An Investigation of Unidimensional Testing Procedures under Latent Trait Theory using Principal Component Analysis

Michael T. McGill

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Educational Research and Evaluation

Approved By
Edward W. Wolfe
Gary Skaggs
Yasuo Miyazaki

Disapproved By
J.P. Morgan

October 28, 2009
Blacksburg, VA

Keywords: Unidimensionality, Dimensionality, Factor Analysis, Principal Component Analysis, Rasch, Item Response Theory
An Investigation of Unidimensional Testing Procedures under Latent Trait Theory using Principal Component Analysis

Michael T. McGill

ABSTRACT

There are several generally accepted rules for detecting unidimensionality, but none are well tested. This simulation study investigated well-known methods, including but not limited to, the Kaiser (k>1) Criterion, Percentage of Measure Validity (greater than 50%, 40%, or 20%), Ratio of Eigenvalues, and Kelley method, and compares these methods to each other and a new method proposed by the author (McGill method) for assessing unidimensionality. After applying principal component analysis (PCA) to the residuals of a Latent Trait Test Theory (LTTT) model, this study was able to address three purposes: determining the Type I error rates associated with various criterion values, for assessing unidimensionality; determining the Type II error rates and statistical power associated with various rules of thumb when assessing dimensionality; and, finally, determining whether more suitable criterion values could be established for the methods of the study by accounting for various characteristics of the measurement context. For those methods based on criterion values, new modified values are proposed. For those methods without criterion values for dimensionality decisions, criterion values are modeled and presented. The methods compared in this study were investigated using PCA on residuals from the Rasch model. The sample size, test length, ability distribution variability, and item
distribution variability were varied and the resulting Type I and Type II error rates of each method were examined.

The results imply that certain conditions can cause improper diagnoses as to the dimensionality of instruments. Adjusted methods are suggested to induce a more stable condition relative to the Type I and Type II error rates. The nearly ubiquitous Kaiser method was found to be biased towards signaling multidimensionality whether it exists or not. The modified version of the Kaiser method and the McGill method, proposed by the author were shown to be among the best at detecting unidimensionality when it was present. In short, methods that take into account changes in variables such as sample size, test length, item variability, and person variability are better than methods that use a single, static criterion value in decision making with respect to dimensionality.
Dedication

This is dedicated to My Parents, Brother, Family, Friends and Leigh.

Thanks for being there.
Acknowledgements

I would like to thank my committee members.

acta est fabula plaudite
# Table of Contents

List of Figures vii

List of Tables xiii

Chapter 1: Introduction 1

Chapter 2: Review of Relevant Literature Methods for Assessing Unidimensionality 4

Chapter 3: Method 24

Chapter 4: Results and Discussion 30

Chapter 5: Summary 220

References 231
List of Figures

Figure Name                                                      Page
Figure 1. Graph of Rasch Model with Difficulty $\delta = 0$        9
Figure 2. Kaiser Type I Error Rates for Person and Item Variance 32
Figure 3. Kaiser Type I Error Rates for Person and Sample Size    34
Figure 4. Percentage of Measure Variance Type I Error Rates for Person and Item Variance 39
Figure 5. Percentage of Measure Variance Type I Error Rates for Person and Sample Size 41
Figure 6. Ratio of Eigenvalues Type I Error Rates for Person and Item Variance 47
Figure 7. Ratio of Eigenvalues Type I Error Rates for Person and Sample Size 49
Figure 8. Kelley Method 5th Percentiles for Person and Item Variance 56
Figure 9. Kelley Method 5th Percentiles for Person and Sample Size 58
Figure 10. Divgi Method 5th Percentiles for Person and Item Variance 64
Figure 11. Divgi Method 5th Percentiles for Person and Sample Size 66
Figure 12. McGill Method 95th Percentiles for Person and Item Variance 72
Figure 13. McGill Method 95th Percentiles for Person and Sample Size 73
Figure 14. Average Type I Errors for Methods relative to Sample Size 79
Figure 15. Average Type I Errors for Methods relative to Test Length 80
Figure 16. Average Type I Errors for Methods relative to Person Variance 82
Figure 17. Average Type I Errors for Methods relative to Item Variance 84
Figure 18. Power rates of Kaiser with Correlation = 0.00.           87
Figure 19. Power rates of Kaiser with Correlation = 0.25.           89
Figure 20. Power rates of Kaiser with Correlation = 0.50.           91
Figure 21. Power rates of Kaiser with Correlation = 0.75.           92
Figure 22. Average Power rates for Kaiser method relative to sample size and correlation. 96
Figure 23. Average Power rates for Kaiser method relative to test length and correlation 97
Figure 24. Average Power rates for Kaiser method relative to Person Variance and correlation. 97
Figure 25. Average Power rates for Kaiser method relative to Item Variance and Correlation. 97
Figure 26. Average Power rates for Kaiser method relative to Item Variance, Person Variance, with correlation of 0.0. 98
Figure 27. Average Power rates for Kaiser method relative to Item Variance, Person Variance, with correlation of 0.25. 98
Figure 28. Average Power rates for Kaiser method relative to Item Variance, Person Variance, with correlation of 0.5. 98
Figure 29. Average Power rates for Kaiser method relative to Item Variance, Person Variance, with correlation of 0.75. 99
Figure 30. Power rates of Modified Kaiser with Correlation = 0.00. 100
Figure 31. Power rates of Modified Kaiser with Correlation = 0.25. 102
Figure 32. Power rates of Modified Kaiser with Correlation = 0.50. 104
Figure 33. Power rates of Modified Kaiser with Correlation = 0.75. 106
Figure 34. Average Power rates for Modified Kaiser method relative to Sample Size and correlation. 108
Figure 35. Average Power rates for Modified Kaiser method relative to Test Length and correlation. 109
Figure 36. Average Power rates for Modified Kaiser method relative to Person Variance and correlation. 109
Figure 37. Average Power rates for Modified Kaiser method relative to Item Variance and correlation. 109
Figure 38. Average Power rates for Modified Kaiser method relative to Item Variance, Person Variance, with correlation of 0.0. 110
Figure 39. Average Power rates for Modified Kaiser method relative to Item Variance, Person Variance, with correlation of 0.25. 110
Figure 40. Average Power rates for Modified Kaiser method relative to Item Variance, Person Variance, with correlation of 0.5. 110
Figure 41. Average Power rates for Modified Kaiser method relative to Item Variance, Person Variance, with correlation of 0.75. 111
Figure 42. Power rates of 20% Measure Variability with Correlation = 0.00. 112
Figure 43. Power rates of 20% Measure Variability with Correlation = 0.25. 114
Figure 44. Power rates of 20% Measure Variability with Correlation = 0.50. 116
Figure 45. Power rates of 20% Measure Variability with Correlation = 0.75. 118
Figure 46. Average Power rates for 20% Measure Variability method relative to Sample Size and correlation. 121
Figure 47. Average Power rates for 20% Measure Variability method relative to Test Length and correlation 121
Figure 48. Average Power rates for 20% Measure Variability method relative to Person Variance and correlation. 121
Figure 49. Average Power rates for 20% Measure Variability method relative to Item Variance and correlation. 122
Figure 50. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.0 122
Figure 51. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.25 122
Figure 52. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.5. 123
Figure 53. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.75. 123
Figure 54. Power rates of 40% Measure Variability with Correlation = 0.00. 124
Figure 55. Power rates of 40% Measure Variability with Correlation = 0.25. 126
Figure 56. Power rates of 40% Measure Variability with Correlation = 0.50 128
Figure 57. Power rates of 40% Measure Variability with Correlation = 0.75. 130
Figure 57. Average Power rates for 40% Measure Variability method relative to Sample Size and correlation. 132
Figure 58. Average Power rates for 40% Measure Variability method relative to Test Length and correlation. 133
Figure 59. Average Power rates for 40% Measure Variability method relative to Person Variance and correlation. 133
Figure 60. Average Power rates for 40% Measure Variability method relative to Item Variance and correlation. 133
Figure 61. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.0. 134
Figure 62. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.25. 134
Figure 63. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.5. 134
Figure 64. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.75. 135
Figure 65. Power rates of 50% Measure Variability with Correlation = 0.00. 135
Figure 66. Power rates of 50% Measure Variability with Correlation = 0.25. 138
Figure 67. Power rates of 50% Measure Variability with Correlation = 0.50 140
Figure 68. Power rates of 50% Measure Variability with Correlation = 0.75. 142
Figure 69. Average Power rates for 50% Measure Variability method relative to Sample Size and correlation. 144
Figure 70. Average Power rates for 50% Measure Variability method relative to Test Length and correlation. 145
Figure 71. Average Power rates for 50% Measure Variability method relative to Person Variance and correlation. 145
Figure 72. Average Power rates for 50% Measure Variability method relative to Item Variance and correlation. 145
Figure 73. Average Power rates for 50% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.0. 146
Figure 74. Average Power rates for 50% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.25. 146
Figure 75. Average Power rates for 50% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.5. 146
Figure 76. Average Power rates for 50% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.75. 147
Figure 77. Power rates of Modified Percentage Measure Variability with Correlation = 0.00. 147
Table 78. Power rates of Modified Percentage Measure Variability with Correlation = 0.25. 150
Table 79. Power rates of Modified Percentage Measure Variability with Correlation = 0.50. 152
Table 80. Power rates of Modified Percentage Measure Variability with Correlation = 0.75. 154
Figure 81. Average Power rates for Modified Proportion of Measure Variability method relative to Sample Size and correlation. 157
Figure 82. Average Power rates for Modified Proportion of Measure Variability method relative to Sample Size and correlation. 157
Figure 83. Average Power rates for Modified Proportion of Measure Variability method relative to Person Variance and correlation. 157
Figure 84. Average Power rates for Modified Proportion of Measure Variability method relative to Item Variance and correlation. 158
Figure 85. Average Power rates for Modified Proportion of Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.0. 158
Figure 86. Average Power rates for Modified Proportion of Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.25. 158
Figure 87. Average Power rates for Modified Proportion of Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.5. 159
Figure 88. Average Power rates for 5 Modified Proportion of Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.75. 159
Figure 89. Power rates Ratio of Eigenvalues with Correlation = 0.00. 160
Figure 90. Power rates Ratio of Eigenvalues with Correlation = 0.25. 162
Figure 91. Power rates Ratio of Eigenvalues with Correlation = 0.50. 164
Figure 92. Power rates Ratio of Eigenvalues with Correlation = 0.75. 166
Figure 93. Average Power rates for Ratio of Eigenvalues method relative to Sample Size and correlation. 168
Figure 94. Average Power rates for Ratio of Eigenvalues method relative to Test Length and correlation. 169
Figure 95. Average Power rates for Ratio of Eigenvalues method relative to Person Variance and correlation. 169
Figure 96. Average Power rates for Ratio of Eigenvalues method relative to Item Variance and correlation. 169
Figure 97. Average Power rates for Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.0. 170
Figure 98. Average Power rates for Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.25. 170
Figure 99. Average Power rates for Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.5. 170
Figure 100. Average Power rates for Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.75. 171
Figure 101. Power rates of Modified Ratio of Eigenvalues with Correlation = 0.00. 172
Figure 102. Power rates of Modified Ratio of Eigenvalues with Correlation = 0.25. 174
Figure 103. Power rates of Modified Ratio of Eigenvalues with Correlation = 0.50. 176
Figure 104. Power rates of Modified Ratio of Eigenvalues with Correlation = 0.75. 178
Figure 105. Average Power rates for Modified Ratio of Eigenvalues method relative to Sample Size and correlation. 181
Figure 106. Average Power rates for Modified Ratio of Eigenvalues method relative to Test Length and correlation. 181
Figure 107. Average Power rates for Modified Ratio of Eigenvalues method relative to Person Variance and correlation. 181
Figure 108. Average Power rates for Modified Ratio of Eigenvalues method relative to Item Variance and correlation. 182
Figure 109. Average Power rates for Modified Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.0. 182
Figure 110. Average Power rates for Modified Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.25. 182
Figure 111. Average Power rates for Modified Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.5. 183
Figure 112. Average Power rates for Modified Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.75. 183
Figure 113. Power rates of Kelley Method with Correlation = 0.00. 184
Figure 114. Power rates of Kelley Method with Correlation = 0.25. 186
Figure 115. Power rates of Kelley Method with Correlation = 0.50. 188
Figure 116. Power rates of Kelley Method with Correlation = 0.75. 190
Figure 117. Average Power rates for Kelley method relative to Sample Size and correlation. 192
Figure 118. Average Power rates for Kelley method relative to Test Length and correlation. 193
Figure 119. Average Power rates for Kelley method relative to Person Variance and correlation. 193
Figure 120. Average Power rates for Kelley method relative to Item Variance and correlation. 193
Figure 121. Average Power rates for Kelley method relative to Item Variance, Person Variance, with correlation of 0.0. 194
Figure 122. Average Power rates for Kelley method relative to Item Variance, Person Variance, with correlation of 0.25. 194
Figure 123. Average Power rates for Kelley method relative to Item Variance, Person Variance, with correlation of 0.5. 194
Figure 124. Average Power rates for Kelley method relative to Item Variance, Person Variance, with correlation of 0.75. 195
Figure 125. Power rates of Divgi Method with Correlation = 0.00. 196
Table 126. Power rates of Divgi Method with Correlation = 0.25. 198
Table 127. Power rates of Divgi Method with Correlation = 0.50. 201
Table 128. Power rates of Divgi Method with Correlation = 0.75. 202
Figure 129. Average Power rates for Divgi method relative to Sample Size and correlation. 204
Figure 130. Average Power rates for Divgi method relative to Test Length and correlation. 205
Figure 131. Average Power rates for Kelley method relative to Person Variance and correlation. 205
Figure 132. Average Power rates for Kelley method relative to Item Variance and correlation. 205
Figure 133. Average Power rates for Divgi method relative to Item Variance, Person Variance, with correlation of 0.0. 206
Figure 134. Average Power rates for Divgi method relative to Item Variance, Person Variance, with correlation of 0.25. 206
Figure 135. Average Power rates for Divgi method relative to Item Variance, Person Variance, with correlation of 0.5. 206
Figure 136. Average Power rates for Divgi method relative to Item Variance, Person Variance, with correlation of 0.75. 207
Figure 137. Power rates of McGill Method with Correlation = 0.00. 208
Figure 138. Power rates of McGill Method with Correlation = 0.25. 210
Figure 139. Power rates of McGill Method with Correlation = 0.50. 212
Figure 140. Power rates of McGill Method with Correlation = 0.75. 214
Figure 141. Average Power rates for McGill method relative to Sample Size and correlation. 216
Figure 142. Average Power rates for McGill method relative to Test Length and correlation. 217
Figure 143. Average Power rates for McGill method relative to Person Variance and correlation. 217
Figure 144. Average Power rates for McGill method relative to Item Variance and correlation. 217
Figure 145. Average Power rates for McGill method relative to Item Variance, Person Variance, with correlation of 0.0. 218
Figure 146. Average Power rates for McGill method relative to Item Variance, Person Variance, with correlation of 0.25. 218
Figure 147. Average Power rates for McGill method relative to Item Variance, Person Variance, with correlation of 0.5. 218
Figure 148. Average Power rates for McGill method relative to Item Variance, Person Variance, with correlation of 0.5. 219
Figure 149. Average Type I errors for Methods relative to Sample Size 222
Figure 150. Average Type I error rates versus test length 223
Figure 151. Average Power of methods by sample size. 225
Figure 152. Average Power error of methods by test length. 226
Figure 153. Average Power of methods by correlation. 227
List of Tables

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1. Results of Stepwise Regression for Modified Kaiser Method</td>
<td>36</td>
</tr>
<tr>
<td>Table 2. Sample Critical Values for the Modified Kaiser Method Based on Independent Variables</td>
<td>37</td>
</tr>
<tr>
<td>Table 3. Modified Kaiser Method VIF values for the parameters</td>
<td>38</td>
</tr>
<tr>
<td>Table 4. Results of Stepwise Regression for Modified Percentage of Variance Method</td>
<td>43</td>
</tr>
<tr>
<td>Table 5. Sample Critical Values for the Modified Percentage of Variance Method Based on Independent Variables</td>
<td>44</td>
</tr>
<tr>
<td>Table 6. Proportion of Variability Method VIF Values for the Parameters of Full Model</td>
<td>45</td>
</tr>
<tr>
<td>Table 7. Proportion of Variability Method VIF Values for the Parameters of Simple Model</td>
<td>46</td>
</tr>
<tr>
<td>Table 8. Proportion of Variability Method VIF Values for the Parameters of Model with Interactions</td>
<td>46</td>
</tr>
<tr>
<td>Table 9. Results of Stepwise Regression for Modified Ratio Method</td>
<td>52</td>
</tr>
<tr>
<td>Table 10. Sample Critical Values for the Modified Ratio Method Based on Independent Variables</td>
<td>53</td>
</tr>
<tr>
<td>Table 11. Modified Ratio of Eigenvalues Method VIF Values for the Parameters of Full Model</td>
<td>53</td>
</tr>
<tr>
<td>Table 12. Modified Ratio of Eigenvalues Method VIF Values for the Parameters of the Simple Model</td>
<td>54</td>
</tr>
<tr>
<td>Table 13. Modified Ratio of Eigenvalues Method VIF Values for the Parameters of the Simple Model with Interactions</td>
<td>54</td>
</tr>
<tr>
<td>Table 14. Results of Stepwise Regression for Kelley Method</td>
<td>60</td>
</tr>
<tr>
<td>Table 15. Sample Critical Values for the Kelley Method Based on Independent Variables</td>
<td>61</td>
</tr>
<tr>
<td>Table 16. Kelley Method VIF Values for the Parameters of Full Model</td>
<td>62</td>
</tr>
<tr>
<td>Table 17. Kelley Method VIF Values for the Parameters of Simple Model</td>
<td>62</td>
</tr>
<tr>
<td>Table 18. Kelley Method VIF Values for the Parameters of Simple Model with Interactions</td>
<td>63</td>
</tr>
<tr>
<td>Table 19. Results of Stepwise Regression for Divgi Method</td>
<td>68</td>
</tr>
<tr>
<td>Table 20. Sample Critical Values for the Divgi Method Based on Independent Variables</td>
<td>69</td>
</tr>
<tr>
<td>Table 21. Divgi Method VIF Values for the Parameters of Full Model</td>
<td>69</td>
</tr>
<tr>
<td>Table 22. Divgi Method VIF Values for the Parameters of the Simple Model</td>
<td>70</td>
</tr>
<tr>
<td>Table 23. Divgi Method VIF Values for the Parameters of the Simple Model with Interactions</td>
<td>70</td>
</tr>
<tr>
<td>Table 24. Results of Stepwise Regression for McGill Method</td>
<td>75</td>
</tr>
<tr>
<td>Table 25. Sample Critical Values for the McGill Method Based on Independent Variables</td>
<td>76</td>
</tr>
<tr>
<td>Table 26. McGill Method VIF values for the parameters of full model</td>
<td>76</td>
</tr>
</tbody>
</table>
Table 27. Average Type I Errors for Methods
Table 28. Average Type I Errors for Methods relative to Test Length
Table 29. Average Type I Errors for Methods relative to Person Variance
Table 30. Average Type I Errors for Methods relative to Item Variance
Table 31. Average Type I error rates for Methods.
Table 32. Average Power of methods by sample size.
Table 33. Average Power error of methods by test length
Table 34. Average Power of methods by correlation.
CHAPTER 1

INTRODUCTION

The concept of unidimensionality is important in testing applications for several reasons. Stout (1987), in particular, noted three primary concerns. First, varying abilities should not place an influence on the measure of the construct that is the basis of the assessment as they do in the case of multidimensional measures. Second, when ordering persons based on a construct, it is necessary to have only one dimension to keep from confounding the rankings of persons on the qualities being measured. Third, the requirement of the unidimensionality assumption is necessary to avoid potential bias in the estimates of item and person parameters (Stout 1987; Reckase 1979; Harrison 1986). While there are several methods for evaluating claims about unidimensionality, there is no commonly agreed upon criterion for deducing the dimensionality of measures. After a comprehensive review of the relevant literature, Hattie (1985) noted that “there are still no known satisfactory indices” (p. 158). Therefore, Hattie concluded, it is necessary to investigate the various indices under known conditions using simulations in order to find consistency among the methods.

In LTTT (Latent Trait Test Theory) applications, one method for evaluating dimensionality focuses on the difference between model-based expected values and observed item responses (i.e., residuals). Residuals from a unidimensional model can be subjected to Principal Component Analysis (PCA) to determine whether covariability exists among the residuals. If it does, by definition, multidimensionality exists. Hence, PCA of latent trait unidimensional model residuals may be useful in determining whether data satisfy the unidimensionality assumption that is typically invoked in applications of those models.
Several “Rules of Thumb” for interpreting evidence relating to unidimensionality in the context of PCA have been noted, but to date studies have failed to reveal which of these methods is best suited for identifying unidimensionality (Hattie, 1987). Thus, one purpose of the proposed study is to determine the Type I error rates associated with various criterion values associated with applications of PCA to LTTT residuals for the purpose of evaluating a unidimensionality assumption. A second purpose of this study is to determine the Type II error rate and statistical power associated with various rules of thumb that are utilized in applications of PCA to LTTT residuals. As is true with any rule of thumb, exceptions to those rules exist. Therefore, a third purpose of this study is to determine whether more suitable criterion values can be established by taking into account various characteristics of the measurement context in applications of PCA to the residuals from LTTT models. Finally, because unmodeled error exists in all applications of LTTT and because such error may influence the performance of various methods for assessing unidimensionality, a fourth purpose of this study is to determine the impact of the Type I and Type II errors and the statistical power associated with criterion values established for applications of PCA to the residuals from LTTT models.

**Overview of Study**

This study evaluates the appropriateness of several rules of thumb for defining criterion values in applications of PCA to LTTT model residuals by determining the rates with which they diagnose or misdiagnose unidimensionality and multidimensionality in cases in which the underlying item responses are dichotomous in nature. Improvements to current rules of thumb are also proposed by applying regression analysis to the results obtained through simulation methods. The validity of inferences that are based on the adjusted critical
values is examined in a second simulation study that focuses on the impact of common model-to-data misfit encountered in applications of LTTT models.
CHAPTER 2

Review of Relevant Literature Methods for Assessing Unidimensionality

Methods used to evaluate dimensionality can be categorized into those based on answer patterns, reliability, principal components, factor analysis, or latent trait models (Hattie, 1985). *Answer pattern* indices are based on the difference between the observed response pattern and theoretical scale pattern, where a scale pattern is one such that a scale score of \( n \) indicates that the \( n \) easiest questions were answered correctly, sometimes referred to as a Guttman response pattern. *Reliability methods* focus on the behavior of Cronbach’s alpha, which depicts the typical interitem correlation on an instrument. The value of alpha increases as the pool of items becomes more homogenous, meaning that alpha increases as the measures becomes more unidimensional (Hattie, 1985).

*Factor Analysis* (FA) and *Principal Component Analysis* methods are similar in their overall approach, but the underlying assumptions differentiate them (Stevens, 2002). FA assumes that there can be error due to variables where as PCA does not. FA also allows components to be correlated while PCA does not. The data analytic process is similar for both FA and PCA in that latent components are identified, each defined by a relatively homogeneous yet distinct set of items (typically referred to as “variables”). Each component is defined by a set of weights (or “loadings”) that denote the degree to which items are correlated with the latent component. Undimensionality, in either FA or PCA, can be defined by the existence of a single dominant component that explains the covariation among the items on a measurement instrument.

The *latent trait methods* are based on a family of measurement models that depict the probability of a particular response to an item as a function of characteristics of the
respondent and characteristics of the item. An underlying assumption of many LTTT models is unidimensionality, and LTTT methods for assessing unidimensionality employ fit values that are based on the difference between the observed and model-based expected responses (i.e., residuals). Unfortunately, most of the latent trait methods for assessing dimensionality lack a theoretical basis and rationale for their suggested critical values that are used for decision making (Hattie, 1985).

Hattie (1985) investigated over thirty methods for assessing unidimensionality from these five broad categories, choosing the most popular methods employed by researchers at that time. This study focuses on several PCA methods reviewed by Hattie, all applied to the assessment of unidimensionality in the context of LTTT models. In the sections that follow, LTTT models are described along with PCA methods for assessing dimensionality. Finally, the results of a pilot study are summarized, and research questions for the dissertation are presented.

**Latent Trait Test Theory**

Latent Trait Test Theory (LTTT), also referred to as Item Response Theory (IRT), constitutes a mathematical model of the process through which persons interact with test items, depicting the probability of various outcomes that may result from that interaction. The simplest of these models is commonly referred to as the Rasch model, has the following form:

\[ \pi_{ni} = \frac{\exp(\theta_n - \delta_i)}{1 + \exp(\theta_n - \delta_i)} \]  \hfill (1)
Where $\pi_{ni}$ is the probability of a correct response by person $n$ of item $i$,

$\theta_n$ is the ability of the person labeled $n$ and

$\delta_i$ is the difficulty associated with item $i$.

Wright and Masters (1982) suggest the following generalized equation, depicting an observed score on a test item as a function of the process depicted by the Rasch model:

$$Y_{ni} = \pi_{ni} \pm \epsilon_{ni}$$

or

$$Y_{ni} = \frac{e^{\theta_n - \delta_i}}{1 - e^{\theta_n - \delta_i}} \pm \epsilon_{ni}$$

$$Y_{ni} \sim \left( \frac{1}{X_{ni}} \right)^{\pi_{ni}} (1 - \pi_{ni})^{1 - x_{ni}}$$

Where $Y_{ni}$ is the observed response by person $n$ on item $i$.

$\pi_{ni}$ is the probability that person $n$ answers item $i$ correctly.

$\epsilon_{ni}$ is the error associated with person $n$ on item $i$.

Where $Y_{ni}$ is response being either 0 for an incorrect response or 1 for a correct response for $n$th person on the $i$th question when taking into account error from the Rasch model. The error term relative to person $n$ with respect to item $i$ is denoted as $\epsilon_{ni}$. The expected value of $n$th person on the $i$th question is denoted as $\pi_{ni}$.

The error term $\epsilon_{ni}$ is distributed as a Binomial with $n=1$ and $p=\pi_{ni}$, or in other words a Bernoulli with $p=\pi_{ni}$. Thus,

$$\epsilon_{ni} \sim \pi_{ni}^{x_{ni}} (1 - \pi_{ni})^{1 - x_{ni}}$$

where $\epsilon_{ni}$ is the error associated with person $n$ on item,

$\pi_{ni}$ is the probability that person $n$ answers item $i$ correctly,
and $x_{ni}$ is the number observed success for the Rasch model consisting of either a 1 for a correct response or 0 for an incorrect response for person $n$ on item $i$.

Since the distribution of $\varepsilon_{ni}$ is known the mean is computed as $\pi_{ni}$ and the standard deviation is $\sqrt{\pi_{ni}(1-\pi_{ni})}$. Due to the fact that one test-taker does not influence another and that in theory a subjects answer on an item should not effect answers on other items, thus implying between and with-in subject independence the $\varepsilon_{ni}$’s are independent. Since the error terms are independent it the follows that the $Y_{ni}$’s are independent as well.

An alternative version of this idea for the errors is found when investigating

$$Y_{ni} = \pi_{ni} + \varepsilon_{ni} \quad (6)$$

or

$$Y_{ni} = \frac{e^{\theta_i - \beta_i}}{1 - e^{\theta_i - \beta_i}} + \varepsilon_{ni} \quad (7)$$

Where $Y_{ni}$ is the item score (0 for an incorrect response and 1 for a correct response) for $n$th person on the $i^{th}$ item. $Y_{ni}$ is distributed as Binomial with probability $\pi_{ni}$ and $n=1$ which is a special case which reduces the binomial to a Bernoulli with a probability of $\pi_{ni}$.

When $\pi_{ni}$ is subtracted from both sides of Equation 6 it is clear that

$$\varepsilon_{ni} = Y_{ni} - \pi_{ni} \quad (8)$$

Thus, with $Y_{ni}$ denoted as 0 for an incorrect response and 1 for a correct response

$$\varepsilon_{ni} = \begin{cases} 1 - \pi_{ni}, & \text{if } Y_{ni} = 1 \\ -\pi_{ni}, & \text{if } Y_{ni} = 0 \end{cases} \quad (9)$$

This implies that the errors $\varepsilon_{ni}$’s are distributed as a Bernoulli that is shifted by $\pi_{ni}$.
Due to the fact that one test-taker does not influence another and that in theory a
subject's answer on an item should not effect answers on other items, thus implying between
and within subject independence the $\varepsilon_{ni}$'s are independent. Since the error terms are
independent it then follows that the $Y_{ni}$'s are independent as well.

Graphically, the Rasch model is depicted by an item information curve (ICC). An
ICC associated with the Rasch model is shown in Figure 1. The x-axis displays the values
person abilities ($\theta_n$), while the y-axis displays the probability that a person with the specified
ability will answer the item in question correctly ($\pi_{ni}$). The point of inflection of the ICC for
a particular item, which is located at $\pi_{ni} = .50$ for the Rasch model, is the point on the x-axis
that designates the difficulty of the item ($\delta_i$). A key point of the Rasch model is that the
slopes of the ICCs are modeled to be equivalent and the lower asymptote is defined to be
zero for all items. The only possible change between items is the difficulty, which would
appear as a shift in the location of the ICC along the x-axis.
Figure 1. Graph of Rasch Model with a difficulty $\delta=0$.

Parameter Estimation

Parameter estimation for the Rasch model is commonly conducted using software such as Winsteps (Linacre, 2009). In the case of Winsteps, parameter estimation is conducted via a two-step approach starting with a Normal Approximation (PROX) estimation algorithm to determine initial values that are used in the second step of the process, which is Joint Maximum Likelihood Estimation (JMLE). During the PROX estimation, a scaling factor is applied to the logit associated with each observed total score (e.g., $R_n$ for examinees) in order to approximate a normal ogive distribution. The initial ability estimate is computed as

$$B_n = \mu_n + \sqrt{1 + \sigma_n^2 / 2.9 \log e \left( \frac{R_n}{N_n - R_n} \right)} \quad (10)$$

where $B_n$ is the initial (PROX) ability estimate for person $n$, $\mu_n$ is the average difficulty of all items taken by the $n$th person, $\sigma_n$ is the standard deviation of the item difficulties for items taken by the $n$th person, $R_n$ is the observed raw score for person
\( n \) (i.e., the sum of the \( x_{ni} \) for that person) and \( N_n \) is maximum score possible for the \( n \)th person. Initial item parameter values are established with analogous equations.

In the second step, these initial estimates of the parameters are utilized as starting values in the JMML procedures. During JMML, expected values (\( \pi_{ni} \)) are computed based on all pairings of those initial ability and difficulty estimates, resulting in an \( n \times i \) matrix of expected values (E). The marginal sums of the expected item and expected person raw scores are then computed. Similarly, marginal sums of the observed scores are computed based on the \( n \times i \) observed score matrix (X), which contains the scores obtained by each examinee in response to each item. The marginal expected and observed scores are then compared to one another, and person or item marginal expected scores are adjusted to be more consistent with the observed scores.

The Newton-Raphson method is used to accomplish this adjustment via solution of \( \theta_n' \) in the following equation.

\[
\theta_n' = \theta_n + \frac{(R_n - \Pi_n)}{mv} \quad (11)
\]

where \( \theta_n \) is a current ability estimate for the \( n \)th person, \( R_n \) is the observed raw score for the \( n \)th person, \( P_n \) is the Rasch expected raw score the \( n \)th person based on the current ability and difficulty estimates, and \( mv \) is the modeled variance, which takes the following form:

\[
mv(\theta_n, \delta_i) = \frac{1}{\sqrt{\sum (\pi_{ni}(1 - \pi_{ni}))}} \quad (12)
\]

where \( \theta_n \) is the ability estimate of the \( n \)th person, \( \delta_i \) is the difficulty estimate of the \( i \)th item, and \( \pi_{ni} \) is the probability of a correct response by the \( n \)th person on the \( i \)th item with respect to the Rasch model.
The process is repeated in an analogous manner for to adjust the item difficulty estimates. The JMLE iterations are discontinued when the largest difference between and observed and expected marginal score or between adjusted estimates on consecutive iterations is less than an arbitrarily defined convergence criteria, which is typically an absolute value of .0001.

**Principal Component Analysis**

Principal Component Analysis is a data-driven method sometimes applied to the assessment of unidimensionality assumptions in LTTT applications. Conceptually, the PCA procedure groups items into sets that correlate with one another but are relatively independent of other internally consistent subsets of items. The purpose of PCA is to explain the covariability between items using a linear combination of the item scores. The objective of creating these linear combinations is often data reduction, but, in the context of dimensionality assessment, the purpose is to improve the interpretability of the measures.

In the current application PCA is conducted on model-based residuals rather than on observed scores. Consider a matrix $X$, an $n \times i$ matrix of observed outcomes relative to $n$ individuals answering $i$ items. The values of the observed matrix, $x_{ni}$ are the dichotomous outcomes—either a 0 for and incorrect answer or a 1 to denote a correct answer.

\[
X_{ni} = \begin{bmatrix}
x_{11} & \cdots & x_{1i} \\
\vdots & & \vdots \\
x_{n1} & \cdots & x_{ni}
\end{bmatrix}
\]  

(13)

The matrix of expected values, defined by Equation 1, are the $\pi_{ni}$ values computed from the estimated ability of person $n$ ($\theta_n$) and the estimated difficulty of item $i$ ($\delta_i$). This $n \times i$ matrix is denoted as $E$. 
A matrix of residuals ($R_{ni}$) is created by taking the difference between $X$ and $E$.

$$
R_{ni} = \begin{bmatrix}
    r_{n1} & \cdots & r_{ni} \\
    \vdots & \ddots & \vdots \\
    r_{ni} & \cdots & r_{ni}
\end{bmatrix} = \begin{bmatrix}
    x_{n1} & \cdots & x_{ni} \\
    \vdots & \ddots & \vdots \\
    x_{ni} & \cdots & x_{ni}
\end{bmatrix} - \begin{bmatrix}
    \hat{\pi}_{n1} & \cdots & \hat{\pi}_{ni} \\
    \vdots & \ddots & \vdots \\
    \hat{\pi}_{ni} & \cdots & \hat{\pi}_{ni}
\end{bmatrix} \quad (15)
$$

A variance-covariance matrix can then be computed between the residuals for all pairs of items by computing the correlation between all pairings of the columns of $R$ and accumulating these correlations into an $i \times i$ matrix denoted $\rho$, which represents the standardized covariation among the residuals from the Rasch model.

$$
\rho = \begin{bmatrix}
    \hat{\rho}_{11} & \cdots & \hat{\rho}_{ii} \\
    \cdots & \ddots & \cdots \\
    \hat{\rho}_{ii} & \cdots & \hat{\rho}_{ii}
\end{bmatrix} \quad (16)
$$

From the correlation matrix, eigenvalues and eigenvectors are created. An eigenvalue is a scalar value denoted as $\lambda_1 \ldots \lambda_i$ relative to an $i \times i$ matrix, in this case the correlation matrix of the $i$ items denoted $\rho$. This scalar value is such that its product with an identity matrix ($I$) of size $i$ subtracted from the correlation matrix results in a value of zero once the determinant is taken (Johnson & Wichern, 1998).

$$
| \rho - \lambda I | = 0 \quad (17)
$$

The results of this determinant is a polynomial equation of order $i$ that when solved will provide values for all $\lambda$’s. The $\lambda$ that is the largest is the eigenvalue related to the first residual dimension, will be denoted as $\lambda_2$ from this point forward. The next largest $\lambda$ will be the associated eigenvalue for the second residual dimension, denoted as $\lambda_3$, and so on. This
method continues for all subsequent eigenvalues in this pattern of decreasing magnitude relative to increasing dimensions.

To apply PCA to unidimensionality assessment in the context of LTTT, the following steps are followed. First, the LTTT model is fit to the observed item response data. McDonald (1985) indicates that this process is equivalent to extracting the first principal component from the observations—from this point forward, the variance accounted for by that first principal component is referred to as $\lambda_1$. Specifically, he explains that the Rasch model, from a factor analytic point of view, “is a nonlinear transformation of the special case of the Spearman single-factor model in which the factor loadings are required to be all the same.” (p. 211) In fact, McDonald indicates that “common factor analysis is really a special case of latent trait theory, based on the principle of local independence.” (p. 203). In the current context, the principle of local independence states that, for any location in the space of variation of the latent trait, the observations are mutually independent. As an indication of the degree to which local independence is evident in the observed data, McDonald recommends examining the covariance matrix of item residuals to see if the items are indeed conditionally uncorrelated.

Hence, the second step of dimensionality assessment is to determine whether the covariance matrix of the item residuals from the Rasch model depart from an idealized covariance matrix in which the off diagonal elements are zero. If the local independence assumption is met, then the resulting residuals from the model-based expectations should be uncorrelated. To accomplish this, a PCA is conducted on the item residual correlations. The focus of this study is the criterion by which one determines whether the amount of variance
accounted for by the residual components is sufficient to conclude that the local independence assumption is violated (i.e., criterion values for the PCA).

One minor issue associated with such an application of PCA is the fact that the amount of variance associated with the component extracted through the LTTT parameter estimation process, will not be on the same scale as the variance components that result from a PCA of the item residuals. Specifically, the total variance in a PCA equals the number of items (or variables) in the analysis. By applying a PCA to the residuals from LTTT, the residual variance (not the total variance) will equal the number of items. The total variance will equal the residual variance plus the variance associated with the first principal component, which is set arbitrarily during the LTTT parameter estimation process. Hence, in order to employ criterion values associated with typical applications of PCA (see below for a detailed discussion of criterion values), one needs to linearly transform the total variance (i.e., the primary component, which is extracted through the LTTT parameter estimation process, and the residual components, which are extracted through a PCA of the LTTT item residuals) to have a variance equal to the number of items on the instrument.

**Criterion Values for PCA**

Hattie (1985) described eight methods for defining criterion values when assessing unidimensionality via PCA. These rules can be separated into two groups. The first group is concerned with the residual values derived from the observed response to an item versus the expected response based on an individual’s ability relative to an items difficulty, while the second group deals directly with the eigenvalues formed from a PCA.

The residual-based methods focus on the (a) sum of the residuals, (b) the frequency of residuals with values less than 0.01, and (c) correlations of raw and factor scores. It is
noteworthy that application of these methods to LTTT model residuals would require special software or additional programming on the part of the analyst. On the other hand, the eigenvalue methods such as (a) percent of variance accounted for by the measure dimension, (b) magnitude of the eigenvalues, (c) ratio of measure eigenvalues to first residual component eigenvalues, (d) ratio of the difference of subsequent component eigenvalues, and (e) ratio of eigenvalues to variance can be readily produced by commercially available LTTT software such as Winsteps (Linacre, 2008). Hence, the focus of this study is limited to methods of assessing unidimensionality based on eigenvalues obtained through PCA of LTTT model residuals because these methods are more likely to be accessible to applied researchers.

The **percent of variance** rule of thumb is based on the notion that, when data are unidimensional, the primary component in a PCA should account for a large portion of the variance and that the remaining components should each account for about the same amount of the remaining variance. In the context of PCA of item residuals from LTTT models, the percentage of variance is computed as

\[
Percentage\_Measure\_Variability = \frac{\lambda_1}{\lambda_1 + \sum_{i=2}^{I} \lambda_i} \quad (18)
\]

Where \(\lambda_1\) is the variance associated with the principal component extracted through the Rasch model parameter estimation, and the \(\lambda_i\)'s are the variances associated with the \(I\) residual components.

Carmines and Zeller (1979) believed that at least 40% of the variability should be attributed to the measure dimension when unidimensionality is present. The Winsteps manual’s “tentative guidelines” (Linacre, 2008, p. 376) recommend a similar value--at least 50% of the variance being accounted for by the first component in order to support the
unidimensionality assumption. Reckase (1979) suggested the smallest percentage of variance, at 20%. No applied studies have evaluated the suitability of these three criterion values for the percent of variance approach, nor has a rationale been provided to suggest that any one of these criterion values should be more appropriate than another. Hence, it is unclear what a suitable criterion value would be for the percent of variance approach to PCA of LTTT model residuals in assessing unidimensionality.

The eigenvalue magnitude method, originally presented by Kaiser (1970), is one of the most commonly applied criterion for dimensionality assessment (Glorfield, 1995). In a typical application of PCA, the Kaiser criterion specifies that components for which the eigenvalue is less than 1.00 should be omitted from consideration because these components account for less variance than a single variable (i.e., an item) contributes to the total variance. In the present context, if the value of $\lambda_1$ is greater than 1.00 and all of the residual components produce values of $\lambda_i$ that are less than 1.00, then the data support the assumption of unidimensionality. On the other hand, if any of the residual $\lambda_i$ values are greater than 1.00, the assumption of unidimensionality is not supported. The Kaiser criterion is commonly applied to dimensionality analyses even though some researchers are critical of its use (Cudeck, 2000; Fabrigar et al., 1999; Steger, 2006; Tabachnick and Fidell, 2001; Thompson and Daniel, 1996; Zwick and Velicer, 1986). Zwick and Velicer (1986) compared the Kaiser criteria to four other methods–namely parallel analysis, Velicer’s minimum average partial (MAP), Cattell’s scree test, and Bartlett’s chi-square test of evaluating dimensionality. The results from the comparison of methods revealed that the Kaiser criterion generally selected over-dimensionalized models, while parallel analysis and Velicer’s minimum average partial
procedure performed the most consistently across all treatment levels (Zwick and Velicer, 1986).

The ratio of the eigenvalues associated with the first principal component ($\lambda_1$) and the second principal component ($\lambda_2$) was first suggested by Lumsden (1959), although he did not provide guidelines or theoretical reasoning for its use (Hattie, 1985). In the context of LTTT models, the Ratio of Eigenvalues would compare the relative values of eigenvalues associated with the first (measures) principal component ($\lambda_1$) and the first residual component ($\lambda_2$):

$$\text{Ratio} = \frac{\lambda_1}{\lambda_2} \quad (19)$$

Hutton (1979) suggested that a large ratio implies unidimensionality while a low value gave reason to suspect multidimensionality. Lumsden did note that, given that this ratio has no maximum, it is unclear how this value would react to errors of sampling or measurement (Hattie, 1985). On the other hand, the Winsteps manual in its “tentative guidelines,” suggest that if the measures account for four or more times the variance accounted for by the first residual dimension then there is “good” evidence of unidimensionality (Linacre, 2008).

Lord (1980) took the ratio of eigenvalues criterion one step further and suggested that one should confirm that the first residual component’s eigenvalue was not appreciably larger than the eigenvalues for the remaining residual components. Similarly, Divgi (1980) suggested that the difference between the eigenvalues for the first and second components divided by the difference between those for the third and forth components, referred to here as the difference ratio method, could be a useful criterion measure. In the present context, the Divgi Method compares the difference between the eigenvalues associated with the first
(measured) and second (first residual) principal components to those of the third and fourth components, which correspond to the second and third residual components, respectively.

\[ \text{Divgi} = \frac{\hat{\lambda}_1 - \hat{\lambda}_2}{\hat{\lambda}_3 - \hat{\lambda}_4} \quad (20) \]

The principle behind the difference ratio method is that if the bulk of the variance is accounted for by the measures and the remaining variance is fairly uniformly distributed among the remaining components, then the measures exhibit strong evidence of unidimensionality. The difference ratio method implies that a large value for the ratio is reason to suspect unidimensionality, yet there is no specific critical value for dimensionality assessment (Divgi, 1980). Hattie (1985) notes that it is quite simple to construct examples using this approach in which the ratio will be unduly large, implying unidimensionality even when data are multidimensional.

The ratio of eigenvalues to variance was presented by Kelley (1942), who offered the name “coefficient of coherence” as a measure of singleness of purpose. In the present context, this index is computed by dividing the eigenvalue of the measured component (\( \lambda_1 \)) by the sum of all item variances.

\[ \text{Kelley} = \frac{\lambda_1}{\sum_{i=1}^{I} \sigma_{ii}} \quad (21) \]

The rationale behind this index is that if all items belong to one dimension then the variances beyond the first (measured) component should be relatively small. Thus, minimal item variances added together will create a larger overall value when used as the denominator, as opposed to the a situation where the item variances are more wide-ranging for multiple
dimensions and the sum used for a denominator, would result in an overall value that is smaller. The smaller the computed Kelley value the more homogeneous an instrument.

In this study, a new method proposed by the author (referred to here as the McGill method) also falls under the umbrella of *ratio of eigenvalues to variance* methods. The McGill method for testing dimensionality is a combination and modification of the ideas presented by Kaiser and Kelley. In the McGill method, the first residual dimension eigenvalue ($\lambda_2$) is divided by the sum of the standard errors for the estimates of the item difficulties based on the Rasch model.

$$\text{McGill} = \frac{\lambda_2}{\sum_{i=1}^{I} SE_i}$$

(22)

The equation for the McGill method is similar to that described by Kelly with the exception that the denominator of the ratio contains the total item variance in the case of Kelly and only the item error variance in the case of McGill. The rationale is that the standard errors are the variability found in the estimation of every single item difficulty. Generally, the square of inverse of the standard error is used as a measure of information relative to an item, where the larger the information the better an item. This method looks at the sum of the squared information for the test and compares it to the variability relative to the residual dimension. This method arose out of empirical inspection that resulted from an earlier pilot study (McGill & Wolfe 2008). In most all cases the sum of the standard errors is greater when the underlying instrument is know to be unidimensional. When keeping all other variables constant the sum of standard errors is decreased when more than one dimension is present. Therefore, the value in the McGill method will be smaller in the unidimensional case than would its equivalent value in a multidimensional case.
Generation of Data Consistent with a Unidimensional Rasch Model

Unidimensional data that is consistent with the Rasch model can be generated through the following process. First, ability parameter values are randomly generated from an arbitrarily chosen distribution. Second, item difficulty values are randomly generated from an arbitrarily chosen distribution. Third, a matrix of expected values for each person-by-item combination is constructed by computing the value of each pairing via the Rasch model equation ($\pi_{ni}$). That is, an $E_{ni}$ matrix is computed [Equation 13]. Fourth, a matrix of values is generated from a random uniform distribution, denoted $U_{ni}$. Finally, a simulated score matrix ($X_{ni}$) that is consistent with the Rasch unidimensional dichotomous model is created by comparing the values in $E_{ni}$ to $U_{ni}$. If the value of $u_{ni}$ was found to be less than or equal to the value of $\pi_{ni}$, $x_{ni}$ was defined to be 1. Otherwise, $x_{ni}$ was defined to be 0. In a real world context, a 1 would represent a correct response and a 0 would represent an incorrect response to a dichotomously scored test item.

Pilot Study

Previous research conducted by this author (McGill & Wolfe, 2008) compared three of these methods for assessing unidimensionality via an application of PCA to the residuals from LTTT models. In that study, item response vectors were generated using the Conquest software (Wu, Adams, Wilson, & Haldane, 2007). Person ability estimates were sampled from $N(\mu,1)$ distributions, while item difficulties were sampled from a $U(-2,2)$ distribution. Unidimensional dichotomous item responses were simulated, varying the number of items (25, 50, 75) and sample size (200, 500, 1000) between datasets. In addition, the researchers simulated an offset of the ability and item parameter means with differences of −1.00, -0.50, 0.00, 0.50, or 1.00. For example, an offset of -0.50 was accomplished by sampling person
abilities from a N(-0.50,1.00) distribution and item difficulties from a U(-2,2) distribution. Hence, the independent variables were number of items (3 levels), sample size (3 levels), and offset (5 levels). Each of the 45 cells of the experiment was replicated 100 times, and each simulated data set was scaled to the unidimensional dichotomous Rasch model.

Parameters for each simulated data set were estimated using the Winsteps software (Linacre, 2008), and residuals from that model were output so that a PCA of the residuals could be conducted. Eigenvalues for the measures (i.e., the dominant dimension) and for the first residual component were computed for each data set. Because Winsteps reports eigenvalues on the residual scale, it was necessary to transform those eigenvalues from the residual scale to the total score scale by dividing each eigenvalue by the total variance units and multiplying by the number of items. This step allowed the researchers to evaluate the first residual component with respect to the Kaiser criterion.

The percentage of variance accounted for by the measures was computed by dividing its eigenvalues by the total variance (i.e., the number of items), and the first criterion recommended in the Winsteps manual (Linacre, 2008) (i.e., the 60% rule) was applied so that dominant dimensions that accounted for less than 60% of the total variance were flagged as exhibiting evidence of multidimensionality. Finally, the second criterion recommended in the Winsteps manual (i.e., the 5% rule) was applied so that when the first residual component accounted for more than 5% of the residual variance, the data set was flagged for evidence of multidimensionality.

The results of that study suggest that these three rules for dimensionality determination are not very stable across the studied independent variables. The Type I error rate, with any of the methods, changed with respect to the number of items and/or the sample
size. A key point to be taken from that pilot study is that sample size and test length play a major role in the performance of the critical values in applications of PCA to LTTT residuals. The difference in the average values of the item and person distributions was not an important independent variable.

The simulations did support the notion that the measure of dispersion related to the item and person distributions may be an important variable. Through varying values of either person or item variance the error was found to change regardless of the method. Thus, due to changing error rates related to changing levels of variability, the variance of distributions was a relevant factor in the performance of existing methods of dimensionality testing.

**Hypotheses**

There exist several methods for evaluating unidimensionality claims in the context of PCA. Unfortunately, defensible critical values have not been established for most of those methods, and their suitability for application to evaluation of residuals in latent trait models is unknown. Compounding this problem is the fact that analysts are not given clear and precise rules from which to make decisions regarding dimensionality. Thus, conclusions drawn based on the results of dimensionality analyses conducted via applications of PCA to LTTT model residuals may depend on which set of rules that an analyst chooses to implement.

The proposed study focuses on the following research questions:

1. What are the Type I and Type II error rates and statistical power of current rule of thumb critical values for PCA methods when applied to LTTT model residuals, particularly as a function of characteristics of the underlying
population distributions such as sample size, test length, and distributional variability?

2. Can adjusted critical values be computed that take into account the characteristics of the underlying population distributions in order to improve the Type I and Type II error rates and statistical power of dimensionality assessments that are based on PCA of LTTT model residuals?

3. Can suitable critical values that are based on population characteristics be determined for PCA of LTTT model residual methods which do not currently have rules of thumb?
Chapter 3

Method

The purpose of this conducted study was to determine the validity of decisions concerning dimensionality of each of five methods for establishing criterion values in the context of PCA of residuals from a latent trait measurement model that were discussed by Hattie (1985). To this end, simulations were run to investigate how well the methods with proposed rules of thumb (Kaiser, Ratio of Eigenvalues, 20% Measure Variance, 40% Measure Variance, and 50% Measure Variance) performed in terms of Type I and Type II error rates as well as statistical power. For all methods, empirically-based critical values were derived from Monte Carlo procedures by identifying the empirical value of each index associated with the most extreme 5% of cases simulated to exhibit unidimensionality. In the sections that follow, simulation and data analytic methods associated with assessment of Type I errors and with assessment of Type II errors.

Type I error and Criteria Evaluation

To assess Type I error rates, unidimensional person and item response vectors were generated using the Conquest software (Wu, Adams, Wilson, & Haldane, 2007). Person ability estimates were sampled from a Normal distribution \( N(0, \sigma_1^2) \), while item difficulties were sampled from a Normal distribution \( N(0, \sigma_2^2) \) distribution. Item responses were simulated varying the number of items (20, 40, 60) and sample size (100, 250, 500, 750, 1000) between datasets. In addition, item and person parameter dispersion was varied by the following levels of distributional variance (0.0625, 0.25, 0.50, 0.75, 1, 1.25 1.5, 2.25). Hence, the independent variables were number of items (3 levels), sample size (5 levels), item parameter dispersion (8 levels), person parameter dispersion (8 levels). Each of the
960 cells of the experiment were replicated 1000 times. Parameters for each simulated data set were estimated using the Winsteps software (Linacre, 2008), and residuals from that model were output so that a PCA of the residuals could be conducted. Eigenvalues for the measures (i.e., the dominant dimension), and the first and second residual component were computed for each data set.

Existing rules of thumb were applied to determine whether each simulated data set exhibits a unidimensional or multidimensional structure. For the **eigenvalue magnitude** method, data sets were flagged for unidimensionality if the eigenvalue for the first residual dimension was less than 1.00 and was flagged for multidimensionality if that eigenvalue was greater than or equal to 1.00. For the **percent of variance** method, data sets were flagged for unidimensionality if more than 20, 40, or 50% (i.e., three rules were used) of the variance was accounted for by the dominant dimension and were flagged for multidimensionality if the dominant dimension accounted for 20, 40, or 50% (rule depending) or more of the total variance. In the **ratio of values** method, the person measure eigenvalue was divided by first residual dimension eigenvalue. When the ratio was found to be less than 4.00, the data set was flagged for multidimensionality, and it was flagged for unidimensionality when this ratio was greater than or equal to 4.00, as recommended in the Winsteps manual (Linacre, 2008).

There are three other PCA methods that do not have documented criterion values. The first is the **ratio of the differences** of subsequent component eigenvalues, proposed by Divgi (1980) which is found by subtracting the eigenvalue for the second component from the eigenvalue for the first component and then dividing that value by the difference between the second component eigenvalue after and the eigenvalue related to the third component. The other method, suggested by Kelley (1942), is the **ratio of eigenvalues to variance**, which is
simply the “measures” eigenvalue divided by the total variance related to the individual instrument questions. Lastly, the McGill method put forth by the author investigates the ratio of the eigenvalue for the first residual dimension and the sum of the item standard errors.

For all eight PCA methods, bootstrap critical values for declaring unidimensionality were computed. In the case of the three methods for which criterion values have not been established, these bootstrap critical values served as initial rules of thumb for applications of these PCA methods to LTTT model residuals. For the three methods for which rules of thumb have been established, the bootstrap critical values served as a value against which established rules of thumb were compared. Bootstrap critical values were determined by identifying either the 5th or 95th percentile of the relevant statistic across the 1000 replications within an experimental design cell under the unidimensional data generation condition. The reason for both percentile values is that dependent on the method whether it be a modified version of a known method or the rule for method which was not earlier known, the appropriate value must be used in the decision. Given a computed value must be larger than a critical value to signal unidimensionality the 5th percentile was used as the base for the rule, on the other hand if a computed value must be smaller than a critical value to signal unidimensionality then the 95th percentile was used.

For each iteration, the standard deviation for both the item and person distributions were computed. The average of these 1000 item and person standard deviations was used as an approximation to the true standard deviation under the other underlying conditions of sample size and test length. Stepwise Regression procedures in the form of Backwards Elimination was then used to predict these bootstrap critical values as a function of the
number of items, sample sizes, and item parameter dispersion. Care was taken to ensure that the rules of model hierarchy were upheld.

The empirically based equations were then subjected to a test for multicolinearity to investigate the relationship between the predictor variables. This was necessary since linear regression models, which exhibit high multicollinearity, have a tendency to be useful only for predicting scenarios from within the data set they were created (Neter, Kutner, Nachtshem, and Wasserman 1996). A predictor variable is deemed to be multicollinear if its Variance Inflation Factor (VIF) is much greater than 10. This of course is a subjective rule as there are no real cut off values to determine multicolinearity (Neter et al 1996).

The goal of these analyses is to determine whether more suitable critical values can be identified for the methods for which rules of thumb have been proposed and to identify suitable critical values for methods for which rules of thumb have not yet, prior to this study, been proposed.

For rule of thumb critical values and bootstrap critical values, Type I error rates were computed as the proportion of times that each method flags unidimensional data sets as being multidimensional (i.e., rejects the null hypothesis stating that the data is unidimensional).

**Type I error Cross-Validation**

To assess Type I error rates for all methods in order to have a comparison, unidimensional person and item response vectors were generated using the Conquest software (Wu, Adams, Wilson, & Haldane, 2007) using the same specifications upon which the bootstrap values were created. Each of the 960 cells of the experiment were replicated 200 times. Parameters for each simulated data set were estimated using the Winsteps software (Linacre, 2008), and residuals from that model were output in order that a PCA of
the residuals could be conducted. Eigenvalues for the measures (i.e., the dominant dimension), the first and second residual component were computed for each data set. Eleven total methods including methods with existing rules, modified versions, and methods without known rules were inspected relative to Type I error. The error rates were calculated from all methods performance is investigated under varying conditions as well as comparison between methods.

**Type II error**

To assess Type II error rates and statistical power, multidimensional data was simulated. The person and item response vectors were generated using the Conquest software (Wu, Adams, Wilson, & Haldane, 2007). Person ability estimates were sampled from a multivariate normal distribution MN(0,σ₁), while item difficulties were sampled from a Normal distribution N(0,σ²) distribution. The multivariate distributions were simulated to exhibit a between-dimension correlation (0.0, 0.25, 0.5, and 0.75). Item responses were simulated varying the number of items (20, 40, 60) and sample size (100, 250, 500, 750, 1000) between datasets. In addition, the item and person parameter dispersion utilized the following values of distributional variance (0.50, 1, and 1.5). Hence, the independent variables were number of items (3 levels), sample size (5 levels), item parameter dispersion (3 levels), person parameter dispersion (3 levels), and correlations between dimensions (4 levels). Each of the 540 cells of the experiment was replicated 1000 times.

Parameters for each simulated data set were estimated using the Winsteps software (Linacre, 2008), and residuals from that model were output so that a PCA of the residuals could be conducted. Eigenvalues for the measures (i.e., the dominant dimension) and the first and second residual component were computed for each data set. Rule of thumb and
bootstrap critical values (as described in the Type I error section) were applied to these multidimensional simulated data results. The Type II error rates were computed as the proportion of times out of the 1000 iterations that the methods fail to reject the hypothesis of unidimensionality, given that the data was truly multidimensional, and statistical power was computed as 1 – Type II error rate. Typically a “good” test is defined as having statistical power of 80% or greater.
Chapter 4

Results and Discussion

This particular chapter is presented in the order in which the simulations were conducted. The standard practice of a Results section followed by a Discussion section holds true. The difference is that Result/Discussion is not given as an overall but as a piecemeal structure. The first subset is the Result/Discussion relative to Type I error rates for the methods with know rules and the steps taken to find rules for methods which were not presently published. The second Result/Discussion subset is a validation of Type I error rates for methods with prior rules and those methods with newly created rules. This also provided a basis for comparison as all examined rules were examined in the same manor. The last subset of Result/Discussion investigated the Power for all methods to determine how well they signaled multidimensionality when it was actually present.

Part I - Type I error

The simulation results for each level combination of variables were recorded through 1000 iterations. This included three levels of test lengths, five levels of sample sizes, eight levels of item variability, and eight levels of person variability. When a method actually possessed a rule the Type I error rates were recorded. If a method was without a standard rule the critical values were kept in order to help determine a rule which will then be used to determine dimensionality. This was done in order to see how all the methods reacted under varying conditions of certain variables.

Kaiser Method

Figure 2 displays the Type I error rates for the three levels of sample size two levels of test length. From this graph, it is clear that as item variability and person variability
increase, the Type I error rate decreases. This pattern emerges regardless of test length or sample size. It is also clear that increases in test length increase the Type I error rate and that increases in sample size decrease Type I error rate, all other things held constant. The values upon which each graph is based are shown in Appendix Kaiser. Clearly, the accuracy of dimensionality depictions using the Kaiser criterion is influenced by test length, sample size, item variance, and person variance.
Figure 2. Kaiser Type I error rates for Person and Item Variance.

Ni = 20

Ni = 60

N = 100

N = 500
Figure 3 shows similar results for the Kaiser method, that as the sample size increases the Type I error decreases. This makes sense, due to the fact that more people provide better estimates and therefore reduce the error rate. However, it must be noted, though, that insufficient sample size could cause unduly high Type I error rates and therefore result in an erroneous dimension classification. An increase in the test length increases the error rate, unlike when one increases the sample size.
Figure 3. Kaiser Type I error rates for Person and Sample Size.

Ni = 20

Ni = 60

Item Variance = 0.0625

Item Variance = 1.25

Item Variance = 2.25
Clearly, error is affected by all independent variables such as item and person variability, sample size, and test length. On a subtler note, there seems to be interactions among all of these variables and their levels. For example is the error rates can be rather similar when there is a low sample size and high item variability as when there is a large sample size and low item variability.

The eigenvalues of the second dimension (i.e. the first residual dimension) were recorded for each of the 1,000 iterations. Ordering these 1,000 eigenvalues enables the identification of a bootstrapped confidence interval. In theory, the optimal Type I error would be 5%, regardless of the value of any variables. For every iteration, the standard deviation can be computed for a set of item difficulties and person abilities, which results in a set of 1,000 standard deviation values for persons and items. The average of the 1,000 standard deviation values were used as the best approximation for the item and person standard deviations in the new Kaiser rule which takes into account variable changes.

As noted, the Type I error rate in many cases exceeded the theoretical 0.05 standard. The change in the Type I error rate appears to be a direct result of the changes in levels of the independent variables. Thus, it is obvious that a rule comprised of a static value for determining unidimensionality under varying conditions is not optimal.

In order to improve Kaiser’s rule, a formula using the independent variables found to cause pattern changes in the error rates such as sample size, test length, item
variance, person variance, and all associated interactions was developed. The basis of the Kaiser rule is that an eigenvalue relative to the 2nd overall dimension, that is, the first residual dimension, less than 1 signals unidimensionality therefore a new version of the Kaiser rule will require the use of the 95th percentile value as the dependent variable in order to hold a 5% error rate. From this point on, the formula that takes into account changes in the levels of sample size, test length, item variance, person variance will be referred to as the Modified Kaiser rule.

Backwards Elimination was used on all independent and dependent variables to find the best model, which uses the 95th percentile value from the bootstrapped confidence interval as the dependent variable. The following variables and interactions were used: sample size, test length, item deviance, person deviance, test length x sample size, test length / sample size, test length x item deviance, sample size x person deviance, (test length x item deviance)/(sample size x person deviance), person deviance/item deviance, person deviance x item deviance. Table 1 below shows the R-Square values of the model upon removal the least significant independent. After removing the four variables the rest were deemed statistically significant and left in the model, as any further removal resulted in much larger drops in the model’s R-square value.

Table 1. Results of Stepwise Regression for Modified Kaiser Method.

<table>
<thead>
<tr>
<th>Round of Removal</th>
<th>Variables Removed</th>
<th>R-Square Value for the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>None</td>
<td>0.9651</td>
</tr>
<tr>
<td>Second</td>
<td>person deviance x item deviance</td>
<td>0.9649</td>
</tr>
<tr>
<td>Third</td>
<td>sample size x person deviance</td>
<td>0.9641</td>
</tr>
<tr>
<td>Forth</td>
<td>(test length x item deviance)/(sample size x person deviance)</td>
<td>0.9641</td>
</tr>
<tr>
<td>Fifth</td>
<td>Sample size x test length</td>
<td>0.9638</td>
</tr>
<tr>
<td>Sixth</td>
<td>Test length x item deviance</td>
<td>0.9469</td>
</tr>
</tbody>
</table>
The following is the optimal model for the Modified Kaiser value as was found using SAS and using the Backward Elimination method of Stepwise Regression:

\[
\text{ModifiedKaiser} = 1.9239 - 0.00015283(\text{samplesize}) + 0.00351(\text{testlength})
\]

\[
- 0.30889(\text{itemdeviation}) - 0.5832(\text{persondeviation}) + 1.57838(\text{testlength} / \text{samplesize}) - 0.00511(\text{testlength} \times \text{itemdeviation})
\]

\[
- 0.01721(\text{persondeviation} / \text{itemdeviation})
\]

(23)

where the Modified Kaiser value is that which a dimensional eigenvalue must be greater than in order to imply a useful dimension. In this formula, Item Deviation and Person Deviation are the sample standard deviations computed from the item and person distributions respectively for any new instrument in which the dimensionality will be tested. Table 2 provides some sample critical values based on varying levels of sample size, test length, person and item variance.

This model for the Modified Kaiser value has an R-squared value of 0.9638; the model explains 96.38% of the variability in finding the value of the 95th percentile under varying conditions. Thus, this model will be quite good at keeping a Type I error rate close to the desired 5% level. While this level of error will never be perfectly achieved, it is an improvement on the original Kaiser rule with respect to Type I error, which is dependent on certain variables.

Table 2. Sample Critical Values for the Modified Kaiser Method based on Independent Variables.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test Length</th>
<th>Person Variance</th>
<th>Item Variance</th>
<th>Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>20</td>
<td>0.0625</td>
<td>0.25</td>
<td>2.028711</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.0625</td>
<td>0.25</td>
<td>2.749363</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>1.221828</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>1.53368</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.0625</td>
<td>2.25</td>
<td>0.724631</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.0625</td>
<td>2.25</td>
<td>1.445283</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>0.040131</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>0.351983</td>
</tr>
</tbody>
</table>
As noted earlier the Modified Kaiser method fits the data very well. Below in Table 3 the Variance Inflation Factors (VIF) are recorded for all predictor variables.

Using the rule that a VIF must be much greater than 10 to signal multicollinearity, it is found that only one predictor may be a problem. The interaction term of Test Length and Item Deviance result in a VIF value of 11.3 which is not “much” greater than 10. Thus, the equation for the Modified Kaiser method displays minimal if any multicollinearity in its construction.

Table 3. Modified Kaiser Method VIF values for the parameters

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>2.48991</td>
</tr>
<tr>
<td>Test Length</td>
<td>6.05048</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>7.94397</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>1.10773</td>
</tr>
<tr>
<td>Test Length / Sample Size</td>
<td>2.90192</td>
</tr>
<tr>
<td>Test Length X Item Deviance</td>
<td>11.30261</td>
</tr>
<tr>
<td>Person Deviance/ Item Deviance</td>
<td>1.96822</td>
</tr>
</tbody>
</table>

**Percentage of Variance Explained by Measures**

Three different rules for the percentage of variance explained by measures exist in determining unidimensionality. They are 20%, 40%, and 50% of the variability should be accounted for by the Measure dimension while the rest is attributed to the residuals.

There is little need to show the results for the 40% and 50% cases, as their patterns are
similar to the 20% case, but with the lower error rates, so the error rates for the 40% and 50% rule converged to zero more quickly, with the 50% converging the fastest.

The graph of sample size 100 and 20 questions in Figure 4 shows an interesting trend in the Type I error rate associated with the 20% Measure variability rule. When the item variance is set to 0.0625, 0.25, or 0.5, the error rate decreases as the item variance increases. For item variances greater than 0.5, if the person variance is increased, the error rates begin to increase. Therefore, throughout Figure 4, there is an apparent inconsistency. The condition is most likely parabolic in nature, as an interaction of both item and person variabilities occurs where an error rate will decrease down to zero and then start increasing again. This parabolic pattern can be seen best in the graph where sample size 500 and 60 questions in Figure 3 when the item variance is 0.5.

Figure 4. Percentage of Measure Variance Type I error rates for Person and Item Variance.

Ni = 20

Ni = 60

N = 100
Similar non-simplistic trends, such as parabolic, are evidenced for the effect of both sample size and test length on error rates. This is shown in Graphs where the test length is 20 and 60 with an Item variance of 1.25 in Figure 5, an increase in sample size will either increase or decrease the error depending on person variance, sample size, and test length; their true response is not obvious without accounting for the interaction of the two other variables.
As noted the Type I error rate in many cases exceeded the 5% standard. Clearly, a rule using of a single, static value for the percentage of measure variability to determine unidimensionality under varying conditions is suboptimal.
The measure variability percentages were recorded for each of the 1,000 iterations of each simulation; these values can be ordered to find a bootstrapped confidence interval. The probability of a Type I error should be 0.05, regardless of the value of any variables. For every single iteration, the standard deviation can be computed for that set of item difficulties and person abilities. This results in a set of 1,000 standard deviation values for person and items. The average of these 1,000 standard deviation values was used as the best estimate for the true standard deviation in the model.

In order to improve the rule for the percentage of variability relative to the measure dimension, a new rule based on regression analysis was developed, using the variables that affect Type I error rates, including sample size, test length, item variance, person variance, and all associated interactions as independent variables. The rationale behind the percentage of measure dimension variability is based on the notion that the main dimension should explain most of the variability. Clearly the 50% rule easily defines most as opposed to either the 20 or 40% version of the methods.

Unidimensionality is signaled by a percentage of variability attributed to the measure dimension greater than either 20, 40, or 50% depending on the rule it will be required that the 5th percentile value be kept from every 1,000 iterations. That is, 5% of the minimum percentage values will erroneously signal a case of multidimensionality. This in turn will provide a minimum percentage of measure variability, such that any percentage greater will denote a unidimensional case and a lower percentage will signal multidimensionality with a desired 0.05 probability of Type I error.

Backwards Elimination was used on all independent and dependent variables to find the best model, which uses the 5th percentile value from the bootstrapped confidence
interval as the dependent variable. The following variables and interactions were used: sample size, test length, item deviance, person deviance, test length x sample size, test length / sample size, test length x item deviance, sample size x person deviance, (test length x item deviance)/(sample size x person deviance), person deviance/item deviance, person deviance x item deviance, item deviance^2, person deviance^2, test length x item deviance^2, sample size x person deviance^2. Table 4 below shows the R-Square values of the model upon removal the least significant independent. After removing the five variables the rest were deemed statistically significant and left in the model, as any further removal resulted in much larger drops in the model’s R-square value.

<table>
<thead>
<tr>
<th>Round of Removal</th>
<th>Variables Removed</th>
<th>R-Square Value for the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>None</td>
<td>0.5374</td>
</tr>
<tr>
<td>Second</td>
<td>Sample size x test length deviance</td>
<td>0.5373</td>
</tr>
<tr>
<td>Third</td>
<td>test length x item deviance^2</td>
<td>0.5372</td>
</tr>
<tr>
<td>Forth</td>
<td>sample size x person deviance^2</td>
<td>0.5368</td>
</tr>
<tr>
<td>Fifth</td>
<td>sample size x person deviance</td>
<td>0.5367</td>
</tr>
<tr>
<td>Sixth</td>
<td>person deviance/item deviance</td>
<td>0.5359</td>
</tr>
<tr>
<td>Seventh</td>
<td>test length / sample size</td>
<td>0.5287</td>
</tr>
</tbody>
</table>

The following is the optimal model for the Modified Percentage of Variance value as was found using SAS and using the Backward Elimination method of Stepwise Regression:

\[
Modified\Percentage = -0.64407 + 0.00003092\text{(samplesize)} + 0.00316\text{(testlength)} + 1.13291\text{(ItemDeviance)} + 0.86982\text{(PersonDeviance)} - 0.20508\text{(Testlength/Samplesize)} - 0.00654\text{(TestLength*ItemDeviance)} + 0.13628\text{(Testlength*Itemdeviance)/(Samplesize*Persondeviance)} - 0.29386\text{(ItemDeviance^2)} - 0.1784\text{(PersonDeviance^2)} - 0.43313\text{(ItemDeviance*PersonDeviance)}
\]  

(24)
where the Modified Percentage value is the minimum value of measure dimension percentage of variability that indicates unidimensionality. Table 5 provides some sample critical values based on varying levels of sample size, test length, person and item variance.

From this point on, the formula that takes into account variable changes will be referred to as the Modified Percentage Rule. Using the percentile value from the ordered simulation results as the dependent variable, stepwise regression found the following linear model to be optimal, using SAS: Stepwise regression was used on all independent and dependent variables to find the best model.

This model for the Modified percentage value has an $R^2$ value of 0.5359, explaining 53.59% of the variability in finding the value of the 5th percentile under varying conditions. Thus, this model will not be very effective at keeping a Type I error rate close to the desired 0.05 level. While this level of error will be less than adequate, it may be better than the original 20, 40, or 50% variability rules with respect to Type I error, which is dependent on certain variables.

Table 4. Sample Critical Values for the Modified Proportion of Variability Method based on Independent Variables.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test Length</th>
<th>Person Variance</th>
<th>Item Variance</th>
<th>Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>0.180142</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>0.146662</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>0.356261</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>0.302637</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>0.426705</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>0.41173</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>0.963232</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>1.452952</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>0.13993</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>0.081682</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>0.200229</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>0.221113</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>0.488722</td>
</tr>
</tbody>
</table>
As noted earlier the Modified Percentage method does not fit the data well. Below in Table 6 the Variance Inflation Factors (VIF) are recorded for all predictor variables. Using the rule that a VIF must be much greater than 10 to signal multicollinearity, it is found that there seven predictors may be a problem. Thus, it seems likely that Modified Percentage Method may suffer from multicollinearity. This is a moot point as the lack of fit evident from the R-Square value makes this method unlikely to determine a good predicted cutoff value with respect to independent variables.

Table 6. Proportion of Variability Method VIF values for the parameters of full model

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>7.03484</td>
</tr>
<tr>
<td>Test Length</td>
<td>11.44593</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>32.71168</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>36.89930</td>
</tr>
<tr>
<td>Test Length X Sample Size</td>
<td>12.65236</td>
</tr>
<tr>
<td>Test Length / Sample Size</td>
<td>6.11572</td>
</tr>
<tr>
<td>Test Length X Item Deviance</td>
<td>12.29510</td>
</tr>
<tr>
<td>(Test Length x Item Deviance)/(Sample Size x Person Deviance)</td>
<td>4.33797</td>
</tr>
<tr>
<td>Item Deviance^2</td>
<td>15.77823</td>
</tr>
<tr>
<td>Person Deviance^2</td>
<td>33.15191</td>
</tr>
<tr>
<td>Person Deviance X Item Deviance</td>
<td>16.65203</td>
</tr>
</tbody>
</table>

Table 7 depicts the parameters for the most simple model while Table 8 shows a slightly more complex model interaction terms. The increase in independent variables clearly increased the R-Square value implying better fit. Also of note is the intercept term, it was only positive in the most simplistic case. As the number of variables increased, and fit of the model increased, the intercept became more negative. This was due to the original data set of percentages that are very close to zero. Therefore to fit the model well a negative intercept was needed to better approximate the points. Though it is
counter-intuitive to have a negative intercept which implies a negative percentage at the
lowest end of the spectrum relative to the percentage of variance. Thus, a likely
conclusion is that a linear model approach may not have been an appropriate method.

Table 7. Proportion of Variability Method VIF Values for the Parameters of Simple
Model with R-Square = 0.1702

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter Estimate</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>0.33464</td>
<td>0</td>
</tr>
<tr>
<td>Sample Size</td>
<td>0.00004839</td>
<td>1.00128</td>
</tr>
<tr>
<td>Test Length</td>
<td>-0.00289</td>
<td>1.02874</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>-0.0684</td>
<td>1.00238</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>0.07620</td>
<td>1.02893</td>
</tr>
</tbody>
</table>

Table 8. Proportion of Variability Method VIF Values for the Parameters of Model with
Interactions with R-Square = 0.4021

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter Estimate</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.30653</td>
<td>0</td>
</tr>
<tr>
<td>Sample Size</td>
<td>-0.00000753</td>
<td>17.57431</td>
</tr>
<tr>
<td>Test Length</td>
<td>0.00207</td>
<td>8.62410</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>0.66898</td>
<td>18.43678</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>0.50101</td>
<td>9.12973</td>
</tr>
<tr>
<td>Sample Size X Test Length</td>
<td>0.000000354807</td>
<td>9.82328</td>
</tr>
<tr>
<td>Test Length X Item Deviance</td>
<td>-0.00570</td>
<td>11.62770</td>
</tr>
<tr>
<td>Sample Size X Person Deviance</td>
<td>0.00003925</td>
<td>11.63708</td>
</tr>
<tr>
<td>Item Deviance X Person Deviance</td>
<td>-0.48688</td>
<td>15.41442</td>
</tr>
</tbody>
</table>

**Ratio of Eigenvalues**

When using the ratio of eigenvalues to identify unidimensionality, the magnitude of the
Measure dimension must be at least four times greater than the first residual dimension.

Figure 6 indicates that, when variables such as sample size, test length, and person
variance are held constant while simultaneously increasing item variance, the Type I
error first decreases and then increases. Similarly, increasing person variance while keeping everything else constant results in a similar parabolic trend in the Type I error rate.

Figure 6. Ratio of Eigenvalues Type I error rates for Person and Item Variance.
The sample size and test length also have a parabolic effect on error rates, with increasing then decreasing error rates dependent on situation. This is rather clear in Figure 7 where a constant test length and item variance while letting the sample size and person variance vary. The pattern for all the associated variables will be related to individual variable effects, along with variable interactions. Of note is that the Type I error trend that starts at 100% and decrease to 0% while the item variance increases will actually occur faster as the test length is increased. In other words, a set range of errors relative to item variability will decrease as the number of questions increases. The graph in Figure 6 where item variance is 2.25 and test length is 60 is especially interesting because it gives the notion of an almost cubic relationship of variables once the item variance becomes large enough.
Figure 7. Ratio of Eigenvalues Type I error rates for Person and Sample Size.

$N_i = 20$

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>Person Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0625</td>
<td>0</td>
</tr>
<tr>
<td>1.25</td>
<td>2</td>
</tr>
<tr>
<td>2.25</td>
<td>5</td>
</tr>
</tbody>
</table>

$N_i = 60$

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>Person Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0625</td>
<td>0</td>
</tr>
<tr>
<td>1.25</td>
<td>2</td>
</tr>
<tr>
<td>2.25</td>
<td>5</td>
</tr>
</tbody>
</table>
Clearly, error is affected by the item and person variability, sample size, and test length individually. On a subtler note, there seems to be interactions among all of these variables and their levels as noted by the shift in the patterns relative to levels.

For every simulation the eigenvalues for the first and second dimension, (i.e. the measure and the first residual dimension), were recorded for each of the 1,000 iterations. The ratio value was created by dividing the first eigenvalue by the second eigenvalue. By first ordering the resulting 1,000 ratios, one can isolate a bootstrapped confidence interval. Theoretically, the optimal Type I error rate should be 5%, regardless of the value of any variables.

For every iteration, the standard deviation can be computed for the set of item difficulties and person abilities, resulting in a set of 1,000 standard deviations for persons and items. The average of these 1,000 standard deviation values were used as the best approximation for the standard deviation in the model.

As noted, the Type I error rate in many cases exceeded the 5% standard, especially when either the item, person, or both variabilities were small. The Type I error improved as the variances increased, up to a point. Thus, a rule comprised of a static value for determining unidimensionality under varying conditions is less than optimal.

In order to improve the rule for the ratio of eigenvalues, a formula will be made using the variables found to cause pattern changes such as sample size, test length, item variance, person variance, and all associated interactions as independent variables. The basis of the rule for the ratio of eigenvalues is that, in the unidimensional case, the eigenvalue of the first dimension should be greater than the second eigenvalue, such that
the variability associated with the first dimension is at least four times greater than the variability found in the second dimension.

Therefore, a modification of the ratio of the eigenvalue rule, which would keep a constant 5% error rate, requires the use of the 5th percentile of the bootstrapped eigenvalues. This percentile is used since all of the recorded ratios are relative to unidimensional cases where the first eigenvalue should be of a certain magnitude greater than the second eigenvalue. Therefore the lowest 5% of the values, the 5th percentile, should be the set of false positives where multidimensionality would be chosen. Using the percentile value as the dependent variable stepwise regression found the following model to be optimal. From this point on the formula that takes into account variable changes will be known as the Modified Ratio rule.

Backwards Elimination was used on all independent and dependent variables to find the best model, which uses the 5th percentile value from the bootstrapped confidence interval as the dependent variable. The following variables and interactions were used: sample size, test length, item deviance, person deviance, test length x sample size, test length / sample size, test length x item deviance, sample size x person deviance, (test length x item deviance)/(sample size x person deviance), person deviance/item deviance, person deviance x item deviance, item deviance^2, person deviance^2, test length x item deviance^2, sample size x person deviance^2. Table 9 below shows the R-Square values of the model upon removal the least significant independent. After removing the four variables the rest were deemed statistically significant and left in the model, as any further removal resulted in much larger drops in the model’s R-square value.
Table 9. Results of Stepwise Regression for Modified Ratio Method.

<table>
<thead>
<tr>
<th>Round of Removal</th>
<th>Variables Removed</th>
<th>R-Square Value for the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>None</td>
<td>0.4786</td>
</tr>
<tr>
<td>Second</td>
<td>person deviance/item deviance</td>
<td>0.4786</td>
</tr>
<tr>
<td>Third</td>
<td>item deviance^2</td>
<td>0.4785</td>
</tr>
<tr>
<td>Forth</td>
<td>person deviance^2</td>
<td>0.4780</td>
</tr>
<tr>
<td>Fifth</td>
<td>test length x item deviance</td>
<td>0.4746</td>
</tr>
</tbody>
</table>

The following is the optimal model for the Modified Percentage of Variance value as was found using SAS and using the Backward Elimination method of Stepwise Regression:

\[
ModifiedRatio = -26.98522 + 0.01413(Samplesize) + 0.27262(TestLength)
+ 33.31281(ItemDeviance) + 20.29164(PersonDeviance)
+ 0.00009305(TestLength * SampleSize) - 8.93989(TestLength / SampleSize)
+ 0.0244(SampleSize * PersonDeviance) - 19.11722(ItemDeviance * PersonDeviance)
- 0.20405(TestLength * ItemDeviance^2) - 0.00868(Samplesize * PersonDeviance^2)
\]  

where the Modified Ratio value is minimum value that would indicate a useful second dimension. Table 10 provides some sample critical values based on varying levels of sample size, test length, person and item variance.

This model for the Modified Ratio of Eigenvalues value has an R-squared value of 0.4746, explaining 47.46% of the variability in finding the value of the 5\(^{th}\) percentile under varying conditions. Thus, this model would not be very effective at keeping a Type I error rate close to the desired 5% level. While this level of error is undesirable, it may yet be an improvement over the eigenvalues rule with respect to Type I error which was shown to be dependent on certain variables.
Table 10. Sample Critical Values for the Modified Ratio of Eigenvalues Method based on Independent Variables.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test Length</th>
<th>Person Variance</th>
<th>Item Variance</th>
<th>Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>9.214724</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>2.023805</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>27.44729</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>6.171795</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>22.34995</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>29.54086</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>17.45692</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>51.07601</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>11.78811</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>25.54719</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>48.45012</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>21.3992</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>48.21278</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>61.97186</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>8.405906</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>18.64501</td>
</tr>
</tbody>
</table>

As noted earlier the Ratio of Eigenvalue method does not fit the data well reflected in the R-Square value of 0.4746. Below in Table 11 the Variance Inflation Factors (VIF) are recorded for all predictor variables. Using the rule that a VIF must be much greater than 10 to signal multicollinearity, it is found that all but two predictors may be a problem, with an obvious case of multicollinearity relative to the Sample Size X Person Deviance interaction. Thus, it seems likely that the Ratio of Eigenvalues Method may suffer from multicollinearity. This is irrelevant as the lack of fit evident from the R-Square value makes this method unlikely to determine a good predicted cutoff value with respect to the independent variables.

Table 11. Modified Ratio of Eigenvalues Method VIF Values for the Parameters of Full Model

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>29.49982</td>
</tr>
<tr>
<td>Test Length</td>
<td>7.78936</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>15.25124</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>9.04221</td>
</tr>
<tr>
<td>Test Length X Sample Size</td>
<td>12.92582</td>
</tr>
<tr>
<td>Test Length / Sample Size</td>
<td>3.838607</td>
</tr>
</tbody>
</table>
Table depicts the parameters for the most simple model while Table shows a slightly more complex model interaction terms. The increase in independent variables clearly increased the R-Square value implying better fit. Also of note is the intercept term, was only close to being positive in the most simplistic case. As the number of variables increased, and fit of the model increased, the intercept became more negative. This was due to the original data set of values that approach zero. Therefore to fit the model well a negative intercept was needed to better approximate the points. Though it is counter-intuitive to have a negative Ratio of Eigenvalues value since it is created by the sum of item variances divided by the eigenvalue of the measure dimension. Thus, a likely conclusion is that a linear model approach may not have been an appropriate method.

Table 12. Modified Ratio of Eigenvalues Method VIF Values for the Parameters of the Simple Model with R-Square = 0.1943

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter Estimate</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.03392</td>
<td>0</td>
</tr>
<tr>
<td>Sample Size</td>
<td>0.00740</td>
<td>1.00128</td>
</tr>
<tr>
<td>Test Length</td>
<td>0.08555</td>
<td>1.02878</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>-0.73901</td>
<td>1.00243</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>6.12640</td>
<td>1.02893</td>
</tr>
</tbody>
</table>

Table 13. Modified Ratio of Eigenvalues Method VIF Values for the Parameters of the Simple Model with Interactions with R-Square = 0.3956

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter Estimate</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-25.81387</td>
<td>0</td>
</tr>
<tr>
<td>Sample Size</td>
<td>-0.00589</td>
<td>17.57436</td>
</tr>
<tr>
<td>Test Length</td>
<td>0.27649</td>
<td>8.63783</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>31.91497</td>
<td>18.43671</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>20.70664</td>
<td>9.12981</td>
</tr>
<tr>
<td>Sample Size X Test Length</td>
<td>0.00016775</td>
<td>9.82805</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>-----</td>
<td>-------</td>
</tr>
<tr>
<td>Test Length X Item Deviance</td>
<td>-0.30719</td>
<td>11.61717</td>
</tr>
<tr>
<td>Sample Size X Person Deviance</td>
<td>0.00622</td>
<td>11.63708</td>
</tr>
<tr>
<td>Item Deviance X Person Deviance</td>
<td>-19.50689</td>
<td>15.42322</td>
</tr>
</tbody>
</table>

**Kelley Method**

The pattern of the graph in Figure 8 for the 5th percentiles of the Kelley method when the sample size is 100 and the test length is 20 shows that when person variance is held constant at 0.0625 while the item variance is increased, the 5th percentile of the 1,000 Kelley values decreases. When the item variances are lower than 2.25 the error rates increase as the person variance increase up until the person variance reaches about 1 to 1.25. After that point the errors seem to decrease with increasing person variance as in most of the graphs of Figure 8. Yet in graphs where the test length is increased to 60 questions, after reaching a person variance around 1 or 1.25 the error rates appear to oscillate.

In Figure 8 where the graphs are depicting sample size of 20 questions the general trend is that error increase with an increase in person variance and at times then decrease. This parabolic pattern is evident in the aforementioned graphs though the actual values for the error rates shift slightly with changes in sample sizes and test length. Also, it seems that error rates are the greatest when the item variability is at its lowest.

When comparing increases in sample size it is interesting to note that the patterns are similar given the same test length, but that the percentile values are increase with increasing sample sizes. It is also interesting to observe that, start to decrease it is an abrupt drop to close to zero and not a gradual progression. Due to the lack of more levels
of variability in the current study, a more in-depth look at the pattern in error rates after the abrupt drop is not possible at this time.

Figure 8. Kelley method 5th Percentiles for Person and Item Variance.

Ni = 20

Ni = 60

N = 100

N = 500
When the variability of the item difficulties is kept constant, the patterns relative to test length looks clear, as in the graph of item variance 0.0625 and test length 20 found in Figure 8 where error rate increases with person variability regardless of the sample. This pattern changes after either or both test length and item variance are increased as in all graphs other than the one previously mentioned in Figure 9. This new pattern shows that percentile values will increase when person variability is low and then begins to decrease at a certain point dependent upon an interaction of the test length and the variability of the item difficulties. Thus, in this case the 5th percentile values are directly proportional to test length when the person variance is small, but this relationship changes to inverse proportionality when the person variability becomes large enough. The simplest result from the graphs in Figure 9 is that sample size is directly related to the increase in the 5th percentile.
Figure 9. Kelley method $5^{th}$ Percentiles for Person and Sample Size.

- **Ni = 20**
  - Item Variance $= 0.0625$
  - Item Variance $= 1.25$
  - Item Variance $= 2.25$

- **Ni = 60**
  - Item Variance $= 0.0625$
  - Item Variance $= 1.25$
  - Item Variance $= 2.25$

The Kelley method does not have any sort of rule for a dimensionality choice. Therefore, there was no ability to show performance of the method under Type I error.
conditions. Regardless, noting the patterns of the 5\textsuperscript{th} percentiles found under changing levels of variables, a rule using Kelley’s method for identifying unidimensionality would not work well using a static value.

In order to develop a rule for the Kelley method, a formula will be developed that uses the variables found to cause pattern changes, such as sample size, test length, item variance, person variance, and all associated interactions as independent variables. The Kelley value is found by dividing the sum of all item standard errors, as determined by the Rasch method, by the eigenvalue of the measure dimension. One theory is that if all questions belong to one dimension then the standard errors should be at a minimum.

Thus, minimal standard errors added together will create a larger overall value when used as the denominator, as opposed to the situation where the standard errors are wider ranging for multiple dimensions and the sum used for a denominator would result in an overall value that is smaller. Thus, the 5\textsuperscript{th} percentile of Kelley values for the 1,000 iterations are used for the cutoff where the remaining lower 5\% of values may be mistaken for the multidimensional case.

Backwards Elimination was used on all independent and dependent variables to find the best model, which uses the 5\textsuperscript{th} percentile value from the bootstrapped confidence interval as the dependent variable. The following variables and interactions were used: sample size, test length, item deviance, person deviance, test length x sample size, test length / sample size, test length x item deviance, sample size x person deviance, (test length x item deviance)/(sample size x person deviance), person deviance/item deviance, person deviance x item deviance, item deviance\textsuperscript{2}, person deviance\textsuperscript{2}, test length x item deviance\textsuperscript{2}, sample size x person deviance\textsuperscript{2}. Table 14 below shows the R-Square values of
the model upon removal the least significant independent. After removing the three
variables the rest were deemed statistically significant and left in the model, as any
further removal resulted in much larger drops in the model’s R-square value.

Table 14. Results of Stepwise Regression for Kelley Method.

<table>
<thead>
<tr>
<th>Round of Removal</th>
<th>Variables Removed</th>
<th>R-Square Value for the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>None</td>
<td>0.6387</td>
</tr>
<tr>
<td>Second</td>
<td>Test length x item deviance²</td>
<td>0.6387</td>
</tr>
<tr>
<td>Third</td>
<td>person deviance²</td>
<td>0.6383</td>
</tr>
<tr>
<td>Forth</td>
<td>person deviance/ item deviance</td>
<td>0.6368</td>
</tr>
<tr>
<td>Fifth</td>
<td>test length/sample size</td>
<td>0.6296</td>
</tr>
</tbody>
</table>

The following is the optimal model for the Modified Percentage of Variance value
as was found using SAS and using the Backward Elimination method of Stepwise
Regression:

\[
Kelley = -5.79128 - 0.00001822(SampleSize) + 0.04882(Testlength) \\
+ 10.99199(ItemDeviance) + 4.0468(PersonDeviance) \\
- 0.00003246(SampleSize \times Testlength) - 1.29116(Testlength / SampleSize) \\
- 0.05729(Testlength \times ItemDeviance) + 0.00717(SampleSize \times PersonDeviance) \\
- 2.76945(ItemDeviance²) - 4.30756(ItemDeviance \times PersonDeviance) \\
- 0.00292(SampleSize \times PersonDeviance²)
\]

where the Kelley value is a minimum value that would indicate a useful second
dimension. Table 15 provides some sample critical values based on varying levels of
sample size, test length, person and item variance.

This model for the Kelley value has an R-squared value of 0.6368, explaining
63.68% of the variability in finding the value of the 5th percentile of the Kelley values for
all 1000 iterations under varying conditions. Thus, this model would not be very effective
at keeping a Type I error rate close to the desired 5% level. Thus, this model would be
mildly adequate in determining unidimensionality. Therefore, the prospect of obtaining
the desired 0.05 probability of Type I error using this model is unlikely, due to the poor level of model fit with respect to changing variables.

Table 15. Sample Critical Values for the Kelley Method based on Independent Variables.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test Length</th>
<th>Person Variance</th>
<th>Item Variance</th>
<th>Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>1.94792</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>1.214324</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>1.74343</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>2.106174</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>3.9659</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>4.699496</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>9.57299</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>13.42259</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>0.867189</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>0.837335</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>2.824161</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>1.729185</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>4.812631</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>4.842485</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>8.726259</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>13.27961</td>
</tr>
</tbody>
</table>

As noted earlier the Kelley method does not fit the data well reflected in the R-Square value of 0.6368. Below in Table 16 the Variance Inflation Factors (VIF) are recorded for all predictor variables. Using the rule that a VIF must be much greater than 10 to signal multicollinearity, it is found that all but two predictors may be a problem, with a obvious case of multicollinearity relative to the Sample Size X Person Deviance interaction. Thus, it seems likely that Kelley Method may suffer from multicollinearity. This is irrelavent as the lack of fit evident from the R-Square value makes this method unlikely to determine a good predicted cutoff value with respect to the independent variables.
Table 16. Kelley Method VIF Values for the Parameters of Full Model

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>29.56095</td>
</tr>
<tr>
<td>Test Length</td>
<td>11.31965</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>32.52135</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>9.15849</td>
</tr>
<tr>
<td>Test Length X Sample Size</td>
<td>12.93020</td>
</tr>
<tr>
<td>Test Length / Sample Size</td>
<td>3.83615</td>
</tr>
<tr>
<td>Test Length X Item Deviance</td>
<td>11.63505</td>
</tr>
<tr>
<td>Sample Size X Person Deviance</td>
<td>104.02147</td>
</tr>
<tr>
<td>Item Deviance²</td>
<td>15.61825</td>
</tr>
<tr>
<td>Person Deviance X Item Deviance²</td>
<td>15.49755</td>
</tr>
<tr>
<td>Sample Size X Person Deviance²</td>
<td>50.32327</td>
</tr>
</tbody>
</table>

Table 17 depicts the parameters for the most simple model while Table 18 shows a slightly more complex model interaction terms. The increase in independent variables clearly increased the R-Square value implying better fit. Also of note is the intercept term, it was only positive in the most simplistic case. As the number of variables increased, and fit of the model increased, the intercept became more negative. This was due to the original data set of values that approach zero. Therefore to fit the model well a negative intercept was needed to better approximate the points. Though it is counter-intuitive to have a negative Kelley value since it is created by the sum of item variances divided by the eigenvalue of the measure dimension. Thus, a likely conclusion is that a linear model approach may not have been an appropriate method.

Table 17. Kelley Method VIF Values for the Parameters of Simple Model with R-Square = 0.3688

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter Estimate</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.80299</td>
<td>0</td>
</tr>
<tr>
<td>Sample Size</td>
<td>0.00299</td>
<td>1.00128</td>
</tr>
<tr>
<td>Test Length</td>
<td>-0.02374</td>
<td>1.02878</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>-0.71954</td>
<td>1.00243</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>0.665517</td>
<td>1.02893</td>
</tr>
</tbody>
</table>
Table 18. Kelley Method VIF Values for the Parameters of Simple Model with Interactions with R-Square = 0.4021

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter Estimate</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-4.40664</td>
<td>0</td>
</tr>
<tr>
<td>Sample Size</td>
<td>0.00275</td>
<td>17.57436</td>
</tr>
<tr>
<td>Test Length</td>
<td>0.03688</td>
<td>8.63783</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>6.24463</td>
<td>18.43671</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>4.29988</td>
<td>9.12981</td>
</tr>
<tr>
<td>Sample Size X Test Length</td>
<td>-0.00002175</td>
<td>9.82805</td>
</tr>
<tr>
<td>Test Length X Item Deviance</td>
<td>-0.05462</td>
<td>11.61717</td>
</tr>
<tr>
<td>Sample Size X Person Deviance</td>
<td>0.00105</td>
<td>11.63708</td>
</tr>
<tr>
<td>Item Deviance X Person Deviance</td>
<td>-4.56796</td>
<td>15.42322</td>
</tr>
</tbody>
</table>

**Divgi Method**

The Divgi method, like the Kelley method, also lacks a single, set value for determine dimensionality. The Divgi value is the difference between the first and second eigenvalue divided by the difference in the third and forth eigenvalue. The rationale behind the value is that larger values indicate unidimensionality, since the first (measure)dimension should contain most of the loadings; the remaining dimensions should have minimal loadings if the instrument is truly unidimensional. A small denominator would yield a large overall Divgi value. Without a standard rule, actual values are used from the 5th percentiles of the Divgi values for all 1,000 iterations since small values should indicate a multidimensional situation.

In Figure 10 the graphs where the test length is 20 and the changes in item and person variance are shown, generally the value of the percentiles increase as the item or person variability increases. When the test is made longer, in this case 60 questions, the effect on the value of the 5th percentile is parabolic in relation to increasing person
variability as can been seen in Figure 9 graphs where the sample size is increased to 60, yet the pattern remains similar.

Figure 10. Divgi method 5th Percentiles for Person and Item Variance.

Ni = 20

Ni = 60

N = 100

N = 500

N = 500
The graph in Figure 11 with item variance of 0.0625 and test length of 20 indicates a clear increase in percentiles as the sample size increases when the item variance is 0.0625 and the instrument has 20 questions. In the graph of Figure 11 when the test length is increased to 60 and item variance is increased to 0.0625 the trend changes from just increasing to a parabolic relationship based on person variability.

The other graphs in Figure 11 with 60 question instruments with the only difference being that they are holding item variability constant at 1.25 and 2.25 show similar patterns. In both cases the percentile values peak and then abruptly drop to minimal values with the only difference being a modal shift due to an interaction of item and person variance. Overall the larger the sample size creates increased values of the 5\textsuperscript{th} percentile.
Figure 11. Divgi method 5th Percentiles for Person and Sample Size.

Ni = 20

Item Variance = 0.0625

Ni = 60

Item Variance = 1.25

Item Variance = 2.25
Since the Divgi method does not have a hard and fast rule for indicating
dimensionality, there was no way to show performance of the method under Type I error
conditions. Regardless, noting the patterns of the 5th percentiles found under changing
levels of variables, a standard rule using a static value under Divgi’s method would not
work well.

In order to develop a rule for the Divgi method a formula was developed using the
variables found to cause pattern changes such as sample size, test length, item variance,
person variance, and all associated interactions as independent variables. The 5th
percentile of Divgi values for the 1000 iterations are used for the cutoff where the
remaining lower 5% of values may be mistaken for the multidimensional case.

Backwards Elimination was used on all independent and dependent variables to
find the best model, which uses the 5th percentile value from the bootstrapped confidence
interval as the dependent variable. The following variables and interactions were used:
sample size, test length, item deviance, person deviance, test length x sample size, test
length / sample size, test length x item deviance, sample size x person deviance, (test
length x item deviance)/(sample size x person deviance), person deviance/item deviance,
person deviance x item deviance, item deviance^2, person deviance^2, test length x item
deviance^2, sample size x person deviance^2. Table 19 below shows the R-Square values of
the model upon removal the least significant independent. After removing the five
variables the rest were deemed statistically significant and left in the model, as any
further removal resulted in much larger drops in the model’s R-square value.
Table 19. Results of Stepwise Regression for Divgi Method.

<table>
<thead>
<tr>
<th>Round of Removal</th>
<th>Variables Removed</th>
<th>R-Square Value for the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>None</td>
<td>0.5406</td>
</tr>
<tr>
<td>Second</td>
<td>person deviance^2</td>
<td>0.5406</td>
</tr>
<tr>
<td>Third</td>
<td>person deviance/ item deviance</td>
<td>0.5403</td>
</tr>
<tr>
<td>Forth</td>
<td>Test length x item deviance</td>
<td>0.5387</td>
</tr>
<tr>
<td>Fifth</td>
<td>item deviance^2</td>
<td>0.5378</td>
</tr>
<tr>
<td>Sixth</td>
<td>Test length/sample size</td>
<td>0.5341</td>
</tr>
<tr>
<td>Seventh</td>
<td>Test length x sample size</td>
<td>0.5174</td>
</tr>
</tbody>
</table>

The following is the optimal model for the Modified Percentage of Variance value as was found using SAS and using the Backward Elimination method of Stepwise Regression:

\[
\text{Divgi} = -633.92193 - 0.32016(\text{SampleSize}) + 4.42618(\text{TestLength}) \\
+ 716.95292(\text{ItemDeviance}) + 378.53878(\text{PersonDeviance}) \\
+ 0.00485(\text{TestLength} \times \text{SampleSize}) + 0.65714(\text{SampleSize} \times \text{PersonDeviance}) \\
- 391.00613(\text{ItemDeviance} \times \text{PersonDeviance}) - 4.31946(\text{SampleSize} \times \text{ItemDeviance}^2) \\
- 0.216966(\text{TestLength} \times \text{PersonDeviance}^2)
\]

(27)

where the Divgi value is the value such the difference between the measure dimension eigenvalue and first residual dimensional eigenvalue is divided by the difference of the second and third residual dimension eigenvalues must be greater than in order to imply unidimensionality. Table 20 provides some sample critical values based on varying levels of sample size, test length, person and item variance.

This model for the Divgi value has an R-square value of 0.5341, explaining 53.41% of the variability in finding the value of the 5th percentile under varying conditions. Thus, this model would not be very effective at keeping a Type I error rate close to the desired 5% level; the model would be only mildly useful determining dimensionality. Therefore, the prospect of obtaining the desired 0.05 probability of Type I error using this model is unlikely, due to the poor model fit.
Table 20. Sample Critical Values for the Divgi Method based on Independent Variables.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test Length</th>
<th>Person Variance</th>
<th>Item Variance</th>
<th>Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>328.4062</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>131.0766</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>1249.733</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>1052.404</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>342.8997</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>496.8361</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>2142.452</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>1988.516</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>624.3633</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>252.4338</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>20983.26</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>20611.33</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>1229.795</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>1558.331</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>20693.13</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>20364.59</td>
</tr>
</tbody>
</table>

As noted earlier the Divgi method does not fit the data well reflected in the R-Square value of 0.5341. Below in Table the Variance Inflation Factors (VIF) are recorded for all predictor variables. Using the rule that a VIF must be much greater than 10 to signal multicollinearity, it is found that all but two predictors may be a problem, with an obvious case of multicollinearity relative to the Sample Size X Person Deviance interaction. Thus, it seems likely that the Divgi Method may suffer from multicollinearity. This is irrelevant as the lack of fit evident from the R-Square value makes this method unlikely to determine a good predicted cutoff value with respect to the independent variables.

Table 21. Divgi Method VIF Values for the Parameters of Full Model

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>29.45152</td>
</tr>
<tr>
<td>Test Length</td>
<td>5.06443</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>15.24812</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>9.04251</td>
</tr>
<tr>
<td>Test Length X Sample Size</td>
<td>9.78598</td>
</tr>
<tr>
<td>Sample Size X Person Deviance</td>
<td>103.82746</td>
</tr>
<tr>
<td>Person Deviance X Item Deviance</td>
<td>15.18653</td>
</tr>
</tbody>
</table>
Table 22 depicts the parameters for the most simple model while Table 23 shows a slightly more complex model interaction terms. The increase in independent variables clearly increased the R-Square value implying better fit. Also of note is the intercept term, was only close to being positive in the most simplistic case. As the number of variables increased, and fit of the model increased, the intercept became more negative. This was due to the original data set of values that approach zero. Therefore to fit the model well a negative intercept was needed to better approximate the points. Though it is counter-intuitive to have a negative Ratio of Eigenvalues value since it is created by the sum of item variances divided by the eigenvalue of the measure dimension. Thus, a likely conclusion is that a linear model approach may not have been an appropriate method.

Table 22. Divgi Method VIF Values for the Parameters of the Simple Model with R-Square = 0.3224

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter Estimate</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.03392</td>
<td>0</td>
</tr>
<tr>
<td>Sample Size</td>
<td>0.00740</td>
<td>1.00139</td>
</tr>
<tr>
<td>Test Length</td>
<td>0.08555</td>
<td>1.02833</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>-0.73901</td>
<td>1.00264</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>6.12640</td>
<td>1.02848</td>
</tr>
</tbody>
</table>

Table 23. Divgi Method VIF Values for the Parameters of the Simple Model with Interactions R-Square = 0.4817

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Parameter Estimate</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-614.13646</td>
<td>0</td>
</tr>
<tr>
<td>Sample Size</td>
<td>-0.11231</td>
<td>17.53607</td>
</tr>
<tr>
<td>Test Length</td>
<td>6.09933</td>
<td>8.63588</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>691.99255</td>
<td>18.42616</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>389.07731</td>
<td>9.12791</td>
</tr>
<tr>
<td>Sample Size X Test Length</td>
<td>0.00482</td>
<td>9.79196</td>
</tr>
<tr>
<td>Test Length X Item</td>
<td>-6.56523</td>
<td>11.59167</td>
</tr>
<tr>
<td>Deviance</td>
<td>Sample Size X Person Deviance</td>
<td>0.20134</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>Item Deviance X Person Deviance</td>
<td>-400.54766</td>
<td></td>
</tr>
</tbody>
</table>

**McGill Method**

In theory, this value should a small percentage, since the variability due to the questions should outweigh the variability due to the residuals. Therefore, the values of the 95th percentiles for the McGill values shall be used as a cutoff, since a large value should indicate the case where the eigenvalue of the first residual dimension is large, relative to the standard errors, thereby indicating a multidimensional case.

There is a very clear pattern that has emerged from all the cases presented in Figure 12. While holding all variables constant, the value of the 95th percentiles increases as the variance of the item difficulties is increased. Yet, as the person increases, the percentile values decrease. As both the variances increase the value for the percentile decreases. The percentile values are dependent on the distributional variances, sample size, and test lengths in a simple relationship, i.e. not parabolic. Considering the general pattern remains the same except for shifts there may be an interaction relative to the sample size, and variability in both item difficulties and person abilities.
Figure 12. McGill method 95th Percentiles for Person and Item Variance.

- \( N = 100 \)
- \( N = 500 \)
- \( N = 1000 \)
- \( N_i = 20 \)
- \( N_i = 60 \)
All the graphs in Figure 13 show the pattern of the percentiles when person variance and sample size are changed, regardless of the other variables, the increase in the 95th percentile is related to the increase in the sample size. Not quite as obvious, this noted increase does not occur as quickly when the test length is increased. When increasing the test length and holding everything else constant as is the case of comparing the graphs when item variance is 0.0625 and the number of items are either 20 or 60, the percentile values decrease for an increased number of items.

Figure 13. McGill method 95th Percentiles for Person and Sample Size.

Ni = 20

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>750</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0625</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Ni = 60

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>100</th>
<th>250</th>
<th>500</th>
<th>750</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0625</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>1.25</th>
<th>2.5</th>
<th>5.0</th>
<th>10.0</th>
<th>20.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0625</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>1.25</th>
<th>2.5</th>
<th>5.0</th>
<th>10.0</th>
<th>20.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0625</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
</tbody>
</table>
In order to develop a rule for the McGill method a regression analysis was conducted using the variables found to cause pattern changes such as sample size, test length, item variance, person variance, and all associated interactions as independent variables. The 95th percentile of McGill values for the 1000 iterations are used for the cutoff where the remaining lower 5% of values may be mistaken for the multidimensional case.

Backwards Elimination was used on all independent and dependent variables to find the best model, which uses the 95th percentile value from the bootstrapped confidence interval as the dependent variable. The following variables and interactions were used: sample size, test length, item deviance, person deviance, test length x sample size, test length / sample size, test length x item deviance, sample size x person deviance, (test length x item deviance)/(sample size x person deviance), person deviance/item deviance, and person deviance x item deviance. Table 24 below shows the R-Square values of the model upon removal the least significant independent. After removing the four variables the rest were deemed statistically significant and left in the model, as any further removal resulted in much larger drops in the model’s R-square value.
Table 24. Results of Stepwise Regression for McGill Method.

<table>
<thead>
<tr>
<th>Round of Removal</th>
<th>Variables Removed</th>
<th>R-Square Value for the Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>First</td>
<td>None</td>
<td>0.9278</td>
</tr>
<tr>
<td>Second</td>
<td>Test length x item deviance</td>
<td>0.9276</td>
</tr>
<tr>
<td>Third</td>
<td>person deviance x item deviance</td>
<td>0.9275</td>
</tr>
<tr>
<td>Forth</td>
<td>Test length/sample size</td>
<td>0.9268</td>
</tr>
<tr>
<td>Fifth</td>
<td>Sample size x person deviance</td>
<td>0.9258</td>
</tr>
<tr>
<td>Sixth</td>
<td>Test length/sample size</td>
<td>0.8880</td>
</tr>
</tbody>
</table>

The following is the optimal model for the Modified Percentage of Variance value as was found using SAS and using the Backward Elimination method of Stepwise Regression:

\[
\begin{align*}
McGill & = 0.75262 + 0.00048248(\text{SampleSize}) - 0.00612(\text{TestLength}) \\
& - 0.19386(\text{ItemDeviance}) - 0.21183(\text{PersonDeviance}) \\
& - 0.00000566(\text{Samplesize} \times \text{Testlength}) - 0.00398(\text{PersonDeviance} \times \text{ItemDeviance}) \\
& - 0.00008421(\text{SampleSize} \times \text{ItemDeviance} \times \text{PersonDeviance}) \\
& + 0.00232(\text{TestLength} \times \text{ItemDeviance} \times \text{PersonDeviance})
\end{align*}
\]  

(23)

where the McGill value is the maximum value of the ratio between the eigenvalue of the first residual dimension and the sum of the item standard to indicate unidimensionality.

Table 25 provides some sample critical values based on varying levels of sample size, test length, person and item variance.

This model for the McGill value has an R-squared value of 0.9258, explaining 92.58% of the variability in finding the value of the 95\textsuperscript{th} percentile under varying conditions. Thus, this model would be useful for assessing unidimensionality while keeping a Type I error rate close to the desired 5\% level. While this level of error will not be perfect, it will not often exceed the desired 0.05 probability of obtaining a Type I error.
Table 25. Sample Critical Values for the McGill Method based on Independent Variables.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test Length</th>
<th>Person Variance</th>
<th>Item Variance</th>
<th>Critical Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>0.567802</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>0.306162</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>0.197082</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>0.018158</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>0.161142</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>0.054098</td>
</tr>
<tr>
<td>100</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>0.073583</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>2.25</td>
<td>2.25</td>
<td>0.128777</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>0.25</td>
<td>0.894986</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>0.25</td>
<td>0.429586</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>0.25</td>
<td>0.486371</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>2.25</td>
<td>0.25</td>
<td>0.067371</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>0.25</td>
<td>2.25</td>
<td>0.450431</td>
</tr>
<tr>
<td>1000</td>
<td>60</td>
<td>0.25</td>
<td>2.25</td>
<td>0.031431</td>
</tr>
<tr>
<td>1000</td>
<td>20</td>
<td>2.25</td>
<td>2.25</td>
<td>0.125344</td>
</tr>
</tbody>
</table>

As noted earlier the McGill method fits the data very well. Below in Table the Variance Inflation Factors (VIF) are recorded for all predictor variables. Using the rule that a VIF must be much greater than 10 to signal multicollinearity, it is found that only one predictor may be a problem. The interaction term of Person Variance and Item Deviance result in a VIF value of 20.76 which is obviously greater than 10, but it may not be that much greater. Thus, the equation for the McGill method displays some possible if any multicollinearity in its construction.

Table 26. McGill Method VIF values for the parameters of full model

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Variance Inflation Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>10.82513</td>
</tr>
<tr>
<td>Test Length</td>
<td>6.51841</td>
</tr>
<tr>
<td>Item Deviance</td>
<td>9.59170</td>
</tr>
<tr>
<td>Person Deviance</td>
<td>6.24458</td>
</tr>
<tr>
<td>Test Length X Sample Size</td>
<td>9.68809</td>
</tr>
<tr>
<td>Person Deviance X Item Deviance</td>
<td>20.75787</td>
</tr>
<tr>
<td>Test Length X Person Deviance X Item Deviance</td>
<td>8.45371</td>
</tr>
<tr>
<td>Sample Size X Person Deviance X Item Deviance</td>
<td>6.17523</td>
</tr>
</tbody>
</table>
Discussion

All the methods that had publicized rules were less than ideal in their Type I error performance. The error rates were affected by changes in variables such as the variability in examinee abilities, the variability of item difficulties, test length, sample size and in some cases their interactions. Thus, it is rather notable that the use of a single static value for determining dimensionality is not the most efficient way to test instruments. The rules are much more complex than they were ever thought in the past.

In almost every single case where a rule already exists, as the number of items on an instrument increases the rate of Type I error increases. This is due to the idea that as the number of items increases the likelihood of them measuring one and only one dimension diminishes. Intuitively, there should only be so many ways in which to ask question relative to one construct. The more questions there are the more likely there are other constructs which may surface, even if they are not the intended main concept.

For the methods which did not previously have rules, the percentile values also appeared dependent on the conditions of variables such as the variability in examinee abilities, the variability of item difficulties, test length, sample size and in some cases their interactions. Most notably the Modified Kaiser

Part II - Cross Validation

This Result/Discussion section summarizes the results of a cross-validation of the Type I errors derived in the regression analyses conducted in the previous sections of the report. A simulation was conducted for each combination of the same variables as previously reported using 200 iterations per cell of the experimental design. This included three levels of test lengths, five levels of sample sizes, eight levels of item
variability, and eight levels of person variability. This simulation provides both a check on the Type I error rates associated with the derived critical values as well as providing a basis for comparison for the various methods.

**Cross-Validation Results**

Table 27 shows the average of the Type I errors across all the variables. Overall, the Modified Kaiser method has the smallest average error of any method presented. Even though it contains the least amount of error it is below the standard level of a 5% Type I error. Every method other than the Modified Kaiser has average error rates greater than 5%. The next best methods are the Ratio of Eigenvalues, Divgi, and the Modified Percentage which all have an average error rate less than 20%. Three other methods have average error rates less than 25%. These are the Kelley, McGill, and Modified Ratio of Eigenvalues method which correctly determines unidimensionality when it exists about 3 out of every 4 times.

Table 27. Average Type I errors for Methods

<table>
<thead>
<tr>
<th>Method</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified Kaiser</td>
<td>0.00099</td>
</tr>
<tr>
<td>Kaiser</td>
<td>0.377323</td>
</tr>
<tr>
<td>Modified Percentage</td>
<td>0.17874</td>
</tr>
<tr>
<td>20% Percentage</td>
<td>0.168896</td>
</tr>
<tr>
<td>40% Percentage</td>
<td>0.472396</td>
</tr>
<tr>
<td>50% Percentage</td>
<td>0.664741</td>
</tr>
<tr>
<td>Modified Ratio of Eigenvalues</td>
<td>0.23999</td>
</tr>
<tr>
<td>Ratio of Eigenvalues</td>
<td>0.123948</td>
</tr>
<tr>
<td>McGill</td>
<td>0.224823</td>
</tr>
<tr>
<td>Kelley</td>
<td>0.212271</td>
</tr>
</tbody>
</table>
The average Type I error with respect to each method for every sample size level is found in Figure 14. The Modified Kaiser method has the lowest level of error regardless of sample size. For a sample of 100 examinees the original Kaiser method has the most amount of error, erroneously signaling multidimensionality about 2/3rds of the time when an instrument is actually unidimensional. Besides the Modified Kaiser the next best method seems to be Divgi, but it has more error at the extremes of sample size. The other methods which consistently keep a Type I error rate less than 20 % are the Modified Percentage of Variance, 20% Percentage of Variance, and the Modified Ratio of Eigenvalues rule.

Figure 14. Average Type I errors for Methods relative to Sample Size

The average Type I error across test length is displayed in Figure 15 for all the methods. The lowest average error is found in all cases with the Modified Kaiser method.
Many of the methods increase the amount of error as the test length increases. With a sample size of 20, the 40% and 50% Percentage of Variance along with the Modified Ratio of Eigenvalues methods have the highest levels of error being greater that 30%. While the second best method is the 20% Percentage of Variance method. With a 40-item instrument the best methods, other than Modified Kaiser, are the Divgi and Kelley. While the most error is found when the 50% Percentage of Variance method is used. When test length is increased to 60 questions the second best method is the Modified Percentage of Variance method and the worst is the static 50% Percentage of Variance method.

Figure 15. Average Type I errors for Methods relative to Test Length

Table 28. Average Type I errors for Methods relative to Test Length

<table>
<thead>
<tr>
<th># of items</th>
<th>20</th>
<th>40</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaiser</td>
<td>0.000125</td>
<td>0.000281</td>
<td>0.002563</td>
</tr>
<tr>
<td>Modified Kaiser</td>
<td>0.226906</td>
<td>0.397344</td>
<td>0.507719</td>
</tr>
<tr>
<td>Modified</td>
<td>0.197313</td>
<td>0.152906</td>
<td>0.186</td>
</tr>
</tbody>
</table>
The average Type I error across the levels of person variance are found in Figure 16. The general trend shows that as the variability in the distribution of examinee ability increases the error rates tend to decrease. The method with the least amount of error regardless of the level of the person ability variance is the Modified Kaiser. While the most error is found using the 50% Percentage of Variance method across all variance levels. The next best methods seem to be Ratio of Eigenvalues, Divgi, Kelley, McGill, and Modified Percentage of Variance method which have an error rate around or below 20% except in some cases where the variability is rather small.
Figure 16. Average Type I errors for Methods relative to Person Variance

Table 29. Average Type I errors for Methods relative to Person Variance

<table>
<thead>
<tr>
<th>Person Variance</th>
<th>0.5</th>
<th>0.75</th>
<th>1.0</th>
<th>1.25</th>
<th>1.5</th>
<th>2.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kaiser</td>
<td>0.005</td>
<td>0.0001670</td>
<td>0.001833</td>
<td>0</td>
<td>0.0005</td>
<td>0</td>
</tr>
<tr>
<td>Modified</td>
<td>0.787583</td>
<td>0.5860587667</td>
<td>0.341</td>
<td>0.34025</td>
<td>0.163583</td>
<td>0.1675</td>
</tr>
<tr>
<td>Modified</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>0.339417</td>
<td>0.3155830</td>
<td>0.1598330</td>
<td>0.095167</td>
<td>0.09225</td>
<td>0.082667</td>
</tr>
<tr>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>0.433167</td>
<td>0.222667</td>
<td>0.155250</td>
<td>0.064083</td>
<td>0.0225</td>
<td>0.103917</td>
</tr>
<tr>
<td>40%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>0.846333</td>
<td>0.644250</td>
<td>0.622583</td>
<td>0.433250</td>
<td>0.354917</td>
<td>0.2635</td>
</tr>
<tr>
<td>50%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage</td>
<td>0.95775</td>
<td>0.7935830</td>
<td>0.8324250</td>
<td>0.6804170</td>
<td>0.617083</td>
<td>0.518417</td>
</tr>
<tr>
<td>Modified Ratio</td>
<td>0.288333</td>
<td>0.3565830</td>
<td>0.3555830</td>
<td>0.2753330</td>
<td>0.242667</td>
<td>0.11425</td>
</tr>
</tbody>
</table>
The average Type I error associated with the changing levels of the variability for the distribution of item difficulties for each method is in Figure 17. As in all the other case the Modified Kaiser method has the least amount of error across the range of values. A similar trend to the person variance is that as the variability increases the error generally decreases. Some of the next best methods are Divgi and the Ratio of Eigenvalues. After the variance passes about 1.0 some of the other best methods are Divgi, Modified Percentage of Variance, Modified Ratio of Eigenvalues, and the McGill method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Ratio of Eigenvalues</th>
<th>0.363583</th>
<th>0.112750</th>
<th>0.102833</th>
<th>0.057583</th>
<th>0.042083</th>
<th>0.096</th>
<th>0.033917</th>
<th>0.182833</th>
</tr>
</thead>
<tbody>
<tr>
<td>McGill</td>
<td>0.69575</td>
<td>0.414917</td>
<td>0.2330</td>
<td>0.081833</td>
<td>0.049667</td>
<td>0.08825</td>
<td>0.164333</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kelley</td>
<td>0.15575</td>
<td>0.196250</td>
<td>0.225917</td>
<td>0.201417</td>
<td>0.238833</td>
<td>0.194167</td>
<td>0.228833</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Divgi</td>
<td>0.237</td>
<td>0.181667</td>
<td>0.129583</td>
<td>0.083667</td>
<td>0.079667</td>
<td>0.064917</td>
<td>0.107583</td>
<td>0.10875</td>
<td></td>
</tr>
</tbody>
</table>
Figure 17. Average Type I errors for Methods relative to Item Variance

Table 30. Average Type I errors for Methods relative to Item Variance

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>Modified Kaiser</th>
<th>Kaiser</th>
<th>Modified Percentage</th>
<th>20% Percentage</th>
<th>40% Percentage</th>
<th>50% Percentage</th>
<th>Modified Ratio of Eigenvalues</th>
<th>McGill</th>
<th>Kelley</th>
<th>Divgi</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.007583</td>
<td>0.000333</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.220917</td>
<td>0.343667</td>
<td>0.295583</td>
</tr>
<tr>
<td>0.25</td>
<td>0.755917</td>
<td>0.555417</td>
<td>0.569667</td>
<td>0.349583</td>
<td>0.3635</td>
<td>0.186083</td>
<td>0.0445</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.401583</td>
<td>0.203167</td>
<td>0.177417</td>
</tr>
<tr>
<td>0.75</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.401583</td>
<td>0.203167</td>
<td>0.177417</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.401583</td>
<td>0.203167</td>
<td>0.177417</td>
</tr>
<tr>
<td>1.25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.401583</td>
<td>0.203167</td>
<td>0.177417</td>
</tr>
<tr>
<td>1.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.401583</td>
<td>0.203167</td>
<td>0.177417</td>
</tr>
<tr>
<td>2.25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.401583</td>
<td>0.203167</td>
<td>0.177417</td>
</tr>
</tbody>
</table>

Modified Percentage

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>20%</th>
<th>40%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
<tr>
<td>0.25</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
<tr>
<td>0.5</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
<tr>
<td>0.75</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
<tr>
<td>1</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
<tr>
<td>1.25</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
<tr>
<td>1.5</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
<tr>
<td>2.25</td>
<td>0.17350</td>
<td>0.131083</td>
<td>0.131083</td>
</tr>
</tbody>
</table>

Modified Ratio of Eigenvalues

<table>
<thead>
<tr>
<th>Item Variance</th>
<th>0.396917</th>
<th>0.486167</th>
<th>0.384667</th>
<th>0.230417</th>
<th>0.1485</th>
<th>0.103917</th>
<th>0.106333</th>
<th>0.063</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.396917</td>
<td>0.486167</td>
<td>0.384667</td>
<td>0.230417</td>
<td>0.1485</td>
<td>0.103917</td>
<td>0.106333</td>
<td>0.063</td>
</tr>
<tr>
<td>0.25</td>
<td>0.396917</td>
<td>0.486167</td>
<td>0.384667</td>
<td>0.230417</td>
<td>0.1485</td>
<td>0.103917</td>
<td>0.106333</td>
<td>0.063</td>
</tr>
<tr>
<td>0.5</td>
<td>0.396917</td>
<td>0.486167</td>
<td>0.384667</td>
<td>0.230417</td>
<td>0.1485</td>
<td>0.103917</td>
<td>0.106333</td>
<td>0.063</td>
</tr>
<tr>
<td>0.75</td>
<td>0.396917</td>
<td>0.486167</td>
<td>0.384667</td>
<td>0.230417</td>
<td>0.1485</td>
<td>0.103917</td>
<td>0.106333</td>
<td>0.063</td>
</tr>
<tr>
<td>1</td>
<td>0.396917</td>
<td>0.486167</td>
<td>0.384667</td>
<td>0.230417</td>
<td>0.1485</td>
<td>0.103917</td>
<td>0.106333</td>
<td>0.063</td>
</tr>
<tr>
<td>1.25</td>
<td>0.396917</td>
<td>0.486167</td>
<td>0.384667</td>
<td>0.230417</td>
<td>0.1485</td>
<td>0.103917</td>
<td>0.106333</td>
<td>0.063</td>
</tr>
<tr>
<td>1.5</td>
<td>0.396917</td>
<td>0.486167</td>
<td>0.384667</td>
<td>0.230417</td>
<td>0.1485</td>
<td>0.103917</td>
<td>0.106333</td>
<td>0.063</td>
</tr>
<tr>
<td>2.25</td>
<td>0.396917</td>
<td>0.486167</td>
<td>0.384667</td>
<td>0.230417</td>
<td>0.1485</td>
<td>0.103917</td>
<td>0.106333</td>
<td>0.063</td>
</tr>
</tbody>
</table>
### Discussion

Overall, the best method with respect to Type I error rate is the Modified Kaiser method. It has the absolute lowest error rates under any condition. It is worth noting that, under any condition, the average is less than the standard level of 5% Type I error, and in many cases the error was zero. Thus, either this method worked very well or has a high likelihood of determining unidimensionality regardless of actual dimensionality.

Other rules that performed well at detecting correctly diagnosing unidimensionality are the Divgi, McGill, Ratio of Eigenvalues, and Modified Percentage of Variance methods. These methods keep close to or below 20% error rates most of the time. However, all of these methods produce Type I error rates that are higher than the intended 5% level. An interesting note is that, of all the methods that were investigated, the Kaiser method, which is one of the most widely used methods to date, has the third highest average of overall Type I error. The other two methods with the highest error rates, namely the 40% Percentage of Variance and 50% Percentage of Variance, have a rule with an obvious cutoff value. Thus, there is additional evidence that methods that do not take into account changes in certain variables will be less optimal at determining dimensionality. In almost all cases the modified methods and the methods that were finally given a cutoff criteria have some of lowest Type I error rates.
Part III – Power

This section summarizes analysis of the statistical power of each method evaluated in the previous sections. Multidimensional data was simulated using the Conquest software (Wu, Adams, Wilson, & Haldane, 2007) via the Multidimensional Random Coefficients Multinomial Logit Model (MRCMLM). Person ability estimates were sampled from a multivariate normal distribution $MN(0, \sigma_1)$, and item difficulties were sampled from a Normal distribution $N(0, \sigma_2)$ distribution. The multivariate distributions were simulated to have one of the following levels of correlation (0.00, 0.25, 0.50, and 0.75). Item responses were simulated varying the number of items (20, 40, 60) and sample size (100, 250, 500, 750, 1000) between datasets. In addition, the item and person parameter dispersion was crossed by the following values of distributional variance (0.50, 1, and 1.5). Therefore, the independent variables were the number of items (3 levels), sample size (5 levels), item parameter dispersion (3 levels), person parameter dispersion (3 levels), and correlations between dimensions (4 levels). Each of the 540 cells of the experiment was replicated 1000 times. The discussion of statistical power is organized around the results for each method.

Kaiser Method Results and Discussion

When the correlation between dimensions is zero, the Kaiser method is a very powerful test. That is, the power, or probability of determining whether multidimensional exists, is high, as displayed in Figure 18. For most cases, the power is nearly 1.00, with only a few cases with powers below .90. The power is still relatively high when the correlation increases to 0.25 between dimensions, as illustrated in Figure 19. The only
A noticeable loss in power is when there are a small number of questions and the item variance and sample size are increased simultaneously.

The drop in power rates is even more obvious in Figure 20, where the correlation is increased to 0.5. Power for the shorter instruments begins to decrease at smaller sample sizes than those found at lower correlation level. Finally, when the inter-dimensional correlation is increased to 0.75, the power diminishes even faster for short instruments, yet the remains at a decent level for a large number of questions (Figure 21).

Figure 18. Power rates of Kaiser with Correlation = 0.00.

Ni = 20

Ni = 60

N = 100
Figure 19. Power rates of Kaiser with Correlation = 0.25.

Ni = 20

Ni = 60

N = 100

N = 1000
Figure 20. Power rates of Kaiser with Correlation = 0.50.

Ni = 20

N = 100

Ni = 60

N = 1000
N = 250

N = 500

N = 750
Figure 21. Power rates of Kaiser with Correlation = 0.75.

Ni = 20

Ni = 60

N = 100

N = 1000
On the whole, the power associated with Kaiser method is quite good. Where a “good” test is a power level above 0.80. There were 3 levels of test length, 5 levels of sample size, 3 levels of item variance and 3 levels of person variance culminating into 135 cells variable combinations for each of the correlation levels. A correlation of 0.00 yielded an average power of 0.98 per cell. The average cell power decreased to 0.95 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.86. The largest decrease in the average power is when the correlation between dimensions is set to 0.75, at which point the average power is 0.62.

Figure 22 and Figure 23 look at the power rates for sample size and test length relative to correlation respectively. In both cases the increase in correlation between dimensions diminished the power. When investigating sample size it appears that the most power is found for non-extreme sizes. Except when the relationship between the dimensions is high caused by a case of inverse proportionality between the sample size and the power. The test length displays a case of direct proportionality between the number of items on a test and power.
With respect to the item and person variance, as the correlation increased the power decreased as in Figures 22 and 23. When using the Kaiser method the average power level remained similar across person variance within each level of correlation (Figure 24). On the other hand the greater the item variability the lower the power level especially when interacted with the correlation (Figure 25).

The resulting power for the interaction between the Item and Person variances are found in Figures 26 through 29. The larger the Item variance the lower the power is of the Kaiser test regardless of correlation. The Person Variance has a more interesting trend, as the variability increases the power increases while the correlation is at most 0.5 as in Figures 26 through 29. When the correlation becomes greater than 0.5 as is the case in Figure 29 when the Person Variability is increased the power decreases.

Figure 22. Average Power rates for Kaiser method relative to sample size and correlation.
Figure 23. Average Power rates for Kaiser method relative to test length and correlation.

Figure 24. Average Power rates for Kaiser method relative to Person Variance and correlation.

Figure 25. Average Power rates for Kaiser method relative to Item Variance and correlation.
Figure 26. Average Power rates for Kaiser method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 27. Average Power rates for Kaiser method relative to Item Variance, Person Variance, with correlation of 0.25.

Figure 28. Average Power rates for Kaiser method relative to Item Variance, Person Variance, with correlation of 0.5.
**Modified Kaiser Method Results and Discussion**

For the Modified Kaiser method the power reached its lowest when tests are short and taken by small samples, as in Figure 30. The power increased as the person variance increased, but decreased as the item variance increased. On the whole, longer tests have greater power. An Increase in the correlation between the dimensions caused a decrease in power, as can be seen in Figure 30-33. Since the larger the correlation between the dimensions, the more related the dimensions are: Therefore, the larger the correlation the more likely the dimensions take on the appearance of the unidimensional case, and it will be more difficult to distinguish a true multidimensional case.
Figure 30. Power rates of Modified Kaiser with Correlation = 0.00.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 31. Power rates of Modified Kaiser with Correlation = 0.25.

Ni = 20

<table>
<thead>
<tr>
<th>Person Variance</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Ni = 60

<table>
<thead>
<tr>
<th>Person Variance</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Ni = 100

<table>
<thead>
<tr>
<th>Person Variance</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Ni = 250

<table>
<thead>
<tr>
<th>Person Variance</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Ni = 500

<table>
<thead>
<tr>
<th>Person Variance</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>0.6</td>
<td>0.6</td>
</tr>
<tr>
<td>0.7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 32. Power rates of Modified Kaiser with Correlation = 0.50.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 33. Power rates of Modified Kaiser with Correlation = 0.75.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
When the inter-dimensional correlation is very low the power is at an acceptable level, provided the number of items on the test is relatively large. As the correlation increases the power levels drop quickly. A correlation of 0.00 yielded an average power of 0.75 per cell, which decreases to 0.62 when the correlation was raised to 0.25. Further increases to the correlation, to 0.50 and then 0.75 resulted in average powers of 0.39 and 0.09, respectively. Therefore, the power relative to Modified Kaiser method is not at an acceptable level in almost any case, especially for short tests.
The average Power rates for sample size and test length are shown in Figure 34 and Figure 35 respectively. Overall in either case the power decreases as the correlation increases. The Modified Kaiser method increases the power with increases in sample size (Figure 34). Similarly, an increase in test length causes an increase in power as well.

The relationship of Person variability and correlation in Figure 36 implies that the power is directly proportional to an increase in variability. The opposite holds true for the Item variability. As seen in Figure 37 as the variability in the distribution of item difficulties increased as the average power rates decreased.

The typical trend is that as the person variance increases the power increase across the levels of correlation, as noted in Figure 38 through 41. The rates of increasing power look similar across the levels of item variability until reaching a correlation greater than 0.5. Once the correlation is greater than 0.5 the rates of increasing power across increasing person variability diminishes when the item variability is either 1 or 1.5 (Figure 41).

Figure 34. Average Power rates for Modified Kaiser method relative to Sample Size and correlation.
Figure 35. Average Power rates for Modified Kaiser method relative to Test Length and correlation.

Figure 36. Average Power rates for Modified Kaiser method relative to Person Variance and correlation.

Figure 37. Average Power rates for Modified Kaiser method relative to Item Variance and correlation.
Figure 38. Average Power rates for Modified Kaiser method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 39. Average Power rates for Modified Kaiser method relative to Item Variance, Person Variance, with correlation of 0.25.

Figure 40. Average Power rates for Modified Kaiser method relative to Item Variance, Person Variance, with correlation of 0.5.
Figure 41. Average Power rates for Modified Kaiser method relative to Item Variance, Person Variance, with correlation of 0.75.

20% Measure Variability Results and Discussion

As seen in Figure 42, the power is low when using a cutoff value of 20% measure variability to signal a multidimensional case. In the best-case scenario, where the correlation between the dimensions is zero, most of the power values are nearly zero. The increase to a correlation of 0.25 and 0.50 as in Table 43 and Table 44, respectively, show even more decrease in the power relative to the variables. The odd outcome is that at the highest level of correlation namely 0.75 in this case, Table 45, the power actually increased a slight degree over the correlation of 0.50.
Figure 42. Power rates of 20% Measure Variability with Correlation = 0.00.

Ni = 20

N = 100

Ni = 60

N = 250

N = 500
$N = 750$

$N = 1000$
Figure 43. Power rates of 20% Measure Variability with Correlation = 0.25.

Ni = 20

N = 100

Ni = 60

N = 250

N = 500
Figure 44. Power rates of 20% Measure Variability with Correlation = 0.50.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 45. Power rates of 20% Measure Variability with Correlation = 0.75.

Ni = 20

N = 100

Ni = 60

N = 250

N = 500
There are only a handful of cases where the power is at an acceptable level. Even as the correlation increases the power levels remain rather constant. A correlation of 0.00 yielded an average power of 0.06 per cell. The average cell power decreased to 0.02 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.01. The lowest average power is 0.02 when correlation is 0.75 between the dimensions. Therefore, the power relative to 20% Measure Variability method is not at an acceptable level.
The increased correlation in Figure 46 results in power that decreased and then increased for almost all levels of the sample size. If large values of correlation were used it would probably be obvious that all samples sizes possessed this parabolic trend. The levels of test length show that power decreased as the correlation increased in Figure 47. Similarly the test lengths display a parabolic trend, similar to the sample size relationship, when the number of questions is sufficiently large.

The patterns for the average power rates are nearly identical for the person and item variance across the correlation levels, note Figures 48 and 49. When the variability is at the lowest level, in this case 0.50, as the correlation increased the power decreased. Yet, as the variability increased the trend changed such that power and correlation were directly proportional.

When the correlation was held constant the power appeared to decrease as the person variance increased, especially when the item variance was small, specifically 0.5. This trend is obvious in Figures 50 and 51. This trend is departed when correlation is increased to be greater than 0.25 where a large item variance caused power rates to increase proportionately as person variability increased as in Figure 40. Then the trend changed again for a correlation of 0.75 where the power increased and then decreased for the highest level of item variance given increasing person variance.
Figure 46. Average Power rates for 20% Measure Variability method relative to Sample Size and correlation.

Figure 47. Average Power rates for 20% Measure Variability method relative to Test Length and correlation.

Figure 48. Average Power rates for 20% Measure Variability method relative to Person Variance and correlation.
Figure 49. Average Power rates for 20% Measure Variability method relative to Item Variance and correlation.

Figure 50. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 51. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.25.
40% Measure Variability Results and Discussion

The power when using the rule of 40% measure variability is greater than that for 20% measure variability in almost every situation. Regardless of the inter-dimensional correlation, as the item variance increases the power decreases (Tables 54 through 57). Also, regardless of correlation between dimensions, as the person variance increases the power decreases. There is little change in the observed power when changing the levels of either sample size or test length.
Figure 54. Power rates of 40% Measure Variability with Correlation = 0.00.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 55. Power rates of 40% Measure Variability with Correlation = 0.25.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
N = 750

N = 1000
Figure 56. Power rates of 40% Measure Variability with Correlation = 0.50.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 57. Power rates of 40% Measure Variability with Correlation = 0.75.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance

0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance
The power for the 40% Measure variance method is generally quite high and reasonable, i.e. greater than .80. A problem with maintaining high power arises when increasing variance in either the person ability or item difficulty, which causes a noticeable decrease in the power. On the other hand, as the correlation increases, the power levels change drastically. A correlation of 0.00 yielded an average power of 0.70 per cell. The average cell power is decreased to 0.57 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.44. The lowest average power is 0.33 when correlation is 0.75 between the dimensions. Therefore the power relative to
40% Measure Variability method is somewhat decent, more so at lower correlation levels.

Figures 57 and 58 present decreased power when the correlation is increased with respect to either test length or sample size. When the sample size was investigated it was found that the largest and smallest levels had the lowest average power. With respect to test length Figure 58 shows that more test items related to greater power.

The lowest levels of item and person variability result in larger levels of power as in Figures 59 and 60. As the correlation increased the power rates decreased, as was also the case with sample size and test length.

When holding the correlation constant as in Figures 61 through 64 the patterns are rather similar. As the person variability increased the power decreased. The low level of item variance provided the largest average level of power. The gradient of the decreased power became steeper as the correlation increased with respect to the item and person variability.

Figure 57. Average Power rates for 40% Measure Variability method relative to Sample Size and correlation.
Figure 58. Average Power rates for 40% Measure Variability method relative to Test Length and correlation.

Figure 59. Average Power rates for 40% Measure Variability method relative to Person Variance and correlation.

Figure 60. Average Power rates for 40% Measure Variability method relative to Item Variance and correlation.
Figure 61. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 62. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.25.

Figure 63. Average Power rates for 20% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.5.
50% Measure Variability Results and Discussion

The power in each cell is slightly higher than the power found in the 40% or 20% measure variability cases, and exhibits patterns similar to those lower level percentage methods. For instance, the increase in either person or item variance is generally paired with a decrease in the power. A slight difference can be seen in Figure 65, in the form of higher power relative to shorter test lengths. The power also seems to hold steady and is not so not highly affected by the change in dimensional correlations.

Figure 65. Power rates of 50% Measure Variability with Correlation = 0.00.
Figure 66. Power rates of 50% Measure Variability with Correlation = 0.25.

Ni = 20

N = 100

N = 250

N = 500
Figure 67. Power rates of 50% Measure Variability with Correlation = 0.50.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 68. Power rates of 50% Measure Variability with Correlation = 0.75.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
The 50% Measure variance method is the most powerful of the three static percentage of variance methods. Increasing the inter-dimensional correlation does cause wide changes in power when increasing from one level to the next. A correlation of 0.00 yielded an average power of 0.93 per cell. The average cell power is decreased to 0.86 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.76. The lowest average power is 0.66 when correlation is 0.75 between the dimensions. Therefore the power of detecting multidimensionality when using the 50% Measure
Variability method is rather high, and is the preferred method over the other percentage of variance methods when maximizing power.

The power for this method did not decrease very rapidly over the range of presented correlations. Specifically for sample size in Figure 69 and test length in Figure 70 the power slope is not steep while correlation was increased.

Figures 71 and 72 show the trend for person and item variance respectively. These trends are similar in that they decreased as correlation was increased. Also power was inversely proportional to the variability, i.e. lower variability was associated with higher power.

Figures 73 through 74 display almost the same trends as mentioned above. As the as person or item variance increased the power decreased as well as their interaction when both variabilities are large and the power dropped faster. As the correlation increased the degradation of power occurred at a much steeper rate for higher levels of variabilities than what noted for the low correlation level (Graph 76).

Figure 69. Average Power rates for 50% Measure Variability method relative to Sample Size and correlation.
Figure 70. Average Power rates for 50% Measure Variability method relative to Test Length and correlation.

Figure 71. Average Power rates for 50% Measure Variability method relative to Person Variance and correlation.

Figure 72. Average Power rates for 50% Measure Variability method relative to Item Variance and correlation.
Figure 73. Average Power rates for 50% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 74. Average Power rates for 50% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.25.

Figure 75. Average Power rates for 50% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.5.
Figure 76. Average Power rates for 50% Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.75.

### Modified Percentage Measure Variability Results and Discussion

The power rates for the Modified Percentage method are not adequate for a usable method. In Figure 77, when the dimensions are uncorrelated, the power decreased as item variance increased, while at other times the power will increased and then decreased. Regardless of the correlation, as seen in Tables 77 through 80, the power will decrease as a result of the increase in person variability.

Figure 77. Power rates of Modified Percentage Measure Variability with Correlation = 0.00.

- **Ni = 20**
- **Ni = 60**
- **N = 100**
N = 1000
Table 78. Power rates of Modified Percentage Measure Variability with Correlation = 0.25.

Ni = 20

Ni = 60

N = 100

N = 250
Table 79. Power rates of Modified Percentage Measure Variability with Correlation = 0.50.

Ni = 20

Ni = 60

N = 100

N = 250
Table 80. Power rates of Modified Percentage Measure Variability with Correlation = 0.75.

<table>
<thead>
<tr>
<th>Ni = 20</th>
<th>Ni = 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 100</td>
<td>N = 250</td>
</tr>
</tbody>
</table>

![Graphs showing power rates for different conditions](image-url)
Of all the methods that use the percentage of variance accounted for by the measure dimension as a cutoff for choosing the presence of unidimensionality, this one is among the worst, being only slightly better than the 20% method. A correlation of 0.00 yielded an average power of 0.22 per cell. The average cell power decreases to 0.15 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.11. The lowest average power is 0.09 when correlation is 0.75 between the dimensions. Therefore, the power relative to Modified Percentage Measure Variability method is rather low, higher power can be found in either the method with a 40% or 50% cutoff.

Clearly in the method of Modified Proportions of Measure Variability the correlation decreased as the correlation increased. Figure 81 denotes this trend in terms of sample size. Even though power decreased across the correlations, the lowest power was found for the extreme levels of sample size namely 100 and 1000 examinees. Relative to test length as in Figure 82 the decreasing power trend is obvious along with greater power being associated with shorter tests.

Both the item and person variance, found in Figures 82 and 83 respectively, show an inverse relationship for power across the correlation levels. For person and item variance the lowest levels of variability resulted in the more power than situations with greater variability in the distributions of person abilities or item difficulties. These similar trends are echoed in Figures 84 through 87 where the correlation is held constant and the variances are increased causing power to decrease.
Figure 81. Average Power rates for Modified Proportion of Measure Variability method relative to Sample Size and correlation.

Figure 82. Average Power rates for Modified Proportion of Measure Variability method relative to Sample Size and correlation.

Figure 83. Average Power rates for Modified Proportion of Measure Variability method relative to Person Variance and correlation.
Figure 84. Average Power rates for Modified Proportion of Measure Variability method relative to Item Variance and correlation.

Figure 85. Average Power rates for Modified Proportion of Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 86. Average Power rates for Modified Proportion of Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.25.
Figure 87. Average Power rates for Modified Proportion of Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.5.

![Graph](image1)

Figure 88. Average Power rates for 5 Modified Proportion of Measure Variability method relative to Item Variance, Person Variance, with correlation of 0.75.

![Graph](image2)

**Ratio of Eigenvalues Results and Discussion**

When using the Ratio of Eigenvalues method, multiple dimensions should be present when the eigenvalue for the measure dimension is less than four times the eigenvalue of the first residual dimension. In most of the cells, the power is very low, implying that this method has a tendency to signal unidimensionality when multidimensionality is actually present. Figure 89 shows that the power can be acceptable when the sample size is at its lowest level. As the correlation between the dimensions is increased (Figurees 90 through 92), the power drops quickly. The power also decreases as the item and person variance increases.
Figure 89. Power rates Ratio of Eigenvalues with Correlation = 0.00.

Ni = 20

Ni = 60

N = 100

N = 250
Figure 90. Power rates Ratio of Eigenvalues with Correlation = 0.25.

Ni = 20

N = 100

Ni = 60

N = 250

N = 500
Figure 91. Power rates Ratio of Eigenvalues with Correlation = 0.50.

- **Ni = 20**
  - N = 100
  - N = 250
  - N = 500

- **Ni = 60**
  - N = 100
  - N = 250
  - N = 500
Figure 92. Power rates Ratio of Eigenvalues with Correlation = 0.75.

\(Ni = 20\)

\(N = 100\)

\(Ni = 60\)

\(N = 250\)

\(N = 500\)
The Ratio of Eigenvalues method has some of the lowest average power levels per cell of any of the methods. A correlation of 0.00 yielded an average power of 0.05 per cell. The average cell power is decreased to 0.02 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.01. The lowest average power is 0.01 when correlation is 0.75 between the dimensions. With power levels this low the usefulness of the Ratio of Eigenvalues method is questionable in almost every combination of variables.
The general trend relative to sample size found in Figure 93 is that as correlation increased the power decreased for each level of examinee. The only slight deviation is the lowest level of sample size where the power increased slightly from a correlation of 0.50 to 0.75.

When sample sizes are less than 60 the power decreased with increases in correlation (Figure 94). The oddity occurs at a test length of 60 questions where the power increased form correlation values of 0.25 up to 0.75.

Figures 95 and 96 which denote the relationship of person and item variance to correlation respectively show similar trends. Given the lower levels of variance, in these cases less than 1.5, the power decreased with an increase in correlation. While high levels of variability actually increased power while the correlation increased.

In Figures 97 and 98 where the correlation is held constant at 0.0 and 0.5 respectively the trend is decreasing power as the person variance increased. After the correlation is 0.5 or greater as in Figures 99 and 100 the trend changed for the item variance of 1.5 such that as the person variance increased the power decreased.

Figure 93. Average Power rates for Ratio of Eigenvalues method relative to Sample Size and correlation.
Figure 94. Average Power rates for Ratio of Eigenvalues method relative to Test Length and correlation.

Figure 95. Average Power rates for Ratio of Eigenvalues method relative to Person Variance and correlation.

Figure 96. Average Power rates for Ratio of Eigenvalues method relative to Item Variance and correlation.
Figure 97. Average Power rates for Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 98. Average Power rates for Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.25.

Figure 99. Average Power rates for Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.5.
Figure 100. Average Power rates for Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.75.

**Modified Ratio of Eigenvalues Results and Discussion**

As shown in Figure 101, when the correlation between dimensions is 0.00, the Modified Ratio of Eigenvalues method has high power under varying conditions. The power generally increases as the person variance increases. An obvious pattern does not emerge, however, when increasing item difficulties: at times the power increases and other times it decreases, but on the whole it does not vary by a great amount.

In Figure 102, when the correlation increases to 0.25, the item variance pattern of power levels drop faster than those found with lower correlations. Looking at all the correlations levels in Tables 45 through 48 simultaneously, an interesting trend occurs: the power levels for the lowest level of item variability is usually large, regardless of the correlation. Thus, this method seems well suited when test questions are of similar difficulty.
Figure 101. Power rates of Modified Ratio of Eigenvalues with Correlation = 0.00.

Ni = 20

N = 100

Ni = 60

N = 250

N = 500
Figure 102. Power rates of Modified Ratio of Eigenvalues with Correlation = 0.25.
Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 103. Power rates of Modified Ratio of Eigenvalues with Correlation = 0.50.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 104. Power rates of Modified Ratio of Eigenvalues with Correlation = 0.75.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
The Modified Ratio of Eigenvalues Method produces power levels that are not ideal but seem to be reasonable, given that most of the power levels are above .80 and are higher for correlations of 0.5 and less. A correlation of 0.00 yielded an average power of 0.86 per cell. The average cell power is decreased to 0.80 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.62. The lowest average power is 0.37 when correlation is 0.75 between the dimensions. Thus, this method seems reasonable in terms of power for choosing multidimensionality when it is actually present in the instrument.
Regardless of sample size as the correlation was increased the level of power decreased as in Figure 105. This graph also shows that for this method the more examinees that take a test cause the power to be increased at any correlation level. Test length is rather similar in that the power decreased with the increase across the levels of correlation (Figure 106). This Modified Ratio of Eigenvalues assigns greater power to those test which are shorter in length.

Both the item and person variance, in Figures 107 and 108, display similar results with respect to correlation. The relationship is one of inverse proportionality where increased correlation causes drops in power for the two types of variability.

Figures 109 through 112 show the power for different correlation levels when noting item and person variance. In all cases the lowest level of variability for either item or person variability had the greatest power. As the item variance increased the power decreased, the power also decreased faster as the correlation value increased. Also the power was usually shown to increase as a result of the increased person variance across most correlation levels. The trend change occurred after the correlation was greater than 0.5, this is where the power began to decrease as a function of an increased person variance. Yet in for all correlation levels when the item variance is at the lowest level the power is very good i.e. above 0.8.
Figure 105. Average Power rates for Modified Ratio of Eigenvalues method relative to Sample Size and correlation.

Figure 106. Average Power rates for Modified Ratio of Eigenvalues method relative to Test Length and correlation.

Figure 107. Average Power rates for Modified Ratio of Eigenvalues method relative to Person Variance and correlation.
Figure 108. Average Power rates for Modified Ratio of Eigenvalues method relative to Item Variance and correlation.

![Figure 108](image1)

Figure 109. Average Power rates for Modified Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.0.

![Figure 109](image2)

Figure 110. Average Power rates for Modified Ratio of Eigenvalues method relative to Item Variance, Person Variance, with correlation of 0.25.

![Figure 110](image3)
Kelley Method Results and Discussion

The Kelley method uses the eigenvalue from the measure dimension and the sum of the item variances to determine dimensionality. The power is best when the number of questions and the number of test takers are small. This can be seen in Figure 113 through 116, where in all of the cases of with short tests and small samples, the power is high.

Regardless of the inter-dimensional correlation, the power drops quickly as result of sample size and test length. For example, the power drop dramatically after 250 people when the correlation is 0.00, but when the correlation is raised to 0.25 the drop in power
starts to occur after 100 persons. In all cases, the power also decreases as a result of either decreasing item or person distributional variance.

Figure 113. Power rates of Kelley Method with Correlation = 0.00.
Figure 114. Power rates of Kelley Method with Correlation = 0.25.

Ni = 20

N = 100

Ni = 60

N = 250

N = 500

Person Variance

0.4 0.6 0.8 1 1.2 1.4 1.6

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Power

Person Variance

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Power

Person Variance

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Power

Person Variance

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Power

Person Variance

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

Power
Figure 115. Power rates of Kelley Method with Correlation = 0.50.

**Ni = 20**

**Ni = 60**

**N = 100**

**N = 250**

**N = 500**
Figure 116. Power rates of Kelley Method with Correlation = 0.75.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
The Kelley method will only be useful in a handful of cases due to its low overall power. The best time to use it is when the sample size is small and there is low correlation between the dimensions. A correlation of 0.00 yielded an average power of 0.38 per cell which is rather low. The average cell power is decreased to 0.24 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.16. The lowest average power is 0.12 when correlation is 0.75 between the dimensions. Thus, this is not a good method due to the low level of power.
In Figure 117 the power is shown relative to the level of correlation for each of the five levels of sample size. The power decreased across the range of increased correlation. Power was found to be inversely related to the number of examinees. Figure 118 denotes the power related to correlation and the test length of instruments. The power decreased as the level of correlation was increased. The number of items is related inversely to the power such that as the test length increased the power decreased.

Similar patterns are found in Figures 119 and 130 where the item and person variance are found to cause a decrease in power as the variability is increased. This pattern is consistent across the levels of correlation. Figures 131 through 134 show that as both item and person variability were increased the power was decreased.

Figure 117. Average Power rates for Kelley method relative to Sample Size and correlation.
Figure 118. Average Power rates for Kelley method relative to Test Length and correlation.

Figure 119. Average Power rates for Kelley method relative to Person Variance and correlation.

Figure 120. Average Power rates for Kelley method relative to Item Variance and correlation.
Figure 121. Average Power rates for Kelley method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 122. Average Power rates for Kelley method relative to Item Variance, Person Variance, with correlation of 0.25.

Figure 123. Average Power rates for Kelley method relative to Item Variance, Person Variance, with correlation of 0.5.
**Divgi Method Results and Discussion**

When testing for multidimensionality, values for power using the Divgi method are rather consistent in their patterns and are overall too low to recommend this method. For all four levels of correlation, longer tests result in a higher level of power. Figures 125 through 128 all show that as the person variability increases, the power will decrease. Conversely, increases in the item difficulty variability result in an increase in the power. For the most part, the patterns and overall power look similar, regardless of the correlation between the dimensions.
Figure 125. Power rates of Divgi Method with Correlation = 0.00.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Table 126. Power rates of Divgi Method with Correlation = 0.25.

<table>
<thead>
<tr>
<th>Ni = 20</th>
<th>Ni = 60</th>
</tr>
</thead>
</table>

| N = 100 | N = 250 | N = 500 |

<table>
<thead>
<tr>
<th>Person Variance</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.40</td>
</tr>
<tr>
<td>0.10</td>
<td>0.45</td>
</tr>
<tr>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td>0.30</td>
<td>0.55</td>
</tr>
<tr>
<td>0.40</td>
<td>0.60</td>
</tr>
<tr>
<td>0.50</td>
<td>0.65</td>
</tr>
<tr>
<td>0.60</td>
<td>0.70</td>
</tr>
<tr>
<td>0.70</td>
<td>0.75</td>
</tr>
<tr>
<td>0.80</td>
<td>0.80</td>
</tr>
<tr>
<td>0.90</td>
<td>0.85</td>
</tr>
<tr>
<td>1.00</td>
<td>0.90</td>
</tr>
</tbody>
</table>

Graphs showing the relationship between Person Variance and Power for different values of N and Ni.
N = 750

N = 1000
Table 127. Power rates of Divgi Method with Correlation = 0.50.

**Ni = 20**

**Ni = 60**

**N = 100**

**N = 250**

**N = 500**
N = 750

N = 1000
Table 128. Power rates of Divgi Method with Correlation = 0.75.

<table>
<thead>
<tr>
<th>Ni</th>
<th>20</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>100</td>
<td>500</td>
</tr>
<tr>
<td>Power</td>
<td>Person Variance</td>
<td></td>
</tr>
</tbody>
</table>

- Ni = 20
- Ni = 60
- N = 100
- N = 250
- N = 500
On the whole, the average power for the cells is rather low. A correlation of 0.00 yielded an average power of 0.31 per cell. The average cell power is decreased to 0.23 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.18. The lowest average power is 0.13 when correlation is 0.75 between the dimensions. Since the Divgi method is not very powerful for any level of correlation, it would not be an appropriate method for detecting multidimensionality when it actually exists.

With respect to sample size in Figure 129 as the correlation between dimensions was increased the power decreased. The extreme levels in this study, 100 and 1000
examinees, were found to have the greatest degree of power. Across the test lengths, as the correlation was increased the power was decreased (Graph 130). The shorter the test the greater the power for the Divgi method, regardless of correlation level.

The patterns in Figures 131 and 132 are very similar. Both item and person variance displayed decreased power as a function of increased correlation. The smaller the variance the greater the power regardless of correlation.

Across the correlation levels as in Figures 133 through 136 the power decreased with either increased item or person variance. Also the interaction of both the distributional variabilities caused a decreased rate of power in the Divgi method with the increase in both variances.

Figure 129. Average Power rates for Divgi method relative to Sample Size and correlation.
Figure 130. Average Power rates for Divgi method relative to Test Length and correlation.

Figure 131. Average Power rates for Kelley method relative to Person Variance and correlation.

Figure 132. Average Power rates for Kelley method relative to Item Variance and correlation.
Figure 133. Average Power rates for Divgi method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 134. Average Power rates for Divgi method relative to Item Variance, Person Variance, with correlation of 0.25.

Figure 135. Average Power rates for Divgi method relative to Item Variance, Person Variance, with correlation of 0.5.
McGill Method Results and Discussion

Using the method developed by the author for this study, the power found when the correlation is less than 0.50, as noted in Figures 137 and 138, is generally large. Problems occur when the sample size is small and, to a lesser extent, when the number of questions is small. In general, however, as the item variability increases the power decreases. Yet, the power is directly proportional to person variability.

When the dimensions are perfectly uncorrelated, small sample sizes of 100 yield low power. Low power levels were also observed in sample sizes of 250 and lower when the correlation of dimensions was raised to 0.25 or even 0.50. Thus, the power is related to the sample size and the correlation between instrument dimensions.
Figure 137. Power rates of McGill Method with Correlation = 0.00.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
Figure 138. Power rates of McGill Method with Correlation = 0.25.

Ni = 20

Ni = 60

N = 100

N = 250

N = 500
N = 750

N = 1000
Figure 139. Power rates of McGill Method with Correlation = 0.50.

Ni = 20

N = 100

Ni = 60

N = 250

N = 500
N = 750

N = 1000
Figure 140. Power rates of McGill Method with Correlation = 0.75.

Ni = 20

Ni = 60

Ni = 100

Ni = 250

Ni = 500

Power vs. Person Variance

0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance

0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance

0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance

0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance

0.4 0.6 0.8 1 1.2 1.4 1.6

Power

Person Variance
Overall, the average cellular power is high for correlations less than 0.75. A correlation of 0.00 yielded an average power of 0.90 per cell. The average cell power is decreased to 0.87 when the correlation was 0.25. A correlation increase to 0.50 showed an average power of 0.75. The lowest average power is 0.47 when correlation is 0.75 between the dimensions. The McGill method should be relatively good in most cases, care would need to be taken when using few examinees on instruments with large item variance.
Figure 141 displays the sample size and correlation results. The power is the greatest for the largest sample size across all the levels of correlation. The results for test length are in Figure 142 where power decreased as correlation increased. The most power for the McGill method is found when the test length is larger.

Power related to Person and Item variance are found in Figures 143 and 144 respectively. In both cases as the correlation increased the power decreased. The power for person variance increased as the variability increased. As opposed to when item variability was increased the power for the test decreased.

Figures 145 through 146 show the interaction of person and item variability on power. When person variability increased and item variability increased the result is higher power.

Figure 141. Average Power rates for McGill method relative to Sample Size and correlation.
Figure 142. Average Power rates for McGill method relative to Test Length and correlation.

Figure 143. Average Power rates for McGill method relative to Person Variance and correlation.

Figure 144. Average Power rates for McGill method relative to Item Variance and correlation.
Figure 145. Average Power rates for McGill method relative to Item Variance, Person Variance, with correlation of 0.0.

Figure 146. Average Power rates for McGill method relative to Item Variance, Person Variance, with correlation of 0.25.

Figure 147. Average Power rates for McGill method relative to Item Variance, Person Variance, with correlation of 0.5.
Figure 148. Average Power rates for McGill method relative to Item Variance, Person Variance, with correlation of 0.5.
Chapter 5

Summary

Type I Error

To date there have been no studies to identify the best decision criterion for evaluating unidimensionality claims in the context of principal component analysis of latent trait model residuals. In this chapter, I summarize the results from the Type I error analysis first to identify the best functioning method relative to false-positives. After identifying the method with the best Type-I error rate, I present the results of the power analyses from all methods to identify the most appropriate method of finding multidimensional conditions when they actually exist. Finally, I suggest which method seems to be best overall, and the conditions under which various alternatives best operate.

In the first part of the study, I manipulated four variables in order to investigate the effects on each method’s accuracy in identifying unidimensionality. There were five levels of sample size, three levels of test length, and eight levels each of item and person variability, which resulted in 960 different variable combinations. The average Type I error rate based on my simulations is reported in Table 31 for every method for which a criterion value has been suggested. Those methods having an average error rate closest to the desired 5% level are the most suitable. In this case, the method that far outperformed the others with respect to Type I error is the Modified Kaiser, with a rate of 0.001. Indeed, the Modified Kaiser is the only method to achieve a Type I error rate less than 0.05. The Ratio of Eigenvalues rule and the Divgi method were the next best with error rates of 0.123 and 0.124, respectively. The methods accounting for 40 and 50 percent of the variance relative to measure dimension have the highest error rates. These are
followed by the Kaiser method as one of the least accurate procedures for detecting unidimensionality (i.e., a method that indicates multidimensionality when the data are truly unidimensional).

The rules which were modified or were created for the first time in this paper did not always have the best Type I error or power. This was due to poor fit of the equation relative to the independent variables of test length, sample size, item variance, and person variance. The likely reason for their poor performance is the lack of some other important independent variable which has not been taken into account at this time.

<table>
<thead>
<tr>
<th>Method</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaiser</td>
<td>0.377</td>
</tr>
<tr>
<td>Modified Kaiser</td>
<td>0.001</td>
</tr>
<tr>
<td>20% Variance</td>
<td>0.169</td>
</tr>
<tr>
<td>40% Variance</td>
<td>0.472</td>
</tr>
<tr>
<td>50% Variance</td>
<td>0.665</td>
</tr>
<tr>
<td>Modified Percentage of Variance</td>
<td>0.179</td>
</tr>
<tr>
<td>Ratio of Eigenvalues</td>
<td>0.123</td>
</tr>
<tr>
<td>Modified Ratio</td>
<td>0.240</td>
</tr>
<tr>
<td>Kelley</td>
<td>0.212</td>
</tr>
<tr>
<td>Divgi</td>
<td>0.124</td>
</tr>
<tr>
<td>McGill</td>
<td>0.225</td>
</tr>
</tbody>
</table>

Figure 149 summarizes the Type I error rates for each of the levels of sample size. The Modified Kaiser is the method with uniformly minimal error at all sample sizes.

Kelly’s method is the next best for the smallest sample size, while Divgi’s method is the next best for the sample size of 250, The Ratio of Eigenvalues performs second best at sample sizes greater than 500. It is notable that the Percentage of Variance rules did not perform well overall and that the Kaiser method performed poorly, particularly with smaller sample sizes.
In almost every single case, the Type I error rate increased as the test length increased (Figure 150). Again the 50% Proportion of Variance method shows the most error followed by the 40% Proportion of Variance method. The Modified Kaiser method again achieved the lowest Type I error rate across all sample sizes, matched only by Kelley’s method for a test length of 40. The next best methods are Divgi, Ratio of Eigenvalues, 20% Proportion of Variance, and the McGill method as the error is generally less than 25% and somewhat constant.
With respect to Type I error rates, therefore, the best method is the Modified Kaiser. The Percentage of Variance, Divgi, and Ratio of Eigenvalues, and Modified Percentage of Variance methods have Type I error rates less than 20%. The remaining methods,—the Modified Ratio of Eigenvalues, Kelley, and McGill methods—have Type I error rates between 20% and 25%.

**Power**

In the second part of the study the five variables were varied to investigate effects on method accuracy due to changing conditions in order to investigate how well the methods functioned when multidimensionality was present. There were five levels of sample size, three levels of test length, and three levels each of item and person variability along with four levels of correlation between the dimensions resulting in 540 different combinations.

The average power of the different methods when varying sample size and holding all other variables constant can be found in Table 32 and Figure 151. The original
Kaiser method is the only one with a power greater than 80% regardless of the sample size, though it does decrease slightly as the size increases. The 50% Variance rule also is also pretty powerful, except at the largest and smallest sample sizes, where it drops below 80%. With a large enough sample size, however, the McGill method has the greatest power of any of the investigated methods, above 85%, for large sample. Generally, larger samples resulted in greater power.

Table 32. Average Power of methods by sample size.

<table>
<thead>
<tr>
<th>Method</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100</td>
</tr>
<tr>
<td>Kaiser</td>
<td>0.898083</td>
</tr>
<tr>
<td>Modified Kaiser</td>
<td>0.266435</td>
</tr>
<tr>
<td>20% Variance</td>
<td>0.03038</td>
</tr>
<tr>
<td>40% Variance</td>
<td>0.468528</td>
</tr>
<tr>
<td>50% Variance</td>
<td>0.757442</td>
</tr>
<tr>
<td>Modified % Variance</td>
<td>0.139157</td>
</tr>
<tr>
<td>Ratio of Eigenvalues</td>
<td>0.085713</td>
</tr>
<tr>
<td>Modified Ratio of Eigenvalues</td>
<td>0.392741</td>
</tr>
<tr>
<td>Kelley</td>
<td>0.56662</td>
</tr>
<tr>
<td>Divgi</td>
<td>0.226694</td>
</tr>
<tr>
<td>McGill</td>
<td>0.4345</td>
</tr>
</tbody>
</table>
When I increased the number of items, almost every method became more powerful (Table 33 and Figure 152). For short tests consisting of 20 questions, the best method was the Modified ratio of Eigenvalue rule, with a power of 75%. It should be noted that none of the methods achieved a power for 20 questions above 80%. When I increased the test length to 40 questions, the Kaiser method displayed greater power than other methods. At the longest test length of 60 items, both the Kaiser and McGill methods had power levels. It is therefore difficult to pick an overall best method based on test length since some methods work well with few questions and other work better for longer tests.

Table 33. Average Power error of methods by test length.

<table>
<thead>
<tr>
<th>Method</th>
<th>20</th>
<th>40</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaiser</td>
<td>0.667489</td>
<td>0.914017</td>
<td>0.980733</td>
</tr>
<tr>
<td>Modified Kaiser</td>
<td>0.202944</td>
<td>0.5042</td>
<td>0.6786</td>
</tr>
<tr>
<td>20% Variance</td>
<td>0.007856</td>
<td>0.026533</td>
<td>0.046739</td>
</tr>
<tr>
<td>40% Variance</td>
<td>0.437861</td>
<td>0.522594</td>
<td>0.572744</td>
</tr>
<tr>
<td>50% Variance</td>
<td>0.735389</td>
<td>0.81805</td>
<td>0.856493</td>
</tr>
<tr>
<td>Modified % Variance</td>
<td>0.187683</td>
<td>0.123922</td>
<td>0.109911</td>
</tr>
<tr>
<td>Ratio of Eigenvalues</td>
<td>0.036667</td>
<td>0.014572</td>
<td>0.019756</td>
</tr>
<tr>
<td>Modified Ratio of Eigenvalues</td>
<td>0.749722</td>
<td>0.614928</td>
<td>0.614517</td>
</tr>
</tbody>
</table>
As seen in the results from Table 34 and Figure 153, as the correlation between dimensions increased, the power in detecting separate dimensions simultaneously decreased. This is a rather intuitive result, since the larger the correlation between two dimensions, the more the dimensions look similar and thus appear to be unidimensional. In almost every case, regardless of the method used, the largest drop in power occurred when changing the correlation of the dimensions from 0.5 to 0.75. This dramatic drop implies that there is some critical level of correlation bounded by these two values that results in a case where the probability of unidimensional claim in a multidimensional scenario becomes highly likely.

Table 34. Average Power of methods by correlation.

<table>
<thead>
<tr>
<th>Method</th>
<th>R=0.0</th>
<th>R=0.25</th>
<th>R=0.5</th>
<th>R=0.75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kaiser</td>
<td>0.983852</td>
<td>0.951733</td>
<td>0.858622</td>
<td>0.621022</td>
</tr>
<tr>
<td>Modified Kaiser</td>
<td>0.75403</td>
<td>0.61843</td>
<td>0.388778</td>
<td>0.086422</td>
</tr>
<tr>
<td>20% Variance</td>
<td>0.055333</td>
<td>0.024748</td>
<td>0.010911</td>
<td>0.017178</td>
</tr>
</tbody>
</table>
When the dimensions are perfectly uncorrelated, the Kaiser, 50% Variance, and McGill methods all have power over 90%. Increasing the correlation to 0.25 or 0.5 shows that the Kaiser and McGill methods are more powerful than the remaining tested methods. At the highest level of correlation of 0.75, both the Kaiser and McGill methods are the most powerful, but their powers are below 80%, indicating that they don’t work very well when the dimensions are highly correlated.

**Overall Performance**

When investigating dimensionality, as this study does, the overall performance of a method is not easy to define. For example, if one is interested in minimizing Type I
error rate, the Modified Kaiser method seems best. On the other hand, if one is interested in maximizing power, the best methods seem to be the Kaiser, 50% Percentage of Variance, Modified Ratio of Eigenvalues, and McGill methods. Thus, a compromise is in order. For simultaneously reducing the Type I error rate and maximizing the power, both the Modified Ratio of Eigenvalues and the McGill methods work well.

The Modified Kaiser rule has the best Type I error rates of any method that had a known cutoff value prior to the adjustment equation relative to the independent variables. In terms of statistical power, however, it has some of the lowest values of all the methods. Thus, it seems that the reason this method works so well at determining unidimensionality when it exists is because even when multidimensionality exists this method will signal unidimensionality much of the time. The Modified Kaiser method produces the best Type I error, by design of its model, for the methods where a new more suitable criterion value was suggested. After examination of its power levels, however, it seems to be inadequate; most of the time the power is less than 50%.

The famous and widely used Kaiser method had some of the highest power levels of any method investigated. Thus, this method is quite good at signaling multidimensionality when it exists. The downfall comes in the form of Type I error where, on average, it correctly identifies a unidimensional case less than 40% of the time. Therefore, it seems that the Kaiser method favors multidimensional decisions, regardless of whether multidimensionality actual exists.

The Modified Ratio of Eigenvalues method has an overall Type I error rate that is slightly less than 25%; additionally, its power is generally good for most levels of sample size and test length. The McGill method, similarly, has a Type I error rate that is less than
25%, which is equivalent to falsely overdimensionalizing a unidimensional test one out of every four times. The promise in the McGill method is that its power is among the best for all methods studied. The power of the McGill method increased with sample size and remained constant with an increase in correlation between the dimensions. The main difference between the McGill and Modified Ratio of Eigenvalues methods is that the average power for the Modified Ratio of Eigenvalues decreased faster than the McGill method as the correlation levels were increased.

**Future Study**

The focus of this study was twofold: to determine the optimality of general “rules of thumb” for determining dimensionality under Latent Trait Test Theory (LTTT) using Principal Component Analysis, and to establish rules of thumb for methods for which such rules do not exist. To that end, I simulated data consistent with the LTTT model being used to scale the data and upon which the residuals, which are the data subjected to the PCA, are based even though, in realistic applications of these procedures, it is likely that model-to-data misfit will occur.

A useful follow up to this study would follow the same general design, with one exception. The focus would be on the performance of the rule of thumb and bootstrap critical values in the face of data that were not generated to be consistent with the scaling model. For example, the unidimensional data could be generated according to a three-parameter latent trait model for dichotomous data. The three-parameter unidimensional LTTT model can be extended to the multidimensional case by replacing the item difficulty values with a multidimensional difficulty parameter.
One last point to be investigated, before too much in-depth exploration, is the estimation of the item and person variabilities; the use of the average of the standard deviations across iterations as in this study may not be a good estimate. It is my belief that the reason the Type I error rates for both the Modified ratio of Eigenvalues and McGill methods is not closer to 5% is that the estimated value for the standard deviation contains enough error to influence the computed critical values. This error will probably have a greater influence on the McGill method since it is a value that is less than one, where as the Modified Ratio of Eigenvalues creates a number greater than one.
References


