, this approach simultaneously estimated separate cadet virtue abilities, $\theta_{D_i}$, across each latent dimension. The ConQuest command file executed for fitting the Rasch rating scale model using Monte Carlo estimation to the CMV data to implement the multidimensional approach is provided in Figure 30. A graphical representation of the multidimensional approach is illustrated in Appendix N.
The analysis produced estimates for 84 parameters—including the means and variances of $\theta_{Di}$ through $\theta_{D8}$, 37 item difficulty parameters (e.g., one parameter was constrained per dimension for model identification), three step parameters (e.g., one parameter was constrained for model identification), and 28 unique elements of the variance-covariance matrix—with a model fit statistic, $G^2 = 24454.58$, and a degree of parsimony fit index, $AIC = 24622.58$. The reliability of the $\theta_{Di}$ estimates are displayed in Table 27.
Table 27
CMV Eight-Factor Model Multidimensional Reliabilities

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1-6</td>
<td>0.78</td>
</tr>
<tr>
<td>Accountability</td>
<td>7-10</td>
<td>0.90</td>
</tr>
<tr>
<td>Humility</td>
<td>11-16</td>
<td>0.85</td>
</tr>
<tr>
<td>Duty</td>
<td>17-20</td>
<td>0.80</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21-26</td>
<td>0.86</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27-31</td>
<td>0.87</td>
</tr>
<tr>
<td>RfHD</td>
<td>32-38</td>
<td>0.86</td>
</tr>
<tr>
<td>Methodical</td>
<td>39-45</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Note. RfHD = Respect for Human Dignity.

*Model comparisons.* Model fit results from the three dimensionality evaluation approaches are reproduced in Table 28. Since the multidimensional approach is nested in the composite approach, the likelihood ratio test was computed as the difference in deviance, which approximates a $\chi^2$ distribution with degrees of freedom equal to the difference in estimated parameters between the models: $\chi^2 = 2488.77, df = 35, p < 0.001$. Since the difference in deviance between the nested models was statistically significant, evidence existed that the multidimensional model fit the data significantly better than the composite model. However, on the basis of comparison between the non-nested models (i.e., multidimensional versus consecutive), the AIC value from the consecutive model was lower than the multidimensional model indicating an overall preference for the consecutive approach.
Reliability estimates from the three dimensionality evaluation approaches are reproduced in Table 29. Multidimensional reliability enhancement (Allen & Wilson, 2006), which was expected to occur on all dimensions, was only present in the courage, accountability, and care for others subscales (e.g., the courage multidimensional reliability exceeded its consecutive reliability). Since the reliability estimates for the respect for human dignity and methodical dimensions were equivalent and the consecutive approach produced greater reliabilities for the humility, duty, and self-control dimensions, no clear distinction could be made for a model preference based on reliability estimates alone.

Table 29
CMV Eight-Factor Model Reliabilities

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Consecutive Reliability</th>
<th>Multidimensional Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1-6</td>
<td>0.76</td>
<td>0.78</td>
</tr>
<tr>
<td>Accountability</td>
<td>7-10</td>
<td>0.81</td>
<td>0.90</td>
</tr>
<tr>
<td>Humility</td>
<td>11-16</td>
<td>0.86</td>
<td>0.85</td>
</tr>
<tr>
<td>Duty</td>
<td>17-20</td>
<td>0.87</td>
<td>0.80</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21-26</td>
<td>0.83</td>
<td>0.86</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27-31</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td>RfHD</td>
<td>32-38</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>Methodical</td>
<td>39-45</td>
<td>0.84</td>
<td>0.84</td>
</tr>
</tbody>
</table>

*Note.* RfHD = Respect for Human Dignity. Composite Reliability = 0.86.
Consecutive approach and multidimensional approach correlations are provided in Table 30. Each of the multidimensional approach correlations were greater than (or equal to between methodical and respect for human dignity) the consecutive approach correlations by 1.5 to 2.0 times. The higher overall correlations between the dimensions of the multidimensional approach illustrated the influence of the interrelatedness across the eight hypothetical variables of the CMV. These higher associated correlations provided some support for an overall preference for the multidimensional approach.

In summary, comparison of the three dimensionality approaches from the eight-factor model based on model fit, reliabilities, and estimated correlations led the researcher to select the consecutive model as the one for further comparison with the nine-factor selection. While the evidence from the examination of the reliabilities was inconclusive, the model fit testing was determined to be more influential than the inspection of the estimated correlations.
Table 30

*CMV Eight-Factor Consecutive and Multidimensional Correlation Matrix*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Courage</th>
<th>Account</th>
<th>Humility</th>
<th>Duty</th>
<th>CFO</th>
<th>SC</th>
<th>RfHD</th>
<th>Methodical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1.00</td>
<td>0.27</td>
<td>0.12</td>
<td>0.21</td>
<td>0.37</td>
<td>0.34</td>
<td>0.20</td>
<td>0.23</td>
</tr>
<tr>
<td>Account</td>
<td>0.44</td>
<td>1.00</td>
<td>0.37</td>
<td>0.19</td>
<td>0.37</td>
<td>0.33</td>
<td>0.39</td>
<td>0.07</td>
</tr>
<tr>
<td>Humility</td>
<td>0.22</td>
<td>0.55</td>
<td>1.00</td>
<td>0.13</td>
<td>0.34</td>
<td>0.32</td>
<td>0.45</td>
<td>0.04</td>
</tr>
<tr>
<td>Duty</td>
<td>0.40</td>
<td>0.35</td>
<td>0.25</td>
<td>1.00</td>
<td>0.17</td>
<td>0.34</td>
<td>0.08</td>
<td>0.43</td>
</tr>
<tr>
<td>CFO</td>
<td>0.53</td>
<td>0.53</td>
<td>0.44</td>
<td>0.31</td>
<td>1.00</td>
<td>0.29</td>
<td>0.47</td>
<td>0.14</td>
</tr>
<tr>
<td>SC</td>
<td>0.48</td>
<td>0.46</td>
<td>0.42</td>
<td>0.56</td>
<td>0.41</td>
<td>1.00</td>
<td>0.28</td>
<td>0.33</td>
</tr>
<tr>
<td>RfHD</td>
<td>0.29</td>
<td>0.57</td>
<td>0.59</td>
<td>0.16</td>
<td>0.61</td>
<td>0.36</td>
<td>1.00</td>
<td>0.02</td>
</tr>
<tr>
<td>Methodical</td>
<td>0.35</td>
<td>0.13</td>
<td>0.07</td>
<td>0.62</td>
<td>0.16</td>
<td>0.48</td>
<td>0.02</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note.* Account = Accountability; CFO = Care for Others; SC = Self-Control; RfHD = Respect for Human Dignity. Consecutive approach correlations are given above the diagonal; multidimensional approach correlations are given below the diagonal.
Nine-factor theoretical model. The nine-factor theoretical model, based on Rosebush’s (2011) EFA results, contained 45 items forming a multidimensional construct consisting of the following factors: courage, accountability, humility, duty, care for others, self-control, respect for human dignity, attention to detail, and excellence.

Composite approach. The composite approach was applied as a means of comparison with the consecutive and multidimensional approaches. The total score, based on responses to each item on the CMV, was treated as the indicator of a single estimate (i.e., unidimensional) of a cadet’s perception of their overall virtue (i.e., \( \theta \)). The ConQuest command file executed for fitting the Rasch rating scale model using Gauss-Hermite Quadrature estimation to the CMV data to implement the composite approach is provided in Figure 31. A graphical representation of the composite approach is illustrated in Appendix L.

```
title Rating scale analysis;
datafile C:\Users\David\Desktop\Dissertation\IRT\CMV\Composite\CMV_45_items_n=253_Appendix_A_order.dat;
format responses 1-4;
codes 0,1,2,3,4;
recode (1,2,3,4,5) (0,1,2,3,4);
model item + step; /*rating scale*/
estimate;
show estimates-latent >> c:\users\david\desktop\dissertation\irt\cmv\composite\composite.shw;
itana >> c:\users\david\desktop\dissertation\irt\cmv\composite\composite.itn;
```

Figure 31. CMV Nine-Factor Composite Model Specification (ConQuest 3.0)

The analysis produced estimates for 49 parameters—including the mean and variance of \( \theta \), 44 item difficulty parameters (e.g., one parameter was constrained for model identification), and three step parameters (e.g., one parameter was constrained for model identification)—with a model fit statistic, \( G^2 = 26943.35 \), and a degree of parsimony fit index, AIC = 27041.35. The reliability of the perceived cadet virtue estimates was 0.86.
**Consecutive approach.** The consecutive approach was applied as a means of comparison with the composite and multidimensional approaches. This approach modeled each theoretical CMV subscale separately as unidimensional constructs which produced independent $\theta_D$ estimates and standard errors for each dimension. The ConQuest command file executed for fitting the Rasch rating scale model using Gauss-Hermite Quadrature estimation to the CMV data to implement the consecutive approach for the *courage* subscale is provided in Figure 32 (command files for the other subscales were similar). A graphical representation of the consecutive approach is illustrated in Appendix O.

```
title rating scale analysis;
datafile c:\users\David\Desktop\Dissertation\IRT\CMV9\Consecutive\courage\CMV_courage_items_n=253_Appendix_A_order.dat;
format responses 1-6;
codes 0,1,2,3,4;
recode (1,2,3,4,5) (0,1,2,3,4);
model item + step; /"Rating Scale"/
estimate;
show estimates-laten >> c:\users\David\Desktop\Dissertation\IRT\CMV9\Consecutive\courage\courage_consecutive_shw;
titanal >> c:\users\David\Desktop\Dissertation\IRT\CMV9\Consecutive\courage\courage_consecutive_SHAN;
```

**Figure 32. CMV Nine-Factor Consecutive Model Specification (**_Courage_** Subscale Only—**_Other_** Subscales Similar) (ConQuest 3.0)**

Results of these independent analyses to include the model fit statistics, the number of parameters, the degree of parsimony fit indices, and the reliability of the $\theta_D$ estimates are displayed in Table 31. The analyses produced estimates for 81 parameters—including the means and variances of each $\theta_D$, 36 item difficulty parameters (e.g., one parameter was constrained per model for identification), and 27 step parameters (e.g., one parameter was constrained per model for identification).
Table 31
*CMV Nine-Factor Model Consecutive Approach Fit Results*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>$G^2$</th>
<th>Parameters</th>
<th>AIC</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1-6</td>
<td>3545.78</td>
<td>10</td>
<td>3565.78</td>
<td>0.76</td>
</tr>
<tr>
<td>Accountability</td>
<td>7-10</td>
<td>2074.48</td>
<td>8</td>
<td>2090.48</td>
<td>0.81</td>
</tr>
<tr>
<td>Humility</td>
<td>11-16</td>
<td>3484.15</td>
<td>10</td>
<td>3504.15</td>
<td>0.86</td>
</tr>
<tr>
<td>Duty</td>
<td>17-20</td>
<td>1804.59</td>
<td>8</td>
<td>1820.59</td>
<td>0.87</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21-26</td>
<td>3078.21</td>
<td>10</td>
<td>3098.21</td>
<td>0.83</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27-31</td>
<td>2697.16</td>
<td>9</td>
<td>2715.16</td>
<td>0.89</td>
</tr>
<tr>
<td>RfHD</td>
<td>32-38</td>
<td>3888.29</td>
<td>11</td>
<td>3910.29</td>
<td>0.86</td>
</tr>
<tr>
<td>AtD</td>
<td>39-41</td>
<td>1641.41</td>
<td>7</td>
<td>1655.41</td>
<td>0.83</td>
</tr>
<tr>
<td>Excellence</td>
<td>42-45</td>
<td>2215.68</td>
<td>8</td>
<td>2231.68</td>
<td>0.80</td>
</tr>
</tbody>
</table>

*Note.* RfHD = Respect for Human Dignity and AtD = Attention-to-Detail.

**Multidimensional approach.** The multidimensional approach was applied as a means of comparison with the composite and consecutive approaches. By incorporating the correlations between the dimensions, this approach simultaneously estimated separate cadet *virtue* abilities, $\theta_D$, across each latent dimension. The ConQuest command file executed for fitting the Rasch rating scale model using Monte Carlo estimation to the CMV data to implement the multidimensional approach is provided in Figure 33. A graphical representation of the multidimensional approach is illustrated in Appendix P.
Figure 3.3. CMV Nine-Factor Multidimensional Model Specification (ConQuest 3.0)

The analysis produced estimates for 93 parameters—including the means and variances of $\theta_{D1}$ through $\theta_{D9}$, 36 item difficulty parameters (e.g., one parameter was constrained per dimension for model identification), three step parameters (e.g., one parameter was constrained for model identification), and 36 unique elements of the variance-covariance matrix—with a model fit statistic, $G^2 = 24342.63$, and a degree of parsimony fit index, $AIC = 24528.63$. The reliability of the $\theta_{Di}$ estimates are displayed in Table 32.
Table 32  
*CMV Nine-Factor Model Multidimensional Reliabilities*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1-6</td>
<td>0.81</td>
</tr>
<tr>
<td>Accountability</td>
<td>7-10</td>
<td>0.73</td>
</tr>
<tr>
<td>Humility</td>
<td>11-16</td>
<td>0.74</td>
</tr>
<tr>
<td>Duty</td>
<td>17-20</td>
<td>0.72</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21-26</td>
<td>0.82</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27-31</td>
<td>0.82</td>
</tr>
<tr>
<td>RfHD</td>
<td>32-38</td>
<td>0.81</td>
</tr>
<tr>
<td>AtD</td>
<td>39-41</td>
<td>0.80</td>
</tr>
<tr>
<td>Excellence</td>
<td>42-45</td>
<td>0.86</td>
</tr>
</tbody>
</table>

*Note.*  RfHD = Respect for Human Dignity and AtD = Attention-to-Detail.

**Model comparisons.** Model fit results from the three dimensionality evaluation approaches are reproduced in Table 33. Since the multidimensional approach is nested in the composite approach, the likelihood ratio test was computed as the difference in deviance, which approximates a $\chi^2$ distribution with degrees of freedom equal to the difference in estimated parameters between the models: $\chi^2 = 2600.72$, $df = 44$, $p < 0.001$. Since the difference in deviance between the nested models was statistically significant, evidence existed that the multidimensional model fit the data significantly better than the composite model. Moreover, on the basis of comparison between the non-nested models (i.e., multidimensional versus consecutive), the AIC value from the multidimensional model was less than the consecutive model indicating a preference for the multidimensional model.
Table 33
*CMV Nine-Factor Model Comparisons*

<table>
<thead>
<tr>
<th>Approach</th>
<th>Parameters</th>
<th>$G^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Composite</td>
<td>49</td>
<td>26943.35</td>
<td>27041.35</td>
</tr>
<tr>
<td>Consecutive</td>
<td>81</td>
<td>24429.75</td>
<td>24591.75</td>
</tr>
<tr>
<td>Multidimensional</td>
<td>93</td>
<td>24342.63</td>
<td>24528.63</td>
</tr>
</tbody>
</table>

Reliability estimates from the three dimensionality evaluation approaches are reproduced in Table 34. Multidimensional reliability enhancement (Allen & Wilson, 2006), which was expected to occur on all dimensions, was only present in the *courage* and *excellence* subscales (e.g., the *courage* multidimensional reliability exceeded its consecutive reliability). Since the consecutive approach produced greater reliabilities than the multidimensional approach for the remaining seven dimensions, a model preference toward the consecutive approach was demonstrated.

Table 34
*CMV Nine-Factor Model Reliabilities*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Consecutive Reliability</th>
<th>Multidimensional Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1-6</td>
<td>0.76</td>
<td>0.81</td>
</tr>
<tr>
<td>Accountability</td>
<td>7-10</td>
<td>0.81</td>
<td>0.73</td>
</tr>
<tr>
<td>Humility</td>
<td>11-16</td>
<td>0.86</td>
<td>0.74</td>
</tr>
<tr>
<td>Duty</td>
<td>17-20</td>
<td>0.87</td>
<td>0.72</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21-26</td>
<td>0.83</td>
<td>0.82</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27-31</td>
<td>0.89</td>
<td>0.82</td>
</tr>
<tr>
<td>RfHD</td>
<td>32-38</td>
<td>0.86</td>
<td>0.81</td>
</tr>
<tr>
<td>AtD</td>
<td>39-41</td>
<td>0.83</td>
<td>0.80</td>
</tr>
<tr>
<td>Excellence</td>
<td>42-45</td>
<td>0.80</td>
<td>0.86</td>
</tr>
</tbody>
</table>

*Note.* RfHD = Respect for Human Dignity; AtD = Attention to Detail. Composite Reliability = 0.86.

Consecutive approach and multidimensional approach correlations are provided in Table 35. Each of the multidimensional approach correlations were greater than the consecutive approach correlations by approximately 1.5 to 2.0 times. The higher overall
correlations between the dimensions of the multidimensional approach illustrated the influence of the interrelatedness across the nine theoretical variables of the CMV. These higher associated correlations provided some support for an overall preference for the multidimensional approach.

In summary, comparison of the three dimensionality approaches from the nine-factor model based on model fit, reliabilities, and estimated correlations led the researcher to select the multidimensional model as the one for further comparison with the eight-factor selection. While the evidence from the examination of the reliabilities favored the consecutive approach, model fit testing and inspection of the estimated correlations was determined to be more influential in the model selection process.
### CMV Nine-Factor Consecutive and Multidimensional Correlation Matrix

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Courage</th>
<th>Account</th>
<th>Humility</th>
<th>Duty</th>
<th>CFO</th>
<th>SC</th>
<th>RfHD</th>
<th>AtD</th>
<th>Excellence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1</td>
<td>0.27</td>
<td>0.12</td>
<td>0.21</td>
<td>0.37</td>
<td>0.34</td>
<td>0.20</td>
<td>0.16</td>
<td>0.23</td>
</tr>
<tr>
<td>Account</td>
<td>0.43</td>
<td>1</td>
<td>0.37</td>
<td>0.19</td>
<td>0.37</td>
<td>0.33</td>
<td>0.39</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>Humility</td>
<td>0.23</td>
<td>0.54</td>
<td>1</td>
<td>0.13</td>
<td>0.34</td>
<td>0.32</td>
<td>0.45</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>Duty</td>
<td>0.34</td>
<td>0.40</td>
<td>0.30</td>
<td>1</td>
<td>0.17</td>
<td>0.34</td>
<td>0.08</td>
<td>0.33</td>
<td>0.42</td>
</tr>
<tr>
<td>CFO</td>
<td>0.47</td>
<td>0.54</td>
<td>0.45</td>
<td>0.32</td>
<td>1</td>
<td>0.29</td>
<td>0.47</td>
<td>0.07</td>
<td>0.17</td>
</tr>
<tr>
<td>SC</td>
<td>0.52</td>
<td>0.46</td>
<td>0.47</td>
<td>0.57</td>
<td>0.41</td>
<td>1</td>
<td>0.28</td>
<td>0.21</td>
<td>0.37</td>
</tr>
<tr>
<td>RfHD</td>
<td>0.30</td>
<td>0.58</td>
<td>0.60</td>
<td>0.20</td>
<td>0.64</td>
<td>0.39</td>
<td>1</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>AtD</td>
<td>0.30</td>
<td>0.16</td>
<td>0.14</td>
<td>0.50</td>
<td>0.13</td>
<td>0.33</td>
<td>0.10</td>
<td>1</td>
<td>0.56</td>
</tr>
<tr>
<td>Excellence</td>
<td>0.31</td>
<td>0.15</td>
<td>0.12</td>
<td>0.64</td>
<td>0.20</td>
<td>0.51</td>
<td>0.02</td>
<td>0.81</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* Account = Accountability; CFO = Care for Others; SC = Self-Control; RfHD = Respect for Human Dignity; AtD = Attention to Detail. Consecutive approach correlations are given above the diagonal; multidimensional approach correlations are given below the diagonal.
**Overall best model fit.** Results from the best fitting approaches from the nine-factor theoretical model and the eight-factor hypothetical model are reproduced in Table 36. On the basis of comparison between the two non-nested models (i.e., multidimensional nine-factor model versus consecutive eight-factor model), the AIC value from the best fitting multidimensional nine-factor model was less than the best fitting consecutive eight-factor model indicating an overall preference for the nine-factor multidimensional theoretical model.

Table 36

*CMV Overall Best Model Fit Comparisons*

<table>
<thead>
<tr>
<th>Approach</th>
<th>Parameters</th>
<th>$G^2$</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eight-Factor Consecutive</td>
<td>77</td>
<td>24439.43</td>
<td>24593.43</td>
</tr>
<tr>
<td>Nine-Factor Multidimensional</td>
<td>93</td>
<td>24342.63</td>
<td>24528.63</td>
</tr>
</tbody>
</table>

Item difficulty parameter estimates and item fit statistics are provided for the overall best fitting nine-factor multidimensional model in Table 37.

Table 37

*Item Difficulty Estimates for CMV Nine-Factor Multidimensional Model*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>MNSQ</th>
<th>t</th>
<th>MNSQ</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1</td>
<td>-1.81</td>
<td>&lt;0.01</td>
<td>1.12</td>
<td>1.3</td>
<td>1.12</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1.73</td>
<td>&lt;0.01</td>
<td>1.19</td>
<td>2.0</td>
<td>1.17</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-0.32</td>
<td>0.01</td>
<td>1.38</td>
<td>4.1</td>
<td>1.41</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-1.40</td>
<td>&lt;0.01</td>
<td>1.18</td>
<td>1.9</td>
<td>1.17</td>
<td>1.8</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.43</td>
<td>&lt;0.01</td>
<td>1.08</td>
<td>1.0</td>
<td>1.11</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-1.18</td>
<td>0.01</td>
<td>1.27</td>
<td>2.9</td>
<td>1.30</td>
<td>3.1</td>
</tr>
<tr>
<td>Accountability</td>
<td>7</td>
<td>-0.87</td>
<td>&lt;0.01</td>
<td>1.15</td>
<td>1.6</td>
<td>1.21</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-1.72</td>
<td>&lt;0.01</td>
<td>0.95</td>
<td>-0.6</td>
<td>0.92</td>
<td>-0.9</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-1.44</td>
<td>&lt;0.01</td>
<td>1.11</td>
<td>1.2</td>
<td>1.11</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>-2.43</td>
<td>&lt;0.01</td>
<td>0.84</td>
<td>-1.8</td>
<td>0.80</td>
<td>-2.5</td>
</tr>
<tr>
<td>Humility</td>
<td>11</td>
<td>-1.26</td>
<td>&lt;0.01</td>
<td>1.19</td>
<td>2.1</td>
<td>1.23</td>
<td>2.4</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>-1.09</td>
<td>&lt;0.01</td>
<td>1.02</td>
<td>0.2</td>
<td>1.05</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>-1.28</td>
<td>&lt;0.01</td>
<td>1.41</td>
<td>4.1</td>
<td>1.43</td>
<td>4.3</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>-2.64</td>
<td>0.01</td>
<td>1.00</td>
<td>0.1</td>
<td>0.94</td>
<td>-0.7</td>
</tr>
</tbody>
</table>
difficulty parameters were estimated from the model. Post hoc were removed such that all item difficulty parameters were estimated.

Post hoc model modifications. Based on the examination of the infit and outfit mean-square t-statistics from Table 37 (i.e., non-fitting items with the largest absolute values for t were removed from the model), the following 27 items were retained while maintaining three items in each dimension in accordance with the research design: item1,
The ConQuest command file executed for fitting the Rasch rating scale model using Monte Carlo estimation to the CMV data to implement the modified nine-factor multidimensional model is shown in Figure 34. A graphical representation of the *post hoc* nine-factor multidimensional model is illustrated in Figure 35.

**Figure 34.** CMV Nine-Factor Modified Multidimensional Model Specification (ConQuest 3.0)
The analysis produced estimates for 75 parameters—including the means and variances of $\theta_{D1}$ through $\theta_{D9}$, 18 item difficulty parameters (e.g., one parameter was constrained per dimension for model identification), three step parameters (e.g., one parameter was constrained for model identification), and 36 unique elements of the variance-covariance matrix—with a model fit statistic, $G^2 = 14501.16$, and a degree of
parsimony fit index, AIC = 14651.16. The reliability of the \( \theta_{Di} \) estimates are displayed in Table 38.

### Table 38

**CMV Nine-Factor Post Hoc Model Multidimensional Reliabilities**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1,4,5</td>
<td>0.76</td>
</tr>
<tr>
<td>Accountability</td>
<td>7-9</td>
<td>0.80</td>
</tr>
<tr>
<td>Humility</td>
<td>11,12,14</td>
<td>0.77</td>
</tr>
<tr>
<td>Duty</td>
<td>17,18,20</td>
<td>0.79</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21,22,26</td>
<td>0.76</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27,29,30</td>
<td>0.78</td>
</tr>
<tr>
<td>RfHD</td>
<td>32,33,36</td>
<td>0.72</td>
</tr>
<tr>
<td>AtD</td>
<td>39-41</td>
<td>0.77</td>
</tr>
<tr>
<td>Excellence</td>
<td>43-45</td>
<td>0.83</td>
</tr>
</tbody>
</table>

*Note.* RfHD = Respect for Human Dignity and AtD = Attention-to-Detail.

Item difficulty parameter estimates and item fit statistics are given for the modified nine-factor multidimensional model in Table 39. Item difficulty estimates ranged from -0.55 logits to -3.37 logits. With the mean of each \( \theta_{Di} \) latent dimension constrained to zero such that all item difficulty parameters were estimated, each parameter estimate was a negative value indicating the item’s relative ease of positive endorsement (i.e., ease in responding to “like me” or “very much like me” on the rating scale). Infit and outfit mean-squares ranged from 0.82 to 1.44 and their \( t \) statistics ranged from -2.1 to 4.2, respectively. According to de Ayala (2009), a cutoff criteria between 0.5 and 1.5 for infit and outfit mean-squares is considered acceptable for fit adequacy; however, according to Bond and Fox (2007), the expected value for the \( t \) statistic is zero with a cutoff criteria outside the range \(-2.0 \leq t \leq 2.0\).
Table 39
*Item Difficulty Estimates for CMV Nine-Factor Post Hoc Multidimensional Model*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Infit MNSQ</th>
<th>t</th>
<th>Outfit MNSQ</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Courage</td>
<td>1</td>
<td>-2.18</td>
<td>&lt;0.01</td>
<td>1.30</td>
<td>3.0</td>
<td>1.30</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-1.70</td>
<td>&lt;0.01</td>
<td>1.25</td>
<td>2.6</td>
<td>1.27</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-0.55</td>
<td>&lt;0.01</td>
<td>1.07</td>
<td>0.9</td>
<td>1.09</td>
<td>1.0</td>
</tr>
<tr>
<td>Accountability</td>
<td>7</td>
<td>-1.10</td>
<td>&lt;0.01</td>
<td>1.26</td>
<td>2.7</td>
<td>1.28</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-2.12</td>
<td>&lt;0.01</td>
<td>1.00</td>
<td>0.0</td>
<td>0.95</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-1.79</td>
<td>&lt;0.01</td>
<td>1.17</td>
<td>1.8</td>
<td>1.16</td>
<td>1.8</td>
</tr>
<tr>
<td>Humility</td>
<td>11</td>
<td>-1.45</td>
<td>&lt;0.01</td>
<td>1.35</td>
<td>3.6</td>
<td>1.35</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>-1.26</td>
<td>&lt;0.01</td>
<td>1.28</td>
<td>2.9</td>
<td>1.32</td>
<td>3.3</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>-3.06</td>
<td>0.01</td>
<td>1.27</td>
<td>2.7</td>
<td>1.16</td>
<td>1.7</td>
</tr>
<tr>
<td>Duty</td>
<td>17</td>
<td>-2.15</td>
<td>&lt;0.01</td>
<td>0.90</td>
<td>-1.1</td>
<td>0.86</td>
<td>-1.6</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>-2.16</td>
<td>&lt;0.01</td>
<td>1.02</td>
<td>0.3</td>
<td>1.00</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>-2.80</td>
<td>&lt;0.01</td>
<td>0.98</td>
<td>-0.1</td>
<td>0.96</td>
<td>-0.4</td>
</tr>
<tr>
<td>Care for Others</td>
<td>21</td>
<td>-3.18</td>
<td>&lt;0.01</td>
<td>1.44</td>
<td>4.2</td>
<td>1.37</td>
<td>3.8</td>
</tr>
<tr>
<td></td>
<td>22</td>
<td>-3.37</td>
<td>&lt;0.01</td>
<td>1.04</td>
<td>0.5</td>
<td>0.99</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td>26</td>
<td>-2.19</td>
<td>&lt;0.01</td>
<td>1.22</td>
<td>2.3</td>
<td>1.16</td>
<td>1.7</td>
</tr>
<tr>
<td>Self-Control</td>
<td>27</td>
<td>-0.72</td>
<td>&lt;0.01</td>
<td>0.99</td>
<td>-0.1</td>
<td>0.98</td>
<td>-0.2</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>-0.96</td>
<td>&lt;0.01</td>
<td>1.03</td>
<td>0.4</td>
<td>1.04</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>-0.66</td>
<td>&lt;0.01</td>
<td>0.97</td>
<td>-0.3</td>
<td>1.00</td>
<td>0.0</td>
</tr>
<tr>
<td>Respect for Human</td>
<td>32</td>
<td>-2.01</td>
<td>&lt;0.01</td>
<td>1.27</td>
<td>2.8</td>
<td>1.26</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>-2.03</td>
<td>&lt;0.01</td>
<td>1.17</td>
<td>1.9</td>
<td>1.15</td>
<td>1.6</td>
</tr>
<tr>
<td>Dignity</td>
<td>36</td>
<td>-2.23</td>
<td>&lt;0.01</td>
<td>1.22</td>
<td>2.3</td>
<td>1.21</td>
<td>2.2</td>
</tr>
<tr>
<td>Attention</td>
<td>39</td>
<td>-2.10</td>
<td>&lt;0.01</td>
<td>1.17</td>
<td>1.8</td>
<td>1.15</td>
<td>1.7</td>
</tr>
<tr>
<td>to Detail</td>
<td>40</td>
<td>-2.44</td>
<td>&lt;0.01</td>
<td>0.82</td>
<td>-2.1</td>
<td>0.83</td>
<td>-1.9</td>
</tr>
<tr>
<td></td>
<td>41</td>
<td>-1.94</td>
<td>&lt;0.01</td>
<td>1.36</td>
<td>3.6</td>
<td>1.26</td>
<td>2.8</td>
</tr>
<tr>
<td>Excellence</td>
<td>43</td>
<td>-2.01</td>
<td>&lt;0.01</td>
<td>1.31</td>
<td>3.2</td>
<td>1.26</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>44</td>
<td>-1.66</td>
<td>&lt;0.01</td>
<td>0.93</td>
<td>-0.8</td>
<td>0.96</td>
<td>-0.5</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>-2.44</td>
<td>&lt;0.01</td>
<td>1.05</td>
<td>0.5</td>
<td>0.99</td>
<td>-0.1</td>
</tr>
</tbody>
</table>

*Note.* An alternative constraint of setting the means of each latent dimension to zero was used such that all item difficulty parameters were estimated.

The item-person map, in which the mean of each $\theta_D$ latent dimension was constrained to zero such that all item difficulty parameters were estimated, is illustrated in Figure 36. This map, from left to right, provides visual estimates of cadet perceived virtue abilities on each dimension (i.e., the latent ability distributions for each dimension are annotated by groupings of ‘x’ where each ‘x’ represents 2.5 cases) followed by the
item difficulties relative to each dimension. Since the items are plotted based on their estimated difficulty logit positions (e.g., the numeral ‘3’ representing item 5 is plotted at its difficulty estimate of -0.55 logit), items near the top of the grouping of numerals are more difficult to endorse than those at the bottom.

By constraining the mean of each $\theta_{Di}$ latent dimension to zero (i.e., scaling the item difficulty parameters to the $\theta$ metric), the item-person map clearly revealed two problematic areas: 1) the items are only measuring levels of cadet virtue abilities near and below the latent trait means on each dimension, and 2) that while the items are fairly dispersed from near the means of the latent trait and below, there is too much item overlap (i.e., item difficulty redundancy) between the -2.0 and -2.5 logit positions.
Another perspective was gained from review of the item-person map in Figure 37, in which the means of the item difficulty parameters on each dimension were constrained to zero such that the means for each latent ability distribution were estimated. Inspection of the item-person maps, based on the means of each latent trait distribution being greater than the model expected value of zero, revealed that cadets found it easy to endorse the
items relative to each dimension’s mean. Along the range of endorsability, items from the care for others dimension with a mean of 2.80 logits were very easy to endorse while items from the self-control dimension with a mean of 0.71 logit were somewhat easy to endorse.

Figure 37. CMV Nine-Factor Post Hoc Multidimensional Model Latent Variable Item-Person Map (ConQuest 3.0). 1 = item1; 2 = item4; 3 = item5; 4 = item7; 5 = item8; 6 = item9; 7 = item11; 8 = item12; 9 = item14; 10 = item17; 11 = item18; 12 = item20; 13 = item21; 14 = item22; 15 = item26; 16 = item27; 17 = item29; 18 = item30; 19 = item32; 20 = item33; 21 = item36; 22 = item39; 23 = item40; 24 = item41; 25 = item43; 26 = item44; 27 = item45. Courage mean = 1.43 logits; Accountability mean = 1.59 logit; Humility mean = 1.79 logits; Duty mean = 2.25 logits; Care for Others mean = 2.80 logits; Self-Control mean = 0.71 logits; Respect for Human Dignity mean = 2.03 logits; Attention to Detail mean = 2.12 logits; Excellence mean = 1.88 logits.
Research Question Four

The latent factor structures of the September 2012 LMI self-rating and subordinate-rating versions were assessed by IRT techniques on the unidimensional theoretical model (Rosebush, 2012) and on the competing six-factor hypothetical multidimensional model by incorporating Allen and Wilson’s (2006) three phased approach—composite, consecutive, and multidimensional—with ConQuest 3.0 modeling software (Wu et al., 2012).

LMI self-rating model. The unidimensional theoretical model, based on Rosebush’s (2012) EFA results, contained 29 items designed to measure cadet element leader effectiveness. The competing hypothetical model formed a multidimensional construct based on the following six USAF institutional leadership effectiveness sub-competencies: develops and inspires others, takes care of people, builds teams and coalitions, negotiating, vision, and adaptability.

Composite approach. The composite approach was applied as a means of comparison with the consecutive and multidimensional approaches. The total score, based on responses to each item on the LMI, was treated as the indicator of a single estimate (i.e., unidimensional) of a cadet element leader’s perception of their overall effectiveness (i.e., \( \theta \)). The ConQuest command file executed for fitting the Rasch rating scale model using Gauss-Hermite Quadrature estimation to the LMI data to implement the composite approach is provided in Figure 38. A graphical representation of the composite approach is illustrated in Appendix Q.
The analysis produced estimates for 33 parameters—including the mean and variance of $\theta$, 28 item difficulty parameters (e.g., one parameter was constrained for model identification), and three step parameters (e.g., one parameter was constrained for model identification)—with a model fit statistic, $G^2 = 29457.87$, and a degree of parsimony fit index, $AIC = 29523.87$. The reliability of the perceived cadet element leader effectiveness estimates was 1.00.

**Consecutive approach.** The consecutive approach was applied as a means of comparison with the composite and multidimensional approaches. This approach modeled each hypothetical USAF institutional sub-competency subscale separately as unidimensional constructs which produced independent $\theta_D$ estimates and standard errors for each dimension. The ConQuest command file executed for fitting the Rasch rating scale model using Gauss-Hermite Quadrature estimation to the LMI self-rating data to implement the consecutive approach for the *develops and inspires others* subscale is provided in Figure 39 (command files for the other subscales were similar). A graphical representation of the consecutive approach is illustrated in Appendix R.
Results of these independent analyses to include the model fit statistics, the number of parameters, the degree of parsimony fit indices, and the reliability of the $\theta_D$ estimates are displayed in Table 40. The analyses produced estimates for 53 parameters—including the means and variances of each $\theta_D$, 23 item difficulty parameters (e.g., one parameter was constrained per model for identification), and 18 step parameters (e.g., one parameter was constrained per model for identification).

Table 40  
**LMI Self-Rating Model Consecutive Approach Fit Results**

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>$G^2$</th>
<th>Parameters</th>
<th>AIC</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1-6</td>
<td>3746.88</td>
<td>10</td>
<td>3766.88</td>
<td>0.76</td>
</tr>
<tr>
<td>TCoP</td>
<td>7-9</td>
<td>2059.88</td>
<td>7</td>
<td>2073.88</td>
<td>0.54</td>
</tr>
<tr>
<td>BTaC</td>
<td>10-15</td>
<td>3850.53</td>
<td>10</td>
<td>3870.53</td>
<td>0.74</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16-18</td>
<td>2258.39</td>
<td>7</td>
<td>2272.39</td>
<td>0.66</td>
</tr>
<tr>
<td>Vision</td>
<td>19-24</td>
<td>3784.55</td>
<td>10</td>
<td>3804.55</td>
<td>0.77</td>
</tr>
<tr>
<td>Adaptability</td>
<td>25-29</td>
<td>3033.68</td>
<td>9</td>
<td>3051.68</td>
<td>0.84</td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions.

**Multidimensional approach.** The multidimensional approach was applied as a means of comparison with the composite and consecutive approaches. By incorporating the correlations between the dimensions, this approach simultaneously estimated separate cadet element leader effectiveness abilities, $\theta_{Di}$, across each latent dimension. The ConQuest command file executed for fitting the Rasch rating scale model using Monte Carlo estimation to the LMI self-rating data to implement the multidimensional approach.
is provided in Figure 40. A graphical representation of the multidimensional approach is illustrated in Appendix S.

**Figure 40. LMI Self-Rating Multidimensional Model Specification (ConQuest 3.0)**

The analysis produced estimates for 53 parameters—including the means and variances of $\theta_{D1}$ through $\theta_{D6}$, 23 item difficulty parameters (e.g., one parameter was constrained per dimension for model identification), three step parameters (e.g., one parameter was constrained for model identification), and 15 unique elements of the variance-covariance matrix—with a model fit statistic, $G^2 = 17442.30$, and a degree of parsimony fit index, $AIC = 17548.30$. The reliability of the $\theta_{D1}$ estimates are displayed in Table 41.
Table 41
LMI Self-Rating Model Multidimensional Reliabilities

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1-6</td>
<td>0.89</td>
</tr>
<tr>
<td>TCoP</td>
<td>7-9</td>
<td>0.88</td>
</tr>
<tr>
<td>BTaC</td>
<td>10-15</td>
<td>0.90</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16-18</td>
<td>0.78</td>
</tr>
<tr>
<td>Vision</td>
<td>19-24</td>
<td>0.89</td>
</tr>
<tr>
<td>Adaptability</td>
<td>25-29</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Note. DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions.

Model comparisons. Model fit results from the three dimensionality evaluation approaches are reproduced in Table 42. Since the multidimensional approach is nested in the composite approach, the likelihood ratio test was computed as the difference in deviance, which approximates a $\chi^2$ distribution with degrees of freedom equal to the difference in estimated parameters between the models: $\chi^2 = 12015.57$, $df = 20$, $p < 0.001$. Since the difference in deviance between the nested models was statistically significant, evidence existed that the multidimensional model fit the data significantly better than the composite model. Moreover, on the basis of comparison between the non-nested models (i.e., multidimensional versus consecutive), the AIC value from the multidimensional model was less than the consecutive model indicating a preference for the multidimensional model.
Reliability estimates from the three dimensionality evaluation approaches are reproduced in Table 43. Multidimensional reliability enhancement (Allen & Wilson, 2006), which was expected to occur on all dimensions, was present in all subscales except for adaptability (e.g., the multidimensional reliability exceeded consecutive reliability in all cases except the last dimension). Since the multidimensional approach produced greater reliabilities than the consecutive approach for all but one dimension, a model preference toward the multidimensional approach was demonstrated.

Table 43
LMI Self-Rating Model Reliabilities

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Consecutive Reliability</th>
<th>Multidimensional Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1-6</td>
<td>0.76</td>
<td>0.89</td>
</tr>
<tr>
<td>TCoP</td>
<td>7-9</td>
<td>0.54</td>
<td>0.88</td>
</tr>
<tr>
<td>BTaC</td>
<td>10-15</td>
<td>0.74</td>
<td>0.90</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16-18</td>
<td>0.66</td>
<td>0.78</td>
</tr>
<tr>
<td>Vision</td>
<td>19-24</td>
<td>0.77</td>
<td>0.89</td>
</tr>
<tr>
<td>Adaptability</td>
<td>25-29</td>
<td>0.84</td>
<td>0.82</td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions. Composite Reliability = 1.00.

Consecutive approach and multidimensional approach correlations are provided in Table 44. Each of the multidimensional approach correlations exceeded the consecutive approach correlations. The higher overall correlations between the dimensions of the multidimensional approach illustrated the influence of the interrelatedness across the six
hypothetical variables of the LMI. These higher associated correlations provided some support for an overall preference for the multidimensional approach.

In summary, comparison of the three dimensionality approaches based on model fit, reliabilities, and estimated correlations led the researcher to assess the multidimensional model as the most adequate and one for further post hoc modification. Item difficulty parameter estimates and item fit statistics are provided for the overall best fitting multidimensional model in Table 4.5.

Table 4
LMI Self-Rating Consecutive and Multidimensional Correlation Matrix

<table>
<thead>
<tr>
<th>Subscale</th>
<th>DaIO</th>
<th>TCoP</th>
<th>BTaC</th>
<th>Negot</th>
<th>Vision</th>
<th>Adapt</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td></td>
<td>0.62</td>
<td>0.77</td>
<td>0.58</td>
<td>0.81</td>
<td>0.50</td>
</tr>
<tr>
<td>TCoP</td>
<td>0.93</td>
<td>1</td>
<td>0.63</td>
<td>0.50</td>
<td>0.62</td>
<td>0.56</td>
</tr>
<tr>
<td>BTaC</td>
<td>0.97</td>
<td>0.95</td>
<td>1</td>
<td>0.57</td>
<td>0.73</td>
<td>0.53</td>
</tr>
<tr>
<td>Negot</td>
<td>0.82</td>
<td>0.82</td>
<td>0.81</td>
<td>1</td>
<td>0.55</td>
<td>0.55</td>
</tr>
<tr>
<td>Vision</td>
<td>0.98</td>
<td>0.94</td>
<td>0.96</td>
<td>0.78</td>
<td>1</td>
<td>0.52</td>
</tr>
<tr>
<td>Adapt</td>
<td>0.69</td>
<td>0.84</td>
<td>0.73</td>
<td>0.76</td>
<td>0.69</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions; Negot = Negotiating; Adapt = Adaptability. Consecutive approach correlations are given above the diagonal; multidimensional approach correlations are given below the diagonal.

Table 45
Item Difficulty Estimates for LMI Self-Rating Multidimensional Model

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>MNSQ</th>
<th>t</th>
<th>MNSQ</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1</td>
<td>-2.90</td>
<td>&lt;0.01</td>
<td>1.01</td>
<td>0.1</td>
<td>0.99</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-2.92</td>
<td>&lt;0.01</td>
<td>0.85</td>
<td>-2.0</td>
<td>0.84</td>
<td>-2.2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-3.08</td>
<td>&lt;0.01</td>
<td>0.94</td>
<td>-0.8</td>
<td>0.96</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-2.66</td>
<td>&lt;0.01</td>
<td>0.94</td>
<td>-0.8</td>
<td>0.92</td>
<td>-1.1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-2.72</td>
<td>&lt;0.01</td>
<td>0.98</td>
<td>-0.3</td>
<td>0.97</td>
<td>-0.4</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-4.39</td>
<td>&lt;0.01</td>
<td>1.12</td>
<td>1.5</td>
<td>1.07</td>
<td>1.0</td>
</tr>
<tr>
<td>TCoP</td>
<td>7</td>
<td>-2.50</td>
<td>&lt;0.01</td>
<td>0.88</td>
<td>-1.5</td>
<td>0.88</td>
<td>-1.6</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-4.40</td>
<td>0.01</td>
<td>1.18</td>
<td>2.4</td>
<td>1.28</td>
<td>3.5</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-2.78</td>
<td>&lt;0.01</td>
<td>1.44</td>
<td>4.9</td>
<td>1.42</td>
<td>5.0</td>
</tr>
<tr>
<td>BTaC</td>
<td>10</td>
<td>-2.97</td>
<td>&lt;0.01</td>
<td>1.20</td>
<td>2.4</td>
<td>1.17</td>
<td>2.2</td>
</tr>
</tbody>
</table>
The ConQuest command file executed for fitting the Rasch rating scale model using Monte Carlo estimation to the LMI self-rating data to implement the modified six-factor multidimensional model is shown in Figure 41. A graphical representation of the *post hoc* six-factor multidimensional model is illustrated in Figure 42.
Figure 41. LMI Self-Rating Six-Factor Modified Multidimensional Model Specification (ConQuest 3.0)
Figure 42. LMI Self-Rating Six-Factor Post Hoc Multidimensional Model Specification (Amos Version 18)

The analysis produced estimates for 42 parameters—including the means and variances of $\theta_{D1}$ through $\theta_{D6}$, 12 item difficulty parameters (e.g., one parameter was constrained per dimension for model identification), three step parameters (e.g., one parameter was constrained for model identification), and 15 unique elements of the variance-covariance matrix—with a model fit statistic, $G^2 = 11290.33$, and a degree of
parsimony fit index, AIC = 11374.33. The reliability of the $\theta_{Di}$ estimates are displayed in Table 46.

Table 46

*LMI Self-Rating Six-Factor Post Hoc Model Multidimensional Reliabilities*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1,3,5</td>
<td>0.82</td>
</tr>
<tr>
<td>TCoP</td>
<td>7-9</td>
<td>0.83</td>
</tr>
<tr>
<td>BTaC</td>
<td>11,12,15</td>
<td>0.86</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16-18</td>
<td>0.81</td>
</tr>
<tr>
<td>Vision</td>
<td>22-24</td>
<td>0.85</td>
</tr>
<tr>
<td>Adaptability</td>
<td>26,27,29</td>
<td>0.79</td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions.

Item difficulty parameter estimates and item fit statistics are given for the modified six-factor multidimensional model in Table 47. Item difficulty estimates ranged from -1.84 logits to -4.68 logits. With the mean of each $\theta_{Di}$ latent dimension constrained to zero such that all item difficulty parameters were estimated, each parameter estimate was a negative value indicating the item’s relative ease of positive endorsement (i.e., ease in responding to “like me” or “very much like me” on the rating scale). Infit and outfit mean-squares ranged from 0.83 to 1.40 and their $t$ statistics ranged from -2.4 to 4.7, respectively. According to de Ayala (2009), a cutoff criteria between 0.5 and 1.5 for infit and outfit mean-squares is considered acceptable for fit adequacy; however, according to Bond and Fox (2007), the expected value for the $t$ statistic is zero with a cutoff criteria outside the range $-2.0 \leq t \leq 2.0$.  

165
Table 47  
*Item Difficulty Estimates for LMI Self-Rating Modified Multidimensional Model*

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Infit</th>
<th>MNSQ</th>
<th>t</th>
<th>MNSQ</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1</td>
<td>-3.15</td>
<td>0.02</td>
<td>1.00</td>
<td>0.0</td>
<td>0.98</td>
<td>-0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-3.33</td>
<td>0.01</td>
<td>0.95</td>
<td>-0.6</td>
<td>0.97</td>
<td>-0.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-2.97</td>
<td>0.01</td>
<td>0.98</td>
<td>-0.2</td>
<td>0.96</td>
<td>-0.5</td>
<td></td>
</tr>
<tr>
<td>TCoP</td>
<td>7</td>
<td>-2.75</td>
<td>0.01</td>
<td>0.89</td>
<td>-1.5</td>
<td>0.89</td>
<td>-1.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-4.68</td>
<td>0.01</td>
<td>1.14</td>
<td>1.9</td>
<td>1.22</td>
<td>2.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-3.03</td>
<td>&lt;0.01</td>
<td>1.40</td>
<td>4.7</td>
<td>1.37</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>BTaC</td>
<td>11</td>
<td>-4.27</td>
<td>0.01</td>
<td>0.90</td>
<td>-1.5</td>
<td>0.88</td>
<td>-1.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>-3.88</td>
<td>0.01</td>
<td>1.02</td>
<td>0.3</td>
<td>1.01</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>-3.90</td>
<td>0.02</td>
<td>0.98</td>
<td>-0.3</td>
<td>0.94</td>
<td>-0.8</td>
<td></td>
</tr>
<tr>
<td>Negotiating</td>
<td>16</td>
<td>-2.62</td>
<td>0.01</td>
<td>0.94</td>
<td>-0.7</td>
<td>0.95</td>
<td>-0.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>-2.23</td>
<td>0.02</td>
<td>0.98</td>
<td>-0.3</td>
<td>0.97</td>
<td>-0.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>-1.84</td>
<td>0.02</td>
<td>1.36</td>
<td>4.3</td>
<td>1.37</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>Vision</td>
<td>22</td>
<td>-2.63</td>
<td>&lt;0.01</td>
<td>0.93</td>
<td>-1.0</td>
<td>0.93</td>
<td>-0.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>-3.31</td>
<td>0.02</td>
<td>0.89</td>
<td>-1.5</td>
<td>0.89</td>
<td>-1.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>24</td>
<td>-2.75</td>
<td>0.01</td>
<td>1.12</td>
<td>1.6</td>
<td>1.10</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Adaptability</td>
<td>26</td>
<td>-3.00</td>
<td>0.01</td>
<td>1.14</td>
<td>1.8</td>
<td>1.13</td>
<td>1.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>-2.69</td>
<td>0.01</td>
<td>0.83</td>
<td>-2.4</td>
<td>0.85</td>
<td>-2.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>-2.61</td>
<td>&lt;0.01</td>
<td>0.89</td>
<td>-1.5</td>
<td>0.89</td>
<td>-1.6</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions. An alternative constraint of setting the means of each latent dimension to zero was used such that all item difficulty parameters were estimated.

The item-person map, in which the mean of each $\theta_{D_i}$ latent dimension was constrained to zero such that all item difficulty parameters were estimated, is illustrated in Figure 43. This map, from left to right, provides visual estimates of perceived cadet element leader *effectiveness* abilities on each dimension (i.e., the latent ability distributions for each dimension are annotated by groupings of ‘x’ where each ‘x’ represents 2.9 cases) followed by the item difficulties relative to each dimension. Since the items are plotted based on their estimated difficulty logit positions (e.g., the numeral ‘3’ representing item5 is plotted at its difficulty estimate of -2.97 logits), items near the top of the grouping of numerals are more difficult to endorse than those at the bottom.
By constraining the mean of each $\theta_{di}$ latent dimension to zero (i.e., scaling the item difficulty parameters to the $\theta$ metric), the item-person map clearly revealed two problematic areas: 1) the items are only measuring levels of perceived cadet element leader *effectiveness* abilities much below the latent trait means on each dimension, and 2) that while the items are fairly dispersed from much below the means of the latent traits, there is too much item overlap (i.e., item difficulty redundancy) between the -2.7 and -3.0 logit positions.
Another perspective was gained from review of the item-person map in Figure 44, in which the means of the item difficulty parameters on each dimension were constrained.
to zero such that the means for each latent ability distribution were estimated. Inspection of the item-person maps, based on the means of each latent trait distribution being greater than the model expected value of zero, revealed that cadet element leaders found it easy to endorse the items relative to each dimension’s mean. Along the range of endorsability, items from the builds teams and coalitions dimension with a mean of 4.16 logits were very easy to endorse while items from the negotiating dimension with a mean of 2.35 logits were somewhat easy to endorse.
**Figure 44.** LMI Self-Rating Six-Factor Post Hoc Multidimensional Model Latent Variable Item-Person Map (ConQuest 3.0). 1 = item1; 2 = item3; 3 = item5; 4 = item7; 5 = item8; 6 = item9; 7 = item11; 8 = item12; 9 = item15; 10 = item16; 11 = item17; 12 = item18; 13 = item22; 14 = item23; 15 = item24; 16 = item26; 17 = item27; 18 = item29. 

Develops and Inspires Others mean = 3.28 logits; Takes Care of People mean = 3.60 logits; Builds Teams and Coalitions mean = 4.16 logits; Negotiating mean = 2.35 logits; Vision mean = 3.02 logits; Adaptability mean = 2.90 logits.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>TCoP</th>
<th>BTAc</th>
<th>Negot</th>
<th>Vision</th>
<th>Adapt</th>
<th>+item</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>6</td>
<td>xx</td>
<td>xx</td>
<td>xxx</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>5</td>
<td>xxx</td>
<td>xxx</td>
<td>xxx</td>
<td>x</td>
<td>xxx</td>
<td>xxx</td>
</tr>
<tr>
<td>4</td>
<td>xxxxx</td>
<td>xxxxx</td>
<td>xxxxxx</td>
<td>xxx</td>
<td>xxxxx</td>
<td>xxxxx</td>
</tr>
<tr>
<td>3</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
</tr>
<tr>
<td>2</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
</tr>
<tr>
<td>1</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
<td>xxxxxx</td>
</tr>
<tr>
<td>0</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*each 'x' represents 2.8 cases*
**LMI subordinate-rating model.** The unidimensional theoretical model, based on Rosebush’s (2012) EFA results, contained 29 items designed to measure cadet element leader *effectiveness* from the perspective of cadet subordinates. The competing hypothetical model formed a multidimensional construct based on the following six USAF institutional *leadership effectiveness* sub-competencies: *develops and inspires others, takes care of people, builds teams and coalitions, negotiating, vision, and adaptability.*

**Composite approach.** The composite approach was applied as a means of comparison with the consecutive and multidimensional approaches. The total score, based on responses to each item on the LMI, was treated as the indicator of a single estimate (i.e., unidimensional) by cadet subordinates on their respective element leader’s overall *effectiveness* (i.e., \( \theta \)). The ConQuest command file executed for fitting the Rasch rating scale model using Gauss-Hermite Quadrature estimation to the LMI data to implement the composite approach is provided in Figure 45. A graphical representation of the composite approach is illustrated in Appendix Q.

```
title rating scale analysis;
datafile C:\Users\David\Desktop\Dissertation\IRT\LMIsubs\Composite\LMI_subs_t1_29_items_R=1777_Appendix_R.order.dat;
format responses 1-29;
codes 0.1.2.3.4;
recode (1.2.3.4.5) (0.1.2.3.4);
model item + step:  /*rating scale*/
estimates nodes=25;
show estimates-latent >> C:\Users\David\Desktop\Dissertation\IRT\LMIsubs\Composite\composite.shw;  
fitanal >> C:\Users\David\Desktop\Dissertation\IRT\LMIsubs\Composite\composite.fit;
```

*Figure 45. LMI Subordinate-Rating Composite Model Specification (ConQuest 3.0)*

The analysis produced estimates for 33 parameters—including the mean and variance of \( \theta \), 28 item difficulty parameters (e.g., one parameter was constrained for model identification), and three step parameters (e.g., one parameter was constrained for
model identification)—with a model fit statistic, $G^2 = 90105.92$, and a degree of parsimony fit index, $AIC = 90171.92$. The reliability of the element leader effectiveness estimates was 1.00, as rated by cadet subordinates.

**Consecutive approach.** The consecutive approach was applied as a means of comparison with the composite and multidimensional approaches. This approach modeled each hypothetical USAF institutional sub-competency subscale separately as unidimensional constructs which produced independent $\theta_D$ estimates and standard errors for each dimension. The ConQuest command file executed for fitting the Rasch rating scale model using Gauss-Hermite Quadrature estimation to the LMI subordinate-rating data to implement the consecutive approach for the *develops and inspires others* subscale is provided in Figure 46 (command files for the other subscales were similar). A graphical representation of the consecutive approach is illustrated in Appendix R.

![ConQuest 3.0 Consecutive Model Specification](image)

**Figure 46.** LMI Subordinate-Rating Consecutive Model Specification (*Develops and Inspires Others* Subscale Only—Other Subscales Similar) (ConQuest 3.0)

Results of these independent analyses to include the model fit statistics, the number of parameters, the degree of parsimony fit indices, and the reliability of the $\theta_D$ estimates are displayed in Table 48. The analyses produced estimates for 53 parameters—including the means and variances of each $\theta_D$, 23 item difficulty parameters (e.g., one parameter was constrained per model for identification), and 18 step parameters (e.g., one parameter was constrained per model for identification).
Table 48

LMI Subordinate-Rating Model Consecutive Approach Fit Results

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>$G^2$</th>
<th>Parameters</th>
<th>AIC</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1-6</td>
<td>24218.71</td>
<td>10</td>
<td>24238.71</td>
<td>1.00</td>
</tr>
<tr>
<td>TCoP</td>
<td>7-9</td>
<td>9299.78</td>
<td>7</td>
<td>9313.78</td>
<td>0.78</td>
</tr>
<tr>
<td>BTaC</td>
<td>10-15</td>
<td>16659.67</td>
<td>10</td>
<td>16679.67</td>
<td>0.84</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16-18</td>
<td>10451.84</td>
<td>7</td>
<td>10465.84</td>
<td>0.82</td>
</tr>
<tr>
<td>Vision</td>
<td>19-24</td>
<td>17141.71</td>
<td>10</td>
<td>17161.71</td>
<td>0.87</td>
</tr>
<tr>
<td>Adaptability</td>
<td>25-29</td>
<td>20680.52</td>
<td>9</td>
<td>20698.52</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions.

**Multidimensional approach.** The multidimensional approach was applied as a means of comparison with the composite and consecutive approaches. By incorporating the correlations between the dimensions, this approach simultaneously estimated separate cadet element leader *effectiveness* abilities based on subordinate ratings, $\theta_{Di}$, across each latent dimension. The ConQuest command file executed for fitting the Rasch rating scale model using Monte Carlo estimation to the LMI subordinate-rating data to implement the multidimensional approach is provided in Figure 47. A graphical representation of the multidimensional approach is illustrated in Appendix S.
The analysis produced estimates for 53 parameters—including the means and variances of $\theta_{Di}$ through $\theta_{D6}$, 23 item difficulty parameters (e.g., one parameter was constrained per dimension for model identification), three step parameters (e.g., one parameter was constrained for model identification), and 15 unique elements of the variance-covariance matrix—with a model fit statistic, $G^2 = 69937.53$, and a degree of parsimony fit index, $AIC = 70043.53$. The reliability of the $\theta_{Di}$ estimates are displayed in Table 49.
Table 49
*LMI Subordinate-Rating Model Multidimensional Reliabilities*

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1-6</td>
<td>0.99</td>
</tr>
<tr>
<td>TCoP</td>
<td>7-9</td>
<td>0.99</td>
</tr>
<tr>
<td>BTaC</td>
<td>10-15</td>
<td>0.99</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16-18</td>
<td>0.96</td>
</tr>
<tr>
<td>Vision</td>
<td>19-24</td>
<td>0.99</td>
</tr>
<tr>
<td>Adaptability</td>
<td>25-29</td>
<td>0.97</td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions.

**Model comparisons.** Model fit results from the three dimensionality evaluation approaches are reproduced in Table 50. Since the multidimensional approach is nested in the composite approach, the likelihood ratio test was computed as the difference in deviance, which approximates a $\chi^2$ distribution with degrees of freedom equal to the difference in estimated parameters between the models: $\chi^2 = 20168.39$, $df = 20$, $p < 0.001$. Since the difference in deviance between the nested models was statistically significant, evidence existed that the multidimensional model fit the data significantly better than the composite model. Moreover, on the basis of comparison between the non-nested models (i.e., multidimensional versus consecutive), the AIC value from the multidimensional model was less than the consecutive model indicating a preference for the multidimensional model.
Reliability estimates from the three dimensionality evaluation approaches are reproduced in Table 51. Multidimensional reliability enhancement (Allen & Wilson, 2006), which was expected to occur on all dimensions, was present in all subscales except for develops and inspires others and adaptability (e.g., the multidimensional reliability exceeded consecutive reliability in all cases except for two dimensions). Since the multidimensional approach produced greater reliabilities than the consecutive approach for all but two dimensions, a model preference toward the multidimensional approach was demonstrated.

Table 51

LMI Subordinate-Rating Model Reliabilities

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Consecutive Reliability</th>
<th>Multidimensional Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1-6</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>TCoP</td>
<td>7-9</td>
<td>0.78</td>
<td>0.99</td>
</tr>
<tr>
<td>BTaC</td>
<td>10-15</td>
<td>0.84</td>
<td>0.99</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16-18</td>
<td>0.82</td>
<td>0.96</td>
</tr>
<tr>
<td>Vision</td>
<td>19-24</td>
<td>0.87</td>
<td>0.99</td>
</tr>
<tr>
<td>Adaptability</td>
<td>25-29</td>
<td>1.00</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Note. DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions. Composite Reliability = 1.00.

Consecutive approach and multidimensional approach correlations are provided in Table 52. Each of the multidimensional approach correlations exceeded the consecutive approach correlations. The higher overall correlations between the dimensions of the
multidimensional approach illustrated the influence of the interrelatedness across the six hypothetical variables of the LMI. These higher associated correlations provided some support for an overall preference for the multidimensional approach.

In summary, comparison of the three dimensionality approaches based on model fit, reliabilities, and estimated correlations led the researcher to assess the multidimensional model as the most adequate and one for further post hoc modification.

Item difficulty parameter estimates and item fit statistics are provided for the overall best fitting multidimensional model in Table 53.

Table 52
LMI Subordinate-Rating Consecutive and Multidimensional Correlation Matrix

<table>
<thead>
<tr>
<th>Subscale</th>
<th>DaIO</th>
<th>TCoP</th>
<th>BTaC</th>
<th>Negot</th>
<th>Vision</th>
<th>Adapt</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1</td>
<td>0.86</td>
<td>0.91</td>
<td>0.84</td>
<td>0.94</td>
<td>0.83</td>
</tr>
<tr>
<td>TCoP</td>
<td>0.98</td>
<td>1</td>
<td>0.88</td>
<td>0.81</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td>BTaC</td>
<td>0.99</td>
<td>0.99</td>
<td>1</td>
<td>0.83</td>
<td>0.91</td>
<td>0.84</td>
</tr>
<tr>
<td>Negot</td>
<td>0.95</td>
<td>0.96</td>
<td>0.96</td>
<td>1</td>
<td>0.84</td>
<td>0.82</td>
</tr>
<tr>
<td>Vision</td>
<td>0.99</td>
<td>0.99</td>
<td>0.99</td>
<td>0.96</td>
<td>1</td>
<td>0.85</td>
</tr>
<tr>
<td>Adapt</td>
<td>0.93</td>
<td>0.95</td>
<td>0.93</td>
<td>0.93</td>
<td>0.94</td>
<td>1</td>
</tr>
</tbody>
</table>

Note. DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions; Negot = Negotiating; Adapt = Adaptability. Consecutive approach correlations are given above the diagonal; multidimensional approach correlations are given below the diagonal.

Table 53
Item Difficulty Estimates for LMI Subordinate-Rating Multidimensional Model

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Infit MNSQ</th>
<th>Infit t</th>
<th>Outfit MNSQ</th>
<th>Outfit t</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>1</td>
<td>-4.47</td>
<td>&lt;0.01</td>
<td>1.32</td>
<td>7.5</td>
<td>1.19</td>
<td>5.4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-4.74</td>
<td>&lt;0.01</td>
<td>1.02</td>
<td>0.6</td>
<td>0.85</td>
<td>-4.6</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>-4.67</td>
<td>&lt;0.01</td>
<td>1.30</td>
<td>7.2</td>
<td>1.23</td>
<td>6.5</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-4.48</td>
<td>&lt;0.01</td>
<td>1.11</td>
<td>2.7</td>
<td>0.96</td>
<td>-1.1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-4.61</td>
<td>&lt;0.01</td>
<td>1.11</td>
<td>2.8</td>
<td>0.92</td>
<td>-2.6</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>-5.92</td>
<td>&lt;0.01</td>
<td>1.39</td>
<td>9.3</td>
<td>1.18</td>
<td>5.2</td>
</tr>
<tr>
<td>TCoP</td>
<td>7</td>
<td>-4.60</td>
<td>&lt;0.01</td>
<td>0.84</td>
<td>-4.4</td>
<td>0.83</td>
<td>-5.2</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-5.91</td>
<td>&lt;0.01</td>
<td>1.37</td>
<td>8.7</td>
<td>1.15</td>
<td>4.3</td>
</tr>
</tbody>
</table>

177
modified six
using Monte Carlo estimation to the LMI
item
maintai
values for
mean
that all item difficulty parameters were estimated
Coalitions. An alternative constraint of setting the means of each latent dimension to zero was used such
Note
Adaptability
Vision
Negotiating
BTaC

<table>
<thead>
<tr>
<th>Item</th>
<th>Fit</th>
<th>T-Mean</th>
<th>A-Mean</th>
<th>T-Std</th>
<th>A-Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>-4.25</td>
<td>&lt;0.01</td>
<td>1.20</td>
<td>4.8</td>
<td>1.05</td>
</tr>
<tr>
<td>10</td>
<td>-4.85</td>
<td>&lt;0.01</td>
<td>1.30</td>
<td>7.2</td>
<td>1.14</td>
</tr>
<tr>
<td>11</td>
<td>-5.47</td>
<td>&lt;0.01</td>
<td>0.88</td>
<td>-3.2</td>
<td>0.68</td>
</tr>
<tr>
<td>12</td>
<td>-5.15</td>
<td>&lt;0.01</td>
<td>1.34</td>
<td>8.1</td>
<td>1.17</td>
</tr>
<tr>
<td>13</td>
<td>-5.25</td>
<td>&lt;0.01</td>
<td>1.09</td>
<td>2.2</td>
<td>0.91</td>
</tr>
<tr>
<td>14</td>
<td>-4.05</td>
<td>&lt;0.01</td>
<td>1.20</td>
<td>4.9</td>
<td>1.08</td>
</tr>
<tr>
<td>15</td>
<td>-5.28</td>
<td>&lt;0.01</td>
<td>1.28</td>
<td>6.7</td>
<td>1.26</td>
</tr>
<tr>
<td>16</td>
<td>-4.45</td>
<td>&lt;0.01</td>
<td>1.08</td>
<td>2.2</td>
<td>1.00</td>
</tr>
<tr>
<td>17</td>
<td>-3.77</td>
<td>&lt;0.01</td>
<td>1.01</td>
<td>0.3</td>
<td>0.90</td>
</tr>
<tr>
<td>18</td>
<td>-2.98</td>
<td>&lt;0.01</td>
<td>1.41</td>
<td>9.9</td>
<td>1.41</td>
</tr>
<tr>
<td>19</td>
<td>-4.49</td>
<td>&lt;0.01</td>
<td>1.14</td>
<td>3.6</td>
<td>1.04</td>
</tr>
<tr>
<td>20</td>
<td>-4.81</td>
<td>&lt;0.01</td>
<td>1.07</td>
<td>1.8</td>
<td>0.95</td>
</tr>
<tr>
<td>21</td>
<td>-5.63</td>
<td>&lt;0.01</td>
<td>1.12</td>
<td>3.1</td>
<td>0.96</td>
</tr>
<tr>
<td>22</td>
<td>-3.88</td>
<td>&lt;0.01</td>
<td>0.88</td>
<td>-3.4</td>
<td>0.78</td>
</tr>
<tr>
<td>23</td>
<td>-4.99</td>
<td>&lt;0.01</td>
<td>1.25</td>
<td>6.1</td>
<td>1.27</td>
</tr>
<tr>
<td>24</td>
<td>-4.27</td>
<td>&lt;0.01</td>
<td>1.39</td>
<td>9.1</td>
<td>1.34</td>
</tr>
<tr>
<td>25</td>
<td>-4.25</td>
<td>&lt;0.01</td>
<td>0.86</td>
<td>-3.9</td>
<td>0.77</td>
</tr>
<tr>
<td>26</td>
<td>-4.03</td>
<td>&lt;0.01</td>
<td>0.81</td>
<td>-5.4</td>
<td>0.71</td>
</tr>
<tr>
<td>27</td>
<td>-3.87</td>
<td>&lt;0.01</td>
<td>0.82</td>
<td>-5.0</td>
<td>0.76</td>
</tr>
<tr>
<td>28</td>
<td>-4.10</td>
<td>&lt;0.01</td>
<td>0.87</td>
<td>-3.6</td>
<td>0.73</td>
</tr>
<tr>
<td>29</td>
<td>-4.15</td>
<td>&lt;0.01</td>
<td>0.84</td>
<td>-4.5</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Note. DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalition. An alternative constraint of setting the means of each latent dimension to zero was used such that all item difficulty parameters were estimated.

**Post hoc model modifications.** Based on the examination of the infit and outfit mean-square t-statistics from Table 53 (i.e., non-fitting items with the largest absolute values for t were removed from the model), the following 18 items were retained while maintaining three items in each dimension in accordance with the research design: *item2, item4, item5, item7-item9, item10, item13-item14, item16-item21, item25, item27, and item29.*

The ConQuest command file executed for fitting the Rasch rating scale model using Monte Carlo estimation to the LMI subordinate-rating data to implement the modified six-factor multidimensional model is shown in Figure 48. A graphical
representation of the *post hoc* six-factor multidimensional model is illustrated in Figure 49.

![Figure 48. LMI Subordinate-Rating Six-Factor Modified Multidimensional Model Specification (ConQuest 3.0)](image-url)
The analysis produced estimates for 42 parameters—including the means and variances of $\theta_{D1}$ through $\theta_{D6}$, 12 item difficulty parameters (e.g., one parameter was constrained per dimension for model identification), three step parameters (e.g., one parameter was constrained for model identification), and 15 unique elements of the variance-covariance matrix—with a model fit statistic, $G^2 = 45217.48$, and a degree of
parsimony fit index, AIC = 45301.48. The reliability of the \( \theta_{Di} \) estimates are displayed in Table 54.

Table 54

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Items</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO</td>
<td>2,4,5</td>
<td>0.91</td>
</tr>
<tr>
<td>TCoP</td>
<td>7-9</td>
<td>0.90</td>
</tr>
<tr>
<td>BTaC</td>
<td>10,13,14</td>
<td>0.90</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16-18</td>
<td>0.89</td>
</tr>
<tr>
<td>Vision</td>
<td>19-21</td>
<td>0.90</td>
</tr>
<tr>
<td>Adaptability</td>
<td>25,27,29</td>
<td>0.89</td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions.

Item difficulty parameter estimates and item fit statistics are given for the modified six-factor multidimensional model in Table 55. Item difficulty estimates ranged from -3.01 logits to -5.82 logis. With the mean of each \( \theta_{Di} \) latent dimension constrained to zero such that all item difficulty parameters were estimated, each parameter estimate was a negative value indicating the item’s relative ease of positive endorsement (i.e., ease in responding to “like the Leader” or “very much like the Leader” on the rating scale).

Infit and outfit mean-squares ranged from 0.67 to 1.34 and their \( t \) statistics ranged from -11.2 to 9.1, respectively. According to de Ayala (2009), a cutoff criteria between 0.5 and 1.5 for infit and outfit mean-squares is considered acceptable for fit adequacy; however, according to Bond and Fox (2007), the expected value for the \( t \) statistic is zero with a cutoff criteria outside the range \(-2.0 \leq t \leq 2.0\).
Table 55
Item Difficulty Estimates for LMI Subordinate-Rating Modified Multidimensional Model

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Item</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>Infit</th>
<th>Outfit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DaIO</td>
<td>2</td>
<td>-4.84</td>
<td>&lt;0.01</td>
<td>0.98</td>
<td>-0.6</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>-4.58</td>
<td>&lt;0.01</td>
<td>1.04</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>-4.71</td>
<td>&lt;0.01</td>
<td>1.01</td>
<td>0.2</td>
</tr>
<tr>
<td>TCoP</td>
<td>7</td>
<td>-4.56</td>
<td>&lt;0.01</td>
<td>0.77</td>
<td>-6.7</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>-5.82</td>
<td>&lt;0.01</td>
<td>1.29</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>-4.21</td>
<td>&lt;0.01</td>
<td>1.07</td>
<td>1.9</td>
</tr>
<tr>
<td>BTaC</td>
<td>10</td>
<td>-4.85</td>
<td>&lt;0.01</td>
<td>1.16</td>
<td>4.0</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>-5.24</td>
<td>&lt;0.01</td>
<td>1.00</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>-4.05</td>
<td>&lt;0.01</td>
<td>1.02</td>
<td>0.6</td>
</tr>
<tr>
<td>Negotiating</td>
<td>16</td>
<td>-4.50</td>
<td>&lt;0.01</td>
<td>1.02</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>-3.81</td>
<td>&lt;0.01</td>
<td>0.95</td>
<td>-1.4</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>-3.01</td>
<td>&lt;0.01</td>
<td>1.31</td>
<td>8.1</td>
</tr>
<tr>
<td>Vision</td>
<td>19</td>
<td>-4.56</td>
<td>&lt;0.01</td>
<td>1.07</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>-4.88</td>
<td>&lt;0.01</td>
<td>1.02</td>
<td>0.4</td>
</tr>
<tr>
<td></td>
<td>21</td>
<td>-5.69</td>
<td>&lt;0.01</td>
<td>1.06</td>
<td>1.5</td>
</tr>
<tr>
<td>Adaptability</td>
<td>25</td>
<td>-4.29</td>
<td>&lt;0.01</td>
<td>0.89</td>
<td>-3.2</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>-3.91</td>
<td>&lt;0.01</td>
<td>0.86</td>
<td>-4.1</td>
</tr>
<tr>
<td></td>
<td>29</td>
<td>-4.19</td>
<td>&lt;0.01</td>
<td>0.83</td>
<td>-5.0</td>
</tr>
</tbody>
</table>

Note. DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions. An alternative constraint of setting the means of each latent dimension to zero was used such that all item difficulty parameters were estimated.

The item-person map, in which the mean of each $\theta_{Di}$ latent dimension was constrained to zero such that all item difficulty parameters were estimated, is illustrated in Figure 50. This map, from left to right, provides visual estimates of cadet element leader effectiveness abilities as rated by the subordinates on each dimension (i.e., the latent ability distributions for each dimension are annotated by groupings of ‘x’ where each ‘x’ represents 15.3 cases) followed by the item difficulties relative to each dimension. Since the items are plotted based on their estimated difficulty logit positions (e.g., the numeral ‘3’ representing item5 is plotted at its difficulty estimate of -4.71
logits), items near the top of the grouping of numerals are more difficult to endorse than those at the bottom.

By constraining the mean of each $\theta_D$ latent dimension to zero (i.e., scaling the item difficulty parameters to the $\theta$ metric), the item-person map clearly revealed two problematic areas: 1) the items are only measuring levels of subordinate-rated cadet element leader effectiveness abilities much below the latent trait means on each dimension, and 2) that while the items are fairly dispersed from much below the means of the latent traits, there is too much item overlap (i.e., item difficulty redundancy) between the -4.0 and -5.0 logit positions.
Another perspective was gained from review of the item-person map in Figure 51, in which the means of the item difficulty parameters on each dimension were constrained.
to zero such that the means for each latent ability distribution were estimated. Inspection of the item-person maps, based on the means of each latent trait distribution being greater than the model expected value of zero, revealed that cadet subordinates found it easy to endorse the items relative to each dimension’s mean. Along the range of endorsability, items from the takes care of people dimension with a mean of 5.06 logits were very easy to endorse while items from the negotiating dimension with a mean of 3.83 logits were somewhat easy to endorse.
Figure 51. LMI Subordinate-Rating Six-Factor Post Hoc Multidimensional Model Latent Variable Item-Person Map (ConQuest 3.0). 1 = item2; 2 = item4; 3 = item5; 4 = item7; 5 = item8; 6 = item9; 7 = item10; 8 = item13; 9 = item14; 10 = item16; 11 = item17; 12 = item18; 13 = item19; 14 = item20; 15 = item21; 16 = item25; 17 = item27; 18 = item29. Develops and Inspires Others mean = 4.79 logits; Takes Care of People mean = 5.06 logits; Builds Teams and Coalitions mean = 4.82 logits; Negotiating mean = 3.83 logits; Vision mean = 5.03 logits; Adaptability mean = 4.21 logits.
Research Question Five

The typological latent factor structure of the November 2011 CMV was assessed by exploratory LCA techniques on the best fitting post hoc modified model from research question one with Mplus Version 7 (Muthén & Muthén, 2012a) by incorporating Wang and Wang’s (2012) three-step modeling approach: 1) determine the optimal number of latent classes, 2) evaluate the quality of the classification of latent class membership, and 3) define the latent classes.

Optimal number of latent classes. The item responses from the CMV nine-factor post hoc modified model from research question one (see Figure 23) were selected as the input data for the LCA. The 27 items in the model, based on the modified CFA model, were designed to measure character virtues based on the following nine theoretical constructs: courage, accountability, humility, duty, care for others, self-control, respect for human dignity, attention to detail, and excellence. In order to ensure sufficient values in each cell of the contingency table, the rating scale was recoded into dichotomous responses (Collins & Lanza, 2010); for example, the item responses “very much like me” and “like me” were recoded as “like me” with a value of 1, while the item responses “neutral,” “unlike me,” and “very much unlike me” were recoded as “unlike me” with a value of 0. The latent class model analyzed, in which the “boxed” observed categorical indicators along with associated “circled” error components measured the unobserved “circled” categorical latent class variable $c$, is illustrated in Figure 52.
The typical Mplus input file specification for estimating the fit of the $k$-class model to be compared with a series of increasing class number models is displayed in Figure 52.
Figure 53. The results of the specification illustrate that the dependent variables (e.g., item1-item2, item5, item7-item8, item10-item12, item14, item17-item19, item21-item22, item26-item29, item34, item36-item37, item39-item41, and item43-item45), representing the dichotomously scored items, were treated as ordered categorical variables in the model and estimation process of \( k \) classes (e.g., CLASSES ARE c(\( k \)) option under the VARIABLE command) through the Mplus CATEGORICAL option (Muthén & Muthén, 2012b). Additionally, to avoid local maxima of likelihood when greater than two classes were specified, the Mplus input file included the STARTS and STITERATIONS options under the ANALYSIS command to specify random sets of starting values (greater than the defaults) for the initial and final stages of optimization and for the number of iterations in each optimization, respectively (Muthén & Muthén, 2012b, Wang & Wang 2012). Finally, to ensure unbiased BLRT \( p \)-values, the input file also included the LRTBOOTSTRAP and LRTSTARTS options under the ANALYSIS command to increase the number of bootstrap draws and increase the initial stage random starts and final stage optimizations from the default values, respectively (Muthén & Muthén, 2012b, Wang & Wang, 2012). Wang and Wang’s (2012) suggested values for the previously described four options were incorporated in all model specifications.
In order to obtain evidence that model estimation resulted in global maximum of likelihood, two specific random seeds associated with the initial best log-likelihood value from each model were specified after arriving at each initial solution. The OPTSEED option under the ANALYSIS command was set equal to a seed of a random start associated with the best log-likelihood value after setting the STARTS option under the ANALYSIS command to zero (Wang & Wang, 2012). This procedure, recommended by Wang and Wang (2012), ensured each initial best log-likelihood solution was replicated at least twice, providing evidence of global maxima solutions. The typical Mplus input file specification for this type of solution replication is given in Figure 54.

Figure 53. CMV LCA Model Specification (3-Class Model—Other Classes Similar) (Mplus Version 7)
The optimal number of classes was determined by analyzing the fit of a series of increasing class number models by comparing the \( k \)-class model with the \((k-1)\)-class model (Wang & Wang, 2012). The fit statistics and information criterion indices for the models, which ranged from 1 to 6 latent classes, are tabulated in Table 56. Both the LMR LR test \( (p = 0.142) \) and the ALMR LR test \( (p = 0.144) \) were statistically non-significant in the 4-class model; therefore, the test failed to reject the 3-class model in favor for a four or more class model. While the non-decreasing BIC \( (7645.20) \) of the 5-class model supported evidence for the 4-class model, the non-decreasing ABIC \( (7165.14) \) of the 6-class model supported evidence for the 5-class model, and the statistically significant BLRT \( (p < 0.001) \) supported evidence for at least six classes, no single class showed two sources of rejection evidence except for the 4-class solution.
Therefore, the fit of the 3-class model was determined to be adequate and the preferred model for further analysis.

**Table 56**

<table>
<thead>
<tr>
<th>Statistic/Index</th>
<th>1-class</th>
<th>2-class</th>
<th>3-class</th>
<th>4-class</th>
<th>5-class</th>
<th>6-class</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMR LRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>0.026</td>
<td>0.142</td>
<td>0.117</td>
<td>0.656</td>
</tr>
<tr>
<td>ALMR LRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BLRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AIC</td>
<td>8076.55</td>
<td>7535.29</td>
<td>7350.39</td>
<td>7237.72</td>
<td>7154.06</td>
<td>7104.49</td>
</tr>
<tr>
<td>BIC</td>
<td>8171.95</td>
<td>7729.63</td>
<td>7643.67</td>
<td>7629.92</td>
<td>7645.20</td>
<td>7694.56</td>
</tr>
<tr>
<td>ABI</td>
<td>8086.36</td>
<td>7555.27</td>
<td>7380.54</td>
<td>7278.03</td>
<td>7204.54</td>
<td>7165.14</td>
</tr>
</tbody>
</table>

*Note.* LMR LRT = Lo-Mendell-Rubin Likelihood Ratio Test; ALMR LRT = Adjusted Lo-Mendell-Rubin Likelihood Ratio Test; BLRT = Bootstrap Likelihood Ratio Test.

**Quality of the classification.** With the 3-class model determined to be the optimal number of classes based on model fit, the quality of the classification was examined on the basis of the estimated posterior probabilities. While membership of individuals into a latent class is not definitely determined, individuals are assigned into a latent class based on their largest posterior probability; the probability of misclassification is low when an individual’s highest posterior probability is close to 1.0 (Wang & Wang, 2012).

The final class counts and proportions for the latent class patterns, based on the estimated posterior probabilities for a cadet to be partially assigned to each class, are given in Table 57. From the table after rounding, 101 cadets (40.0%) were assigned to Class 1, 95 cadets (37.6%) were assigned to Class 2, and 57 cadets (22.4%) were assigned to Class 3—which yielded adequate size and sample proportion among the classes.
The average latent class posterior probabilities for the most likely latent class membership are reported in Table 58. The probability of correct class membership for cadets assigned to the first class was 0.96, while the probability of misclassification was 0.04. Similarly, for cadets assigned to the second class, the probability of correct class membership was 0.97, while the probability of misclassification was 0.03; for cadets assigned to the third class, the probability of correct class membership was 0.94, while the probability of misclassification was 0.06. These average latent class probabilities for most likely latent class membership well exceeded Nagin’s (2005) criterion for minimum acceptable class membership classification based on an average posterior probability of at least 0.7 for all groups.

Table 57
*CMV Final Latent Class Counts and Proportions*

<table>
<thead>
<tr>
<th>Classes</th>
<th>Counts</th>
<th>Proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>101.19</td>
<td>40.0%</td>
</tr>
<tr>
<td>2</td>
<td>95.15</td>
<td>37.6%</td>
</tr>
<tr>
<td>3</td>
<td>56.66</td>
<td>22.4%</td>
</tr>
</tbody>
</table>

Another criterion to summarize posterior misclassification is based on entropy, a single value summary of the degree of uncertainty or disorder in the model scaled such that large values indicate less classification error (Collins & Lanza, 2010). The entropy
statistic for the 3-class model was 0.90; this is considered a high value according to Clark (2010) and it can be concluded that latent class membership classification quality was adequate.

**Latent classes defined.** The heterogeneity in the sample cadet population was determined by examination of the estimated item-response probability of endorsing “like me” for each of the 27 items. The three latent classes—*strong identification with virtues, moderate identification with virtues,* and *weak identification with virtues*—were defined by the researcher based on the observed pattern of item-response probabilities. The *strong identification with virtues* class, denoted as Class 2 consisting of 95 cadets, had the highest item-response probabilities for the greatest number of items (i.e., had the highest probability of endorsing “like me” on 20 out of 27 items). For example, Class 2 had the highest probabilities of endorsing “like me” on all items except for the three duty items (*item17-item19*), the three attention to detail items (*item39-item41*), and the last excellence item (*item45*) in which this class had the second highest probabilities of endorsing those remaining seven items. In a similar methodology, the researcher defined Class 1 containing 101 cadets as *moderate identification with virtues* and Class 3 containing 57 cadets as *weak identification with virtues*.

The unconditional latent class probabilities and the conditional probabilities for endorsing “like me” are reported by latent class in Table 59. Conditional probability profiles for endorsing “like me” for the 3-Class model are illustrated in Figure 55. By referencing the unconditional probability, the typology of self-perceived cadet *virtue* identification can be better understood. For example, the 37.6% of cadets assigned to
latent Class 2 (i.e., strong identification with virtues) had a high probability of identifying with all of the constructs of the CMV; while this class had the second highest probabilities relative to the other classes of endorsing the duty and attention to detail constructs and the last item of the excellence construct, the probabilities for doing so were still high ranging from 0.76 to 0.92. Regarding the 40.0% of cadets assigned to latent Class 1 (i.e., moderate identification with virtues), the probabilities of endorsing “like me” fell between the probabilities of the other two classes in five of the nine constructs of the CMV; however, this class had the lowest probabilities of endorsing the duty, self-control, attention to detail, and excellence constructs and the second item of the courage construct. Of the remaining 22.4% of cadets assigned to latent Class 3 (i.e., weak identification with virtues) the probabilities of endorsing “like me” were less than the probabilities of the other two classes in five of the nine constructs of the CMV; however, this class had the highest probabilities of endorsing duty, attention to detail, and last item of the excellence construct while the probabilities of endorsing “like me” on the self-control construct, second item of courage, and other excellence items fell between the other two classes. Reference to the conditional probability profiles for endorsing “like me” in the 3-Class model in Figure 55 visually illustrate the heterogeneity in the sample cadet population.
Table 59
**CMV 3-Class LCA Membership Probabilities**

<table>
<thead>
<tr>
<th>Item</th>
<th>Class 1—<em>Moderate ID</em> Probability</th>
<th>Class 2—<em>Strong ID</em> Probability</th>
<th>Class 3—<em>Weak ID</em> Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unconditional Probability</td>
<td>Conditional “Like Me” Probability</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.376</td>
<td>0.224</td>
</tr>
<tr>
<td>Courage <em>i1</em></td>
<td>0.635</td>
<td>0.877</td>
<td>0.589</td>
</tr>
<tr>
<td>Courage <em>i2</em></td>
<td>0.615</td>
<td>0.867</td>
<td>0.658</td>
</tr>
<tr>
<td>Courage <em>i5</em></td>
<td>0.267</td>
<td>0.592</td>
<td>0.188</td>
</tr>
<tr>
<td>Account <em>i7</em></td>
<td>0.453</td>
<td>0.681</td>
<td>0.289</td>
</tr>
<tr>
<td>Account <em>i8</em></td>
<td>0.713</td>
<td>0.886</td>
<td>0.593</td>
</tr>
<tr>
<td>Account <em>i10</em></td>
<td>0.866</td>
<td>0.978</td>
<td>0.747</td>
</tr>
<tr>
<td>Humility <em>i11</em></td>
<td>0.533</td>
<td>0.788</td>
<td>0.248</td>
</tr>
<tr>
<td>Humility <em>i12</em></td>
<td>0.487</td>
<td>0.711</td>
<td>0.230</td>
</tr>
<tr>
<td>Humility <em>i14</em></td>
<td>0.818</td>
<td>0.967</td>
<td>0.586</td>
</tr>
<tr>
<td>Duty <em>i17</em></td>
<td>0.488</td>
<td>0.915</td>
<td>0.980</td>
</tr>
<tr>
<td>Duty <em>i18</em></td>
<td>0.487</td>
<td>0.883</td>
<td>1.000</td>
</tr>
<tr>
<td>Duty <em>i19</em></td>
<td>0.334</td>
<td>0.861</td>
<td>0.852</td>
</tr>
<tr>
<td>CFO <em>i21</em></td>
<td>0.840</td>
<td>0.903</td>
<td>0.619</td>
</tr>
<tr>
<td>CFO <em>i22</em></td>
<td>0.899</td>
<td>0.989</td>
<td>0.668</td>
</tr>
<tr>
<td>CFO <em>i26</em></td>
<td>0.757</td>
<td>0.853</td>
<td>0.552</td>
</tr>
<tr>
<td>Self-Control <em>i27</em></td>
<td>0.151</td>
<td>0.904</td>
<td>0.206</td>
</tr>
<tr>
<td>Self-Control <em>i28</em></td>
<td>0.209</td>
<td>0.958</td>
<td>0.295</td>
</tr>
<tr>
<td>Self-Control <em>i29</em></td>
<td>0.201</td>
<td>0.917</td>
<td>0.202</td>
</tr>
<tr>
<td>RFHD <em>i34</em></td>
<td>0.690</td>
<td>0.748</td>
<td>0.317</td>
</tr>
<tr>
<td>RFHD <em>i36</em></td>
<td>0.725</td>
<td>0.907</td>
<td>0.606</td>
</tr>
<tr>
<td>RFHD <em>i37</em></td>
<td>0.746</td>
<td>0.940</td>
<td>0.601</td>
</tr>
<tr>
<td>AtD <em>i39</em></td>
<td>0.480</td>
<td>0.758</td>
<td>0.906</td>
</tr>
<tr>
<td>AtD <em>i40</em></td>
<td>0.541</td>
<td>0.869</td>
<td>0.981</td>
</tr>
<tr>
<td>AtD <em>i41</em></td>
<td>0.505</td>
<td>0.761</td>
<td>0.927</td>
</tr>
<tr>
<td>Excellence <em>i43</em></td>
<td>0.547</td>
<td>0.889</td>
<td>0.761</td>
</tr>
<tr>
<td>Excellence <em>i44</em></td>
<td>0.360</td>
<td>0.809</td>
<td>0.681</td>
</tr>
<tr>
<td>Excellence <em>i45</em></td>
<td>0.528</td>
<td>0.923</td>
<td>1.000</td>
</tr>
</tbody>
</table>

*Note.* Account = Accountability; CFO = Care for Others; RFHD = Respect for Human Dignity; AtD = Attention to Detail; *i*= *item*; ID = Identification. To calculate the conditional probability of “Unlike Me” for any item, subtract the conditional probability of “Like Me” from 1.
**Research Question Six**

The typological latent factor structures of the September 2012 LMI *self-rating* and *subordinate-rating* versions were assessed by exploratory LCA techniques on the best fitting *post hoc* modified models from research question four with *Mplus* Version 7 (Muthén & Muthén, 2012a) by incorporating Wang and Wang’s (2012) three-step modeling approach: 1) determine the optimal number of latent classes, 2) evaluate the quality of the classification of latent class membership, and 3) define the latent classes. As in Gerber et al. (2009), latent class analyses of the October 2012 LMI *self-rating* and *subordinate-rating* versions were conducted in a confirmatory manner to validate the appropriateness and fit of the exploratory models.

**LMI self-rating exploratory model.** The item responses from the LMI self-rating six-factor *post hoc* modified model from research question four (see Figure 42)
were selected as the input data for the LCA. The 18 items in the model, based on the modified IRT model, were designed to measure cadet element leader *effectiveness* based on the following six USAF institutional *leadership effectiveness* sub-competencies: *develops and inspires others, takes care of people, builds teams and coalitions, negotiating, vision, and adaptability*. In order to ensure sufficient values in each cell of the contingency table, the rating scale was recoded into dichotomous responses (Collins & Lanza, 2010); for example, the item responses “very much like me” and “like me” were recoded as “like me” with a value of 1, while the item responses “neutral,” “unlike me,” and “very much unlike me” were recoded as “unlike me” with a value of 0. The latent class model analyzed, in which the “boxed” observed categorical indicators along with associated “circled” error components measured the unobserved “circled” categorical latent class variable $c$, is illustrated in Figure 56.
Optimal number of latent classes. The typical Mplus input file specification for estimating the fit of the $k$-class model to be compared with a series of increasing class number models is displayed in Figure 57. The results of the specification illustrate that the dependent variables (e.g., item1, item3, item5, item7-item9, item11-item12, item15-item18, item22-item24, item26-item27, and item29), representing the dichotomously scored items, were treated as ordered categorical variables in the model and estimation process of $k$ classes (e.g., CLASSES ARE c(k) option under the VARIABLE command) through the Mplus CATEGORICAL option (Muthén & Muthén, 2012b). Additionally, to
avoid local maxima of likelihood when greater than two classes were specified, the
Mplus input file included the STARTS and STITERATIONS options under the
ANALYSIS command to specify random sets of starting values (greater than the
defaults) for the initial and final stages of optimization and for the number of iterations in
each optimization, respectively (Muthén & Muthén, 2012b, Wang & Wang 2012).
Finally, to ensure unbiased BLRT $p$-values, the input file also included the
LRTBOOTSTRAP and LRTSTARTS options under the ANALYSIS command to
increase the number of bootstrap draws and increase the initial stage random starts and
final stage optimizations from the default values, respectively (Muthén & Muthén, 2012b,
described four options were incorporated in all model specifications.

```
Mplus Version 7
Muthén & Muthén
03/17/2013   8:09 PM
Input instructions
TITLE: LMI self 8-Class LCA model;
DATA:
   FILE IS C:/Users/David/Desktop/Dissertation/LCA/LMISelf  
   /Class/LMISelf_t1_18_items_r-357_Appendix_B_order_recode_0_or_1.dat;
VARIABLE:
   NAMES ARE item1-item8 item3 item7-item9 item11-item12  
   item5-item8 item22-item24 item26-item27 item29;
   USEVARIABLES ARE item1 item3 item5 item7-item9 item11-item12  
   item5-item8 item22-item24 item26-item27 item29;
   CLASSES ARE c(3);
   CATEGORICAL ARE item1 item3 item5 item7-item9 item11-item12  
   item5-item8 item22-item24 item26-item27 item29;
ANALYSIS:
   TYPE=EVENT;
   STARTS = 1 000 250;
   STITERATIONS = 20;
   LRTBOOTSTRAP = 200;
   LRTSTARTS = 20 5 100 25;
MODEL:
   NOREML;
OUTPUT:
   TECH11 TECH14;
PLOT:
   TYPE = plot3;
   SERIES item1 item3 item5 item7-item9 item11-item12  
   item5-item8 item22-item24 item26-item27 item29 (*);
```

Figure 57. LMI Self-Rating LCA Model Specification (3-Class Model—Other Classes Similar) (Mplus Version 7)

In order to obtain evidence that model estimation resulted in global maximum of
likelihood, two specific random seeds associated with the initial best log-likelihood value
from each model were specified after arriving at each initial solution. The OPTSEED option under the ANALYSIS command was set equal to a seed of a random start associated with the best log-likelihood value after setting the STARTS option under the ANALYSIS command to zero (Wang & Wang, 2012). This procedure, recommended by Wang and Wang (2012), ensured each initial best log-likelihood solution was replicated at least twice, providing evidence of global maxima solutions. The typical Mplus input file specification for this type of solution replication is given in Figure 58.

![Mplus Input Code](image)

**Figure 58.** LMI Self-Rating LCA Replication Model Specification (3-Class Replication Model—Other Classes Similar) (Mplus Version 7)

The optimal number of classes was determined by analyzing the fit of a series of increasing class number models by comparing the $k$-class model with the $(k-1)$-class model (Wang & Wang, 2012). The fit statistics and information criterion indices for the models, which ranged from 1 to 4 latent classes, are tabulated in Table 60. Both the LMR LR test ($p = 0.076$) and the ALMR LR test ($p = 0.078$) were statistically non-significant in the 4-class model; therefore, the test failed to reject the 3-class model in
favor for a four or more class model. Additionally, the non-decreasing BIC (5339.63) of the 4-class model also supported evidence for the 3-class model. Therefore, the fit of the 3-class model was determined to be adequate and the preferred model for further analysis.

Table 60
*LMI Self-Rating LCA Model Comparisons*

<table>
<thead>
<tr>
<th>Statistic/Index</th>
<th>1-class</th>
<th>2-class</th>
<th>3-class</th>
<th>4-class</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMR LRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>0.032</td>
<td>0.076</td>
</tr>
<tr>
<td>ALMR LRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>0.033</td>
<td>0.078</td>
</tr>
<tr>
<td>BLRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AIC</td>
<td>5871.50</td>
<td>5173.61</td>
<td>5071.56</td>
<td>5048.80</td>
</tr>
<tr>
<td>BIC</td>
<td>5941.30</td>
<td>5317.09</td>
<td>5288.71</td>
<td>5339.63</td>
</tr>
<tr>
<td>ABIC</td>
<td>5884.20</td>
<td>5199.70</td>
<td>5111.06</td>
<td>5101.70</td>
</tr>
</tbody>
</table>

*Note.* LMR LRT = Lo-Mendell-Rubin Likelihood Ratio Test; ALMR LRT = Adjusted Lo-Mendell-Rubin Likelihood Ratio Test; BLRT = Bootstrap Likelihood Ratio Test.

**Quality of the classification.** With the 3-class model determined to be the optimal number of classes based on model fit, the quality of the classification was examined on the basis of the estimated posterior probabilities. While membership of individuals into a latent class is not definitely determined, individuals are assigned into a latent class based on their largest posterior probability; the probability of misclassification is low when an individual’s highest posterior probability is close to 1.0 (Wang & Wang, 2012).

The final class counts and proportions for the latent class patterns, based on the estimated posterior probabilities for a cadet element leader to be partially assigned to each class, are given in Table 61. From the table after rounding, 141 cadet element leaders (39.4%) were assigned to Class 1, 163 cadet element leaders (45.7%) were
assigned to Class 2, and 53 cadet element leaders (15.0%) were assigned to Class 3—which yielded adequate size and sample proportion among the classes.

Table 61
*LMI Self-Rating Final Latent Class Counts and Proportions*

<table>
<thead>
<tr>
<th>Classes</th>
<th>Counts</th>
<th>Proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>140.55</td>
<td>39.4%</td>
</tr>
<tr>
<td>2</td>
<td>163.03</td>
<td>45.7%</td>
</tr>
<tr>
<td>3</td>
<td>53.42</td>
<td>15.0%</td>
</tr>
</tbody>
</table>

The average latent class posterior probabilities for the most likely latent class membership are reported in Table 62. The probability of correct class membership for cadet element leaders assigned to the first class was 0.92, while the probability of misclassification was 0.08. Similarly, for cadet element leaders assigned to the second class, the probability of correct class membership was 0.91, while the probability of misclassification was 0.09; for cadets assigned to the third class, the probability of correct class membership was 0.93, while the probability of misclassification was 0.07. These average latent class probabilities for most likely latent class membership well exceeded Nagin’s (2005) criterion for minimum acceptable class membership classification based on an average posterior probability of at least 0.7 for all groups.

Table 62
*LMI Self-Rating Average Latent Class Probabilities for Most Likely Latent Class Membership*

<table>
<thead>
<tr>
<th>Classes</th>
<th>Probability of Class 1 Membership</th>
<th>Probability of Class 2 Membership</th>
<th>Probability of Class 3 Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.920</td>
<td>0.080</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>0.055</td>
<td>0.913</td>
<td>0.032</td>
</tr>
<tr>
<td>3</td>
<td>&lt;0.001</td>
<td>0.073</td>
<td>0.927</td>
</tr>
</tbody>
</table>
Another criterion to summarize posterior misclassification is based on entropy, a single value summary of the degree of uncertainty or disorder in the model scaled such that large values indicate less classification error (Collins & Lanza, 2010). The entropy statistic for the 3-class model was 0.81; this is considered a high value according to Clark (2010) and it can be concluded that latent class membership classification quality was adequate.

**Latent classes defined.** The heterogeneity in the sample cadet element leader population was determined by examination of the estimated item-response probability of endorsing “like me” for each of the 18 items. The three latent classes—*highly effective self-rated leaders, moderately effective self-rated leaders, and somewhat effective self-rated leaders*—were defined by the researcher based on the observed pattern of item-response probabilities. The *highly effective self-rated leaders* class, denoted as Class 1 consisting of 141 element leaders, had the highest item-response probabilities for each of the 18 items (i.e., had the highest probability of endorsing “like me”). The researcher defined Class 2, which contained 163 cadets with the second highest item-response probabilities for each of the 18 items, as *moderately effective self-rated leaders*; Class 3, which contained 53 cadets and had the lowest item-response probabilities for each of the 18 items, was defined as *somewhat effective self-rated leaders*.

The unconditional latent class probabilities and the conditional probabilities for endorsing “like me” are reported by latent class in Table 63. Conditional probability profiles for endorsing “like me” for the 3-Class model are illustrated in Figure 59 and visually illustrate the heterogeneity in the sample cadet element leader population.
Table 63
*LMI Self-Rating 3-Class LCA Membership Probabilities*

<table>
<thead>
<tr>
<th>Item</th>
<th>Class 1 Probability</th>
<th>Class 2 Probability</th>
<th>Class 3 Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unconditional</td>
<td>Conditional “Like Me”</td>
<td></td>
</tr>
<tr>
<td>DaIO i1</td>
<td>0.394</td>
<td>0.968</td>
<td>0.829</td>
</tr>
<tr>
<td>DaIO i3</td>
<td>1.000</td>
<td>0.996</td>
<td>0.863</td>
</tr>
<tr>
<td>DaIO i5</td>
<td>0.999</td>
<td>0.990</td>
<td>0.811</td>
</tr>
<tr>
<td>TCoP i7</td>
<td>0.993</td>
<td>0.990</td>
<td>0.773</td>
</tr>
<tr>
<td>TCoP i8</td>
<td>0.093</td>
<td>0.990</td>
<td>0.969</td>
</tr>
<tr>
<td>TCoP i9</td>
<td>0.947</td>
<td>0.990</td>
<td>0.796</td>
</tr>
<tr>
<td>BTaC i11</td>
<td>1.000</td>
<td>0.990</td>
<td>0.990</td>
</tr>
<tr>
<td>BTaC i12</td>
<td>0.974</td>
<td>0.953</td>
<td>0.953</td>
</tr>
<tr>
<td>BTaC i15</td>
<td>1.000</td>
<td>0.923</td>
<td>0.923</td>
</tr>
<tr>
<td>Negotiating i1 6</td>
<td>0.989</td>
<td>0.692</td>
<td>0.692</td>
</tr>
<tr>
<td>Negotiating i1 7</td>
<td>0.955</td>
<td>0.634</td>
<td>0.634</td>
</tr>
<tr>
<td>Negotiating i1 8</td>
<td>0.834</td>
<td>0.466</td>
<td>0.466</td>
</tr>
<tr>
<td>Vision i22</td>
<td>0.977</td>
<td>0.753</td>
<td>0.753</td>
</tr>
<tr>
<td>Vision i23</td>
<td>1.000</td>
<td>0.883</td>
<td>0.883</td>
</tr>
<tr>
<td>Vision i24</td>
<td>0.986</td>
<td>0.692</td>
<td>0.692</td>
</tr>
<tr>
<td>Adaptability i2 6</td>
<td>0.947</td>
<td>0.736</td>
<td>0.736</td>
</tr>
<tr>
<td>Adaptability i2 7</td>
<td>0.972</td>
<td>0.720</td>
<td>0.720</td>
</tr>
<tr>
<td>Adaptability i2 9</td>
<td>0.982</td>
<td>0.704</td>
<td>0.704</td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions; *ii* = *item* 1. To calculate the conditional probability of “Unlike Me” for any item, subtract the conditional probability of “Like Me” from 1.
LMI subordinate-rating exploratory model. The item responses from the LMI subordinate-rating six-factor post hoc modified model from research question four (see Figure 49) were selected as the input data for the LCA. The 18 items in the model, based on the modified IRT model, were designed to measure cadet element leader effectiveness from subordinate ratings based on the following six USAF institutional leadership effectiveness sub-competencies: develops and inspires others, takes care of people, builds teams and coalitions, negotiating, vision, and adaptability. In order to ensure sufficient values in each cell of the contingency table, the rating scale was recoded into dichotomous responses (Collins & Lanza, 2010); for example, the item responses “very much like the Leader” and “like the Leader” were recoded as “like the Leader” with a value of 1, while the item responses “neutral,” “unlike the Leader,” and “very much unlike the Leader” were recoded as “unlike the Leader” with a value of 0. The latent
class model analyzed, in which the “boxed” observed categorical indicators along with associated “circled” error components measured the unobserved “circled” categorical latent class variable $c$, is illustrated in Figure 60.

![Diagram of LMI Subordinate-Rating LCA Model (Amos Version 18)](image)

**Figure 60.** LMI Subordinate-Rating LCA Model (Amos Version 18)

*Optimal number of latent classes.* The typical Mplus input file specification for estimating the fit of the $k$-class model to be compared with a series of increasing class number models is displayed in Figure 61. The results of the specification illustrate that the dependent variables (e.g., item2, item4-item5, item7-item10, item13-item14, item16-item21, item25, item27, and item29), representing the dichotomously scored items, were
treated as ordered categorical variables in the model and estimation process of $k$ classes (e.g., CLASSES ARE c($k$) option under the VARIABLE command) through the Mplus CATEGORICAL option (Muthén & Muthén, 2012b). Additionally, to avoid local maxima of likelihood when greater than two classes were specified, the Mplus input file included the STARTS and STITERATIONS options under the ANALYSIS command to specify random sets of starting values (greater than the defaults) for the initial and final stages of optimization and for the number of iterations in each optimization, respectively (Muthén & Muthén, 2012b, Wang & Wang 2012). Finally, to ensure unbiased BLRT $p$-values, the input file also included the LRTBOOTSTRAP and LRTSTARTS options under the ANALYSIS command to increase the number of bootstrap draws and increase the initial stage random starts and final stage optimizations from the default values, respectively (Muthén & Muthén, 2012b, Wang & Wang, 2012). Wang and Wang’s (2012) suggested values for the previously described four options were incorporated in all model specifications.
In order to obtain evidence that model estimation resulted in global maximum of likelihood, two specific random seeds associated with the initial best log-likelihood value from each model were specified after arriving at each initial solution. The OPTSEED option under the ANALYSIS command was set equal to a seed of a random start associated with the best log-likelihood value after setting the STARTS option under the ANALYSIS command to zero (Wang & Wang, 2012). This procedure, recommended by Wang and Wang (2012), ensured each initial best log-likelihood solution was replicated at least twice, providing evidence of global maxima solutions. The typical Mplus input file specification for this type of solution replication is given in Figure 62.

Figure 61. LMI Subordinate-Rating LCA Model Specification (5-Class Model—Other Classes Similar) (Mplus Version 7)
The optimal number of classes was determined by analyzing the fit of a series of increasing class number models by comparing the \(k\)-class model with the \((k-1)\)-class model (Wang & Wang, 2012). The fit statistics and information criterion indices for the models, which ranged from 1 to 6 latent classes, are tabulated in Table 64. Both the LMR LR test \((p = 0.265)\) and the ALMR LR test \((p = 0.268)\) were statistically non-significant in the 6-class model; therefore, the test failed to reject the 5-class model in favor for a six or more class model. While the non-decreasing BIC (18143.92) of the 5-class model supported evidence for the 4-class model and the statistically significant BLRT \((p < 0.001)\) supported evidence for at least 6 six classes, no single class showed two sources of rejection evidence except for the 6-class solution. Therefore, the fit of the 5-class model was determined to be adequate and the preferred model for further analysis.
Table 64

*LMI Subordinate-Rating LCA Model Comparisons*

<table>
<thead>
<tr>
<th>Statistic/Index</th>
<th>1-class</th>
<th>2-class</th>
<th>3-class</th>
<th>4-class</th>
<th>5-class</th>
<th>6-class</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMR LRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.265</td>
</tr>
<tr>
<td>ALMR LRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.268</td>
</tr>
<tr>
<td>BLRT p-value</td>
<td>N/A</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>AIC</td>
<td>28517.38</td>
<td>19519.04</td>
<td>18013.04</td>
<td>17726.78</td>
<td>17628.55</td>
<td>17562.77</td>
</tr>
<tr>
<td>BIC</td>
<td>28616.07</td>
<td>19721.90</td>
<td>18320.07</td>
<td>18137.98</td>
<td>18143.92</td>
<td>18182.32</td>
</tr>
<tr>
<td>ABIC</td>
<td>28558.89</td>
<td>19604.36</td>
<td>18142.16</td>
<td>17899.71</td>
<td>17845.29</td>
<td>17823.33</td>
</tr>
</tbody>
</table>

*Note.* LMR LRT = Lo-Mendell-Rubin Likelihood Ratio Test; ALMR LRT = Adjusted Lo-Mendell-Rubin Likelihood Ratio Test; BLRT = Bootstrap Likelihood Ratio Test.

**Quality of the classification.** With the 5-class model determined to be the optimal number of classes based on model fit, the quality of the classification was examined on the basis of the estimated posterior probabilities. While membership of individuals into a latent class is not definitely determined, individuals are assigned into a latent class based on their largest posterior probability; the probability of misclassification is low when an individual’s highest posterior probability is close to 1.0 (Wang & Wang, 2012).

The final class counts and proportions for the latent class patterns, based on the estimated posterior probabilities for cadet subordinates to be partially assigned to each class, are given in Table 65. From the table after rounding, 1022 cadet subordinates (57.5%) were assigned to Class 1, 60 cadet subordinates (3.4%) were assigned to Class 2, 243 cadet subordinates (13.7%) were assigned to Class 3, 329 cadet subordinates (18.5%) were assigned to Class 4, and 123 cadet subordinates (6.9%) were assigned to Class 5—which yielded adequate size and sample proportion among the classes.
Table 65
*LMI Subordinate-Rating Final Latent Class Counts and Proportions*

<table>
<thead>
<tr>
<th>Classes</th>
<th>Counts</th>
<th>Proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1022.28</td>
<td>57.5%</td>
</tr>
<tr>
<td>2</td>
<td>60.46</td>
<td>3.4%</td>
</tr>
<tr>
<td>3</td>
<td>243.05</td>
<td>13.7%</td>
</tr>
<tr>
<td>4</td>
<td>328.69</td>
<td>18.5%</td>
</tr>
<tr>
<td>5</td>
<td>122.52</td>
<td>6.9%</td>
</tr>
</tbody>
</table>

The average latent class posterior probabilities for the most likely latent class membership are reported in Table 66. The probability of correct class membership for cadet subordinates assigned to the first and second classes was 0.96, while the probability of misclassification was 0.04. For cadets assigned to the third, fourth, and fifth classes, the probability of correct class membership was 0.88, 0.83, and 0.90, while the probability of misclassification was 0.12, 0.17, and 0.10, respectively. These average latent class probabilities for most likely latent class membership well exceeded Nagin’s (2005) criterion for minimum acceptable class membership classification based on an average posterior probability of at least 0.7 for all groups.

Table 66
*LMI Subordinate-Rating Average Latent Class Probabilities for Most Likely Latent Class Membership*

<table>
<thead>
<tr>
<th>Classes</th>
<th>Class 1 Membership</th>
<th>Class 2 Membership</th>
<th>Class 3 Membership</th>
<th>Class 4 Membership</th>
<th>Class 5 Membership</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.956</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.044</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>&lt;0.001</td>
<td>0.959</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.041</td>
</tr>
<tr>
<td>3</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.881</td>
<td>0.085</td>
<td>0.035</td>
</tr>
<tr>
<td>4</td>
<td>0.066</td>
<td>&lt;0.001</td>
<td>0.102</td>
<td>0.833</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>5</td>
<td>&lt;0.001</td>
<td>0.016</td>
<td>0.087</td>
<td>&lt;0.001</td>
<td>0.897</td>
</tr>
</tbody>
</table>
Another criterion to summarize posterior misclassification is based on entropy, a single value summary of the degree of uncertainty or disorder in the model scaled such that large values indicate less classification error (Collins & Lanza, 2010). The entropy statistic for the 5-class model was 0.87; this is considered a high value according to Clark (2010) and it can be concluded that latent class membership classification quality was adequate.

**Latent classes defined.** The heterogeneity in the sample cadet subordinate population was determined by examination of the estimated item-response probability of endorsing “like the Leader” for each of the 18 items. The five latent classes—see leaders extremely effective, see leaders highly effective, see leaders somewhat effective, see leaders somewhat ineffective, and see leaders highly ineffective—were defined by the researcher based on the observed pattern of item-response probabilities. The see leaders extremely effective class, denoted as Class 1 consisting of 1022 cadet subordinates, had the highest item-response probabilities for each of the 18 items (i.e., had the highest probability of endorsing “like the Leader”). The researcher defined Class 4, which contained 329 cadet subordinates with the second highest item-response probabilities for each of the 18 items, as see leaders highly effective; Class 3, which contained 243 cadet subordinates had the next highest item-response probabilities for each of the 18 items, was defined as see leaders somewhat effective. The final two classes, Class 5 (123 cadets) and Class 2 (60 cadets), had the lowest item-response probabilities and were defined as see leaders somewhat ineffective and see leaders highly ineffective, respectively.
The unconditional latent class probabilities and the conditional probabilities for endorsing “like the Leader” are reported by latent class in Table 67. Conditional probability profiles for endorsing “like the Leader” for the 5-Class model are illustrated in Figure 63 and visually illustrate the heterogeneity in the sample cadet subordinate population.

Table 67
*LMI Subordinate-Rating 5-Class LCA Membership Probabilities*

<table>
<thead>
<tr>
<th>Item</th>
<th>Class 1 Probability</th>
<th>Class 2 Probability</th>
<th>Class 3 Probability</th>
<th>Class 4 Probability</th>
<th>Class 5 Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>DaIO i2</td>
<td>1.000</td>
<td>0.032</td>
<td>0.640</td>
<td>0.955</td>
<td>0.264</td>
</tr>
<tr>
<td>DaIO i4</td>
<td>0.999</td>
<td>0.029</td>
<td>0.598</td>
<td>0.826</td>
<td>0.219</td>
</tr>
<tr>
<td>DaIO i5</td>
<td>0.997</td>
<td>0.000</td>
<td>0.622</td>
<td>0.899</td>
<td>0.311</td>
</tr>
<tr>
<td>TCoP i7</td>
<td>1.000</td>
<td>0.001</td>
<td>0.687</td>
<td>0.937</td>
<td>0.227</td>
</tr>
<tr>
<td>TCoP i8</td>
<td>0.999</td>
<td>0.080</td>
<td>0.832</td>
<td>0.955</td>
<td>0.747</td>
</tr>
<tr>
<td>TCoP i9</td>
<td>0.989</td>
<td>0.000</td>
<td>0.472</td>
<td>0.805</td>
<td>0.207</td>
</tr>
<tr>
<td>BTaC i10</td>
<td>0.996</td>
<td>0.000</td>
<td>0.662</td>
<td>0.889</td>
<td>0.316</td>
</tr>
<tr>
<td>BTaC i13</td>
<td>0.999</td>
<td>0.000</td>
<td>0.807</td>
<td>0.940</td>
<td>0.544</td>
</tr>
<tr>
<td>BTaC i14</td>
<td>0.985</td>
<td>0.000</td>
<td>0.429</td>
<td>0.829</td>
<td>0.129</td>
</tr>
<tr>
<td>Negotiating i16</td>
<td>0.999</td>
<td>0.060</td>
<td>0.625</td>
<td>0.851</td>
<td>0.328</td>
</tr>
<tr>
<td>Negotiating i17</td>
<td>0.994</td>
<td>0.000</td>
<td>0.447</td>
<td>0.717</td>
<td>0.032</td>
</tr>
<tr>
<td>Negotiating i18</td>
<td>0.914</td>
<td>0.049</td>
<td>0.370</td>
<td>0.484</td>
<td>0.063</td>
</tr>
<tr>
<td>Vision i19</td>
<td>0.990</td>
<td>0.062</td>
<td>0.602</td>
<td>0.873</td>
<td>0.359</td>
</tr>
<tr>
<td>Vision i20</td>
<td>0.995</td>
<td>0.000</td>
<td>0.700</td>
<td>0.906</td>
<td>0.254</td>
</tr>
<tr>
<td>Vision i21</td>
<td>0.999</td>
<td>0.105</td>
<td>0.894</td>
<td>0.965</td>
<td>0.665</td>
</tr>
<tr>
<td>Adapt i25</td>
<td>0.997</td>
<td>0.032</td>
<td>0.562</td>
<td>0.846</td>
<td>0.170</td>
</tr>
<tr>
<td>Adapt i27</td>
<td>0.991</td>
<td>0.000</td>
<td>0.392</td>
<td>0.724</td>
<td>0.124</td>
</tr>
<tr>
<td>Adapt i29</td>
<td>0.998</td>
<td>0.000</td>
<td>0.424</td>
<td>0.842</td>
<td>0.170</td>
</tr>
</tbody>
</table>

*Note.* DaIO = Develops and Inspires Others; TCoP = Takes Care of People; BTaC = Builds Teams and Coalitions; Adapt = Adaptability; *i* = *item*. To calculate the conditional probability of “Unlike the Leader” for any item, subtract the conditional probability of “Like the Leader” from 1.
Figure 63. Conditional Probability Profiles of Endorsing “Like the Leader” for 5-Class LMI Subordinate-Rating LCA Model (Mplus Version 7)

**LMI self-rating confirmatory model.** The item responses from the October 2012 LMI self-rating data corresponding to the item responses from the September 2012 LMI self-rating exploratory LCA (see Figure 56) were selected as the input data for the confirmatory LCA. In order to ensure sufficient values in each cell of the contingency table, the rating scale was recoded into dichotomous responses (Collins & Lanza, 2010); for example, the item responses “very much like me” and “like me” were recoded as “like me” with a value of 1, while the item responses “neutral,” “unlike me,” and “very much unlike me” were recoded as “unlike me” with a value of 0.

**Confirmatory LCA model specification.** Confirmatory LCA was conducted with the October 2012 LMI self-rating data by fixing all item-response probabilities in the new
data to those estimated item-response probabilities from the September 2012 LMI self-rating 3-class LCA model (Finch & Bronk, 2011; Gerber et al., 2009; Muthén & Muthén, 2012b). Under the Mplus MODEL command, the September 2012 item-response probabilities were fixed on each class of the October 2012 model specification (e.g., the Mplus statement “%c#k%” permitted k-class deterministic constraints) by inputting the September 2012 item threshold estimates after the Mplus “item#$1@” statement (Muthén & Muthén, 2012b). For example, the September 2012 threshold estimate for item1 of Class 1 was -3.41; therefore, the Mplus statement to fix the item-response probability for item1 in Class 1 of the October 2012 model was [item1$1@-3.410]. The highly constrained Mplus input file specification for confirming the fit of the 3-class model by freely estimating only the independent class sizes is displayed in Figure 64.
Figure 64. LMI Self-Rating Confirmatory LCA 3-Class Model Specification (Mplus Version 7)

Confirmatory LCA results. The 3-class LCA model from the September 2012 LMI self-rating data was confirmed with the October 2012 data. The LMR LR test ($p <$
0.001), the ALMR LR test \( (p < 0.001) \), and the BLRT \( (p < 0.001) \) indicated the 3-class model fit significantly better than a 2-class model. Additionally, as reported in Table 68, the average latent class probabilities for the most likely latent class membership are provided for both the September and October models; these probabilities are interpreted as reliability measures for class assignment (Gerber et al., 2009) and are well above Nagin’s (2005) criterion based on an average posterior probability of at least 0.7 for all groups.

Table 68
*LMI Self-Rating LCA Two Sample Comparisons*

<table>
<thead>
<tr>
<th>Latent Class</th>
<th>September 2012</th>
<th>October 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Class Probability For Most Likely Class Membership</td>
<td>Class Size</td>
</tr>
<tr>
<td>Class 1</td>
<td>0.920</td>
<td>39.4%</td>
</tr>
<tr>
<td>Class 2</td>
<td>0.913</td>
<td>45.7%</td>
</tr>
<tr>
<td>Class 3</td>
<td>0.927</td>
<td>15.0%</td>
</tr>
</tbody>
</table>

**LMI subordinate-rating confirmatory model.** The item responses from the October 2012 LMI subordinate-rating data corresponding to the item responses from the September 2012 LMI subordinate-rating exploratory LCA (see Figure 60) were selected as the input data for the confirmatory LCA. In order to ensure sufficient values in each cell of the contingency table, the rating scale was recoded into dichotomous responses (Collins & Lanza, 2010); for example, the item responses “very much like the Leader” and “like the Leader” were recoded as “like the Leader” with a value of 1, while the item responses “neutral,” “unlike the Leader,” and “very much unlike the Leader” were recoded as “unlike the Leader” with a value of 0.
**Confirmatory LCA model specification.** Confirmatory LCA was conducted with the October 2012 LMI subordinate-rating data by fixing all item-response probabilities in the new data to those estimated item-response probabilities from the September 2012 LMI subordinate-rating 5-class LCA model (Finch & Bronk, 2011; Gerber et al., 2009; Muthén & Muthén, 2012b). Under the Mplus MODEL command, the September 2012 item-response probabilities were fixed on each class of the October 2012 model specification (e.g., the Mplus statement “%c#k%” permitted k-class deterministic constraints) by inputting the September 2012 item threshold estimates after the Mplus “item#$1@” statement (Muthén & Muthén, 2012b). For example, the September 2012 threshold estimate for item2 of Class 1 was -8.180; therefore, the Mplus statement to fix the item-response probability for item2 in Class 1 of the October 2012 model was [item2$1@-8.180]. The highly constrained Mplus input file specification for confirming the fit of the 5-class model by freely estimating only the independent class sizes is displayed in Figure 65.
Figure 65. LMI Subordinate-Rating Confirmatory LCA 5-Class Model Specification (Mplus Version 7)

**Confirmatory LCA results.** The 5-class LCA model from the September 2012 LMI subordinate-rating data was confirmed with the October 2012 data. The LMR LR test ($p < 0.001$), the ALMR LR test ($p < 0.001$), and the BLRT ($p < 0.001$) indicated the 5-class model fit significantly better than a 4-class model. Additionally, as reported in Table 69, the average latent class probabilities for the most likely latent class membership are provided for both the September and October models; these probabilities are
interpreted as reliability measures for class assignment (Gerber et al., 2009) and are well
above Nagin’s (2005) criterion based on an average posterior probability of at least 0.7
for all groups.

Table 69
*LMI Subordinate-Rating LCA Two Sample Comparisons*

<table>
<thead>
<tr>
<th>Latent Class</th>
<th>September 2012</th>
<th>October 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average Class Probability</td>
<td>For Most Likely Class Membership</td>
</tr>
<tr>
<td>Class 1</td>
<td>0.956</td>
<td>57.5%</td>
</tr>
<tr>
<td>Class 2</td>
<td>0.959</td>
<td>3.4%</td>
</tr>
<tr>
<td>Class 3</td>
<td>0.881</td>
<td>13.7%</td>
</tr>
<tr>
<td>Class 4</td>
<td>0.833</td>
<td>18.5%</td>
</tr>
<tr>
<td>Class 5</td>
<td>0.897</td>
<td>6.9%</td>
</tr>
</tbody>
</table>
Chapter Four: Discussion

“The statistician cannot evade the responsibility for understanding the processes he applies or recommends.” - Sir Ronald Aylmer Fisher, 1890-1962; (Fisher, 1937, pp. 1-2)

This chapter presents a summary of the study, major findings according to each research question, integration of the three analysis techniques, limitations of the research, and recommendations for further study.

Summary of the Study

This study introduced the theoretical underpinnings regarding the development of two new measures of the Air Force Academy’s “leader of character” definition, the Character Mosaic Virtues (CMV) and the Leadership Mosaic Inventory (LMI), and articulated the need for additional psychometric analyses and substantive interpretation to strengthen the rigor of the assessment of the latent factor structures. The review of the literature explained the theory and application of confirmatory factor, item response theory, and latent class analyses in assessing the latent factor structures of new measures in social science research. In line with DeVellis’ (2003) scale development guidance, where he stated “validation is a cumulative, ongoing process” (p. 159) in which scales of ordered categorical response formats should be analyzed with interval-based methods, this research furthered Rosebush’s (2011, 2012) initial validation studies that culminated in exploratory analyses of the factor structures of the CMV and LMI.
Confirmatory factor analysis (CFA) provided additional evidence to strengthen the validity claims of both the CMV and LMI, especially regarding construct validity. Special CFA methods involving robust weighted least squares (WLS) estimation were implemented to analyze the rating responses to confirm the previous EFA results, to predict factor structure based on *a priori* hypotheses, to provide statistical criteria regarding model fit to real data, to test and compare alternative models to the data, and to determine the dimensionality of measurement (DeVellis, 2003; Kline, 2011).

An alternative to CFA in the analysis of CMV and LMI item-level data incorporated into the study was the nonlinear approach of item response theory (IRT). This sophisticated technique estimated the probability of a response based on the amount of a latent trait. Individual item difficulty parameters were estimated as the amount of the latent trait required for a cadet to give a particular response to an item. The IRT analyses were extended to explore the underlying dimensions of the CMV and LMI through multidimensional IRT (MIRT).

Another technique, latent class analysis (LCA), permitted the inference of classifying mixtures of unobserved cadet subpopulations based on their responses to the CMV and LMI. The analyses uncovered the meaning and number of underlying subpopulations not evident through the other two traditional factor structure analysis techniques. The creation of the study’s latent class models complemented the more traditional dimensional approaches of structure assessment with an understanding of the unobserved cadet subpopulations which could lead to future targeted cadet developmental
opportunities being applied at organizational levels or groups deficient in certain latent traits.

This study leveraged the complementary nature of CFA, IRT, and LCA to strengthen the rigor and sophistication of evaluation of the newly developed CMV and LMI scales. By exposing more researchers, decision-makers, and other stakeholders to these three advanced psychometric evaluation methods, the goal of this study was to benefit the fields of moral development and leadership development, especially at the nation’s service academies.

Major Findings

This section discusses the major findings of the study based on each research question. Interpretations of the results are provided based on the review of the literature.

Research question one. This question asked if analysis of the November 2011 CMV data using CFA techniques would support a dimensional structure consistent with Rosebush’s (2011) EFA results and demonstrate desirable psychometric properties of acceptable model fit, construct reliability, and construct validity. To address this question, a competing a priori (Bandolos & Finney, 2010) eight-factor hypothetical model was analyzed and compared with the nine-factor theoretical model (Rosebush, 2011).

The fit of the CMV data to the eight-factor hypothetical model was assessed as marginally adequate since only two of the five global model-fit criteria for categorical outcomes were met. Additionally, eight items out of forty-five did not have sufficiently
high enough factor loadings to enable their factors to explain the majority of the variability of each measured item (Kline, 2011).

The fit of the CMV data to the nine-factor theoretical model was also assessed as marginally adequate since only two of the five global model-fit criteria for categorical outcomes were met. Similar to the competing model, eight items out of forty-five did not have sufficiently high enough factor loadings to enable their factors to explain the majority of the variability of each measured item (Kline, 2011).

However, when comparing the fit of the two models, the nine-factor theoretical model was a statistically significant improvement over the fit of the eight-factor hypothetical model. This result provided statistical evidence that analysis of the CMV data using CFA techniques supported a dimensional structure consistent with Rosebush’s (2011) EFA results.

*Post hoc* modification of the nine-factor model demonstrated the desirable psychometric properties of acceptable model fit, construct reliability, and construct validity. In accordance with this study’s methodology, the modified model retained the three items from each factor of the theoretical model with the highest standardized factor loadings. The fit of the CMV data to the nine-factor modified model was assessed as adequate since four of the five global model-fit criteria for categorical outcomes were met and only two items out of twenty-seven did not have sufficiently high enough factor loadings to enable their factors to explain the majority of the variability of each measured item (Kline, 2011). This result provided statistical evidence that analysis of the CMV
data using CFA techniques demonstrated the desirable psychometric property of acceptable model fit.

The modified model also demonstrated the desirable psychometric properties of construct reliability and construct validity. In addition to the evidence for construct validity provided by the significant factor loadings on each subscale (Anderson & Gerbing, 1988), construct reliabilities for each of the nine-factors in the modified model well exceeded the cutoff value and construct average variance extracted (AVE) exceeded the variance due to measurement error for each factor; both results provided evidence for adequate convergent validity (Fornell & Larcker, 1981). Discriminant validity was evaluated by comparing factor AVE with the shared variance between each pair of factors (Fornell & Larcker, 1981); in all cases, the factor AVE of the CMV modified model exceeded the shared variance between each pair of factors.

**Research question two.** This question asked if analyses of the September 2012 LMI data using CFA techniques would support a dimensional structure consistent with Rosebush’s (2012) EFA results and demonstrate desirable psychometric properties of acceptable model fit, construct reliability, and construct validity. To address this question, a competing a priori (Bandalos & Finney, 2010) six-factor hypothetical model based on the six USAF institutional sub-competencies was analyzed and compared with the unidimensional theoretical model (Rosebush, 2012) for both the self-rating and subordinate-rating versions.

The fit of the self-rating and subordinate-rating LMI data to the six-factor hypothetical model was found to be inadmissible due to several of the model estimated
correlations being greater than or equal to one between the subscales. According to Muthén (2006), when the estimated correlations between two latent variables are greater than or equal to one, the solutions are not admissible since the respective factors are not statistically distinguishable (msg. 15). Since no other a priori hypothetical multidimensional models were justifiable based on the theoretical underpinnings of Rosebush’s (2012) newly developed LMI scales, further analyses of the LMI data focused on the unidimensional assessments.

The fit of the LMI self-rating data to the unidimensional theoretical model was assessed as inadequate since none of the five global model-fit criteria for categorical outcomes were met. Additionally, twenty-six items out of twenty-nine did not have sufficiently high enough factor loadings to enable their factors to explain the majority of the variability of each measured item (Kline, 2011). Consecutive analyses of the unidimensional model by collapsing the rating scales from 5-points to 2-points, in which slight model improvement was not realized until the analysis of the dichotomous rating scale, demonstrated a possible diagnosis of misfit due to small overall sample size relative to the number of response categories (Byrne, 2012).

The fit of the LMI subordinate-rating data to the unidimensional theoretical model was assessed as adequate since two of the five global model-fit criteria for categorical outcomes well exceeded the cutoff values. Additionally, all twenty-nine items had sufficiently high enough factor loadings to enable their factors to explain the majority of the variability of each measured item (Kline, 2011). This result provided statistical
evidence that analysis of the LMI subordinate-rating data using CFA techniques supported a dimensional structure consistent with Rosebush’s (2012) EFA results.

Post hoc modification of the subordinate-rating unidimensional model demonstrated the desirable psychometric properties of acceptable model fit, construct reliability, and construct validity. In accordance with this study’s methodology, the modified model retained the three items from each of the USAF institutional sub-competencies with the highest standardized factor loadings. The fit of the LMI subordinate-rating data to the unidimensional modified model was assessed as adequate since three of the five global model-fit criteria for categorical outcomes were met and all twenty-seven items had sufficiently high enough factor loadings to enable their factor to explain the majority of the variability of each measured item (Kline, 2011). This result provided statistical evidence that analysis of the LMI subordinate-rating data using CFA techniques demonstrated the desirable psychometric property of acceptable model fit.

The modified subordinate-rating model also demonstrated the desirable psychometric properties of construct reliability and construct validity. In addition to the evidence for construct validity provided by the significant factor loadings on the single subscale (Anderson & Gerbing, 1988), the construct reliability in the modified model well exceeded the cutoff value and construct average variance extracted (AVE) exceeded the variance due to measurement error; both results provided evidence for adequate convergent validity (Fornell & Larcker, 1981).

Research question three. This question asked if analysis of the November 2011 CMV data using IRT techniques would support a dimensional structure consistent with
the CFA results and demonstrate desirable psychometric properties of acceptable model fit, item fit, and reliability. To address this question, the competing \textit{a priori} eight-factor hypothetical model was analyzed and compared with the nine-factor theoretical model (Rosebush, 2011) by means of composite, consecutive, and multidimensional approaches (Allen & Wilson, 2006).

The best fit of the CMV data to the eight-factor hypothetical model based on model fit, reliabilities, and estimated correlations of the three dimensionality approaches was through the consecutive approach. This approach modeled each hypothesized CMV subscale separately as unidimensional constructs which produced independent estimates for a cadet’s \textit{virtue} ability along with standard errors for each dimension (Briggs & Wilson, 2003).

The best fit of the CMV data to the nine-factor theoretical model based on model fit, reliabilities, and estimated correlations of the three dimensionality approaches was through the multidimensional approach. By incorporating the correlations between the dimensions, this approach simultaneously estimated separate cadet \textit{virtue} abilities across each latent dimension (Allen & Wilson, 2006).

On the basis of comparison between the best fitting approaches from the nine-factor theoretical model and the eight-factor hypothetical model, the multidimensional nine-factor theoretical model was a statistically significant improvement over the fit of the consecutive eight-factor hypothetical model. This result provided statistical evidence that analysis of the CMV data using IRT techniques supported a dimensional structure with model fit consistent with the CFA results.
Post hoc modification of the best fitting multidimensional nine-factor model demonstrated the desirable psychometric property of acceptable item fit. In accordance with this study’s methodology, the modified model retained the three items from each dimension of the theoretical model with the lowest infit and outfit mean-square $t$-statistics. As a result, the infit and outfit mean-squares fell within de Ayala’s (2009) cutoff criteria for acceptable fit adequacy. The first item-person map, which provided visual estimates of cadet perceived virtue abilities on each dimension followed by the item difficulties relative to each dimension, revealed two problematic areas: 1) the items were only measuring levels of cadet virtue abilities near and below the latent trait means on each dimension (i.e., indicated the item’s relative ease of positive endorsement), and 2) that while the items were fairly dispersed from near the means of the latent trait and below, there was too much item overlap (i.e., item difficulty redundancy) amongst several of the items. The second item-person map revealed that cadets found it easy to endorse all of the items relative to each dimension’s positive mean—endorsability ranged from the care for others dimension being very easy to the self-control dimension being somewhat easy.

The modified model also demonstrated evidence of acceptable reliability. Reliability was measured as model estimated explained variance divided by total person variance (Wu et al., 2007). The reliability of all dimensions of the modified model exceeded the 0.7 cut-off point for adequate reliability (DeVellis, 2003).

Research question four. This question asked if analyses of the September 2012 LMI data using IRT techniques would support a dimensional structure consistent with the
CFA results and demonstrate desirable psychometric properties of acceptable model fit, item fit, and reliability. To address this question, the competing a priori six-factor hypothetical model based on the six USAF institutional sub-competencies was analyzed and compared with the unidimensional theoretical model (Rosebush, 2012) for both the self-rating and subordinate-rating versions by means of composite, consecutive, and multidimensional approaches (Allen & Wilson, 2006).

The best fit of the LMI self-rating data based on model fit, reliabilities, and estimated correlations amongst the three dimensionality approaches was through the multidimensional approach. By incorporating the correlations between the dimensions, this approach simultaneously estimated separate cadet element leader effectiveness abilities across each latent dimension (Allen & Wilson, 2006). This result provided statistical evidence that analysis of the LMI self-rating data using IRT techniques did not support a dimensional structure consistent with the best fit unidimensional CFA results.

Post hoc modification of the best fitting multidimensional six-factor model demonstrated the desirable psychometric property of acceptable item fit. In accordance with this study’s methodology, the modified model retained the three items from each dimension of the hypothetical model with the lowest infit and outfit mean-square $t$-statistics. As a result, the infit and outfit mean-squares fell within de Ayala’s (2009) cutoff criteria for acceptable fit adequacy. The first item-person map, which provided visual estimates of cadet element leader perceived effectiveness abilities on each dimension followed by the item difficulties relative to each dimension, revealed two problematic areas: 1) the items were only measuring levels of perceived cadet element
leader *effectiveness* abilities much below the latent trait means on each dimension (i.e., indicated the item’s relative ease of positive endorsement), and 2) that while the items were fairly dispersed from much below the means of the latent traits, there was too much item overlap (i.e., item difficulty redundancy) amongst several of the items. The second item-person map revealed that cadet element leaders found it easy to endorse all of the items relative to each dimension’s positive mean—endorsability ranged from the *builds teams and coalitions* dimension being very easy to the *negotiating* dimension being somewhat easy.

The modified model also demonstrated evidence of acceptable reliability. Reliability was measured as model estimated explained variance divided by total person variance (Wu et al., 2007). The reliability of all dimensions of the modified model exceeded the 0.7 cut-off point for adequate reliability (DeVellis, 2003).

The best fit of the LMI *subordinate-rating* data based on model fit, reliabilities, and estimated correlations amongst the three dimensionality approaches was also through the multidimensional approach. By incorporating the correlations between the dimensions, this approach simultaneously estimated separate subordinate-rated cadet element leader *effectiveness* abilities across each latent dimension (Allen & Wilson, 2006). This result provided statistical evidence that analysis of the LMI *subordinate-rating* data using IRT techniques did not support a dimensional structure consistent with the best fit unidimensional CFA results.

*Post hoc* modification of the best fitting multidimensional six-factor model demonstrated the desirable psychometric property of acceptable item fit. In accordance
with this study’s methodology, the modified model retained the three items from each dimension of the hypothetical model with the lowest infit and outfit mean-square $t$-statistics. As a result, the infit and outfit mean-squares fell within de Ayala’s (2009) cutoff criteria for acceptable fit adequacy. The first item-person map, which provided visual estimates of subordinate-rated cadet element leader perceived *effectiveness* abilities on each dimension followed by the item difficulties relative to each dimension, revealed two problematic areas: 1) the items were only measuring levels of subordinate-rated cadet element leader *effectiveness* abilities much below the latent trait means on each dimension (i.e., indicated the item’s relative ease of positive endorsement), and 2) that while the items were fairly dispersed from much below the means of the latent traits, there was too much item overlap (i.e., item difficulty redundancy) amongst several of the items. The second item-person map revealed that cadet subordinates found it easy to endorse all of the items relative to each dimension’s positive mean—endorsability ranged from the *takes care of people* dimension being very easy to the *negotiating* dimension being somewhat easy.

The modified model also demonstrated evidence of acceptable reliability. Reliability was measured as model estimated explained variance divided by total person variance (Wu et al., 2007). The reliability of all dimensions of the modified model exceeded the 0.7 cut-off point for adequate reliability (DeVellis, 2003).

**Research question five.** This question asked if analysis of the November 2011 CMV data using LCA techniques yielded a cutoff point for classifying cadet subpopulations as either having a virtuous character or not, and what combinations of
Character Mosaic virtue endorsements distinguish cadets who have a virtuous character versus those who do not. This question was addressed by following Wang and Wang’s (2012) three-step modeling approach to 1) determine the optimal number of latent classes, 2) evaluate the quality of the classification of latent class membership, and 3) define the latent classes.

The item responses from the CMV nine-factor post hoc modified model from research question one were recoded into dichotomous responses to aid in estimation and interpretation (Collins & Lanza, 2010). A 3-class optimal solution was determined by analyzing the fit of a series of increasing class number models by comparing the $k$-class model with the $(k-1)$-class model (Wang & Wang, 2012). The quality of the classification was determined to be acceptable since the class sizes and sample proportions were adequate, the average class probabilities for most likely class membership well exceeded Nagin’s (2005) criterion, and the entropy statistic was high indicating low classification error (Clark, 2010; Collins & Lanza, 2010). The three latent classes—strong identification with virtues, moderate identification with virtues, and weak identification with virtues—were defined based on the observed pattern of item-response probabilities.

While the results of the LCA did not yield a single cutoff point for classifying cadet subpopulations as either having a virtuous character or not, it provided statistical evidence of three distinct unobserved groups underlying the data. The 37.6% of cadets assigned to the strong identification with virtues class (i.e., Class 2) had a high probability of identifying with all of the constructs of the CMV; while this class had the
second highest probabilities relative to the other classes of endorsing the duty and attention to detail constructs and the last item of the excellence construct, the probabilities for doing so were still high ranging from 0.76 to 0.92. Therefore, cadets assigned to Class 2 identified strongly with the following character virtues—courage, accountability, humility, care for others, self-control, respect for human dignity, and excellence. Regarding the 40.0% of cadets assigned to the moderate identification with virtues class (i.e., Class 1), the probabilities of endorsing “like me” fell between the probabilities of the other two classes in five of the nine constructs of the CMV; however, this class had the lowest probabilities of endorsing the duty, self-control, attention to detail, and excellence constructs and the second item of the courage construct. Therefore, cadets assigned to Class 1 identified moderately with the following character virtues—courage, accountability, humility, care for others, and respect for human dignity. Of the remaining 22.4% of cadets assigned to the weak identification with virtues class (i.e., Class 3), the probabilities of endorsing “like me” were less than the probabilities of the other two classes in five of the nine constructs of the CMV; however, this class had the highest probabilities of endorsing duty, attention to detail, and the last item of the excellence construct while the probabilities of endorsing “like me” on the self-control construct, the second item of courage, and other excellence items fell between the other two classes. Therefore, cadets assigned to Class 3 identified weakly with the following character virtues—courage, accountability, humility, care for others, and respect for human dignity.
**Research question six.** This question asked if analysis of the LMI data using LCA techniques yielded either a cutoff point for classifying cadet element leader subpopulations as either being an effective leader or not, or a cutoff point for classifying cadet subordinate subpopulations who view their element leaders as being effective or not. Additionally, this question sought combinations of the USAF institutional sub-competency endorsements in the LMI data that would distinguish cadet element leaders who are effective leaders versus those who are not. This question was first addressed by following Wang and Wang’s (2012) three-step modeling approach with the September 2012 LMI self-rating and subordinate-rating data in an exploratory approach to 1) determine the optimal number of latent classes, 2) evaluate the quality of the classification of latent class membership, and 3) define the latent classes. The research question was extended by conducting latent class analyses on the October 2012 LMI self-rating and subordinate-rating versions in a confirmatory manner to validate the appropriateness and fit of the exploratory models (Gerber et al., 2009).

**Self-rating LCA.** The item responses from the September 2012 LMI self-rating six-factor post hoc modified model from research question four were recoded into dichotomous responses to aid in estimation and interpretation (Collins & Lanza, 2010). A 3-class optimal solution was determined by analyzing the fit of a series of increasing class number models by comparing the $k$-class model with the $(k-1)$-class model (Wang & Wang, 2012). The quality of the classification was determined to be acceptable since the class sizes and sample proportions were adequate, the average class probabilities for most likely class membership well exceeded Nagin’s (2005) criterion, and the entropy
statistic was high indicating low classification error (Clark, 2010; Collins & Lanza, 2010). The three latent classes—highly effective self-rated leaders, moderately effective self-rated leaders, and somewhat effective self-rated leaders—were defined based on the observed pattern of item-response probabilities.

While the results of the self-rating LCA did not yield a single cutoff point for classifying cadet element leader subpopulations as either being effective or not, it did provide statistical evidence of three distinct unobserved groups underlying the data. The 39.4% of cadet element leaders assigned to the highly effective self-rated leaders’ class had the highest probability of identifying with all of the constructs of the LMI. Regarding the 45.7% of cadet element leaders assigned to the moderately effective self-rated leaders’ class, they had the second highest item-response probabilities for each of the six constructs. Of the remaining 15.0% of cadets assigned to the somewhat effective self-rated leaders’ class, the probabilities of endorsing “like me” were less than the probabilities of the other two classes in all six constructs of the LMI.

The item responses from the October 2012 LMI self-rating data corresponding to the item responses from the September 2012 LMI self-rating exploratory LCA were recoded into dichotomous responses to aid in estimation and interpretation for the confirmatory LCA (Collins & Lanza, 2010). By fixing all item-response probabilities in the October 2012 self-rating specification to those estimated item-response probabilities from the September 2012 self-rating LCA model, the results provided evidence that a 3-class representation of self-rated cadet element leader effectiveness was an appropriate and a reliable fit to the data (Gerber et al., 2009).
**Subordinate-rating LCA.** The item responses from the September 2012 LMI subordinate-rating six-factor post hoc modified model from research question four were recoded into dichotomous responses to aid in estimation and interpretation (Collins & Lanza, 2010). A 5-class optimal solution was determined by analyzing the fit of a series of increasing class number models by comparing the \( k \)-class model with the \((k-1)\)-class model (Wang & Wang, 2012). The quality of the classification was determined to be acceptable since the class sizes and sample proportions were adequate, the average class probabilities for most likely class membership well exceeded Nagin’s (2005) criterion, and the entropy statistic was high indicating low classification error (Clark, 2010; Collins & Lanza, 2010). The five latent classes—see leaders extremely effective, see leaders highly effective, see leaders somewhat effective, see leaders somewhat ineffective, and see leaders highly ineffective—were defined based on the observed pattern of item-response probabilities.

While the results of the subordinate-rating LCA did not yield a single cutoff point for classifying cadet subordinate subpopulations as either seeing their element leaders as being effective or not, it did provide statistical evidence of five distinct unobserved groups underlying the data. The 57.5% of cadet subordinates assigned to the see leaders extremely effective class had the highest probability of identifying with all of the constructs of the LMI, while the next 18.5% assigned to the see leaders highly effective class had the second highest probability. Regarding the 13.7% of cadet subordinates assigned to the see leaders somewhat effective class and the 6.9% assigned to the see leaders somewhat ineffective class, they had the third and fourth highest item-response probabilities.
probabilities for each of the six constructs, respectively. Of the remaining 3.4% of cadet subordinates assigned to the *see leaders highly ineffective* class, the probabilities of endorsing “like me” were less than the probabilities of the other four classes in all six constructs of the LMI.

The item responses from the October 2012 LMI *subordinate-rating* data corresponding to the item responses from the September 2012 LMI *subordinate-rating* exploratory LCA were recoded into dichotomous responses to aid in estimation and interpretation for the confirmatory LCA (Collins & Lanza, 2010). By fixing all item-response probabilities in the October 2012 *subordinate-rating* specification to those estimated item-response probabilities from the September 2012 *subordinate-rating* LCA model, the results provided evidence that a 5-class representation of cadet subordinates based on their element leader effectiveness views was an appropriate and a reliable fit to the data (Gerber et al., 2009).

**Integration of the Analysis Techniques**

A strength of this study was the ability to leverage the complementary nature of CFA, IRT, and LCA analyses to underpin the rigor and sophistication of evaluation of the CMV and the LMI. While Rosebush’s (2011, 2012) psychometric assessments provided initial scale development validations, the present study extended the cumulative validation process through interval-based methods (DeVellis, 2003). Moreover, this study blended the orientation of the statistical analyses by first examining CFA’s variable-oriented approach (Collins & Lanza, 2010), followed by IRT’s variable- and
person-oriented approaches (Onwuegbuzie & Combs, 2010), and concluding with LCA’s person-oriented approach (Collins & Lanza, 2010).

With its emphasis on the identification and accounting of the linear relationships between observed variables applied across persons (Collins & Lanza, 2010), CFA’s variable-oriented approach supported Rosebush’s (2011, 2012) results and provided evidence for the factorial validity of the multidimensional CMV and the unidimensional LMI subordinate-rating version. While the approach did not support the latent factor structure of the LMI self-rating version, it did provide valuable insight to inform the domain regarding the sources of misfit (Bandalos & Finney, 2010). The CFA analyses were found to be most informative at the subscale or construct level (Osteen, 2010); for example, rather than individual item fit, the standardized factor loadings provided empirical evidence for the variance accounted for by each construct relative to the variance due to measurement error. The informative nature of the subscale level was also evident in the post hoc modification process in which only those items with the highest standardized factor loadings from each subscale were retained—this subscale level process eliminated redundant items, improved overall model fit, maintained congeneric measurement, and provided a set of reduced items for the CMV and LMI subordinate-rating version which may be used more effectively in future testing.

With its capacity to estimate both person ability and item difficulty, IRT’s variable- and person-oriented approaches provided a second confirmation of the dimensional structure of the CMV and uncovered new evidence for a multidimensional LMI latent factor structure. By considering the possibility of a nonlinear nature of the
underlying data, the IRT analyses were able to overcome the CFA limitation regarding the LMI six-factor hypothetical model’s estimated correlations being greater than or equal to one between the subscales which resulted in inadmissible CFA solutions (Muthén & Muthén, 2012a). In contrast with CFA, the IRT analyses were found to be most informative at the item level, specifically assessing individual item performance (Osteen, 2010), after testing for the best model fit amongst the three dimensionality approaches. For example, rather than eliminating items based on their proportion of variance explained by the latent factor, post hoc modifications were made based on individual item fit statistics that accounted for differences between observed and model expected responses (de Ayala, 2009), resulting in six retained item differences between the CFA and IRT assessments on the CMV. The informative nature of the item level was also evident in the item-person maps of the post hoc modified models—while the models demonstrated statistical evidence that multidimensional structures best fit the data, the mapping of individual item difficulty parameters against latent cadet abilities revealed that only the lower levels of cadet abilities were being measured with the current items and that item difficulty redundancies still existed after the modification process.

With its focus on grouping unobserved subpopulations of individuals based on their patterns of responses (Collins & Lanza, 2010), LCA’s person-oriented approach organized the latent structures of the CMV and LMI data based on the probability of endorsing “like me” (or “like the Leader” in the case of the LMI subordinate-rating version). With an emphasis on the individual, the results of the LCA provided a different perspective from CFA and IRT regarding how to explain the dissimilarities in
the observed item responses and how to summarize the structures in a meaningful way (Geiser, 2013). After substantively interpreting the classifications of the CMV and LMI, evidence for a 3-point rating scale for the CMV and LMI *self-rating* version and for a 5-point rating scale for the LMI *subordinate-rating* version could be inferred from the results to complement future CFA or IRT analyses.

**Limitations of the Research**

The limitations identified for this study included non-probability sampling, questionnaire administration, missing data, CFA and IRT *post hoc* modifications, and LCA *a priori* data transformations.

With the sampling strategy outside the purview of the researcher, the generalizability of the results was a substantive concern since randomization was not in place with the CMV and LMI convenience samples; improvements in sampling strategies would achieve more robust analyses. In addition to the exclusion of random sampling, the absence of demographic data collection on each sample made the verification of the representativeness of the samples to the larger cadet population impossible. One exception to this limitation was the September 2012 LMI *self-rating* data, in which 357 out of 360 cadet element leaders in the population participated in the research.

Test administration and time between repeated administrations of the LMI possibly limited the findings. While the administration of the CMV was standardized in a classroom setting with proctors, the administration of the LMI questionnaire was delivered over email and potentially introduced bias regarding self-selection. Since it was conceivable that only highly opinionated cadets responded to the LMI *subordinate-*
rating questionnaire and other controls were not in place, generalizability across all cadet subordinates is not certain. For the CMV and the LMI self-rating version, possible inflated responses limited generalizability since these cadets knew beforehand that their scores would be later reviewed with them by their mentor. Another threat to internal validity was possible testing effects due to the one-month elapsed time between the September 2012 and October 2012 administrations of the LMI.

The three analyses in this study were conducted on complete case data sets provided by the Center for Character and Leadership Development. Generalizability of the results is affected since knowledge of how any missing data were treated was not provided. According to Tabachnick and Fidell (2007), distortions of the sample data occur when cases with missing values not randomly distributed are deleted.

The study’s design of limiting the CFA and IRT post hoc modifications to three items per construct was a tradeoff. In one regard, having a minimum of three items per construct helped to ensure that the underlying trait was measured; conversely, model fit was affected when CFA items were retained whose proportion of the variance not explained exceeded the variance explained and when IRT items were retained with fit statistics approaching the cutoff values. In the case of the latter, the factorial validity generalization of these particular subscales is limited.

Recoding the LCA input data into dichotomous variables contributed to the simplicity and interpretation of the exploratory analyses. However, the analyses could have been extended to the polytomous rating scales, especially ones conducted in a confirmatory manner (Geiser, 2013).
Recommendations for Further Study

The next phase of CMV and LMI psychometric assessments in alignment with DeVellis’ (2003) “validation is a cumulative, ongoing process” (p. 159) should involve the integration of CFA, IRT, and LCA analyses to test for measurement invariance (i.e., measurement equivalence). According to Kline (2011), measurement invariance occurs when “scores from the operationalization of a construct have the same meaning under different conditions” (p. 251) such as consistency between groups, stability over time, and constancy between modes of test administration. DeVellis (2003) emphasizes this critical property of measurement by stating, “in order to compare two groups directly, one must assume that measures perform identically on both groups and that any differences observed are due only to the attribute of interest” (p. 151). Cadet demographics such as gender, race/ethnicity, religious preference, intercollegiate status, class year, geographical hometown region, first generation college student status, and language other than English at home status should be collected to provide grouping variables for invariance testing.

While techniques are similar for the testing of invariance over time or over methods of test administration, CFA techniques for testing of invariance over groups is especially useful in providing evidence to refute construct bias—the notion that an instrument measures something different for one group than for another (Kline, 2011). The testing strategy for measurement invariance over groups involves the logical ordering of sets of parameters to be tested in an increasingly constrained fashion (Byrne, 2012). For example, the least restrictive configural invariance hypothesis is retained
when a measurement model is adequately fit across groups in which the number of factors and their corresponding loadings are equivalent while all parameters are freely estimated—here the conclusion could be made that the same factor structure is manifested in each group although through different factor loadings (Kline, 2011). Stricter forms of measurement invariance can be similarly tested by logically applying more constraints in the following order: 1) construct-level metric invariance is tested by constraining the factor loadings to be equal across groups; and 2) equivalence of construct variances and covariances is tested by adding the additional constraints of equal factor variances and covariances (Kline, 2011). According to Byrne (2012), an even stricter traditional form of invariance testing regarding the equivalence of residual variances and covariances is now “widely accepted that this test for equivalence not only is of least interest and importance but also may be considered somewhat unreasonable and indeed not recommended” (p. 195). Testing for invariance over groups using CFA techniques will provide additional evidence for construct validity and enable researchers to refute any claims of construct bias.

Although the focus of invariance testing using CFA techniques is at the construct level, IRT techniques provide a complementary approach of distinguishing differences in group membership at the item level. CMV and LMI data should be analyzed for differential item functioning (DIF), which is concerned with items performing differently across groups that should otherwise be equivalent on the latent trait being measured (DeVellis, 2003). According to Bond and Fox (2007), “when an item’s difficulty estimate location varies across samples by more than the modeled error, then prima facie
evidence of DIF exists” (p. 92). The ability to diagnose and eliminate DIF in the CMV and LMI using IRT techniques will support the CCLD in assessing item bias and provide evidence for culturally competent measures (Osteen, 2010) of cadet character virtues and leadership effectiveness.

Establishing measurement invariance of the CMV and LMI using LCA techniques complements the CFA and IRT approaches with the favorable property of maintaining the consistency of interpretation of the latent classes. Under a latent class framework, Collins and Lanza (2010) explain that an instrument establishes measurement invariance across populations “when individuals who belong to the same latent class, but who are from different populations, have the same probability of providing any given observed response pattern” (p. 118). To test whether these conditional item-response probabilities are invariant across groups, the fit of a model with no constraints is compared to the fit of a model with across group equality constraints for the item-response probabilities (Collins & Lanza, 2010). If the constrained model fits the data as well as the unconstrained model, it can be concluded that measurement invariance holds across the groups. Since the conditional item-response probabilities will be considered statistically identical across groups under a measurement invariance hypothesis, the unconditional latent class probabilities across groups may be directly compared as the interpretation of the latent classes remain unchanged from the population as a whole (Collins & Lanza, 2010).
References


Amos (Version 18) [Computer software]. Crawfordville, FL: Amos Development Corporation.


PASW Statistics GradPack (Version 18) [Computer software]. Chicago, IL: SPSS Inc.


Appendix A

Character Mosaic Virtues’ (CMV) Rating Scale, Dimensions, and Items (Rosebush, 2011, pp. 42-43)

__ Very much unlike me
__ Unlike me
__ Neutral
__ Like me
__ Very much like me

Courage

1. I always stand up for my beliefs.
2. I must stand up for what I believe, even if there are negative results.
3. I never hesitate to publicly express an unpopular opinion.
4. I have taken frequent stands in the face of strong opposition.
5. I always speak up in protest when I hear someone say mean things.
6. I call for action while others talk.

Accountability

7. I always admit when I am wrong.
8. I admit mistakes when they are made.
9. I always initiate confessing my mistakes.
10. I hold myself accountable for whatever mistakes I have made.

Humility

11. I never brag about my accomplishments.
12. People are drawn to me because I am humble.
13. No one would ever describe me as arrogant.
14. I am always humble about the good things that have happened to me.
15. I have been told that modesty is one of my most notable characteristics.

16. I do not act as if I am a special person.

**Duty**

17. I follow through with my plans.

18. I make plans and stick to them.

19. When I make plans, I am certain to make them work.

20. I carry out my plans.

**Care for Others**

21. I love to make other people happy.

22. I enjoy being kind to others.

23. I go out of my way to cheer up people who appear down.

24. I really enjoy doing small favors for friends.

25. It is important to me to maintain harmony within my group.

26. I am as excited about the good fortune of others as I am about my own.

**Self-Control**

27. When I am tempted to do something pleasurable that I know is wrong, I always resist the temptation.

28. I always exercise self-control over inappropriate desires.

29. I always turn away from temptations that are harmful to me.

30. I easily resist temptations.

31. I have an exceptionally high level of self-control over attractive (but harmful) impulses.
Respect for Human Dignity

32. I can accept a lot of different perspectives from others.

33. I understand people who think differently than me.

34. I have a high tolerance of those whose views differ from mine.

35. People who have ideas that are different than mine annoy me. [reverse]

36. I accept people as they are.

37. I believe it is worth listening to everyone’s opinion.

38. I can always see the world from someone else’s perspective.

Attention to Detail

39. I have an eye for detail.

40. I pay attention to details.

41. I pay too little attention to details. [reverse]

Excellence

42. I want everything to add up perfectly.

43. I dislike imperfect work.

44. I am exacting in my work.

45. I demand quality.
Appendix B

Leadership Mosaic Inventory (LMI) Rating Scale, Items, and {Corresponding USAF Institutional Sub-Competencies} 
(Rosebush, 2012, pp. 14-15)

Note: The first phrase inside the brackets [] refers to the Element Leader Self-Rating LMI version while the second phrase inside the brackets [] refers to the Element Leader Subordinate Rating LMI version. The items are exactly the same for both LMI versions.

Please describe how much each of the following statements is like how you see [yourself as the leader of the unit; the leader of your unit].

- **Very much unlike [me; the Leader]**
- **Unlike [me; the Leader]**
- **Neutral**
- **Like [me; the Leader]**
- **Very much like [me; the Leader]**

1. Inspires the unit with plans for the future. {Develops and Inspires Others}

2. Ensures excellent quality of work done by the leader and by the unit members.

{Develops and Inspires Others}

3. Suggests ways to improve the unit's performance. {Develops and Inspires Others}

4. Creates enthusiasm for a task to be completed by the unit. {Develops and Inspires Others}

5. Makes unit members feel they share responsibility for the unit's development.

{Develops and Inspires Others}

6. Tells the unit when they perform well. {Develops and Inspires Others}

7. Contributes to the unit members’ satisfaction of their role in the unit. {Takes Care of People}
8. Behaves in a manner that is thoughtful of unit members’ personal needs. {Takes Care of People}

9. Sacrifices the leader’s own interests to meet the unit’s needs. {Takes Care of People}

10. Encourages unit members to have a team spirit in the unit. {Builds Teams and Coalitions}

11. Supports the unit’s efforts. {Builds Teams and Coalitions}

12. Helps develop good relations among the members of the unit. {Builds Teams and Coalitions}

13. Believes that the unit needs to function as a team. {Builds Teams and Coalitions}

14. Encourages members of the unit to solve problems together. {Builds Teams and Coalitions}

15. Unit members have a clear and effective two-way flow of communication in the unit.
   {Builds Teams and Coalitions}

16. Is good at convincing unit members to do things for the unit. {Negotiating}

17. Offers compelling reasons to get the unit members to do things for the unit.
   {Negotiating}

18. Is very persuasive. {Negotiating}

19. Has a clear understanding of where the unit is going. {Vision}

20. Helps the unit focus on its goals. {Vision}

21. Pays attention to the efforts of the members of the unit. {Vision}

22. Is able to get unit members committed to the leader’s vision of the unit’s future.
   {Vision}
23. Seems alert to what's happening in the unit. {Vision}

24. Is always seeking new opportunities for the unit. {Vision}

25. Adapts easily to new situations involving the unit. {Adaptability}

26. Adapting to quickly-changing situations is a strength of the leader. {Adaptability}

27. Members of the unit would agree that the leader adapts well during stressful situations. {Adaptability}

28. Flexibly adapts to sudden changes. {Adaptability}

29. Even when the unit is experiencing unfamiliar situations, unit members are confident that the leader will successfully adapt. {Adaptability}
Appendix C

MEMORANDUM FOR: USAFA IRB (Ms. Giul Rosado)

FROM: Dr. Michael A. Rosebush, USAFA/CWCS

SUBJECT: Request for Research Exemption

1. Administrative Information
   Title of protocol: Advanced Psychometric Analysis, “Character MOSAIC” Archival and “Lifts Others, Elevates Performance” Archival

   Principal investigator: Dr. Michael Rosebush
   Chief, Character and Leadership Coaching
   719-333-1387
   michael.rosebush@usafa.edu

   Associate investigator(s): David Hignbotham, Lt Col, USAF
   AFT/CI PhD Student, University of Denver
   719-219-9803
   davidhignbotham@gmail.com

   Organization: Center for Character and Leadership Development
   719-333-4904

   Category for exemption: 32 CFR 219.101(b) allows for research in the following categories to be exempt.

   (4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

2. Summary of Research
   Request the use of the following de-identified archival datasets for advanced psychometric analysis to fulfill the dissertation requirements for the Doctor of Philosophy degree in Research Methods and Statistics as sponsored by the Department of Mathematical Sciences in concert with the Center for Character and Leadership Development:

   - Character Mosaic Survey, C2015, summer 2011, pre-test (n=626), administered as research protocol FAC201100968H. Only the respondent’s cadet squadron number and raw score
• responses to the researched items (i.e., no personal identifiers) will be provided by Dr. Rosebush to Lt Col Higginbotham.

• Character Monte Carlo Survey, C2015, fall 2011, post-test (n=233), administered as research protocol FAC20110086R. Only the respondent’s cadet squadron number and raw score responses to the researched items (i.e., no personal identifiers) will be provided by Dr. Rosebush to Lt Col Higginbotham.

• Lifts Others, Elevates Performance, C2014.C2015 (n=371), administered as research protocol FAC201420001SE. Only the respondent’s cadet squadron number and raw score responses to the researched items (i.e., no personal identifiers) will be provided by Dr. Rosebush to Lt Col Higginbotham.

The purpose of this dissertation research is to strengthen the rigor and sophistication of the psychometric assessment previously conducted by USAFA/CWC on the data sets above to include comparisons of each measure’s dimensionality using confirmatory factor analysis (CFA) through structural equation modeling (SEM) with multidimensional item response theory (MIRT) analyses. Confirming the factor structure of both measures with the most advanced psychometric analysis techniques to date will provide USAFA/CWC and its stakeholders with rigorous evidence for the validity of their character and leadership development scales. Furthermore, MIRT analyses will generate sample-independent item characteristic estimations with the ability to assess potential bias in survey items in order to ensure that culturally competent measurement is in place. Each of the data sets, whose variables are the respondent’s cadet squadron number and responses to each of the respective survey items, will be analyzed and compared independently using the CFA and MIRT techniques and results. The data sets will not be merged; they will remain independent in order to represent the sample from which they were drawn.

Additionally, the proposed research will include latent class analysis (LCA) on the datasets in an effort to uncover statistical subtypes of related latent classes based on categorical data (e.g., cadet squadron identities) to find examinee subpopulations or types of attitude structures based on responses to items in both character and leadership development survey instruments. The results of LCA will provide USAFA/CWC with cadet subpopulations to target for specialized or increased developmental experiences. Each of the data sets, whose variables are the respondent’s cadet squadron number and responses to each of the respective survey items, will be analyzed independently using LCA techniques. The data sets will not be merged; they will remain independent in order to represent the sample from which they were drawn.

This research is planned to commence as soon as possible with a completion goal of June 2013.

3. Attachments: Please note that your protocol will not be reviewed until all relevant documents have been received by the IRB office.

• Supporting documents:
  • Lt Col Higginbotham resume
  • Lt Col Higginbotham verification of researcher training: Training Certificate from NIH
This protocol emphasizes good experimental design, minimizing the use of and risks to human subjects. I agree that the research is exempt.

Melinda Stevison, PhD
Assistant Director, Scholarship
Center for Character and Leadership Development

The proposed principal investigator is scientifically qualified to conduct the proposed study or is being supervised by a qualified scientist. I will ensure that the research records and data for this protocol will be maintained for a minimum of 3 years. I agree that the research is exempt under 32 CFR 219.101(b)(4).

Joseph E. Sanders III, Col, USAF
Permanent Professor and Director
Center for Character & Leadership Development
MEMORANDUM FOR DR. MIKE ROSEBUSH  
FROM: HQ USAFA/A8N  
11 September 2012

SUBJECT: Protocol FAC20120053E Exempt Status

1. The HQ USAFA Institutional Review Board considered your request for exempt status for FAC20120053E - Advanced Psychometric Analysis, "Character Mosaic" Archival and "Lifts Others, Elevates Performance" Archival. Your request and any required changes were deemed exempt from IRB oversight in accordance with 32 CFR 219.101, paragraph (b)(4) by an IRB member. The IRB agreed that sufficient safeguards were in place to protect research participants. Please note that the USAFA Authorized Institutional Official, HQ USAFA/CA, and the Surgeon General’s Research Oversight & Compliance Division, AFMSA/SGE-C review all USAFA IRB actions and may amend this decision or identify additional requirements. The USAFA’s DoD Assurance Number is 50946, expiration date 17 August 2014. Our Federally Assured number is FWA00019017, expiration date 20 June 2017.

2. If you are conducting a survey for this study you cannot start this study until you have approval from the Survey Program Manager. The protocol will be considered closed, but will be retained in USAFA/A8N for 3 years then sent to permanent storage. As the principal investigator on the study, the Surgeon General’s Research Oversight & Compliance Division requires that you retain your data, reports, etc. for 3 years following completion of the study.

3. If any aspect of your research protocol changes, you must notify the IRB Chair or IRB Administrator immediately. We will advise you on whether additional IRB review is required.

4. Please use tracking number FAC20120053E in any correspondence regarding this protocol. If you have any questions or if I can be of further assistance, please don’t hesitate to contact me at 314-293-3 or the IRB Chair, Col. Paul Prog at 333-3680.

GAIL B. ROSADO  
HQ USAFA IRB Administrator

Developing Leaders of Character
MEMORANDUM FOR LT COL DAVID HIGGINBOTTOM

FROM: USAFA SVC

SUBJECT: 10 Sep & 9 Oct 2012 De-Identified “Lift Others, Elevate Performance” Archival Data

1. Your request for the use of the 10 Sep 2012 and 9 Oct 2012 “Lift Others, Elevate Performance” De-Identified Archival Data for your dissertation research at the University of Denver is approved since sufficient safeguards are in place to protect research participants.

2. Since the data from the 10 Sep 2012 and 9 Oct 2012 administrations were collected as part of academic course requirements and not in conjunction with any institutional research protocol, the Ethical Review Board (ERB) at USAFA SVC, only the person who collects the data is required to have a waiver from the ERB. Only the person who collects the data is required to have a waiver from the ERB.

3. It is understood that the determination of any entity under 32 CFR 219, paragraphs 69-71 for your dissertation research using this data will be determined by the University of Denver ERB. If you have any questions or if you have further questions, please direct them to contact me at 719-555-1389 or Col Joseph Sanders at 719-333-4934.

MICHAEL A. BRUNOIS, PhD
Chief, Character & Leadership Coaching

JOSPEH E. SANDERS III, Colonel, USAF
Ferrante Professor & Director
Center for Character & Leadership Development
Appendix F

University of Denver

Emily Cullen, MA
Manager, Regulatory Research Compliance
Tel: 303-871-4652

Certification of Human Subjects Approval

December 15, 2012

To,
David Higginbotham, MBAMOS

Subject Human Subject Review

TITLE: An Assessment of Character and Leadership Development Latent Factor Structures through Confirmatory Factor, Item Response Theory, and Latent Class Analysis (Archival Data Request 1 of 2)

IRB#: 2012-2418

Dear Higginbotham,

The Institutional Review Board for the Protection of Human Subjects has reviewed the above named project. The project has been confirmed exempt under 45 CFR Section 46 101 for the procedures and subjects described in the protocol effective 12/11/2012.

This approval is effective for a five-year period.

For the duration of your research study, any changes in:
1. experimental design
2. risk level
3. content of the study
4. materials attached to the original application
5. principal investigator
   must be reviewed and approved by the University of Denver IRB before implementation of those changes.

The University of Denver will terminate this project at the end of the five-year period unless otherwise instructed via correspondence with the Principal Investigator. Please submit a completion report if the study is completed before the expiration date or if you are no longer affiliated with the University of Denver. You must submit a new application at the end of the five-year period if you wish to continue this study.

NOTE: Please add the following information to any consent forms, surveys, questionnaires, invitation letters, etc. you will use in your research as follows: This survey (consent, study, etc.) was approved by the University of Denver’s Institutional Review Board for the Protection of Human Subjects in Research on 12/11/2012. This information will be added by the Research Compliance Office if it does not already appear in the form(s) upon approval and continuation.

The Institutional Review Board appreciates your cooperation in protecting subjects and ensuring that each subject gives a meaningful consent to participate in research projects. If you have any questions regarding your obligations under the Assurance, please do not hesitate to contact us.

Sincerely yours,

[Signature]

266
University of Denver

Emily Calder, MA
Manager, Regulatory Research Compliance
Tel: 303-871-4652

Certification of Human Subjects Approval

December 15, 2012
To,
David Higginbotham, M.M.A.M.O.S

Subject: Human Subject Review

TITLE: An Assessment of Character and Leadership Development Latent Factor Structures through Confirmatory Factor, Item Response Theory, and Latent Class Analysis (Archival Data Request 2 of 2)

IRB#: 2012-2419

Dear Higginbotham,

The Institutional Review Board for the Protection of Human Subjects has reviewed the above named project. The project has been confirmed exempt under 45 CFR Section 46.101 for the procedures and subjects described in the protocol effective 12/11/2012.

This approval is effective for a five-year period.

For the duration of your research study, any changes in:
1. experimental design
2. risk level
3. content of the study
4. materials attached to the original application
5. principal investigator
must be reviewed and approved by the University of Denver IRB before implementation of those changes.

The University of Denver will terminate this project at the end of the five-year period unless otherwise instructed via correspondence with the Principal Investigator. Please submit a completion report if the study is completed before the expiration date or if you are no longer affiliated with the University of Denver. You must submit a new application at the end of the five-year period if you wish to continue this study.

NOTE: Please add the following information to any consent forms, surveys, questionnaires, invitation letters, etc. you will use in your research as follows: This survey (consent, study, etc.) was approved by the University of Denver’s Institutional Review Board for the Protection of Human Subjects in Research on 12/11/2012. This information will be added by the Research Compliance Office if it does not already appear in the forms upon approval and continuation.

The Institutional Review Board appreciates your cooperation in protecting subjects and ensuring that each subject gives a meaningful consent to participate in research projects. If you have any questions regarding your obligations under the Assurance, please do not hesitate to contact us.

Sincerely yours,

[Signature]
Appendix H

CMV Eight-Factor Hypothetical Model Specification (Amos Version 18)
Appendix I

CMV Nine-Factor Theoretical Model Specification (Amos Version 18)
Appendix J

LMI Six-Factor Hypothetical Model Specification (Amos Version 18)
Appendix K

LMI Unidimensional Theoretical Model Specification (Amos Version 18)
Appendix L

CMV Composite Model Specification (Amos Version 18)
Appendix M

CMV Eight-Factor Hypothetical Consecutive Model Specification (Amos Version 18)
Appendix N

CMV Eight-Factor Hypothetical Multidimensional Model Specification (Amos Version 18)
Appendix O

CMV Nine-Factor Theoretical Consecutive Model Specification (Amos Version 18)
Appendix P

CMV Nine-Factor Theoretical Multidimensional Model Specification (Amos Version 18)
Appendix Q

LMI Theoretical Composite Model Specification (Amos Version 18)
Appendix R

LMI Hypothetical Consecutive Model Specification (Amos Version 18)
Appendix S

LMI Hypothetical Multidimensional Model Specification (Amos Version 18)
# Appendix T

Abbreviations and Acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABIC</td>
<td>Adjusted Bayesian Information Criterion</td>
</tr>
<tr>
<td>AFDD</td>
<td>Air Force Doctrine Document</td>
</tr>
<tr>
<td>AFM</td>
<td>Air Force Manual</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike’s Information Criterion</td>
</tr>
<tr>
<td>ALMR LR</td>
<td>Adjusted Lo-Mendell-Rubin Likelihood Ratio</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>ARDA</td>
<td>Awareness Reasoning Deciding Acting</td>
</tr>
<tr>
<td>AVE</td>
<td>Average Variance Extracted</td>
</tr>
<tr>
<td>BIC</td>
<td>Bayesian Information Criterion</td>
</tr>
<tr>
<td>BLRT</td>
<td>Bootstrap Likelihood Ratio Test</td>
</tr>
<tr>
<td>CCLD</td>
<td>Center for Character and Leadership Development</td>
</tr>
<tr>
<td>CFA</td>
<td>Confirmatory Factor Analysis</td>
</tr>
<tr>
<td>CFI</td>
<td>Comparative Fit Index</td>
</tr>
<tr>
<td>CMV</td>
<td>Character Mosaic Virtues</td>
</tr>
<tr>
<td>CTT</td>
<td>Classical Test Theory</td>
</tr>
<tr>
<td>DIF</td>
<td>Differential Item Functioning</td>
</tr>
<tr>
<td>$df_M$</td>
<td>Model Degrees of Freedom</td>
</tr>
<tr>
<td>EFA</td>
<td>Exploratory Factor Analysis</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation-Maximization</td>
</tr>
<tr>
<td>EPC</td>
<td>Expected Parameter Change</td>
</tr>
</tbody>
</table>
ICC  Item Characteristic Curves
IPIP  International Personality Item Pool
IRB  Institutional Review Board
IRT  Item Response Theory
LCA  Latent Class Analysis
LMI  Leadership Mosaic Inventory
LRM LR  Lo-Mendell-Rubin Likelihood Ratio
MI  Modification Indices
MIRT  Multidimensional Item Response Theory
ML  Maximum Likelihood
MRCMLM  Multidimensional Random Coefficients Multinomial Logit Model
ODS  Officer Development System
ORF  Option Response Function
RCMLM  Random Coefficients Multinomial Logit Model
RMSEA  Root Mean Square Error of Approximation
RSM  Rating-Scale Model
SEM  Structural Equation Modeling
TLI  Tucker-Lewis Fit Index
ULI  Unit Loading Identification
USAF  United States Air Force
USAFA  United States Air Force Academy
UVI  Unit Variance Identification
VIA     Values In Action
WLS     Weighted Least Squares
WLSMV   Means and Variances Corrected Diagonally Weighted Least Squares
WRMR    Weighted Root Mean Square Residual