

Development of a Hybrid Method for Dimensionality Identification Incorporating an
Angle-Based Approach

by

Ji Zeng

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Education)
in The University of Michigan
2010

Doctoral Committee:

Joseph A. Martineau, Michigan Department of Education, Co-Chair
Professor Joanne F. Carlisle, Co-Chair
Associate Professor Kerby A. Shedden
Professor Stephen W. Raudenbush, University of Chicago

© Ji Zeng
2010

To my loving husband and supportive parents

Acknowledgements

It is really hard for me to find the correct phrases and words to accurately express my appreciation for all the help and support I have received through all these years.

First, I would like to thank my co-chair, Dr. Joseph A. Martineau, who accompanied me through the journey of this dissertation. He got me interested in the dissertation topic and has tried his best to prioritize our dissertation meetings in his busy work schedule to support me both intellectually and spiritually. I could not possibly finish this dissertation without his challenging questions and warm encouragements.

Second, I would like to thank my other co-chair, Dr. Joanne F. Carlisle. I have known her and worked for her for the past seven years. I can still remember the time when she told me that she wanted to know *my* opinion with regard to a statistical issue, even though it was my second year here and I considered myself very inexperienced. It was at that moment that I decided to learn as much as I could, so that *my* opinion would be even more valuable to her. She is also the one who always encouraged me and helped me to convey my ideas and research to the general audience in a clear and organized way, whether in presentations at research conferences or as articles published in peer-reviewed journals.

Next, I would like to thank Dr. Kerby A. Shedden, who introduced me to the statistical computing world. I took two statistical computing classes with him and fell in love with statistical computing. I would also like to thank him for his thoughtful insights

on this dissertation topic and the intellectual challenges and support he provided me along the way.

I would also like to thank Dr. Stephen W. Raudenbush, whose comments and suggestions on my dissertation not only strengthened the argument of the dissertation itself, but also strengthened my skills as a researcher. He not only provided me excellent examples of how to communicate complicated ideas to researchers who are interested in practical applications, but also served as a role model for me as a researcher pursuing high standards and rigorous logical reasoning through his questions and comments on my dissertation.

Finally and most importantly, I would like to thank my loving family members: my husband Fine-Wei Lin, and my parents Yongxin Zeng and Xia Jiang. They never stopped believing in me even when I lost confidence in myself. I would also like to thank my friend Marilyn McKee for being so supportive all along, and those who have helped me along the way: Ruth Isaia, Carla Olivares, Julia Parkinson, Patricia Everett, and Paul Stemmer.

Table of Contents

Dedication	ii
Acknowledgements	iii
List of Figures	vi
List of Tables	vii
List of Appendices	viii
List of Acronyms	ix
Glossary of Key Concepts	x
Key of Notation	xi
Abstract	xiii
Chapter I: Introduction	1
Importance of Correct Dimensionality Identification.....	1
Factor Analysis and Dimensionality Identification	4
Purpose of the Study	5
Significance of the Study	7
Chapter II: Review of the Literature	9
Formal Definitions of Dimensionality	10
Linear Exploratory Factor Analysis.....	13
Dimensionality Assessment with Dichotomous Data.....	44
The RMK Rule.....	54
Research Questions	61
Chapter III: Method	63
Objective Extensions of the RMK Rule.....	63
Data Generation	74
Evaluation Criteria	83
Chapter IV: Results	85
Identifying the Best Performing Weight Combinations for RMK-OEs	86
Methods Comparison and Development of Hybrid Methods	90
Chapter V: Discussion	123
Summary of Key Findings	124
Comparing Key Findings to Common Beliefs or Previous Findings	126
Additional Contributions of the Key Findings.....	131
Strengths and Limitations of the Study, and Possible Future Research	134
Appendices	140
References	161

List of Figures

Figure

2.1	Angle between two manifest variables in two dimensions and one dimension from Martineau and Reckase (2006, p. 25)	56
2.2	Line Graph from Reckase et al. (2000, p. 17).....	58
2.3	Graph for 2 highly correlated true factors and 10 manifest variables from Martineau and Reckase (2006, p. 30)	60
3.1	Graph for 2 highly correlated ($\rho = 0.9$) true factors and 50 manifest variables with the most complex structure simulated in this dissertation	67
3.2	Graph for 5 highly correlated ($\rho = 0.9$) true factors and 50 manifest variables with the most complex structure simulated in this dissertation	68
3.3	RMK-OE Decision Process	74
4.1	Mean RMSE with Correlation among Factors Equals 0.....	103
4.2	Mean RMSE with Correlation among Factors Equals 0.3.....	103
4.3	Mean RMSE with Correlation among Factors Equals 0.6.....	104
4.4	Mean RMSE with Correlation among Factors Equals 0.7.....	104
4.5	Mean RMSE with Correlation among Factors Equals 0.8.....	105
4.6	Mean RMSE with Correlation among Factors Equals 0.9.....	105
4.7	Mean PAI with Correlation among Factors Equals 0	109
4.8	Mean PAI with Correlation among Factors Equals 0.3	109
4.9	Mean PAI with Correlation among Factors Equals 0.6	110
4.10	Mean PAI with Correlation among Factors Equals 0.7	110
4.11	Mean PAI with Correlation among Factors Equals 0.8	111
4.12	Mean PAI with Correlation among Factors Equals 0.9	111
4.13	Mean RMSE with Simple Structure	116
4.14	Mean RMSE with Type 1 Complex Structure	116
4.15	Mean RMSE with Type 2 Complex Structure	117
4.16	Mean RMSE with Type 3 Complex Structure	117
4.17	Mean PAI with Simple Structure	119
4.18	Mean PAI with Type 1 Complex Structure	119
4.19	Mean PAI with Type 2 Complex Structure	120
4.20	Mean PAI with Type 3 Complex Structure	120

List of Tables

Table

3.1	Weight Assignment for the RMK-OE Composite Score	65
3.2	The Number of Dimensions Corresponding to the Two Slopes in Each Pair of Regression Analysis.....	72
3.3	Explanation of the Three Types of Complex Structure	78
3.4	Summary of Simulation Conditions that were Fixed and Allowed to Vary	81
4.1	Values and Rank Orders of Mean and SD of RMSE and PAI across All Simulation Conditions for Each RMK-OE Weight Combination	88
4.2	Mean and SD of RMSE and PAI across All Simulation Conditions for the Two Candidates for the Final RMK-OE Rule.....	89
4.3	Values and Rank Orders of Mean and SD of RMSE and PAI across All Simulation Conditions for Different Methods in Comparison	91
4.4	Mean and SD of RMSE and PAI for Different Hybrid Methods	94
4.5	Values and Rank Orders of Mean and SD of RMSE and PAI across All Simulation Conditions for the Ten Methods in Comparison	96
4.6	Mean Percentage of Each Method in Each Estimation Error Category across All Simulation Conditions	100

List of Appendices

Appendix

A.1	MATLAB Program Code for Bootstrap in RMK-OE	140
A.2	MATLAB Program Code for Decision Process in RMK-OE	144
A.3	MATLAB Program Code for Data Generation.....	145
A.4	MATLAB Program Code for Existing Factor Retention Rules.....	149
A.5	MATLAB Program Code for RMSE Comparison of Various Rules	152
A.6	MATLAB Program Code for PAI Comparison of Various Rules.....	155
B	Mean Percentage of Each RMK-OE Weight Combination in Each Estimation Error Category across All Simulation Conditions	160

List of Acronyms

AIC	Akaike's information criterion
BIC	Bayesian information criterion
EFA	Exploratory factor analysis
EI	Essential independence.
EV1	Eigenvalues-greater-than-one
MAP2	Minimum average partial correlation with each element squared
MAP4	Minimum average partial correlation with each element raised to the fourth power
ML	Maximum likelihood extraction method
PA	Parallel analysis
PCA	Principal component analysis
RMK	The angle-based approach first proposed by Reckase, Martineau, & Kim (2000)
RMK-OE	The objective extension of the RMK rule
SEscree	Standard error scree: an objective extension of the subjective scree test
SLI	Strong (or full) principle of local independence
ULS	Unweighted least squares extraction method
WLI	Weak (or bivariate) principle of local independence

Glossary of Key Concepts

<i>Factor pattern matrix</i>	A matrix of fixed regression coefficients. When underlying factors are orthogonal, it is equivalent to factor structure matrix.
<i>Factor structure matrix</i>	It is the product of factor pattern matrix and factor covariance matrix. It gives covariances between manifest variables and underlying factors.

Key of Notation

Note: Some symbols that only appeared once and with immediate explanation in the text are not included here.

Linear Factor Analysis:

$\boldsymbol{\mu}$	The mean vector of manifest (or observed) variables.
$\boldsymbol{\Lambda}$	True factor pattern matrix.
$\hat{\boldsymbol{\Lambda}}$	Estimated factor pattern matrix.
λ_{ij}	The element of $\hat{\boldsymbol{\Lambda}}$ at i^{th} row and j^{th} column.
\mathbf{f}	A random vector representing unobservable common factors.
$\boldsymbol{\Phi}$	Population covariance matrix of unobservable common factors (i.e., $\text{Cov}(\mathbf{f})$), which is assumed to be a positive-definite matrix.
\mathbf{e}	A random vector representing unobservable unique factors (or errors) affecting each corresponding manifest variables only.
$\boldsymbol{\Psi}$	Population covariance matrix of unobservable unique factors, which is assumed to be a diagonal matrix.
$\hat{\boldsymbol{\Psi}}$	Estimated covariance matrix of unobservable unique factors.
$\boldsymbol{\Sigma}$	Population covariance matrix of manifest variables, which has a decomposition of $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$.
$\boldsymbol{\Omega}$	Represents the part of $\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'$ (true values) in the text for simplicity.
$\hat{\boldsymbol{\Omega}}$	Estimation of $\boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}'$.
\mathbf{S}	Unbiased estimator of the population covariance matrix $\boldsymbol{\Sigma}$.
\mathbf{R}	Observed correlation matrix, and for continuous manifest variables, Pearson product-moment correlations are used.

- I** Identity matrix.
- p Number of manifest variables. Used as subscript sometimes.
- m Number of underlying factors or dimensions.
- h_i^2 Communality coefficient for i^{th} manifest variable. The true value is the corresponding diagonal element of $\Lambda\Phi\Lambda'$ (or Ω). Its estimation is the corresponding diagonal element of $(\mathbf{S} - \Psi)$.

Abstract

Correct dimensionality identification (i.e., a correct decision on the number of factors to retain) is crucial not only in educational and psychological measurement, but also in various fields such as medicine and sociology that use exploratory factor analysis (EFA) in developing theories. However, to date, no single method has been endorsed for accurate dimensionality identification. In addition, past simulation studies comparing various dimensionality identification rules have ignored the scenario where underlying dimensions are highly correlated in the range of 0.6-0.9. This range has been found to be common in educational and psychological measurement.

In this dissertation, I reviewed four commonly used dimensionality identification rules (plus one variation of one of those rules) and three uncommonly used rules developed for maximum likelihood factor analysis. In addition, I described a recently developed angle-based method and further developed this method to obviate the need for subjective graph reading. I also developed and evaluated several hybrid methods using simulation studies, which took into consideration high correlations among underlying dimensions.

The results indicated that the improved angle-based method was an indispensable component of the final hybrid method. The results also demonstrated a tendency of under-estimation of various commonly used dimensionality identification rules when the underlying dimensions were highly correlated. This calls into question the validity of previously developed theories in various fields that involved the use of EFA.

Chapter I

INTRODUCTION

While the subject of this dissertation is a technical topic, this dissertation has been written to be accessible to both technically-inclined and non-technically inclined readers. Chapters I (introduction) and V (discussion) have been written to be accessible to any researcher interested in the impact of dimensionality identification on theories in their fields. Chapter IV (results) has some minor technical elements, but should be accessible to most readers. Technical expertise is required for portions of Chapter II (literature review) and Chapter III (method) to assure that the technical aspects of the topic are adequately covered.

Importance of Correct Dimensionality Identification

Dimensions (or factors)¹ are often considered to be unobservable attributes, latent traits, or underlying constructs² that an instrument is purported to measure (McDonald, 1981; Yen & Fitzpatrick, 2006). In this dissertation, dimensionality identification is defined as the estimation of the number of dimensions underlying a set of observed (manifest)³ variables. In the field of educational measurement, correct dimensionality identification influences the validity of test equating and scaling (explained in detail

¹ In this dissertation, dimensions and factors are used interchangeably.

² In this dissertation, the term “constructs” are sometimes directly used in the text (when describing measurement related concepts) to refer to dimensions or factors.

³ In this dissertation, observed variable and manifest variable are used interchangeably.

below), which in turn affects the accuracy of test scores, and, as a result, impacts the quality of educational decision-making (Kolen & Brennan, 2004).

In many educational decision-making contexts (especially at the institutional and public policy levels), test scores from multiple test administrations are involved (Kolen & Brennan, 2004). For example, students who apply to universities may have taken the SAT at different times during the same year, and thus may have taken alternate test forms due to test security. In order to make scores from alternate forms comparable, (horizontal) equating techniques are used to adjust for differences in item difficulty (Kolen & Brennan, 2004). One of the fundamental requirements for successful equating is to have “equal constructs” across forms (Lord, 1980; Wendler & Walker, 2006), that is, the dimensionality of a set of items must be the same across forms. Therefore, accurately identifying the underlying dimensions is vital for proper equating.

Correct dimensionality identification is even more important in vertical scaling, which is defined as the process of converting several scores from different levels of an achievement test battery to a common cross-level scale (Kolen & Brennan, 2004). Recently, an increasing number of educational researchers have shown interest in using growth models to monitor students’ learning and value-added models for inferring teacher and school effectiveness (Kim & Sunderman, 2005; Linn, 2004; Raudenbush, 2004). However, many researchers have argued that since scale scores used in growth or value-added models unavoidably involve multiple dimensions, the score differences based on such scales are not comparable (Doran, 2003; Martineau, 2006; Reckase, 2004). In order to create and maintain approximately equal-interval scales (so that the scores on the vertical scale are comparable), one prerequisite is to correctly identify the underlying

dimensions of the tests that are involved; as approximately equal-interval scale is possible only when the scaling techniques are conducted for each dimension rather than as a composite across multiple dimensions (Martineau & Reckase, 2006).

Although the above discussion focuses more on test-like data (which are not continuous⁴), the importance of correctly identifying the number of underlying dimensions also pertains to continuous data. In fact, researchers have studied cognitive abilities (with both continuous and dichotomous data) for more than 60 years (Carroll, 1993), with the assumption that their studies correctly identified the number of underlying dimensions. Given the methods researchers have most frequently used in identifying the number of underlying dimensions with continuous data (e.g., Fabrigar, Wegener, MacCallum, & Strahan, 1999) and the way these methods have performed in simulation studies (e.g., Mumford, Ferron, Hines, Hogarty, & Kromrey, 2003), an important question is whether the number of underlying dimensions has been accurately uncovered. In order to enhance the validity of dimensionality assessment (or factor analysis)⁵ results in general, and to have a better foundation for a further investigation with test-like data, it is important to start with a thorough investigation of the performance of various methods with continuous data (the simplest case) regarding accurate estimation of the number of dimensions⁶. Before describing the purpose and significance of this dissertation in more details, I first introduce the relations between factor analysis and dimensionality identification.

⁴ Since test items are usually scored dichotomously or polytomously, the item response data are either binary or ordinal, thus each item response cannot be considered as a continuous variable.

⁵ In this dissertation, dimensionality assessment (or factor analysis, in exploratory sense) is defined as full dimensionality estimation, which involves the estimation of both the number of underlying dimensions and the relationships between observed variables and underlying dimensions.

⁶ Accurate specification of the relationships between underlying dimensions and observed variables is only possible when the number of dimensions to retain is accurately identified.

Factor Analysis and Dimensionality Identification

Factor analysis (linear and non-linear) approaches as well as nonparametric approaches have been widely used for dimensionality assessment (Netemeyer, et al., 2003; Swaminathan, Hambleton, & Rogers, 2007). Non-linear factor analysis approaches such as Multidimensional Item Response Theory (MIRT) have been developed to overcome problems arising from applying linear factor analysis to test-like data (i.e., binary or ordinal item response data) (Swaminathan, et al., 2007). Nevertheless, “many of the statistical estimation procedures used for MIRT are very similar to those used for [linear] factor analysis” (Reckase, 2007, p609). Nonparametric approaches for dimensionality assessment are all based on conditional association (Sijtsma & Meijer, 2007), and measures of association such as correlation or covariance are the building blocks for linear factor analysis. Therefore, a thorough investigation of dimensionality assessment methods, and more specifically, dimensionality identification rules (i.e., factor or dimension retention rules) within the framework of linear factor analysis are necessary and fundamental to the understanding of the general dimensionality identification problem.

There are two general classes of factor analysis—exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). While EFA has typically been considered more appropriate to use when researchers do not have a strong prior theory or knowledge about underlying dimensions, CFA is considered to be more appropriate when researchers have strong expectations about underlying dimensions (Tate, 2002; Thompson, 2004). Nevertheless, researchers can use EFA dimensionality identification techniques to verify whether the expected dimensionality in a CFA model is reasonable.

Therefore, although the focus of this dissertation is on EFA, the method involved here can also be applied to CFA models for dimensionality verification purposes.

When EFA is used to uncover underlying dimensions, the goal is to accurately reveal the relations between each observed variable and underlying dimensions. However, in order to achieve this, researchers need to first correctly identify the *number* of underlying dimensions, or in other words, make an accurate dimension (or factor) retention decision. Estimating the number of dimensions to retain remains one of the major challenges in factor analysis (Yang, 2008). For linear EFA, the eigenvalue-greater-than-one rule and the scree test are commonly used factor retention rules in *practice* (e.g., Carroll, 1993; Kim & Mueller, 1978; Loehlin, 2004; Sharma, 1996), while parallel analysis (PA) and minimum average partial correlation (MAP) have been found to be the best performing methods in *simulation* (e.g., Henson & Roberts, 2006; Zwick & Velicer, 1986). All of these rules tend to under-estimate the number of dimensions when underlying dimensions are highly correlated (e.g., Mumford et al., 2003). Gierl, Leighton, and Tan (2006) stated that many educational and psychological constructs are correlated in the range of 0.6 and 0.9; hence there remains a great need for a dimension retention rule to work well with data arising from highly correlated underlying dimensions.

Purpose of the Study

In previous simulation studies (i.e., Martineau & Reckase, 2006; Zeng & Martineau, 2008), the approach first proposed by Reckase, Martineau, and Kim (2000) (referred to as the RMK rule hereafter, described in detail in Chapter II) has been found to perform much better than the commonly used rules examined in those studies when the underlying dimensions were highly correlated. Unfortunately, the RMK rule, as

developed so far, has two significant disadvantages: (1) it involves subjective graph reading, and (2) it is time consuming. This prevents researchers from having a large number of replications in each combination of their simulation conditions (e.g., Zeng & Martineau, 2008).

In addition, performance of three existing, but not widely used, factor retention rules developed for maximum likelihood factor analysis (i.e., factor analysis with maximum likelihood extraction)—maximum likelihood ratio chi-square test (Jöreskog, 1967), Akaike’s information criterion (AIC) (Akaike, 1987), and Bayesian information criterion (BIC) (Sclove, 1987)⁷—have not yet been compared to other commonly used factor retention rules. This comparison also needs to be carried out. Thus, this dissertation aims to achieve three goals:

- (1) Develop an objective extension of the RMK rule and create related computer programs to obviate the need for subjective graph reading,
- (2) Evaluate the accuracy of the objective extension of the RMK rule and the three above mentioned rules for maximum likelihood factor analysis to those commonly used rules, and
- (3) Develop a hybrid method, which is based on a compilation of results from a small set of the best performing factor retention rules.

The evaluation of these rules especially takes into consideration the following two conditions in data simulation:

- (1) underlying dimensions are highly correlated, and

⁷ The details of maximum likelihood extraction, maximum likelihood ration chi-square test, AIC, and BIC can be found in Chapter II.

(2) the relations between observed variables and underlying dimensions display various complexity (i.e., different structure complexities).

I take high correlations among underlying dimensions into consideration for reasons described above. I also take different structure complexities into consideration here to make the simulated data closer to real data than has been studied in the past.

Besides simulation studies, in Chapter II, I formally introduce all commonly used dimension retention rules (along with the existing but poorly studied maximum likelihood ratio chi-square test, AIC, and BIC rules) for linear factor analysis, non-linear factor analysis, and nonparametric approaches. Through theoretical comparison of the models and assumptions, and with the simulation results of continuous data, I intend to demonstrate the advantages of the RMK rule and the hybrid method in making accurate dimension retention decisions when the underlying dimensions of the data are highly correlated.

Significance of the Study

The significance of the study derives from the importance of correct dimensionality identification, as accurate and comparable scales with desirable characteristics (e.g. linear and interval-scale measurement) can be generated only when underlying dimensions of observed variables (or test items) can be accurately identified. Different conceptualizations of underlying constructs lead to different resulting scales. In education, this then results in the alteration of the rank ordered positions of schools and students, and thus impacts group differences and program effects involving the use of these different scales (e.g., Koretz & Hamilton, 2006). Similar problems also arise in any other field that uses factor analysis in theory development.

No easy-to-implement dimension retention rules have been developed with the particular aim of tackling data with highly correlated underlying dimensions and complex structure. However, such data are common in education and psychology (Gierl et al., 2006).

This dissertation aims to enhance the usability of the RMK rule, which has been found to be promising especially when the underlying dimensions are highly correlated; and to further investigate its performance by comparing it with additional rules. In this dissertation, I also propose different hybrid methods, with or without the use of the enhanced RMK rule. This dissertation not only provides a more comprehensive examination of the performance of various existing factor retention rules and the newly proposed RMK objective extension and hybrid method, but also calls researchers' attention to the importance of accurate dimensionality identification in their theory development if EFA is involved.

Chapter II

REVIEW OF THE LITERATURE

In this chapter, I first review formal definitions of dimensions and factors, as existing techniques in dimensionality assessment are based on at least one of these definitions (Roussos, Stout, & Marden, 1998). I then introduce the model and its assumptions, and the steps and decision-points involved in linear exploratory factor analysis (EFA), as linear EFA plays a fundamental role in the development of various dimensionality assessment methods. Because the intent is for future research to extend the methods developed in this dissertation to dichotomous or polytomous⁸ data, and linear EFA is more appropriately used with continuous data, the methods of dimensionality assessment developed for non-continuous data are also reviewed to set the stage for discussion about future research. As none of the existing methods have been found to perform well when underlying dimensions are highly correlated, the attention is then shifted to the newly proposed RMK rule. Besides the description of the RMK rule, its theoretical advantages over other existing rules are also presented. The chapter ends with a list of research questions that are the focus of this dissertation.

Formal Definitions of Dimensionality

The formal definitions of dimensionality seem to first appear in McDonald (1981). McDonald (1981) defined dimensionality and the number of (common) factors

⁸ Although dichotomous and polytomous data are both mentioned here, when non-continuous data is discussed in this chapter, the focus is on dichotomous data.

based on the concept of local independence. According to McDonald (1981), there are two definitions of local independence (i.e., strong and weak). The strong (or full) principle of local independence (SLI) states that the probabilities of item responses⁹ are mutually statistically independent when conditioned on fixed values of the latent traits (McDonald, 1981; McDonald, 1999). It can be mathematically represented as follows (McDonald, 1997, p. 257; Sijtsma & Meijer, 2007, p. 721):

$$P(\mathbf{X}_i | \boldsymbol{\theta}) = \prod_{j=1}^J P(X_{ij} | \boldsymbol{\theta}), \quad (2.1)$$

where $P(\mathbf{X}_i | \boldsymbol{\theta})$ denotes the conditional probability of \mathbf{X} (item response vector) for person i (thus denoted \mathbf{X}_i) given $\boldsymbol{\theta}$ (the vector indicating latent traits or underlying dimensions). The SLI-based definition of dimensionality is thus the minimum number¹⁰ of abilities or latent traits needed to satisfy the definition of SLI under the additional assumption of monotonicity (Roussos, 1995). The monotonicity assumption means that the probability of getting an item correct is non-decreasing as the quantities representing underlying latent traits increase (Reckase, 1997).

⁹Although definitions of dimensions are given in text based on item responses, these concepts generally apply to continuous data also. When such definitions are applied to continuous data, some concepts of the exploratory factor analysis model are involved. When one common factor is assumed, the regression function of each manifest variable on the common factor can be expressed as (McDonald, 1981, p. 101): $E(X_j | F = f_i) = \mu_j + \lambda_j f_i$, where E represents the expectation function, X_j denotes the j^{th} manifest variable, F indicates the common factor with f_i being a value of F for i^{th} observation, μ_j is the mean value of the j^{th} manifest variable, and λ_j represents the regression weight of the common factor on the j^{th} manifest variable. The assumption here is that given the latent trait value f_i of F , the manifest variables are mutually statistically independent (McDonald, 1981).

¹⁰Note that McDonald did not specifically define the number of underlying dimensions as the “minimum number” of latent traits satisfying SLI or WLI (which is explained in the following paragraph), but just as the number of latent traits satisfying SLI or WLI (see McDonald, 1981, 1999). However, he did mention that researchers “can reject the hypothesis as to the number of latent traits, necessarily in a favor of a larger number” if a specified model does not fit the data adequately (McDonald, 1981, p. 116). This statement implies the use of “minimum number” of latent traits satisfying SLI or WLI. Therefore, I adopt the definition used by Roussos (1995) which specifically states the term “minimum number” here.

The strong principle of local independence implies the weak (or bivariate) principle of local independence (WLI), which states that all possible pairs of items are uncorrelated when conditioned on fixed values of latent traits (McDonald, 1981; McDonald, 1999). Mathematically, it is represented as follows (Embretson & Reise, 2000, p229; Roussos, 1995, p13; Stout et al., 1996, p332):

$$\text{Cov}\{X_j, X_k | \boldsymbol{\theta}\} = 0, \text{ for } j \neq k. \quad (2.2)$$

Here, $\text{Cov}\{ \}$ denotes covariance, and X_j and X_k represent the response vectors (for multiple respondents) on manifest variables j and k respectively. The WLI-based definition of dimensionality is thus the minimum number of abilities or latent traits needed to satisfy the definition of WLI under the additional assumption of monotonicity (Roussos, 1995). In application of linear and non-linear factor analysis, researchers do not use the SLI-based definition of dimensionality as it is almost impossible to test for SLI due to the lack of data; instead they employ the WLI-based definition of dimensionality (McDonald, 1981; Roussos, Stout, & Marden, 1998). According to McDonald (1981, p. 116), “Under assumptions of multivariate normality, the weak form of the principle implies the strong form, as well as conversely.”

Stout (1990) further developed the concept of essential independence (EI), which is an even weaker form of SLI. EI requires that as the test length grows, the magnitude of the covariances on average converges to 0 when conditioned on fixed values of latent traits. Mathematically, it is represented as follows (Sijtsma & Meijer, 2007, p. 723; Stout, 1990, p. 298):

$$\frac{\sum_{1 \leq j < k \leq N} |\text{Cov}(X_j, X_k) | \boldsymbol{\theta}|}{\binom{N}{2}} \rightarrow 0 \text{ as } N \rightarrow \infty, \quad (2.3)$$

where N indicates the number of items in a test. The EI-based definition of dimensionality is thus the minimum number of abilities or latent traits needed for EI to hold under the additional assumption of weak monotonicity (Roussos, 1995; Stout, 1990). Weak monotonicity assumption means that the sum of probabilities of getting each item correct is non-decreasing as latent traits increase; and this implies that some items may not meet the monotonicity assumptions when examined separately (Stout, 1990).

In conclusion, there are three definitions of dimensionality based on different definitions of local independence. Specifically, these are SLI-based, WLI-based, and EI-based definitions of dimensionality. According to Roussos et al. (1998), the most effective techniques in assessing dimensionality are based on at least one of these definitions. Roussos et al. (1998) further categorized these dimensionality assessment methods in two ways: (1) parametric versus nonparametric; and (2) assessing or confirming (lack of) unidimensionality versus full dimensionality estimation (i.e., estimation of the number of underlying dimensions, as well as the estimation of variable-dimension relations).

In this dissertation, I am only concerned with full dimensionality estimation, and in particular, the estimation of the number of underlying dimensions (or the performance of various factor retention rules). Moreover, only continuous data and dichotomous data are considered here, with the emphasis on continuous data as a first step toward a generalized methodology able to address continuous, dichotomous, polytomous, and mixed data. Furthermore, as mentioned in Chapter I, this dissertation only focuses on

exploratory dimensionality assessment, though the methods can be applied to the confirmatory framework. In the following sections, I review both parametric (i.e., linear and non-linear EFA techniques) and nonparametric approaches in estimating number of underlying dimensions. The purpose of this review is not to present a comprehensive list of all existing methods, but rather to demonstrate the details of some commonly used methods that are representative of each research framework.

Linear Exploratory Factor Analysis

In this section, I begin with an introduction on the linear EFA population model and its assumptions, as well as necessary steps and decision-points involved in linear EFA. A brief discussion about indeterminacy problems is also provided. I then present details of some factor retention rules commonly used in linear EFA along with the three uncommonly used (and poorly-studied) rules developed for maximum likelihood factor analysis. Finally, I summarize some simulation studies comparing commonly used factor retention rules.

Population Model and Assumptions

Before presenting the population model and related statistical assumptions for linear EFA, it is useful to state the simple heuristic assumption researchers made when conducting EFA to assess underlying dimensions of a data set: Observed variables are correlated because they measure the same latent traits (or common factors) (Kim & Mueller, 1978; McDonald, 1981), but not “because they are determined by a common cause, or linked together in a causal sequence, or related by some other theoretical mechanism” (McDonald, 1981, p. 107). This assumption applies to both linear and non-

linear EFA when used for dimensionality assessment, and is implied by the mathematical expression of the model.

In matrix notation, the linear EFA model can be written as follows (Yanai & Ichikawa, 2007, p. 261):

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{f} + \mathbf{e}, \quad (2.4)$$

where $\mathbf{x} = (x_1, \dots, x_p)'$ is a random vector of p manifest variables, $\boldsymbol{\mu} = E(\mathbf{x})$ is its mean vector, $\mathbf{f} = (f_1, \dots, f_m)'$ is a random vector representing m unobservable common factors¹¹, $\boldsymbol{\Lambda}$ is a $p \times m$ matrix of fixed regression coefficients (McDonald, 1985; Thompson, 2004) which is more properly called the *factor pattern matrix* (Reymont & Jöreskog, 1993; Thompson, 2004), and $\mathbf{e} = (e_1, \dots, e_p)'$ is a random vector representing p unobservable unique factors (or errors)¹² contributing to the variance of the p manifest variables. Note that Equation (2.4) also implies the assumption that the relations between common factors and manifest variables are linear (Kim & Mueller, 1978; McDonald, 1981; Widaman, 2007)—thus the label *linear factor analysis*.

The main focus in linear EFA is the decomposition of the covariance matrix of manifest variables. Let $\text{Cov}(\mathbf{x}) = \boldsymbol{\Sigma}$ denotes this matrix. If the following assumptions are satisfied:

¹¹ Note that in linear factor analysis, m is assumed to be smaller than p . In fact, McDonald (2000, p. 101) stated that “A useful sufficient condition for identification in a multidimensional item response model (which carries over directly from well-established factor theory, e.g., Bollen, 1989) is that for each trait there are at least two items measuring it that are *factorial simple*, in Thurstone’s (1947) classical terminology. This means that they measure only one trait, having zero loadings on all others.”

¹² This single quantity actually should be more appropriately seen as the sum of two uncorrelated quantities: measurement error and the specific part of the observed variable (Reymont & Jöreskog, 1993). Since up to now, researchers still have not found a way to differentiate the two in an EFA model, I call this quantity unique factor for convenience in this dissertation.

- (1) Unique factors are uncorrelated with each other and have a mean vector of zeroes but may have different variances, or mathematically, $E(\mathbf{e})=\mathbf{0}$ and $\text{Cov}(\mathbf{e})=\mathbf{\Psi}$, where $\mathbf{\Psi}$ is a diagonal matrix with each diagonal element representing the corresponding variance for each unique factor;
- (2) The common factors have a mean vector of zeros and a covariance matrix $\mathbf{\Phi}$ which is a positive-definite matrix, or mathematically, $E(\mathbf{f})=\mathbf{0}$ and $\text{Cov}(\mathbf{f})=\mathbf{\Phi}$;
- (3) Common factors and unique factors are uncorrelated, or mathematically, $\text{Cov}(\mathbf{f}, \mathbf{e}) = \mathbf{0}$;

then (according to Bartholomew, Steele, Moustaki, & Galbraith, 2008; Jennrich, 2007; Jöreskog, 2007; Widaman, 2007; Yanai & Ichikawa, 2007) Σ has a decomposition

$$\Sigma = \Lambda\Phi\Lambda' + \Psi. \quad (2.5)$$

The focus on the covariance matrix and zero correlation among common and unique factors indicates that linear EFA is adopting the WLI-based definition of dimensionality (see also Russos, 1995). Note that the matrix $\Lambda\Phi$ gives the covariances between each manifest variables and underlying common factors, and is termed the *factor structure matrix* (Reyment & Jöreskog, 1993). In the next three sub-sections, I present the steps, discuss indeterminacy problems, and describe decision-points involved in conducting linear EFA.

Steps

Linear exploratory factor analysis is typically done in two steps: *extraction* and *rotation* (Jennrich, 2007; Jöreskog, 2007). In the extraction step, researchers obtain an

arbitrary orthogonal solution for Equation (2.5) which assumes $\Phi = \mathbf{I}$ ¹³. Various extraction methods exist in the literature, and I choose to describe only the *unweighted least squares method* (ULS) and *maximum likelihood method* (ML), as “the majority of factor analyses ... are estimated using ML and ULS” (Jöreskog, 2007, p. 69).

Unweighted Least Squares Method (ULS)

This method looks for a solution which minimizes the following function (Jöreskog, 2007, p52; Yanai & Ichikawa, 2007, p278):

$$F_{ULS}(\Lambda, \Psi) = \frac{1}{2} tr[(\mathbf{S} - \Lambda\Lambda' - \Psi)^2], \quad (2.6)$$

where *tr* stands for trace, and \mathbf{S} is the unbiased estimator of the population covariance matrix Σ , with each element being computed as $s_{ij} = \frac{1}{N-1} \sum_{n=1}^N (x_{ni} - \bar{x}_i)(x_{nj} - \bar{x}_j)$ (note that \bar{x}_i and \bar{x}_j represent the means of i^{th} and j^{th} manifest variable respectively) (Reyment & Jöreskog, 1993, pp. 36-37).¹⁴

A special case of this method is the well-known *principal axis factoring* approach, where the elements of $\hat{\Psi}$ and $\hat{\Lambda}$ are estimated iteratively based on the estimation of each other (i.e., estimate $\hat{\Lambda}$ using $(\mathbf{S} - \hat{\Psi})$ where $\hat{\Psi} = \text{diag}(\mathbf{S} - \hat{\Lambda}\hat{\Lambda}')$), until

¹³ Here since $\Phi = \mathbf{I}$, $\Lambda\Phi\Lambda' = \Lambda\Lambda'$. For convenience, I refer to $\Lambda\Phi\Lambda'$ in Equation (2.5) as Ω hereafter.

¹⁴ Note that the constant “ $\frac{1}{2}$ ” in Equation (2.6) is not necessary, as some books do not have this quantity and only express the function as $F_{ULS} = tr[(\mathbf{S} - \Sigma)^2]$ (Bartholomew & Knott, 1999, p. 53; Basilevsky, 1994, p. 364). Recall that for initial extraction, all factors are orthogonal, thus Σ is decomposed as $\Lambda\Lambda' + \Psi$ (the quantity that is subtracted from \mathbf{S} in Equation (2.6)). Moreover, in general, the function to be minimized can be expressed as $F = \frac{1}{2} tr[(\mathbf{S} - \Sigma)\mathbf{W}^{-1}]^2$, where \mathbf{W} is a weight matrix (Yanai & Ichikawa, 2007). If $\mathbf{W} = \mathbf{I}_p$, then the expression is ULS; and if $\mathbf{W} = \mathbf{S}$, then the function to be minimized is generalized least squares (Yanai & Ichikawa, 2007).

the estimation of $\hat{\Lambda}$ converges (Basilevsky, 1994; Kim & Mueller, 1978; Reymont & Jöreskog, 1993).

If a correlation matrix is used instead of a covariance matrix for factor analysis, then \mathbf{S} in Equation (2.6) is replaced by \mathbf{R} (Jöreskog, 2007, p52). If all manifest variables are continuous, then each element of \mathbf{R} is a Pearson product-moment correlation coefficient, which is computed as $r_{ij} = s_{ij} / s_i s_j$ (note that s_i and s_j represent the standard deviations of i^{th} and j^{th} manifest variable respectively) (Reymont & Jöreskog, 1993, p. 38). A final note about ULS is that “no distributional assumptions are made” (MacCallum, Browne, & Cai, 2007, p. 164).

Maximum Likelihood Method (ML)

Unlike ULS, the ML method assumes that all manifest variables are a random sample from a multivariate normal distribution $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ (Basilevsky, 1994; Jöreskog, 2007; Kim & Mueller, 1978; Reymont & Jöreskog, 1993; Yanai & Ichikawa, 2007). Alternatively, the common factors (\mathbf{f}) and unique factors (\mathbf{e}) (see Equation (2.4)) may be assumed to be from independent normal distributions of $N(\mathbf{0}, \mathbf{I})$ ¹⁵ and $N(\mathbf{0}, \boldsymbol{\Psi})$ respectively (Bartholomew & Knott, 1999). The principle of this method is to find population parameters (under the given hypothesis) that are most likely to produce the observed sample correlation or covariance matrix (Kim & Mueller, 1978). The related hypothesis to be tested is (Basilevsky, 1994; Jöreskog, 2007) $H_0 : \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' + \boldsymbol{\Psi}$ against $H_a : \boldsymbol{\Sigma}$ is any positive definite matrix .

¹⁵ In general, it should be $N(\mathbf{0}, \boldsymbol{\Phi})$, but researchers usually assume $\boldsymbol{\Phi} = \mathbf{I}$ for the sake of convenience and computational efficacy (Basilevsky, 1994).

Different specifications of this method have been proposed in the literature (Basilevsky, 1994), and the most commonly used one was first introduced by Lawley (1940, 1941) (see Basilevsky, 1994; Jöreskog, 2007; Reymont & Jöreskog, 1993; Yanai & Ichikawa, 2007), which aims at maximizing the following log-likelihood function:

$$\log L = -\frac{n}{2}[\ln |\hat{\Sigma}| + \text{tr}(\mathbf{S}\hat{\Sigma}^{-1})], \quad (2.7)$$

where $n = N - 1$ with N representing the sample size, \mathbf{S} is as defined in Equation (2.6), and the maximum likelihood estimate of Σ under the null hypothesis is $\hat{\Sigma} = \hat{\Lambda}\hat{\Lambda}' + \hat{\Psi}$.

Maximizing Equation (2.7) is equivalent to minimizing the following function (Basilevsky, 1994; Jöreskog, 2007):

$$F_{ML} = \ln |\hat{\Sigma}| - \ln |\mathbf{S}| + \text{tr}(\mathbf{S}\hat{\Sigma}^{-1}) - p, \quad (2.8)$$

where p denotes the number of manifest variables.

Equation (2.8) can be approximated by another function where the squared differences between sample covariance matrix and estimated population covariance matrix is weighted by the product of related unique factors, i.e.,

$$F_{ML} \cong \sum_j \sum_k [(s_{jk} - \hat{\sigma}_{jk})^2 / u_j^2 u_k^2], \quad (2.9)$$

where s_{jk} and $\hat{\sigma}_{jk}$ are sample covariance and estimated population covariance between observed variables j and k respectively, and u_j^2 and u_k^2 are sample unique factors of manifest variables j and k (MacCallum et al., 2007).

From Equation (2.9), it is obvious that the ML method assigns smaller weights to squared residuals when corresponding manifest variables have larger unique factors and assigns greater weights to squared residuals when corresponding manifest variables have smaller unique factors (Kim & Mueller, 1978; MacCallum et al., 2007). Recall that for

ULS, all squared residuals are weighted equally. This difference in weighting of squared residuals is the manifestation of different assumptions about the nature of error made by the ML and ULS methods (MacCallum et al., 2007). Specifically, the ML method assumes no model error (i.e., all error is assumed to be normal theory sampling error), while the ULS method makes no explicit differentiation between sampling error and model error. This difference in error assumptions (thus the different weighting of squared residuals) has been identified as the reason ULS performs better than ML in recovering weak common factors (i.e., common factors that produce smaller covariances) (MacCallum et al., 2007). A final note concerning the ML method is that this method has been found to be robust in parameter estimation when data depart somewhat from the multivariate normal assumption (Jöreskog, 2007).

Based on the initial extraction results, a rotation step is then conducted with the aim of achieving a simple structure (Bartholomew et al., 2008; Jöreskog, 2007; Kim & Mueller, 1978). This simple structure criterion was introduced by Thurstone (1947) to gain interpretability of resulting factors, and it means that, for each manifest variable in the factor pattern matrix, the corresponding row elements only have high values (i.e., far from zero) on a few (or preferably one) column(s) (i.e., common factor(s)), and close to zero values on other columns (Kim & Mueller, 1978; Netemeyer et al., 2003; Yanai & Ichikawa, 2007).

In the rotation step, Λ and Φ are estimated from $\hat{\Omega}$ obtained from the extraction step (Jennrich, 2007). Existing rotation methods can be specified as either orthogonal or oblique, with the former keeping the common factors mutually uncorrelated and the latter allowing correlation among common factors (Netemeyer et al., 2003; Thompson, 2004;

Yanai & Ichikawa, 2007). Note that although rotation changes the values in the factor pattern matrix (Λ) and the factor covariance matrix (Φ), the communality coefficients¹⁶ (denoted as h_i^2 for manifest variable i hereafter) for each manifest variable and the reproduced data covariance matrix do not change (Bartholomew et al., 2008; Thompson, 2004). I choose to briefly describe only *varimax* and *promax* rotations here as they are the most commonly used methods for orthogonal rotation and oblique rotation respectively (Netemeyer et al., 2003; Thompson, 2004).

Varimax Rotation

This rotation method was developed by Kaiser (1958). The principle of this method is to maximize the column-wise variance of the squared factor pattern coefficients (Kaiser, 1958; Reymont & Jöreskog, 1993). Since this variance reaches a maximum only when extreme values (far from zero values and close to zero values) rather than values in the middle appear in the column, simple structure for that factor is achieved when this variance is maximized (Kaiser, 1958; Reymont & Jöreskog, 1993). The use of squared factor pattern coefficients instead of the coefficients themselves in the criterion avoids the complication of signs on the coefficients (Reymont & Jöreskog, 1993). The technical details of this method can be found in Kaiser (1958).

Promax Rotation

This method was proposed by Hendrickson and White (1964). It starts with a factor pattern matrix that has been rotated to orthogonal simple structure. The factor

¹⁶ The communality coefficients represent the portion of variance of each manifest variable that is in common with the other manifest variables, and their true values are the diagonal elements of Ω (Reymont & Jöreskog, 1993; Thompson, 2004). Their estimates, however, are the diagonal elements of $\mathbf{S} - \Psi$ (Reymont & Jöreskog, 1993).

The reason for including this statement is to show that different rotations would not change the communality, thus fixing it in the simulation study should not have any differential impact on rotation.

pattern matrix obtained from varimax rotation is usually used as the starting matrix (Hendrickson & White, 1964; Reymont & Jöreskog, 1993). Two steps are then carried out. The first step involves the definition of a target factor pattern matrix and the second step involves the computation of a transformation matrix which transforms the starting matrix into the target factor pattern matrix. The related technical details can be found in Hendrickson & White (1964).

Indeterminacy Problems with EFA

The two frequently mentioned indeterminacy problems in linear EFA are: factor rotational indeterminacy and factor score indeterminacy¹⁷ (see Basilevsky, 1994; Hair, Anderson, Tatham, & Black, 1998; McDonald, 1985; Reymont & Jöreskog, 1993; Sharma, 1996; Steiger & Schönemann, 1978; Yanai & Ichikawa, 2007). Rotational indeterminacy is sometimes called transformational indeterminacy (Reymont & Jöreskog, 1993; Timm, 2002), meaning that when at least two factors are involved, Equation (2.5) holds for any transformation of underlying factors using a nonsingular matrix \mathbf{A} (i.e., transform \mathbf{f} in Equation (2.4) into $\mathbf{A}\mathbf{f}$). The covariance matrix is then changed into $\mathbf{A}\mathbf{\Phi}\mathbf{A}'$ and if the factor pattern matrix $\mathbf{\Lambda}$ in Equation (2.5) is replaced by $\mathbf{\Lambda}\mathbf{A}^{-1}$, then $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{A}^{-1}(\mathbf{A}\mathbf{\Phi}\mathbf{A}')(\mathbf{\Lambda}\mathbf{A}^{-1})' + \mathbf{\Psi} = \mathbf{\Lambda}(\mathbf{A}^{-1}\mathbf{A})\mathbf{\Phi}\mathbf{A}'(\mathbf{A}')^{-1}\mathbf{\Lambda}' + \mathbf{\Psi} = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}' + \mathbf{\Psi}$. Therefore, different estimations of factor pattern and factor covariance matrices can still satisfy Equation (2.5) (see Reymont & Jöreskog, 1993, p. 80).

Factor score indeterminacy means that for a given factor analysis solution (a set of $\mathbf{\Lambda}$, $\mathbf{\Phi}$, and $\mathbf{\Psi}$), infinitely many factor scores can be obtained, which satisfy the model

¹⁷ This dissertation does not concern factor score computation, thus factor score indeterminacy is not described in further detail here.

expression (2.4) and all three model assumptions (Beauducel, 2007; Steiger & Schönemann, 1978; Yanai & Ichikawa, 2007). Note that these indeterminacy problems have been considered as fundamental and unsolvable by some researchers (e.g., Maraun, 1996). Nevertheless, restrictions and criteria have been proposed to obtain unique solutions based on the linear EFA model described above. For example, Thurstone's (1947) idea of "simple structure" was proposed to resolve rotational indeterminacy (Steiger & Schönemann, 1978).

Some researchers also consider the number of factors to retain to be one of the indeterminacy problems in linear EFA (e.g., Elffers, Bethlehem, & Gill, 1978). In my understanding, the problem of the number of factors to retain is different from the two indeterminacy problems mentioned above, as the WLI-based definition of dimensionality (recall that linear EFA adopts WLI-based definition of dimensionality) states that the number of dimensions is the minimum number of latent traits satisfying WLI specified in Equation (2.2); thus indeterminacy does not exist. The use of "minimum number" in the definition actually includes the suggested "minimum rank solution" by Elffers et al. (1978, p. 185) to solve their so called "*m*- indeterminacy" (p. 185) problem. In fact, all factor retention rules described in detail later in this chapter involve the identification of a "minimum number" based on different criteria. However, whether such "minimum numbers" are the correct number of underlying dimensions is another matter, and consequently, a focus of this dissertation. Before exploring the details of these factor retention rules, I describe briefly the decision-points involved in linear EFA below.

Decision-Points Involved in Linear EFA

According to Thompson (2004), five decisions are involved in linear exploratory factor analysis (EFA). However, since the fifth decision (which concerns the computation of factor scores) is only necessary when the EFA is used as a data reduction method (i.e., not as dimensionality assessment method), it is not covered here. The four remaining decisions to be made are choosing: (1) the association matrix to use, (2) the extraction method, (3) the rotation method, and (4) the number of factors to retain (Thompson, 2004). I have described the most commonly used extraction and rotation methods above, and the corresponding choices should be based on the characteristics of the data at hand, the researchers' prior belief, and the pros and cons of the methods I list above. Therefore, I only discuss the other two decisions here.

Choice of Association Matrix

Recall that the main focus in linear EFA is the decomposition of the covariance matrix of manifest variables. Note, however, that the covariance matrix of standardized manifest variables is the correlation matrix \mathbf{R} (Afifi, Clark, & May, 2004). Different factors may be obtained from linear EFA if different matrices of association are used (Thompson, 2004). Specifically, since bivariate covariance is influenced by both the bivariate correlation and the univariate variances of the two manifest variables, when a covariance matrix is used for linear EFA, some resulting factors may be a function of correlations among manifest variables while others may be a function of the variability of individual manifest variables (Thompson, 2004). Therefore, if researchers do not want their resulting factors to be influenced by the variability of manifest variables, a correlation matrix should be the association matrix for analysis.

For continuous data, “[t]he Pearson product-moment bivariate correlation matrix is the matrix of associations most commonly used in EFA” (Thompson, 2004, p. 29). Also, when the correlation matrix is decomposed, the obtained factor structure matrix is in fact the correlations between manifest variables and underlying factors (e.g., Afifi et al., 2004).

Choice of Number of Factors to Retain

Correctly identifying the number of underlying factors is crucial in EFA, as the choice of this number directly affects the estimations of Λ and Φ in Equation (2.5) (note that this number determines the number of columns in Λ , and also the number of columns and rows in Φ), and has indirect influence on the estimation of Ψ . Misspecification of this number can be categorized as either over-extraction¹⁸ (or over-estimation, i.e., when more factors than the true number are retained) or under-extraction (or under-estimation, i.e., when fewer factors than the true number are retained) (Fava & Velicer, 1992; Fava & Velicer, 1996; Wood, Tataryn, & Gorsuch, 1996).

Although the research examining the effects of under- and over-extraction is limited, especially in the sense that only orthogonal underlying factors have been examined in simulations, researchers have found that under-extraction presents a more severe problem than over-extraction (Fava & Velicer, 1996; Wood et al., 1996). Specifically, the problem of under-extraction involves loss of information (Fava & Velicer, 1996; Zwick and Velicer, 1986), and in particular, manifest variables “that should load on unextracted factors may incorrectly show loadings on the extracted factors” and “loadings for [manifest] variables that genuinely load on the extracted

¹⁸ In this dissertation, over-extraction and over-estimation are used interchangeably, and under-extraction and under-estimation are used interchangeably.

factors may be distorted” (Wood et al., 1996, p. 359). When over-extraction takes place, however, less error is associated with estimated factor pattern coefficients with true factors than in the case of under-extraction (Wood et al., 1996). Nevertheless, when over-extraction occurs and varimax rotation is applied, manifest variables may incorrectly load heavily on surplus factors (i.e., the factors that are over-extracted) rather than on the true underlying factors (Comrey, 1978). Furthermore, Wood et al. (1996) found that “When overextraction occurs, loadings on the false (i.e., surplus) factors contain substantially more error than loadings on the true factors” (p. 359).¹⁹

To avoid the problems associated with either under- or over-extraction, various factor retention rules have been proposed in the literature aiming to correctly identify the number of underlying factors. However, “there is no consensus on the appropriate criteria to use” (Hayton, Allen, & Scarpello, 2004, p. 192), and many researchers have proposed to use multiple rules simultaneously and choose a number when it appears to be consistent across multiple rules (Henson & Roberts, 2006; Kim & Mueller, 1978; Netemeyer et al., 2003; Thompson, 2004; Thompson & Daniel, 1996).

The most commonly used factor retention rules in applied research are the eigenvalues-greater-than-one rule and scree test (Carroll, 1993; Hair et al., 1998; Kim & Mueller, 1978; Loehlin, 2004; Sharma, 1996; Velicer, Eaton, & Fava, 2000). Parallel analysis (PA) and minimum average partial correlation (MAP), on the other hand, are seldom used in applied research, but have been frequently employed in simulation studies and have been found to be much more accurate than the two commonly operationally-

¹⁹ Note that the word “loadings” is used in the original paper of these studies, and this word means both factor pattern and factor structure coefficients here, as underlying factors are assumed to be orthogonal in these studies.

used methods mentioned above (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Henson & Roberts, 2006; Velicer et al., 2000; Zwick & Velicer, 1986).

When the originally proposed PA, scree test, and MAP are introduced in the next section, some of their objective extensions and variations are also briefly described. Apart from the above mentioned four rules, I also explain in detail the following methods in determining the number of factors to retain: Bartlett's chi-square test, maximum likelihood ratio chi-square test, Akaike's information criterion (AIC) and Bayesian information criterion (BIC). Both Bartlett's chi-square test and maximum likelihood ratio chi-square test are statistical tests for selecting models (Gorsuch, 1983; Velicer et al., 2000). AIC and BIC are also model selection criteria, but penalize models with more parameters to estimate in order to advantage parsimonious models (Kuha, 2004; Sclove, 1987).

Some of the above mentioned methods were developed for component analysis (which includes principal component analysis [PCA]), not for common factor analysis; for example, eigenvalue-greater-than-1 and Bartlett's test (e.g., Velicer et al., 2000). Although the empirical results of PCA and common factor analysis have sometimes been found to be similar (e.g., Velicer & Jackson, 1990; Ogasawara, 2000), common factor analysis is considered to be more appropriate method when investigating underlying dimensions (e.g., Kim & Mueller, 1998; Widaman, 2007; Yanai & Ichikawa, 2007).

The major difference between the two approaches is that PCA aims to account for as much variance as possible in the data (thus focusing more on finding a representative summarization of the data by the reduced number of composites), while the goal for common factor analysis is to explain the covariance or correlation among manifest

variables using the theory-based model where common underlying factors are assumed to account for all existing covariance or correlation among manifest variables (Kim & Mueller, 1978; Widaman, 2007). Consequently, the components obtained from PCA should not be considered as latent variables or underlying dimensions (Fabrigar et al., 1999; Yanai & Ichikawa, 2007). Despite the fact that PCA should not be used for identifying underlying dimensions of a data set, Velicer et al. (2000, p. 44) has stated that “component analysis can be viewed as a preliminary step, with the value for the number of components providing a guide in specifying the number of factors.”

Details of the Factor Retention Rules to be Studied

In this section, I describe the above listed rules in detail. They are: (1) eigenvalues-greater-than-one, (2) parallel analysis (PA) and its variations, (3) minimum average partial correlation (MAP) and one variation, (4) scree test and its objective extensions, (5) Bartlett’s chi-square test, (6) maximum likelihood ratio chi-square test, and (7) Akaike’s information criterion (AIC) and Bayesian information criterion (BIC).

(1) Eigenvalues-greater-than-one

Guttman (1954) provided three estimations of the lower bound for the number of factors problem using population correlations: (1) the number of eigenvalues greater than or equal to one when the population correlation matrix is decomposed (i.e., the diagonal elements of the matrix are unities), (2) the number of non-negative eigenvalues when the modified population correlation matrix (of which the diagonal elements are squared multiple correlation of each manifest variable with the remaining manifest variables) is decomposed, and (3) the number of non-negative eigenvalues when the modified population correlation matrix (of which the diagonal elements are the largest squared

correlation among non-diagonal elements of each row) is decomposed (see also Gorsuch, 1983; Yanai & Ichikawa, 2007).

Among these three lower bounds, the first one was found to provide the smallest estimation, and the second one was found to provide the largest estimation (Gorsuch, 1983; Guttman, 1954; Yanai & Ichikawa, 2007). Kaiser (1960) reported that when sample correlations instead of population correlations are used, the second rule listed above almost always led to over-extraction (i.e., the estimated number of factors is more than half as many as the number of manifest variables). The first rule (number of eigenvalues greater than or equal to one) was found to estimate the number of components that is between $\frac{1}{6}$ to $\frac{1}{3}$ of the number of manifest variables, which was deemed a more reasonable result by Kaiser (1960). In addition, Kaiser (1960) argued that eigenvalues-greater-than-one rule was a necessary and sufficient condition for the component to have a positive Kuder-Richardson reliability²⁰, thus he advocated the use of this rule in determining the number of components to retain. Another rationale for the use of this rule is that it is reasonable to expect a factor to explain more variance than any single original manifest variable (thus the cutoff criterion of one) (Netemeyer et al., 2003; Velicer et al., 2000).

Note that the use of eigenvalue-greater-than-one rule is associated with the decomposition of a correlation matrix with unities on the diagonal, thus this rule is seen as appropriate only for PCA (Linn, 1968; Tucker, Koopman, & Linn, 1969; Velicer et al., 2000). Application of this rule to common factor analysis is theoretically inappropriate (Hakstian & Mueller, 1973). Nevertheless, researchers do suggest retaining the same

²⁰ Note that Cliff (1988) demonstrated that this argument is flawed, as he found that the size of an eigenvalue is not directly related to the reliability of the corresponding component.

number of factors as the number of components with eigenvalues greater than one (Cliff, 1988; Linn, 1968; Velicer et al., 2000).

(2) Parallel Analysis (PA) and Its Variations

Recall that the eigenvalues-greater-than-one rule was derived by Guttman (1954) based on population correlations. In order to address its inability to reflect sampling error when dealing with observed correlations, Horn (1965) introduced a method which was later termed “parallel analysis” (Humphreys & Ilgen, 1969, p571) in determining the number of factors to retain.

This method involves generating k sets of data (from uncorrelated or independent normally distributed random numbers) with the same number of manifest variables and observations (i.e., subjects) as the “real data” under study; compute and rank order eigenvalues of each set, and then average over the k sets of eigenvalues in each rank position. The obtained average eigenvalues in each rank position are then compared to the rank ordered eigenvalues obtained from the “real data,” and the researchers are advised to retain only the number of factors where the eigenvalues in “real data” are larger than the average eigenvalues in corresponding rank position. Again, since this rule involves the decomposition of a correlation matrix with unities on the diagonal, it is seen as a method invented for PCA (e.g., Velicer et al., 2000). Nevertheless, researchers extended its use to the decomposition of reduced correlation matrices, such as having squared multiple correlations on the diagonal (Humphreys & Ilgen, 1969)²¹.

In order to avoid the trouble of generating random data when implementing parallel analysis (PA), approaches using regression equations and linear interpolation of

²¹ Crawford and Koopman (1973) found that the Humphreys and Ilgen (1969) extension of the parallel analysis is theoretically weak (comparing to the parallel analysis based on PCA) and has the tendency to fail especially when the sample size increases.

tabled eigenvalues were developed (Velicer et al., 2000). For details of the regression equation approach for factor analysis, please see Montanelli and Humphreys (1976). For details of the various regression equation approaches for PCA and their performance evaluations, please see Allen and Hubbard (1986), Lautenschlager, Lance, and Flaherty (1989), Cota, Longman, Holden, and Fekken (1991), Longman, Cota, Holden, and Fekken (1989), Keeling (2000), and Keeling and Pavur (2004). For details of the linear interpolation approach and related evaluation studies, please see Lautenschlager (1989), Cota, Longman, Holden, and Fekken (1993), and Velicer et al. (2000).

The regression equation approaches and the linear interpolation approach can only be applied within the range of the combination of the following two sets of numbers from which the original method was developed: (1) the number of observations, and (2) the number of manifest variables (Cota et al., 1993; Keeling, 2000; Lautenschlager, 1989; Longman et al., 1989). These methods were developed to avoid the need for random data generation, which was a complicated task in the past. However, sufficient computing power is now commonly available (Glorfeld, 1995), and related program code (for SPSS and SAS, the most commonly used statistical software programs) for conducting the random data generation approach of PA were provided by some researchers (e.g., Hayton et al., 2004; O'Connor, 2000). Therefore, it is more reasonable to use the random data generation approach of PA for this dissertation.

Press-Neto, Jackson, and Somers (2004) found that when data are highly correlated, mean eigenvalues of PA are more accurate than the 95th percentile eigenvalues in estimating the number of components to retain. Since this dissertation focuses on the scenario of highly correlated underlying dimensions (thus manifest variables are

unavoidably highly correlated), the mean eigenvalues of random data generation approach is used as the PA representative.

(3) *Minimum Average Partial Correlation (MAP)*

As mentioned above, besides PA, the minimum average partial correlation (MAP) method proposed by Velicer (1976) is another well-performing factor retention rule in simulation studies. According to Velicer (1976), this method was proposed for component analysis (which includes PCA) and seemed to be the most appropriate “if component analysis is employed as an alternative to factor analysis or as a first-stage solution for factor analysis” (p. 326). Specifically, this method utilizes the partial correlation matrix among manifest variables after the variation related to the extracted components have been removed. For PCA, the partial correlation matrix can be represented as follows (pp. 322-323):

$$\mathbf{R}_m^* = \mathbf{D}^{-1/2}(\mathbf{R} - \mathbf{A}\mathbf{A}')\mathbf{D}^{-1/2}, \quad (2.10)$$

where \mathbf{R}_m^* represents the partial correlation matrix among p manifest variables after m components are partialled out from \mathbf{R} (the original correlation matrix among p manifest variables), \mathbf{A} is the $p \times m$ pattern matrix resulting from the PCA (i.e., correlations between manifest variables and components), and \mathbf{D} is the diagonal of $(\mathbf{R} - \mathbf{A}\mathbf{A}')$.

The following two summary statistics are used to determine the number of components to retain (p. 323):

$$\begin{aligned} f_m &= \sum \sum (r_{ij}^*)^2 / (p(p-1)), \text{ for } i \neq j \\ f_0 &= \sum \sum r_{ij}^2 / (p(p-1)), \text{ for } i \neq j \end{aligned} \quad (2.11)$$

where r_{ij}^* and r_{ij} represent the element in row i and column j of the matrices \mathbf{R}_m^* and \mathbf{R} respectively, p denotes the number of manifest variables, f_m is the average of the squared partial correlations after the first m components are partialled out, and f_0 is the average squared correlation. The value of f_m is calculated for $m = 1$ to $(p-1)$ ²², and it ranges between 0 and 1. The rule works as follows: (1) if $f_1 > f_0$, then no component should be extracted; (2) otherwise, when the minimum value of f_m is reached, the corresponding value of m is the number of components that should be retained.

Velicer (1976) also gave an explanation of why this rule works through the examination of the general form of a partial correlation equation (p. 323):

$$r_{ij \square y} = \frac{r_{ij} - r_{iy}r_{jy}}{((1 - r_{iy}^2)(1 - r_{jy}^2))^{1/2}}, \quad (2.12)$$

where i, j represent any two of the p manifest variables, and y represents a component. According to Equation (2.12), $r_{ij \square y}$ will decrease if the numerator decreases faster than the denominator, and will increase if the numerator decreases slower than the denominator (e.g., when r_{iy} is large and r_{jy} is small, or in general, only one manifest variable has a high correlation with the component, and all other manifest variables have near zero correlations with the component). As a result, implementing this rule would ensure that at least two manifest variables have high correlations with each extracted component (Velicer, 1976).

Similar to the PA method, the MAP method was seldom used in applied research due to its computational difficulties and the fact that popular statistical software packages

²² Velicer (1976) commented that for $m = p$, the value of f_m is indeterminate, as the partial correlation matrix has zeros on the diagonal.

such as SPSS and SAS do not have these rules as built-in functions (e.g. Henson & Roberts, 2006; O'Connor, 2000). In order to make it easier to implement, researchers developed separate program code for conducting MAP in SPSS and SAS (see O'Connor, 2000). Different from the PA method, no major variations of the MAP method have been proposed in the literature. However, Velicer et al. (2000) reported that if the squares in Equation (2.11) were replaced by the fourth power, the resulting decision rule would be more accurate than the original MAP rule. In this dissertation, both the original (squared) and the improved MAP (fourth power) methods are used as MAP representatives.

(4) Scree Test and Its Objective Extensions

The rationale for the scree test rests on a method not detailed in the current literature review: percentage of variance extracted²³ (Gorsuch, 1983; Hayton et al., 2004; Hubbard & Allen, 1987). Two types of percentage of variance extracted can be computed, the percentage of *total* variance extracted is computed for PCA, and the percentage of *common* variance extracted is computed for factor analysis (Gorsuch, 1983). Since the denominator is the same across different eigenvalues (the sum of all the variances is the denominator for percentage of total variance extracted computation, and the sum of the communality estimates is the denominator for the percentage of common variance extracted computation), only the eigenvalues themselves need to be examined to get the same information (Gorsuch, 1983).

However, instead of aiming to achieve a fixed percentage of variance extracted, the scree test aims to separate substantive factors from trivial factors, and it has no

²³ This method is not included here due to two reasons. First, according to the literature, various cutoff numbers (i.e., the fixed value of percentage of variance explained or extracted) have been proposed in the literature, varying from 50% to 80% (see Bartholomew et al., 2008; Covert & McNelis, 1988; Hair et al., 1998; Netemeyer et al., 2003). Second, in essence, percentage of variance extracted is similar to Cattell's scree test, as explained in the text.

correspondence to any fixed percentage of variance extracted (Cattell, 1966). Cattell (1966) provided a detailed description and theoretical rationale for the scree test using subjective judgment (or *subjective scree test*). Specifically, this method employs a plot of eigenvalues in decreasing order, with eigenvalues on the vertical axis and the order of eigenvalues on the horizontal axis.

As both PCA and principal axis factoring methods extract components or factors by size (Gorsuch, 1983), the decreasing order of eigenvalues is thus the order of extraction. Typically, the first few eigenvalues drop sharply (or form a steep curve), and the latter eigenvalues form a relatively flat straight line, which is termed “scree” by Cattell (1966). The decision rule for the subjective scree test was first suggested to be “at the uppermost point actually on the scree” (Cattell, 1966, p.252), but was later changed to be the number immediately before this uppermost point (Cattell & Jaspers, 1967; Cattell & Vogelmann, 1977). Moreover, when at least two scree occur in the plot, Cattell (1966) suggested taking the higher scree and ignoring the lower.

Although the subjective scree test can be applied to both PCA and common factor analysis, Cattell (1966) seemed to suggest that application of the subjective scree test to PCA yielded clearer results, as the use of unities instead of communalities on the diagonal “exercise a stabilizing effect in the direction of a steady fall” (p. 266). Detailed graph reading rules were outlined by Cattell and Vogelmann (1977, see pp. 308-314) and Zoski and Jurs (1990, see p. 216).

Gorsuch (1983) pointed out two major limitations of the subjective scree test: the amount of training needed for making reliable judgment is unknown, and the subjective graph reading prevents the method from being programmed into computer code. Various

objective extensions of the scree test have been proposed in the literature²⁴, including the Cattell-Nelson-Gorsuch (CNG), the multiple regression (MR), the *t*-value index, and the standard error scree (SEscree) procedures (Nasser, Benson, & Wisenbaker, 2002), all of which are regression-based. The details of the CNG method can be found in one of the following references: Gorsuch (1983), Nasser et al. (2002), and Zoski and Jurs (1993, 1996). The details for the use of MR with or without the *t*-value index can be found in Zoski and Jurs (1993). The SEscree procedure is described in detail below, as the simulation study conducted by Nasser et al. (2002) (the only one to date that compares the performance of the four above mentioned objective extensions) found that the SEscree is the most promising objective extension of the subjective scree test.

Zoski and Jurs (1996) developed the standard error scree (SEscree) method to overcome a problem found with both the CNG method and the MR method (with or without *t*-value index): they are not applicable when the number of manifest variables is less than 6 or the number of underlying dimensions is less than 3. To implement the SEscree method, the standard error of estimate is computed for each of the $(p - 2)$ regression lines (eigenvalues are regressed on their ordinal numbers), with the first

²⁴ Cattell and Jaspers (1967) mentioned that “the Scree could readily be converted to an algebraic form and arithmetical answer by taking the first differential of the curve and finding at what point it departs significantly from zero” (p. 41). Ferré (1995) considered that “it is sufficient to determine the point [in the scree plot] where the sign of the second-order derivative changes” (p. 675). The accuracy of these two suggestions has never been investigated in the literature. In addition, Bentler and Yuan (1998) proposed a test for linear trend in the smallest eigenvalues of the correlation matrix. This test was regarded as a statistical version of the scree test by Loehlin (2004). However, when more than one scree occur in the data, this test results in the choice of the lower scree rather than the higher scree suggested by Cattell (1966). Moreover, Bentler and Yuan (1998) claimed that the goal of their paper was not to seek an objective extension of the subjective scree test. Another proposed objective method is the combination of bootstrap and a variation of the Cattell-Nelson-Gorsuch (CNG) (e.g., compares the slope of first three eigenvalues to that of eigenvalues from the third to fifth ordinal positions, instead of using the fourth to sixth eigenvalues as proposed in the original CNG method) (Hong, Mitchell, & Harshman, 2006). Since the simulation of the Hong et al. (2006) study only involves the scenario of three underlying dimensions, the performance of their proposed method needs further investigation. In summary, due to various problems associated with these methods, as stated above, I decide not to include them in the formal text.

regression using all p eigenvalues (where p indicates the number of manifest variables), the second regression using all but the first eigenvalues, the third regression using all but the first two eigenvalues, and so on. The last regression thus employs the last three eigenvalues. The standard error of estimate is computed as follows:

$$S = \sqrt{(\sum (Y - \hat{Y})^2) / (N - 2)}, \quad (2.13)$$

where Y denotes the real eigenvalues, \hat{Y} represents the estimated eigenvalue using the regression, and N is the number of eigenvalues involved in the regression computation. Each standard error of estimate is then compared to the reciprocal of the number of manifest variables (i.e., $1/p$), and the number of standard errors exceeding $1/p$ is the number of factors to retain.

As mentioned above, Nasser et al. (2002) found that the SEscree is the most promising objective extension among the four objective extensions. However, these authors reported that the conditions for the SEscree to ensure reliable results are “large sample sizes and/or large salient factor pattern coefficients” (p. 412). Since the simulation design in this dissertation involves large sample sizes and salient factor pattern coefficients, I use SEscree as the representative of the scree test.

(5) Bartlett’s chi-square test

Horn and Engstrom (1979) pointed out the similarities in the rationale for both Cattell’s (1966) scree test and Bartlett’s (1950, 1951) chi-square test: if the last $(p - m)$ eigenvalues are similar to each other in magnitude, then these roots are trivial. Bartlett’s chi-square test was developed for use in PCA (Bartlett, 1950). Specifically, the chi-square statistic for examining the significance of the left-over eigenvalues (i.e., the

eigenvalues from position $(m+1)$ to the last eigenvalue when they are in decreasing order) is computed as follows (Bartlett, 1950, p. 78):

$$\chi^2 = -\left\{N - \frac{1}{6}(2p+5) - \frac{2}{3}m\right\}(\ln R_{p-m}), \quad (2.14)$$

where R_{p-m} is computed as (see Bartlett, 1950, p. 78, p. 83; Horn & Engstrom, 1979, p. 288)

$$R_{p-m} = \frac{(\lambda_{m+1})(\lambda_{m+2})\cdots(\lambda_p)}{\left[\frac{\lambda_{m+1} + \lambda_{m+2} + \dots + \lambda_p}{p-m}\right]^{(p-m)}}, \quad (2.15)$$

and the corresponding degrees of freedom is (Bartlett, 1950, p. 84; Bartlett, 1951, p. 2; Horn & Engstrom, 1979, p. 289)

$$df = \frac{(p-m+2)(p-m-1)}{2}. \quad (2.16)$$

In Equations (2.14)-(2.16), p indicates the number of manifest variables, m represents the number of components being extracted, and λ_{m+1} is the eigenvalue at ordinal position $(m+1)$ when all eigenvalues are in decreasing order. If the last $(p-m)$ eigenvalues are approximately equal, the numerator and denominator in Equation (2.15) should be approximately equal, and then R_{p-m} should result in a value close to 1, which leads Equation (2.14) to be near 0.

Bartlett's chi-square test was often found to result in over-estimation, especially with conventional *alpha* levels at either 0.05 or 0.01 (Gorsuch, 1973; Horn & Engstrom, 1979; Hubbard & Allen, 1987). Gorsuch (1973) attributed the reason for over-estimation to the fact that the method is oriented towards PCA, where common factors and unique factors are not distinguished from one another. Bartlett's chi-square test was thus

suggested for use as an upper bound rather than the “correct” number of factors to retain (Gorsuch, 1983; Hubbard & Allen, 1987), as “If the next residual matrix is nonsignificant by Bartlett’s test, it would certainly be nonsignificant by any test that checked only for common factors” (Gorsuch, 1983, p. 154). For the simulation study in this dissertation, Bartlett’s chi-square test is not examined, as it is developed for PCA and is found to be mathematically similar to the method developed for common factor analysis—the maximum likelihood ratio chi-square test (Horn & Engstrom, 1979). Nevertheless, the detailed description of the method is included here as the majority of past simulation studies for factor retention rule comparison only considered Bartlett’s chi-square test, not the maximum likelihood ratio chi-square test.

(6) Maximum Likelihood (ML) Ratio Chi-Square Test

Horn and Engstrom (1979) presented the similarities of the mathematical expressions between Bartlett’s chi-square test and the maximum likelihood (ML) ratio chi-square test. They stated that Bartlett’s chi-square test “can be seen to be a likelihood ratio test for a model of the residual matrix” (p. 292).

This test can be easily carried out when the ML extraction method is used.²⁵

Recall that the hypothesis being tested is $H_0 : \Sigma = \Lambda\Lambda' + \Psi$ against

$H_a : \Sigma$ is any positive definite matrix. Moreover, recall that the aim of ML extraction is to maximize Equation (2.7) where log indicates natural logarithm:

$$\log L = -\frac{n}{2}[\ln |\hat{\Sigma}| + tr(\mathbf{S}\hat{\Sigma}^{-1})],$$

or equivalently, to minimize Equation (2.8):

²⁵ Jöreskog (1978) mentioned that the likelihood ratio test can be used for assessing model fit when either ML extraction method or generalized least squares (GLS) extraction method are used.

$$F_{ML} = \ln |\hat{\Sigma}| - \ln |\mathbf{S}| + \text{tr}(\mathbf{S}\hat{\Sigma}^{-1}) - p.$$

From Equation (2.7), it is easy to see that $L = e^{-\frac{n}{2}[\ln|\hat{\Sigma}| + \text{tr}(\mathbf{S}\hat{\Sigma}^{-1})]}$ for the null hypothesis, and

for the alternative hypothesis, $L = e^{-\frac{n}{2}[\ln|\mathbf{S}| + p]}$ (Jöreskog, 1967). Therefore, (-2) times the natural logarithm of the likelihood ratio (which is denoted as U_m here) between

$L = e^{-\frac{n}{2}[\ln|\hat{\Sigma}| + \text{tr}(\mathbf{S}\hat{\Sigma}^{-1})]}$ and $L = e^{-\frac{n}{2}[\ln|\mathbf{S}| + p]}$ is n times F_{ML} (i.e., $U_m = nF_{ML}$). The subscript m

indicates the number of columns in Λ (i.e., the factor pattern matrix) in the null hypothesis (Jöreskog, 1967). Note that $n = N - 1$ here, where N denotes the number of observations or subjects. When the sample size is large, U_m follows approximately a chi-square distribution with the following degrees of freedom (Jöreskog, 1967, p. 457):

$$df = \frac{1}{2}[(p - m)^2 - (p + m)]. \quad (2.17)$$

To test the null hypothesis, the value of U_m is compared to the chi-square distribution with degrees of freedom specified in Equation (2.17). If the value is significant at a specified α level, then the null hypothesis is rejected and the conclusion is that at least $(m + 1)$ common factors are needed (Jöreskog, 1967). This approach is in fact testing a specified m . However, researchers usually do not have any idea about the correct value of m when conducting EFA. The common way of using ML ratio chi-square test in EFA is thus to start from $m = 1$ to test U_m . If it is significant, then test U_m with $m = 2$, and so on, until a value of m is reached where the corresponding U_m is not significant at the specified α level (Gorsuch, 1983; Jöreskog, 1967).

This procedure may easily lead to an over-estimation due to the violation of independent events when computing probabilities (like multiple t tests) (Gorsuch, 1983; Jöreskog, 1967). Moreover, Hayashi, Bentler, and Yuan (2007) stated that over-estimation may imply that the assumptions of this method be violated somehow, and they listed four possible ways to violate the assumptions: (a) the sample size is too small to support this large sample size based test; (b) the manifest variables are not distributed as multivariate normal; (c) there is at least one zero on the diagonal of the population unique variance matrix (Ψ); and (d) when the value of m for the chi-square test exceeds the true value of m , rank deficiency and nonidentifiability of the factor pattern matrix will occur, and the consequence of this is that U_m no longer follows a chi-square distribution.

Due to its tendency to over-estimate, Gorsuch (1983) suggested using the result from the ML ratio chi-square test as an upper bound on the number of factors to retain. In this dissertation, however, I am interested in seeing how this procedure commonly seen as “over-estimating” would perform when the underlying dimensions are highly correlated. Therefore, the procedure described above is used.

(7) Akaike’s Information Criterion (AIC) and Bayesian Information Criterion (BIC)

AIC and BIC are the most commonly used penalized model selection criteria (Markon & Krueger, 2004), as both of them include a penalty term which increases with the number of parameters in the model (Sclove, 1987). Specifically, AIC takes the following form in EFA (Akaike, 1987; Markon & Krueger, 2004; Sclove, 1987):

$$AIC = (-2) \times (\text{maximum log-likelihood}) + (2mp + 2p - m^2 + m), \quad (2.18)$$

while BIC, also known as SIC (Schwarz’s information criterion, see Kuha, 2004), is expressed as follows in EFA (Markon & Krueger, 2004; Sclove, 1987):

$$BIC = (-2) \times (\text{maximum log-likelihood}) + (\ln(N) / 2) \times (2mp + 2p - m^2 + m). \quad (2.19)$$

In both Equations (2.18) and (2.19), p denotes the number of manifest variables, m is the number of factors, and N represents the number of observations or subjects. The term $(2mp + 2p - m^2 + m)$ in these equations is equal to twice the number of free parameters in the corresponding EFA model. It is obtained as follows: the total number of free parameters in an EFA when m factors are retained is $(mp + p)$; the condition that the factor pattern matrix Λ has to be column-wise orthogonal (recall that the initial extraction in EFA always assumes factors to be orthogonal) imposes $m(m-1)/2$ constraints (this amount of constraints is chosen for mathematical convenience, see Timm, 2002, pp. 501-502), the number of free parameters is thus $(mp + p - \frac{m^2}{2} + \frac{m}{2})$

(Sclove, 1987). However, the maximum value of m is constrained as the quantity

$(mp + p - \frac{m^2}{2} + \frac{m}{2})$ has to be less than $p(p+1)/2$, which is the total number of elements in the covariance (or correlation) matrix of manifest variables (Sclove, 1987). The choice of number of factors to retain is made when the corresponding model has the minimum AIC or BIC among competing models (Ichikawa, 1988; Song & Belin, 2008).

Akaike (1987) stated that AIC is “defined with parameters estimated by the method of maximum likelihood” (p. 318). Moreover, according to the explanations above, when maximum-likelihood extraction is used in EFA, both AIC and BIC can be easily computed²⁶.

²⁶ As a likelihood ratio test can be used to assess model fit in EFA when generalized least squares (GLS) extraction method is employed (with some distributional assumptions) (Jöreskog, 1978), AIC and BIC can also be computed when GLS extraction method is used and related distributional assumptions are met.

Researchers have argued that comparing to standard statistical tests such as chi-square tests, applications of AIC and BIC are more likely to result in a more parsimonious model as both selection criteria are trying to balance model fit and model complexity (Kuha, 2004). Moreover, according to Equations (2.18) and (2.19), when the sample size is large (> 8), AIC tends to favor a larger or more complex model (i.e., with more parameters) than BIC (Schwarz, 1978; Sclove, 1987). Despite their attractiveness, AIC and BIC have not been commonly used as factor retention rules in EFA, and thus not much investigation has been carried out to compare their performance to other factor retention rules. In this dissertation, I purposefully included these two rules for comparison to fill this void.

Comparison Literature on the Listed Factor Retention Rules

Although some comparisons of factor retention rules have been based on real data (e.g., Costello & Osborne, 2005), I focus on only simulation studies here; as we only know the true model with simulated data, and method comparisons make sense only when performance can be compared to the truth. Moreover, I only cover simulations from the so called “formal model” (i.e., the mathematical model for factor analysis as indicated in Equation (2.4)) when factor analysis model was used. In addition, I include here some results from simulation studies using PCA, as the majority of literature on method comparisons using simulations has been done with PCA.

In general, these studies mainly focus on the comparison of the following rules: the eigenvalues-greater-than-one rule, the subjective scree test, minimum average partial correlation (MAP), and parallel analysis (PA). The eigenvalues-greater-than one rule was found to be inaccurate. Specifically, it could either under- or over-estimate the number of

factors underlying the data (Cattell & Vogelmann, 1977; Cliff, 1988; Zwick & Velicer, 1982). Moreover, it was found to be most variable across different simulation conditions (Velicer et al., 2000; Zwick & Velicer, 1986). The subjective scree test was found to perform better than the eigenvalues-greater-than-one rule, although it was found to have a tendency of over-estimation (Hakstian, Rogers, & Cattell, 1982; Zwick & Velicer, 1986). PA was found to be the most accurate among these methods, followed closely by MAP (Velicer et al., 2000; Zwick & Velicer, 1986). When correlations among factors were taken into consideration in simulations, it was found that the accuracy of all these commonly examined factor retention rules decreased as the correlation among factors increased (Mumford et al., 2003).

The majority of the above mentioned research studies considered the following conditions in data simulation: (1) sample size, (2) number of manifest variables, (3) number of factors or components, (4) component or factor loadings, and/or (5) communality. However, correlations among underlying factors and different factor pattern complexity were often neglected in these data simulations, and none have considered these two conditions simultaneously. The highest correlation among factors considered in the Mumford et al. (2003) study was 0.5. This is lower than the 0.6-0.9 range usually found with educational and psychological constructs (Gierl et al., 2006). Moreover, most simulation studies have only examined the simple structure scenario. When factor pattern complexity was considered (see Hakstian et al. (1982)), the authors manipulated only the low to medium level loadings (rather than high loadings) for varying factor pattern complexity.

Several deficiencies of these simulation studies can thus be inferred. First, model-selection indices, such as AIC, BIC, or maximum-likelihood ratio chi-square test, have rarely been examined in factor retention rule comparison. Second, high correlation among factors, especially in the range of 0.6 to 0.9, has seldom been considered in simulations. Third, the influence of factor pattern complexity has rarely been investigated in combination with high correlations among factors, although such combination would make the simulated data more similar to real data. This dissertation aims at addressing all these issues, and the related details are provided in Chapter III (method).

Dimensionality Assessment with Dichotomous Data

This section first introduces some basics of item response theory (IRT) models in logistic form. As the normal ogive IRT form is often utilized in non-linear factor analysis approaches, the mathematical relations between parameters used in a multidimensional normal ogive IRT models and those in linear factor analysis are then presented. This is followed by a brief description of commonly used dimensionality assessment methods with dichotomous data, including non-linear exploratory factor analysis and nonparametric approaches. Selected software programs are mentioned along with each approach.

Item Response Theory (IRT) Models

“Item response theory (IRT) is a family of statistical models used to analyze test item data” (Yen & Fitzpatrick, 2006, p. 111). Such data include but are not limited to dichotomously scored data. However, only representative models concerning dichotomous data type are presented below.

Unidimensional IRT

The two main assumptions for unidimensional IRT models are unidimensionality and local independence (Kolen & Brennan, 2004; Tong, 2005). Unidimensionality means that one and only one ability is being measured by a certain test, and local independence means that responses to all items on the test are statistically independent after the examinee's ability (the attribute being measured) is taken into account.

The most widely used dichotomous IRT model is the three-parameter logistic model (3PL) (Kolen & Brennan, 2004). Its mathematical formulation is (Kolen & Brennan, 2004; Yen & Fitzpatrick, 2006, p. 114):

$$P(X_i = 1 | \theta) = c_i + \frac{1 - c_i}{1 + \exp[-Da_i(\theta - b_i)]}, \quad (2.20)$$

where $P(X_i = 1 | \theta)$ represents the probability of a person with ability θ getting item i right, where θ is often scaled to be $N(0,1)$;

D is a constant typically set to be 1.7 to make the difference between the logistic item response curve and the normal ogive no more than .01 for all values of θ ;

a_i is the item discrimination parameter for item i , and it is proportional to the slope of the item characteristic curve at the inflexion point;

b_i is the item difficulty or location parameter for item i ;

c_i is the lower asymptote for item i (often regarded as a pseudo-guessing parameter), which represents the probability for an examinee with very low ability (i.e., $\theta = -\infty$) of correctly answering the item by guessing.

If the guessing parameter c_i is set to 0 in Equation (2.20), a two parameter logistic model (2PL) is obtained (Yen & Fitzpatrick, 2006, p. 114):

$$P(X_i = 1 | \theta) = \frac{1}{1 + \exp[-Da_i(\theta - b_i)]}. \quad (2.21)$$

A one-parameter model can be obtained by setting $Da_i = 1$ in Equation (2.21). It is often referred to as the Rasch model (e.g., Embretson & Reise, 2000; Yen & Fitzpatrick, 2006). The mathematical presentation is as follows (Yen & Fitzpatrick, 2006, p. 113):

$$P(X_i = 1 | \theta) = \frac{1}{1 + \exp[-(\theta - b_i)]}. \quad (2.22)$$

Multidimensional IRT (MIRT)

Although the unidimensional models are more common in the literature and in practice, multidimensional models may better capture the reality in test data (Reckase, 2007; Yen & Fitzpatrick, 2006). The two basic forms are *compensatory* MIRT and *non-compensatory* MIRT models (Reckase, 2007). The reason for such names is easy to observe, given their mathematical representations shown below.

When a pseudo-guessing parameter (c) is included, the logistic form of the compensatory MIRT model is expressed as follows (Reckase, 2007, p. 613):

$$P(X_i = 1 | \theta_j) = c_i + (1 - c_i) \frac{e^{a_i \theta_j + d_i}}{1 + e^{a_i \theta_j + d_i}}, \quad (2.23)$$

where θ_j indicates the ability location of person j in the multidimensional ability space, a_i represents a vector of discrimination parameters (with each element corresponding to the discrimination parameter in a particular dimension), c_i is defined in the same fashion as that in Equation (2.20), and d_i is a scalar item parameter that is related to the item

difficulty. Specifically, item difficulty for item i is computed as $\frac{-d_i}{\sqrt{\sum_{k=1}^p a_{ik}^2}}$, where p is

the number of dimensions in the θ -space (Reckase, 1997). Note that the dot product of the discrimination parameter vector and the ability vector in the exponential term in Equation (2.23) make it possible for high ability in one dimension to compensate for low ability in another (to achieve same overall probability), thus the name compensatory model (Reckase, 2007; Yen & Fitzpatrick, 2006).

The mathematical formulation for non-compensatory model is as follows (Reckase, 2007, p. 615):

$$P(X_i = 1 | \boldsymbol{\theta}_j) = c_i + (1 - c_i) \prod_{\ell=1}^m \frac{e^{1.7a_{i\ell}(\theta_{j\ell} - b_{i\ell})}}{1 + e^{1.7a_{i\ell}(\theta_{j\ell} - b_{i\ell})}}, \quad (2.24)$$

where ℓ is the index for coordinate dimensions, $a_{i\ell}$ represents the item discrimination in dimension ℓ , $b_{i\ell}$ indicates the item difficulty in dimension ℓ , and c_i is again a pseudo-guessing parameter. It is clear from Equation (2.24) that low ability in any dimension results in a lower overall probability in correctly answering the item.

Mathematical Relations between Parameters in IRT Models and Linear EFA

For a compensatory normal ogive MIRT model, if the item difficulty is denoted as b_i (a scalar), item discriminations are denoted as \mathbf{a}_i (a vector), elements in factor pattern matrix are denoted as $\boldsymbol{\lambda}_i$ (a vector), and the cutoff criterion (threshold) is denoted as τ_i (a scalar), then the relations between linear factor analysis parameters and MIRT parameters (for dichotomous item responses) can be expressed as follows (Knol & Berger, 1991, p. 461; McDonald, 1997, p. 262):

$$\begin{aligned}\lambda_i &= a_i(1 + a_i'Pa_i)^{-1/2}, \text{ and} \\ \tau_i &= b_i(1 + a_i'Pa_i)^{-1/2},\end{aligned}\tag{2.25}$$

where P represents the correlation matrix of the underlying dimensions.

Linear Exploratory Factor Analysis²⁷

Linear exploratory factor analysis has been commonly used to investigate dimensionality underlying a set of dichotomous data (Swaminathan, Hambleton, & Rogers, 2007; Yen & Fitzpatrick, 2006). Software programs such as MPLUS can be used to fit linear EFA with dichotomous data, where the correlations estimated in these software programs are *tetrachoric* correlations²⁸ (Bartholomew et al., 2008). According to Muthén (1978), a continuous latent variable is assumed to be underlying each observed binary variable, and the dichotomous data is the result of applying some threshold cut as follows (see Muthén, 1978, p. 552):

$$u_i = \begin{cases} 1, & \text{if } \xi_i^* \geq \tau_i \\ 0, & \text{if } \xi_i^* < \tau_i \end{cases},\tag{2.26}$$

where $i = 1, 2, \dots, p$ indicates each observed variable, ξ_i^* represents the underlying continuous latent variable, u_i is the binary manifest variable, and τ_i indicates the threshold cut. If ξ_i^* is assumed to be multivariate normal with a mean of zero, then Equations (2.4) and (2.5) also apply here, with the exception that the expressed relation is now between the theoretical underlying continuous variables (i.e., ξ_i^*) and dimensions

²⁷ Note that although the use of tetrachoric correlation in factor analysis with dichotomous data is commonly regarded as linear factor analysis in literature, I personally agree with McDonald's (1997) argument that the use of tetrachoric correlation itself is "the multidimensional normal ogive model with a distinct parameterization" (p. 263). Nevertheless, I still adopt the categorization in common literature for the flow of this review.

²⁸ The computation of tetrachoric correlation is difficult without a computer program (McDonald, 1999). It is in principle related to the concept of continuous latent variable and threshold cut described later in the text.

(Muthén, 1978). In essence, the software MPLUS uses generalized least squares or weighted least squares estimation when conducting linear factor analysis with dichotomous data (Muthén, 1978; Muthén & Muthén, 2007). Moreover, a mean-adjusted chi-square test statistic that uses a full weight matrix is the default choice for making factor retention decisions when EFA is applied to dichotomous data (Muthén & Muthén, 2007).

Non-Linear Exploratory Factor Analysis

When linear EFA is used with dichotomous data, some problems have been found. First, the use of tetrachoric correlation coefficients tends to under-estimate the inter-item correlations (Swaminathan et al., 2007). Second, a linear relation between factor pattern coefficients and item difficulty was found, indicating the recovery of the so-called “difficulty” factor instead of a meaningful underlying latent trait (Hattie, 1985). Third, the linearity assumption (i.e., the assumption that the relation between underlying dimensions and manifest variables is linear) does not hold with dichotomous data (Swaminathan et al., 2007).

In order to address the non-linear relations between underlying dimensions and manifest dichotomously coded variables, McDonald (1967, 1997) developed a non-linear factor analysis procedure, which is implemented in the program NOHARM (Swaminathan et al., 2007). In essence, McDonald’s approach approximates the multidimensional normal ogive model by Hermite-Tchebycheff polynomials through harmonic analysis (McDonald, 1997). Although item response models are usually presented in the form of logistic functions (as shown above) due to the ease of

mathematical computations, here the normal ogive²⁹ form of item response models is preferred due to its allowance for “fairly straightforward development of the desired harmonic series” (McDonald, 1997, p. 260).

The NOHARM program employs unweighted least squares (ULS) estimation (i.e., minimizing an ULS function of the first- and second-order marginal proportions) in order to handle large data sets; and due to the use of ULS estimation, the NOHARM program does not provide a statistical significance test of the model (McDonald, 1997; Swaminathan et al., 2007). Several statistics were proposed which were assumed to have approximate chi-square distribution (Gessaroli & De Champlain, 1996; Swaminathan et al., 2007). Although these statistics were shown to actually not have the chi-square distribution, researchers still found them useful in recovering the number of underlying dimensions (Swaminathan et al., 2007). A final note on the non-linear factor analysis proposed by McDonald is that this method, like linear EFA, is also WLI-based (McDonald, 1997; Stout et al., 1996).

Another non-linear factor analysis approach is often regarded as the item factor analysis approach (Zhang & Stout, 1999a). It was first introduced by Bock and Aitkin (1981) and is implemented in the TESTFACT program. Unlike MPLUS or NOHARM program approaches, this approach is SLI-based (Roussos, 1995; Stout et al., 1996). This approach assumes a multivariate standard normal distribution of population ability, uses the tetrachoric correlations for starting values, and is implemented by the use of marginal maximum likelihood estimation (Bock, Gibbons, & Muraki, 1988). Although this method increases the computational complexity exponentially with the number of factors (thus its application is limited to five factors), the increase in computations is linear with the

²⁹ Normal ogive is the cumulative normal distribution.

number of items (thus TESTFACT is applicable for 60~100 items) (Bock et al., 1988). Moreover, this method uses the maximum likelihood ratio chi-square test as goodness-of-fit test to select models, and it is carried out as nested model comparison, adding one factor at a time (Bock et al., 1988).

Nonparametric Approaches

Here I describe two nonparametric approaches (or software programs, as each software program corresponds to each approach) for dimensionality assessment: HCA/CCPROX and DETECT. Both methods use estimates of conditional covariances (Stout et al., 1996), and can be considered EI-based (Sijtsma & Meijer, 2007; van Abswoude, van der Ark, & Sijtsma, 2004).

The HCA/CCPROX software program performs an agglomerative hierarchical cluster analysis using the unweighted pair-group of method of averages with the following proximity measure (Roussos, Stout, & Marden, 1998, p. 8; Stout et al., 1996, pp. 337-338):

$$P_{\text{cov}}(U_i, U_l) = -\frac{1}{\sum N_k} \sum_{k=0}^{n-2} N_k \hat{\text{cov}}(U_i, U_l | S_{i,l} = k) + \text{constant}, \quad (2.27)$$

where $S_{i,l}$ is the examinee's number correct score on the remaining $(n-2)$ items, N_k is the number of examinees with $S_{i,l} = k$, and $\hat{\text{cov}}$ denotes the standard maximum likelihood estimate of the covariance.

While the HCA/CCPROX procedure is seen as a sorting algorithm, DETECT is regarded as a specialized estimation procedure by Stout et al. (1996). The DETECT (short for *Dimensionality Evaluation to Enumerate Contributing Traits*, Zhang & Stout, 1999) procedure partitions a set of items into a pre-specified number of distinct clusters

in such a way that items having positive expected conditional covariances are grouped in the same cluster while items having negative expected conditional covariances are placed in different clusters (Finch & Habing, 2005; Roussos & Ozbek, 2006; van Abswoude et al., 2004). Specifically, the current version of the software DETECT tries to maximize the following DETECT index (rather than the theoretical DETECT index, Zhang & Stout, 1999, p. 241):

$$\hat{D}_*(P) = \frac{2}{n(n-1)} \sum_{1 \leq i < l \leq n} \delta_{il}(P) \overline{\text{cov}}_{U_i, U_l}^*, \quad (2.28)$$

where $\delta_{il}(P)$ equals 1 when U_i and U_l are from the same cluster and -1 otherwise, P is any partition of the items, and n indicates the number of items (p. 239). Moreover,

$$\begin{aligned} \overline{\text{cov}}_{U_i, U_l}^* &= \frac{1}{2} \left[\overline{\text{cov}}_{U_i, U_l}(S) + \overline{\text{cov}}_{U_i, U_l}(T) \right], \\ \text{where} \\ \overline{\text{cov}}_{U_i, U_l}(T) &= \sum_{k=0}^n \frac{J_k}{J} \overline{\text{cov}}(U_i, U_l | T = k), \\ \overline{\text{cov}}_{U_i, U_l}(S) &= \sum_{k=0}^{n-2} \frac{J_{U_i, U_l, k}}{J} \overline{\text{cov}}(U_i, U_l | S_{U_i, U_l} = k). \end{aligned} \quad (2.29)$$

In Equation (2.29), J denotes the total number of examinees, J_k is the number of examinees with total score k , $J_{U_i, U_l, k}$ is the number of examinees with remaining score $S_{U_i, U_l} = k$, $\overline{\text{cov}}(U_i, U_l | T = k)$ is the sample covariance in the sub-sample of examinees with total score k , and $\overline{\text{cov}}(U_i, U_l | S_{U_i, U_l} = k)$ is the sample covariance in the sub-sample of examinees with remaining score k .

It is obvious that both the HCA/CCPROX and DETECT procedures are cluster analysis approaches, with the former under the hierarchical method category and the latter under the partition method category. Like all partition methods, the number of

clusters needs to be pre-specified, and the DETECT procedure utilizes hierarchical methods for obtaining this number (Zhang & Stout, 1999). Therefore, results from the HCA/CCPROX procedure can be used as starting point for the DETECT procedure (Stout et al., 1996). A final note about these two procedures is that both have been found to perform well for data with simple structure or approximate simple structure (Roussos et al., 1998; Zhang & Stout, 1999). Simple structure items are items within one cluster that correspond to only one of the underlying dimensions; and approximate simple structure is defined as the situation where all items within one cluster have their highest discrimination on the same single dimension, but have small amounts of discrimination on other dimensions of the test (Roussos et al., 1998).

Comparison Literature of Dimensionality Assessment with Dichotomous Data

Again, I only focus on results from simulation studies here. Despite its theoretical advantage, the TESTFACT program was found to perform worse than common factor analysis with sample tetrachoric correlations (Knol & Berger, 1991). This study, however, did not examine the influence of correlation among underlying dimensions on the recovery of dimensions (particularly the number of dimensions). The simulation studies that examined such influence were focused on nonparametric methods. For example, van Abswoude et al. (2004) found that the performance of both the DETECT and HCA/CCPROX procedures deteriorated when the correlation among underlying dimensions increased. In addition, the DETECT procedure in general was found to perform better than the HCA/CCPROX procedure, but the opposite was observed when discrimination was low and the test was long.

Note that the DETECT procedure was considered as being most useful when data display approximate simple structure (Stout et al., 1996; Zhang & Stout, 1999). When structure with varying complexity was built into data simulation, Gierl et al. (2006) found that “complex structure items are more difficult to identify as the correlation between dimensions increases and sample size decreases” (p. 287). However, this study only considered the scenario of two underlying dimensions. When the performance of the DETECT procedure was compared to that of the NOHARM program (combined with an approximate likelihood ratio χ^2 test), Finch and Habing (2006) found that the DETECT procedure performed better when there were only two underlying dimensions (which echoes the Gierl et al. (2006) study), but performed worse when there were six underlying dimensions (note that the number of dimensions simulated in this study was two and six only). Moreover, the accuracy of both methods diminished when the correlation among underlying dimensions increased.

None of the above mentioned studies have considered high correlations among underlying dimensions with different levels of complex structure simultaneously. In addition, no single method has been endorsed as the method for accurately recovering the number of underlying dimensions. Therefore, researchers continue to develop new methods. One such new development was first proposed by Reckase, Martineau, & Kim (2000), and is referred to as the RMK rule hereafter.

The RMK Rule

In this section, I first describe the RMK rule in detail. I then discuss its advantages over previously mentioned rules (for both continuous data and dichotomous data).

An Angle-Based Approach: The RMK Rule

The RMK rule treats each observed variable as a vector from the origin in a factor space (i.e., a multidimensional space formed by using underlying factors as axes) (Reckase et al., 2000). In any factor space, any two vectors define a plane, and the angle between the two vectors is the angle defined in that plane. There are five scenarios for different angles between any two vectors in a plane defined by the vectors:

- (1) When two manifest variables are positively but not perfectly correlated, the angle between them is larger than 0 and less than 90 degrees.
- (2) If the two manifest variables are negatively but not perfectly correlated, the angle between them is larger than 90 and less than 180 degrees.
- (3) If they are perfectly and positively correlated, the angle is 0 degrees.
- (4) If they are perfectly and negatively correlated, the angle is 180 degrees.
- (5) When they are uncorrelated (i.e., orthogonal to each other), the angle is 90 degrees.

In general, the angle between any two manifest variables changes when the manifest variables are defined in different factor space with a different number of dimensions (Reckase et al., 2000). Moreover, the average angle among vectors increases when the number of dimensions increases (M. D. Reckase, personal communication, December 1, 2009). An obvious example is the change of angle between two vectors in one and two dimensions. When a factor analysis (or dimensionality assessment) results in retaining two factors, the observed variables are represented as vectors in a two-dimensional space; and each element of a vector indicates the relation between each observed variable and each underlying factor (see Panel A in Figure 2.1). If one factor is retained instead, then

the angle between any two vectors becomes 0 (see Panel B in Figure 2.1). Note that the axes (i.e., underlying factors) in Figure 2.1 Panel A are orthogonal to each other.

Nevertheless, they are not required to be orthogonal.

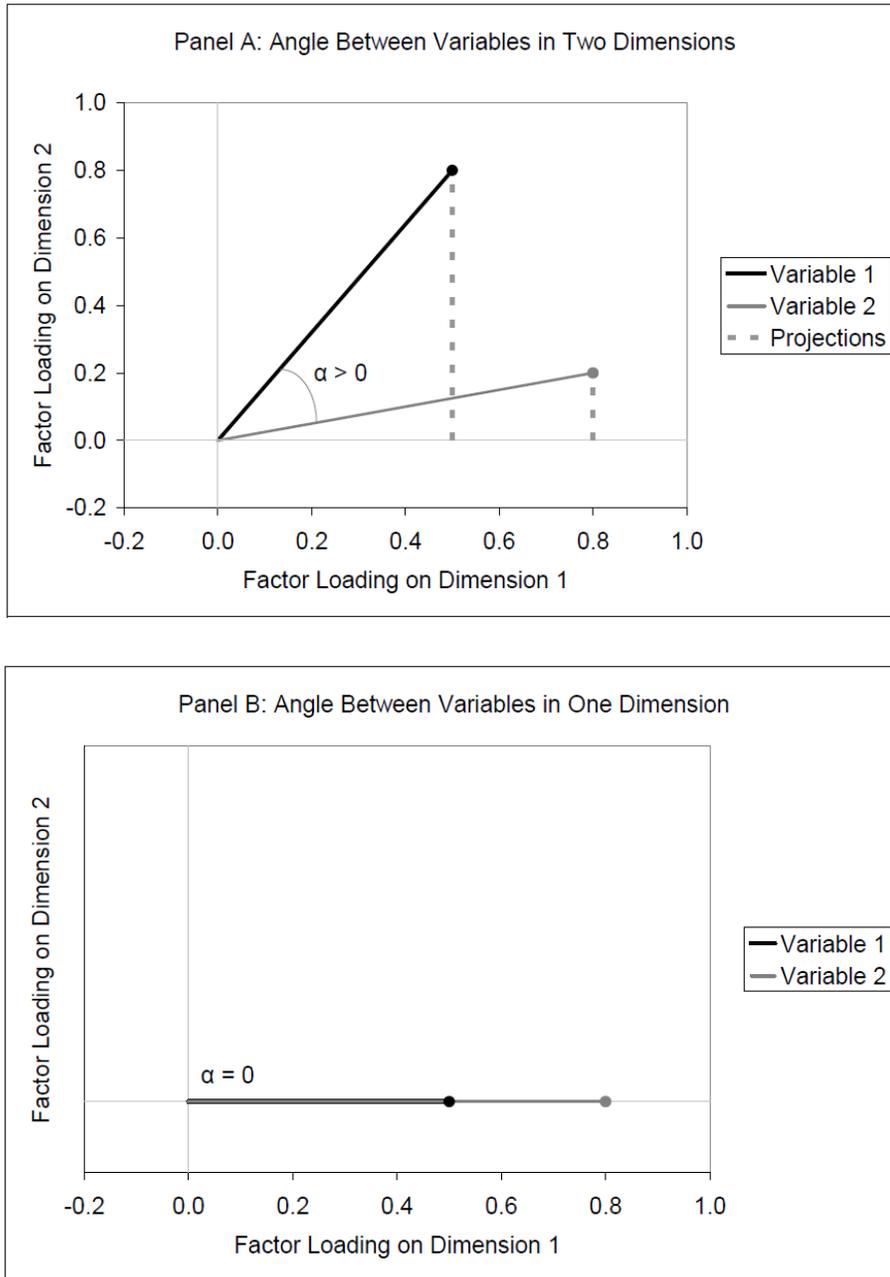


Figure 2.1. Angle between two manifest variables in two dimensions and one dimension from Martineau and Reckase (2006, p. 25)

Due to the increasing nature of average angle between randomly distributed vectors, if too few dimensions are retained, the average angle among vectors is smaller than that of the true dimensions; and if too many dimensions are retained, the average angle among vectors is larger than that of the true dimensions. As a result, when the change between average angles (the difference between the average angle of vectors in the fully specified space and the average angle of vectors in a factor space with lower dimensions) of all pairs of vectors is considered, the smallest number of dimensions that does not demonstrate appreciable angle changes can be “considered as the number of dimensions required to summarize the relationship in the data” (Reckase et al., 2000, p. 2).

Mathematically speaking, the RMK rule treats each manifest variable as a vector defined by each corresponding row in a factor pattern matrix or item discrimination matrix (denoted as A). Due to its use of factor pattern matrix or item discrimination matrix (which are computed from applying either linear factor analysis or non-linear factor analysis), the RMK rule is considered as a parametric approach (in order to be consistent with the categorization in the literature). The angle between the two vectors (A_i and A_j , both start from the origin of the space) is thus (from Reckase et al., 2000):

$$\alpha_{ij} = \arccos \frac{A_i' A_j}{(A_i' A_i)^{1/2} (A_j' A_j)^{1/2}} = \arccos \frac{\sum_{k=1}^m a_{ik} a_{jk}}{\sqrt{\sum_{k=1}^m a_{ik}^2} \sqrt{\sum_{k=1}^m a_{jk}^2}}, \quad (2.30)$$

where a_{ik} and a_{jk} are the elements of the vectors. Note that the component after “arccos” in the above equation is the formula for correlation coefficient computation.

For implementing the RMK rule in making dimension retention decisions, Reckase et al., (2000) suggested the use of a line graph, where the angle differences between the angles based on the full dimension (i.e., the highest number of dimensions that can be investigated) and those at lower dimensions were computed and plotted against the number of dimensions (see Figure 2.2 below, from Reckase et al., 2000, p. 17).

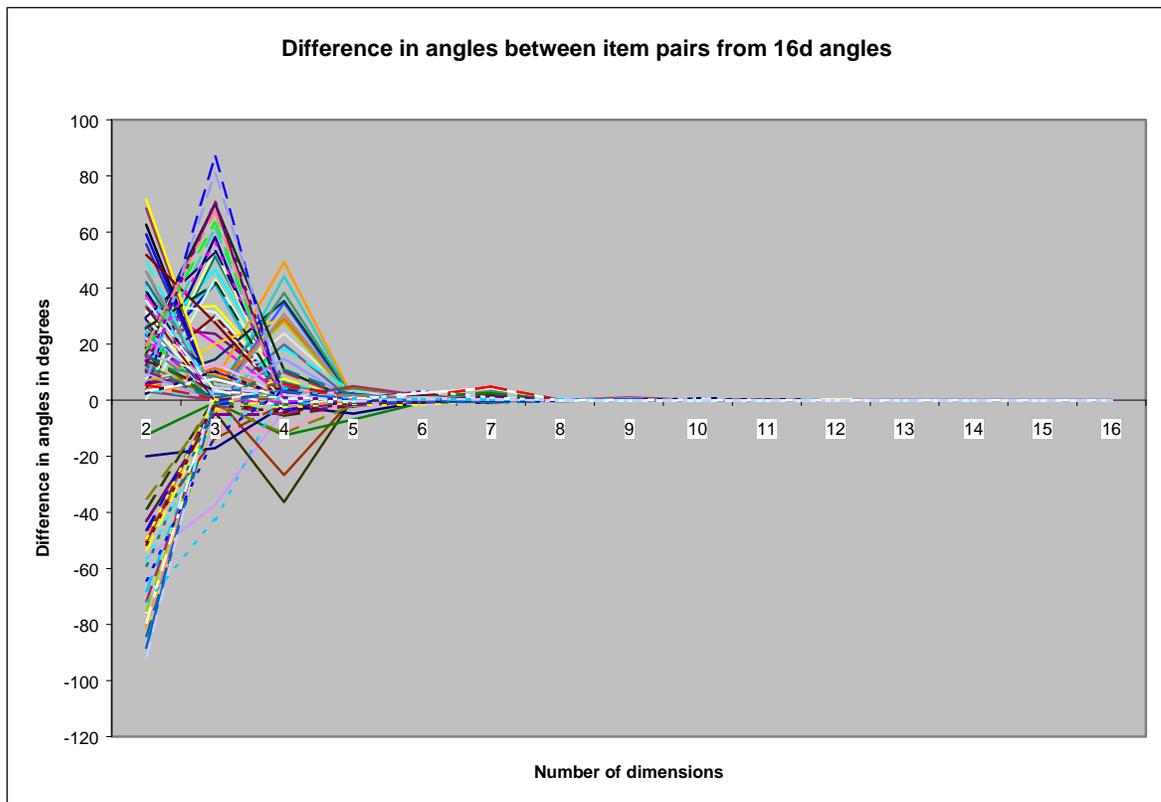


Figure 2.2. Line Graph from Reckase et al. (2000, p. 17)

As shown in Figure 2.2, there is virtually no difference from dimension=8 to dimension=16, and very small differences in angles can be observed from dimension=5 to dimension=7. Therefore, the conclusion from reading this graph is that 5 dimensions should be retained.

There is a drawback of this line graph approach: it is very easy to have too many lines in the graph; and this would make reading the graph difficult. Note that the number of lines (i.e., the number of unique angles between vectors) is $n(n-1)/2$ for n manifest variables (or vectors). In order to overcome this drawback, Martineau and Reckase (2006) proposed the use of a box and whisker plot of the angle changes between the angles computed from m dimensions to those of $(m-1)$ dimensions, and the smallest value for m is 2 (see left lower panel of Figure 2.3). In addition, for a more concise presentation of the information, Martineau and Reckase (2006) also applied the box and whisker plot to the change in average angle of each manifest variable with all others from m dimensions to that of $(m-1)$ dimensions (see right lower panel of Figure 2.3). From these two lower panels in Figure 2.3, there is a large angle change from dimension=1 to dimension=2, but negligible angle changes from dimension=2 to dimension=3 and thereafter. Therefore, the conclusion from reading Figure 2.3 is that two dimensions should be retained.

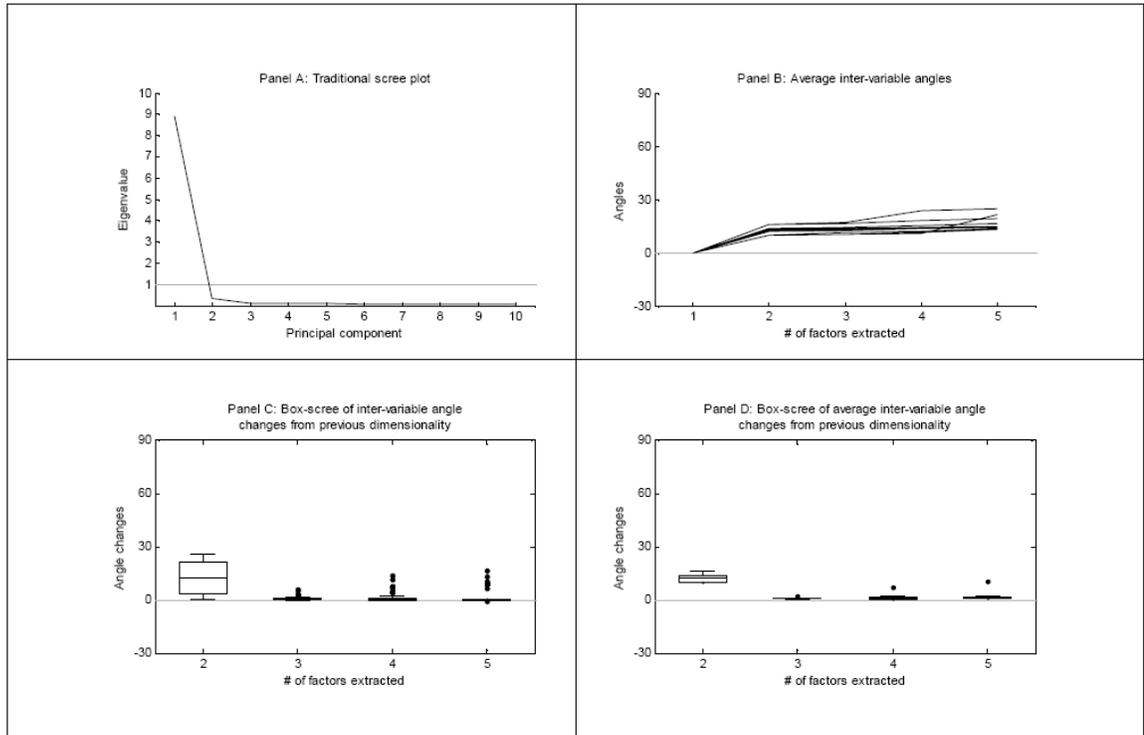


Figure 2.3. Graph for 2 highly correlated true factors and 10 manifest variables from Martineau and Reckase (2006, p. 30)

Advantages of the RMK Rule

Zeng and Martineau (2008) applied the box and whisker plot approach to implement the RMK rule with both varimax and promax rotations, and they compared the subjective graph reading results to other eight commonly used rules, including the eigenvalues-greater-than-one, the scree test, and parallel analysis. The results from implementing the RMK rule was found to be superior to other rules under consideration, especially when the underlying dimensions were highly correlated. Recall that the comparison literature for both continuous data and dichotomous data presented above has shown that commonly used methods all tend to under-estimate the number of dimensions

when correlations among underlying dimensions are high. Therefore, higher accuracy in factor retention decisions when underlying dimensions are highly correlated may be the most obvious advantage of the RMK rule over other commonly used or investigated factor retention rules.

Moreover, as mentioned above, both HCA/CCPROX and DETECT are cluster analysis approaches, thus in essence these two methods tend to group together items that have similar weights on different dimensions. Therefore, it is possible to imagine a case where there are only two underlying dimensions but three clusters. The RMK rule, however, is implemented from the factor analysis perspective. It is thus expected to uncover the true number of underlying dimensions rather than the number of clusters as weighted dimensions.

The third advantage of the RMK rule is its usability. Recall that AIC and BIC are both penalized model selection criteria (thus are theoretically more attractive). However, they have not been commonly used in either practical research or simulation evaluations. This may be because these methods were developed for maximum-likelihood extraction (Akaike, 1987; Sclove, 1987) and can at best be applied to generalized least squares extraction with certain distribution assumptions (Jöreskog, 1978). The application of the RMK rule, however, is not restricted to any extraction method, and can thus be used more widely.

Research Questions

Unfortunately, the RMK rule, as developed so far, has two significant disadvantages: (1) it involves subjective graph reading, and (2) it is time consuming. This prevents researchers from having a large number of replications in each combination of

their simulation conditions (e.g., Zeng & Martineau, 2008). This undermines the external validity of the assertion that the RMK rule is superior. This dissertation thus aims to further develop this promising dimension retention rule into an objective method for educational researchers to use in practice.

To limit this study to a feasible size, dichotomous data are not examined, although the results can theoretically be extended to dichotomous data. Besides evaluating the method for use with continuous data, this dissertation serves as a first step in expanding the evaluation to test-like data (dichotomous and polytomous data) for future research. The general research questions to be addressed in this dissertation include the following:

1. What is a good objective extension based on the subjective graph reading experience involved in implementing the RMK rule with the box and whisker plot approach proposed by Martineau and Reckase (2006)?
2. When correlations among underlying dimensions increase, how does the proposed objective extension of the RMK rule perform compared to other rules under investigation with continuous data?
3. When different levels of complex structure are involved, how does the proposed objective extension of the RMK rule perform compared to other rules under investigation with continuous data?
4. Is it possible to have a hybrid method which involves several best performing methods? If so, what does this hybrid method look like? How much improvement does this hybrid method make in comparison with those best performing methods?

Chapter III

METHOD

In this chapter, I first describe my proposed objective extension of the RMK rule, based on the box and whisker plot approach developed by Martineau and Reckase (2006). Second, I explain the data generation procedure and the data simulation conditions in detail. Third, I list evaluation criteria for comparing the performance of different factor retention rules. For this dissertation, the software program MATLAB was used, and the related program code can be found in Appendix A.

Objective Extensions of the RMK Rule

The objective extension of the RMK rule (referred to as the “RMK-OE” hereafter) is based on an informal meta-cognitive evaluation of the features affecting the subjective reading of the graphs produced in Zeng and Martineau (2008). In recreating the subjective analysis of the graphs, the cognitive tasks carried out were analyzed to determine the important characteristics of the graphs that were considered in coming to a subjective judgment about the appropriate number of factors to retain. The following three quantities played key roles in decision making when box and whisker plots of angle changes (see the lower two panels in Figure 2.3) were examined:

- (1) medians,
- (2) inter-quartile ranges (IQRs), and
- (3) trimmed ranges (i.e., when outliers in box and whisker plots are ignored).

The first quantity, medians of angle changes, was found to be more important than IQRs of angle changes, which in turn was found to be more important than the trimmed ranges of angle changes in factor retention decision making.

Taking all this information into consideration, I implement the RMK-OE on a composite score which is computed as a weighted sum of the three quantities mentioned above, with relatively higher weights given to the first two quantities. Specifically, the weight assignment is restricted to the following conditions:

- (1) the sum of weights always equals 1,
- (2) the weights are always in an increment of 0.1 (for simplicity), and
- (3) the weight of medians \geq the weight of IQRs \geq the weight of trimmed ranges.

As a result, the smallest weight that can be assigned to the medians of angle changes is 0.4. The fourteen different ways of weight assignments are shown in Table 3.1. The order of these weights is the same as in the related program code (see Appendix A.1).

Table 3.1

Weight Assignment for the RMK-OE Composite Score

Contributing Quantities	Weights for Creating RMK-OE Composite		
	Median	IQR	Trimmed Range
	0.8	0.1	0.1
	0.7	0.2	0.1
	0.6	0.3	0.1
Median, IQR, and	0.5	0.4	0.1
Trimmed Range	0.6	0.2	0.2
	0.5	0.3	0.2
	0.4	0.4	0.2
	0.4	0.3	0.3
Median only	1	0	0
	0.9	0.1	0
	0.8	0.2	0
Median and IQR only	0.7	0.3	0
	0.6	0.4	0
	0.5	0.5	0

It is obvious that given one data set, only one composite score (referred to as the “real composite” hereafter) is available for each dimension being investigated. However, it would introduce too much error if decision making is based on this one observation, due to randomness. In order to alleviate this problem, I created a 95% confidence interval

(i.e., CI between 2.5th percentile rank and 97.5th percentile rank) around each real composite based on nonparametric bootstrapped samples from angle changes (or average angle changes) computed from the one data set at hand.

To create the 95% CI, 1000 samples with the size of the original vector of angle changes (or average angle changes) were generated, with each element in each sample selected randomly *with replacement* from the full vector of angle changes (or average angle changes). Medians, IQRs, and trimmed ranges were computed for each sample, and different weighted composites were then calculated. Consequently, there were 1000 composite scores computed from each bootstrapped sample for each different weight combination. I then rank ordered these 1000 numbers, and identified the 25th (2.5th percentile rank) and 975th (97.5th percentile rank) values to construct the 95% CI. The related program code can be found in Appendix A.1³⁰.

To figure out possible rules for RMK-OE, I first examined the resulting plots of the real composites with their bootstrapped 95% CIs for each dimension examined. Two such graphs are shown below (see Figures 3.1-3.2) as examples. These two graphs were created with weights 0.6, 0.2, and 0.2 assigned to medians, IQRs, and trimmed ranges respectively. These graphs were based on data with highly correlated underlying dimensions ($\rho = 0.9$, the highest correlation among dimensions I examined in this dissertation) and the most complex structure³¹ developed for this dissertation. Note that in Figures 3.1 and 3.2, both inter-variable angle changes and average inter-variable angle changes from previous dimensionality are presented for both varimax and promax

³⁰ Note that the presenting order of the program code is not the same as the original order. I actually broke down my original program code into small pieces to be presented in Appendix A for easy reference. The correct order should be Appendix A.3, A.1, A.2, and A.4. Appendix A.5 and Appendix A.6 were run separately.

³¹ Details of different types of complex structures are explained in the data generation section.

rotations. As indicated in both figures, the number of manifest variables is 50. Therefore, $50 \cdot (50-1)/2 = 1,225$ inter-variable angle changes were computed, and 50 average inter-variable angle changes (i.e., the inter-variable angle changes averaged over each variable) were computed.

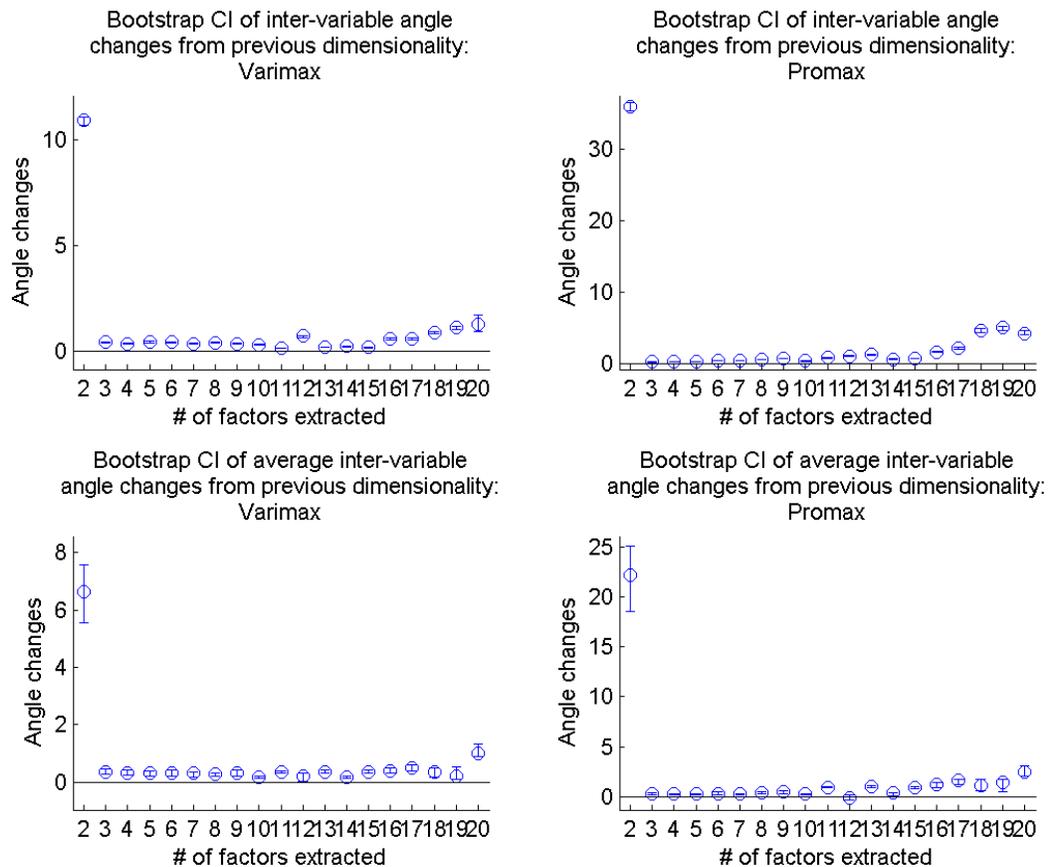


Figure 3.1. Graph for 2 highly correlated ($\rho = 0.9$) true factors and 50 manifest variables with the most complex structure simulated in this dissertation

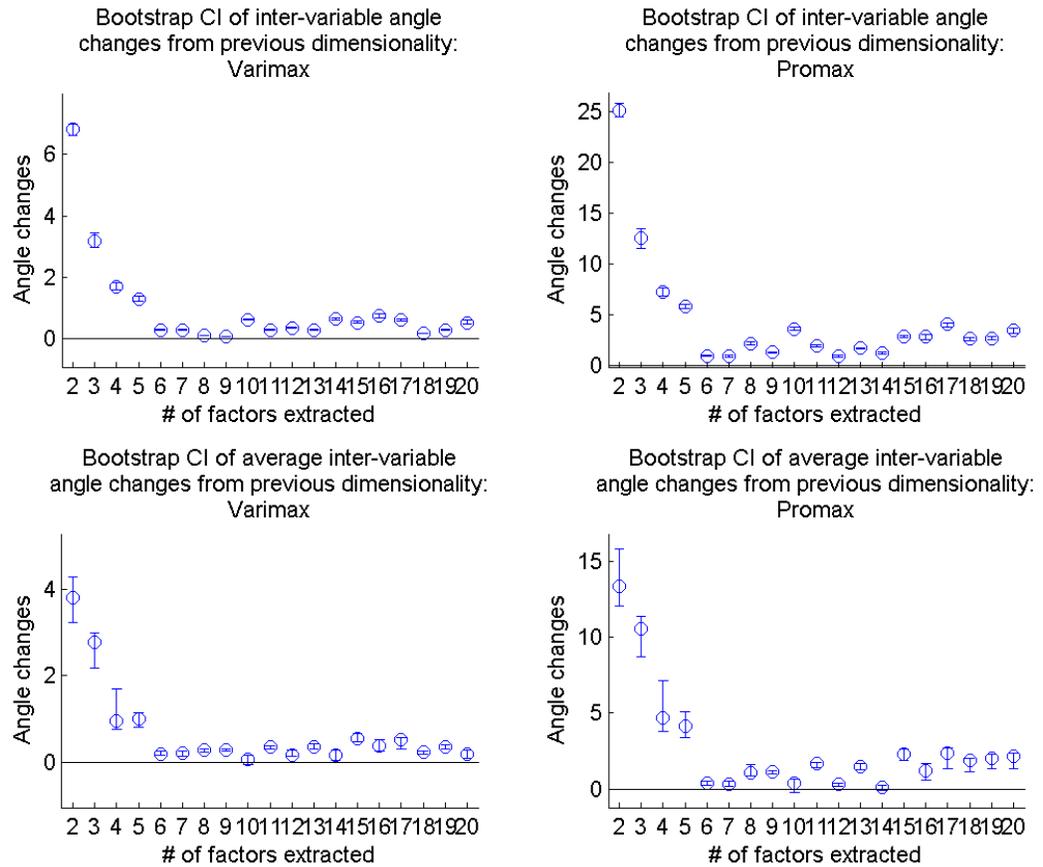


Figure 3.2. Graph for 5 highly correlated ($\rho = 0.9$) true factors and 50 manifest variables with the most complex structure simulated in this dissertation

Note that whereas Martineau and Reckase (2006) created box and whisker plots of angle changes or average angle changes (see the two lower panels in Figure 2.3), I created plots of weighted composites derived from the important characteristics identified in the Martineau and Reckase (2006) graphs (i.e., the median, IQR, and trimmed range of angle changes). Moreover, the weighted composites were plotted together with their corresponding bootstrapped 95% confidence intervals for each possible level of

dimensionality. Compared to the box and whisker plots created by Martineau and Reckase (2006), the graphs as shown in Figures 3.1-3.2 appeared to have more obvious patterns.

Upon examining these graphs, it became clear that this new graph approach is not useful when the number of true underlying dimensions is one (a limitation shared by the box and whisker approach developed by Martineau and Reckase (2006)). This limitation arises because of the nature of the analysis, i.e., the focus on change in angles. Because zero dimensions are never extracted, it is impossible to evaluate the appropriateness of retaining one dimension based on angle changes from retaining zero to retaining one dimension. Even if it were possible to retain zero dimensions, the angles would be undefined at zero dimensions, and all angles at one dimension are equal to zero, making the analysis of angle changes impossible. Therefore, a separate rule is used to determine whether the true underlying dimensionality is equal to one. This rule is explained in detail later.

Under most of the simulation conditions (explained in detail in the data generation section) when the number of true underlying dimensions ≥ 2 , I observed an obvious drop on the left side, a leveling-off trend in the middle, and a small upward trend towards the right side in each panel (see Figures 3.1 and 3.2, for example).

The obvious drop on the left side of each panel appears to be caused by the exhaustion of true underlying dimensionality. The leveling-off trend in the middle appears to be caused by small and random angle changes after the exhaustion of true dimensionality. The upward trend towards the right of each panel appears to be caused by noise when considerably too many dimensions have been retained.

In general, this pattern was most obvious for the average inter-variable (i.e., manifest variable) angle changes with varimax rotation (the panel on the lower left side of each graph shown in either Figure 3.1 or Figure 3.2). The possible reason for varimax rotation demonstrating a clearer overall pattern than promax rotation may be because “[o]blique rotations could shift some of the common error variance into the intercorrelations of factors” (Lee & Comrey, 1979), thus distorting the relations among real underlying dimensions.

Based on the observed patterns mentioned above and some preliminary analysis results of the other methods under investigation, I defined the RMK-OE for each weight combination to be composed of the following two cases: (1) dimension = 1, and (2) dimension ≥ 2 .

The RMK-OE retains only one factor when the following two conditions are met simultaneously:

- (1) the minimum value of the results from maximum-likelihood ratio chi-square test (χ_{ML}^2), AIC, and BIC is 1, and
- (2) the mode of the three (χ_{ML}^2 , AIC, and BIC) is less than or equal to 2.

This works because, according to preliminary data analysis (and confirmed with the full set of replications), χ_{ML}^2 , AIC, and BIC never produced results as dimension=1 when in fact the number of true underlying dimensions is greater than or equal to 2. Other methods under investigation, however, produced such results (i.e., dimension=1) when the underlying dimensions (≥ 2) were highly correlated. When the number of true underlying dimensions is one, the minimum value after applying these three rules is always one. However, sometimes some (but not all) of the three rules would result in

dimension=2. Therefore, both conditions specified above are needed for accurate estimation when the true underlying number of dimensions is one.

If the two conditions described above are met simultaneously, the analysis concludes at dimension=1. However, when the two conditions are not met simultaneously (i.e., when the identification for dimension=1 fails), the analysis continues using a separate set of rules defined for the second case (i.e., dimension \geq 2).

For identifying underlying dimensionality under this second case, the real composite scores (demonstrated as circles in Figures 3.1-3.2) and the corresponding 2.5th and 97.5th percentile rank composite scores from the nonparametric bootstrapped samples (indicated as horizontal short lines in Figures 3.1-3.2) are used as data points in a regression analysis as dependent variables, with dimensionality ($d=2, 3, \dots, n$) as the predictor. In other words, for each weight combination, there are three data points for each dimension under investigation: 2.5th percentile rank composite score, real composite score, and 97.5th percentile rank composite score.

This regression approach in fact examines two slopes at a time, with the two slopes obtained from regression analysis using composite scores based on angle changes as dependent variables, and the corresponding number of dimensions as independent variables. Assuming that the maximum number of true underlying dimensions being considered is m , the number of dimensions corresponding to the two slopes in each pair of regression analyses is summarized in Table 3.2.

Table 3.2

The Number of Dimensions Corresponding to the Two Slopes in Each Pair of Regression Analyses

Retention Decision	Dimensions of Contributing Data Points	
	for Calculating Slope 1	for Calculating Slope 2
Being Tested		
Dimension=2	Dimension=2, 3	Dimension=3, 4, ..., m
Dimension=3	Dimension=2, 3, 4	Dimension=4, 5, ..., m
...
Dimension= $m-2$	Dimension=2, 3, 4 ..., $m-2$	Dimension= $m-2$, $m-1$, m
Dimension= $m-1$	Dimension=2, 3, 4, ..., $m-2$, $m-1$	Dimension= $m-1$, m

Note. This method is for dimensions ≥ 2 , thus the first retention decision being tested is for dimension = 2.

The following RMK-OE rules for dimension=2 to dimension=10 (the range of dimensions examined in this dissertation other than dimension=1) were based on the observation of graphs to arrive at the nearest estimated underlying dimensionality as the true underlying dimensionality simulated in the corresponding data. Based on performance in accurately determining underlying dimensionality, the following quantities (for evaluating dimension= k) are used in setting the rules:

- slope 1_k ,
- slope 2_k ,
- increment in slope2 (slope $2_{k+1} - \text{slope}2_k$), or inc 2_k , and

- absolute value of increment difference of slope2 ($|\text{inc2}_{k+1} - \text{inc2}_k|$), or incdif2_k .

The best estimate of true underlying dimensionality is the smallest dimensionality where either one of the following two conditions is met:

(1) $\text{slope1}_k < -5$ and $\text{slope2}_k \geq -0.00075$

(2) $\text{slope1}_k < 0$ and $\text{slope2}_k \geq -0.00075$ and $\text{inc2}_k < 0.0015$ and $\text{incdif2}_k < 0.0005$

The condition for slope1_k specified in (1) was included here to have an accurate estimation for the dimension=2 scenario. The two conditions concerning inc2_k and incdif2_k specified in (2) were defined to capture the leveling-off pattern (across at least three consecutive numbers of dimensions) found in graph readings as described previously.

The threshold values indicated above were specified based on observations with the simulated data across all conditions³². However, in order to avoid the RMK-OE rule being too data dependent, only threshold values with the last decimal value being 5 or 0 were considered. The related program code can be found in Appendix A.2. The decision process of the RMK-OE is depicted in Figure 3.3.

³² Statistical tests such as t-test for slopes were evaluated, but were found to be useless in defining threshold values.

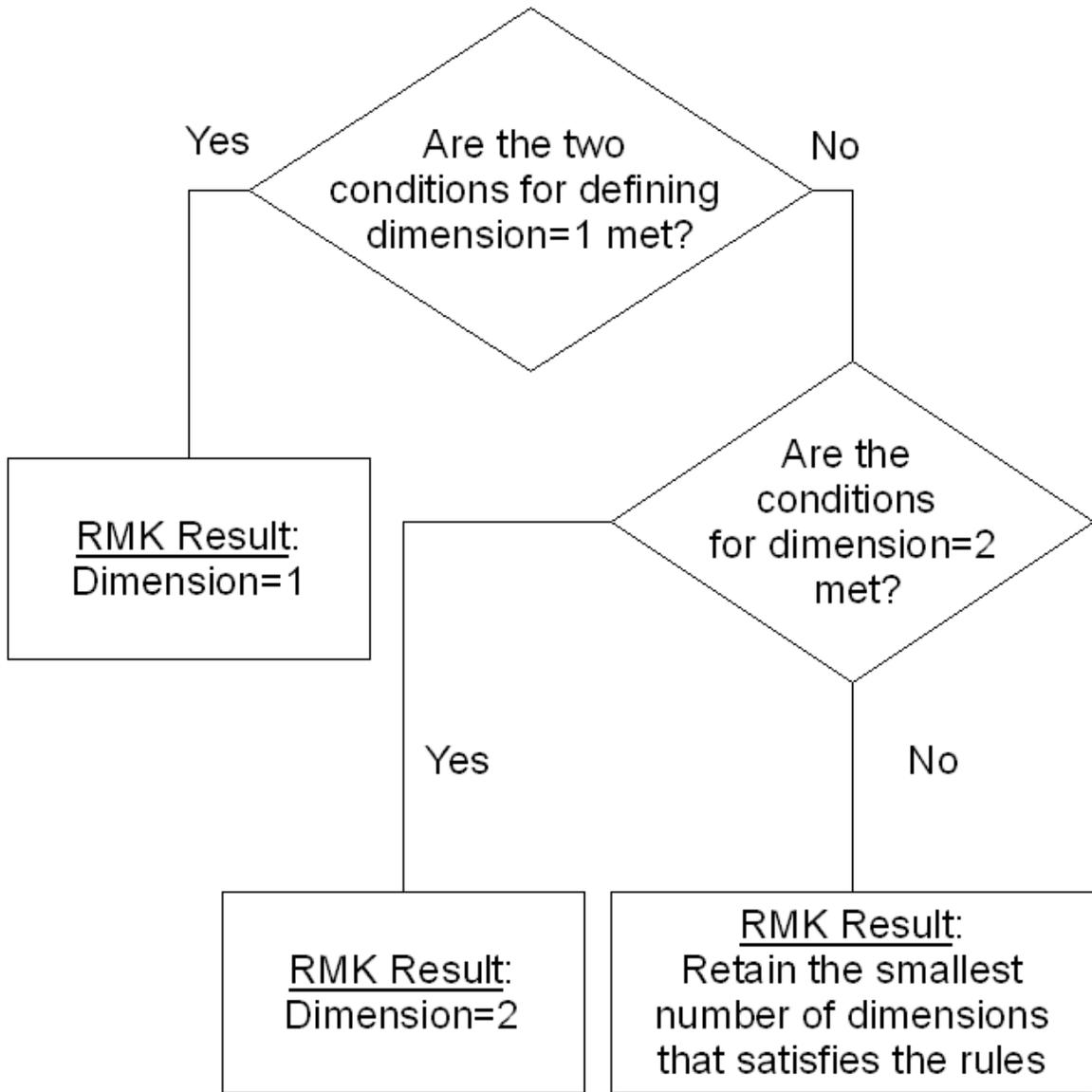


Figure 3.3. RMK-OE Decision Process

Data Generation

In this section, I describe in detail the data generation procedures and simulation conditions. Although dichotomous data is not examined in this simulation study, the continuous data generation in this dissertation did start from a matrix of item

discrimination parameters. These item discrimination parameters are those defined in a compensatory multidimensional IRT (MIRT) model for dichotomous data. How the matrix of item discrimination parameters was used in data generation is described in detail below. This approach was chosen so that a consistent data generation procedure for dichotomous data is possible for future research.

Population Factor Pattern Matrix Specification

According to Gierl et al. (2006), the range of item discrimination parameters used in their study was typical of what could be found in multidimensional tests. Specifically, they defined a high level of item discrimination parameters in the range of .50 to 1.10, and a low level of item discrimination parameters in the range of .05 to .20. I thus started my data generation process with an item discrimination matrix of which all elements were randomly generated from uniform (0.05, 0.20). Some elements in this matrix were then changed into high values that were randomly generated from uniform (0.50, 1.10), based on the complex structure type specified (described in detail below). The resulting matrix was then transformed into a factor pattern matrix that could be used to generate continuous data by applying the first formula in Equation (2.25).

Generation Procedures

For each of the simulation condition combinations (described in detail in the simulation conditions section), the values of the true (underlying) dimensions were randomly generated from a multivariate normal distribution with a zero mean vector and a covariance (equivalent to correlation in this simulation, as all manifest variables were standardized) matrix as specified by the condition. For simplicity, the correlations among

underlying dimensions were assumed to be equal (i.e., the off-diagonal elements of each correlation matrix among underlying factors were the same).

To limit the simulation to a feasible scope, the communality of each manifest variable was fixed at 0.8, and the sample size was fixed at 10,000. The unique (uncorrelated) factors were randomly generated from a multivariate normal distribution with a zero mean vector and a covariance matrix being the diagonal matrix with elements equal to 0.2³³. Continuous manifest variables were then obtained by $\mathbf{x} = \mathbf{\Lambda}\mathbf{f} + \mathbf{e}$, where \mathbf{f} indicates the generated values of true factors, $\mathbf{\Lambda}$ the generated factor pattern matrix, and \mathbf{e} the generated unique factor values³⁴.

Simulation Conditions

To compare the different factor retention rules for continuous data, various conditions were considered and 25 replications were simulated for each combination of the following conditions:

- (1) the number of manifest variables and the corresponding number of true underlying dimensions,
- (2) the correlations among underlying dimensions, and
- (3) simple structure vs. different types of complex structure.

The details of each type of conditions are described below.

³³ Note that the population variance for each unique factor is theoretically $(1 - h^2)$. Since the communality (h^2) is fixed at 0.8 in this simulation, the unique factor variance is consequently fixed at 0.2.

³⁴ This is actually Equation (2.4) with the $\boldsymbol{\mu}$ vector dropped from the equation, as it is a zero vector in this simulation.

(1) Number of Manifest Variables and Number of True Underlying Dimensions

In order to have a comprehensive performance comparison of various factor retention rules, I decided to investigate all levels of true underlying dimensions from 1 to 10. For simplicity, the number of manifest variables was selected to be a multiple of the number of underlying dimensions. Specifically, I chose the number of manifest variables to be 60 for dimension=1, 2, 3, 4, 5, 6, and 10. For dimension=7 and 8, I chose the number of manifest variables to be 56, and for dimension=9, I chose the number of manifest variables to be 54. This approach maximizes the comparability across true dimensionality in terms of the number of manifest variables, but allows for ease of computation.

(2) Correlations among Underlying Dimensions

The magnitude of correlations among underlying dimensions was set to either 0, .3, .6, .7, .8, or .9. The reason for having smaller intervals between 0.6 and 0.9 is that “many educational and psychological constructs are correlated in this range” (Gierl et al., 2006, p. 287). Therefore, the range of correlations between 0.6 and 0.9 are examined most closely.

(3) Simple Structure vs. Different Types of Complex Structure

As mentioned above, the generation of the population factor pattern matrix started from a compensatory MIRT item discrimination matrix. Also, according to Gierl et al. (2006), low item discrimination values are those in the range of (0.05, 0.20), and high item discrimination values are those in the range of (0.50, 1.10). The definitions of simple and complex structure are based on this categorization.

Since the transformation from item discrimination values to factor pattern coefficients is linear in nature, high item discrimination values correspond to high factor pattern coefficients, and low item discrimination values correspond to low factor pattern coefficients. I thus describe the factor pattern structure directly using factor pattern coefficients.

For the simple structure, all manifest variables had high factor pattern coefficients on only one dimension and low factor pattern coefficients on all other dimensions. In this dissertation, three types of complex structure were considered (see Table 3.3). These complex structure types differ in the percentage of manifest variables with high factor pattern coefficients on multiple dimensions (see Table 3.3).

Table 3.3

Explanation of the Three Types of Complex Structure

Complex Structure	Percent of Manifest Variables with High Factor Pattern Coefficients on Multiple Dimensions
Type 1	10
Type 2	30
Type 3	50

In this dissertation, the term “multiple dimensions” is operationally defined as 2 dimensions when the number of true underlying dimensions is 2. When the number of true underlying dimensions is 3 or more, the term is operationally defined as the rounded

value of half of the number of true underlying dimensions. For example, if the true number of underlying dimensions is 3, then the rounded value of half of 3 is 2.

Thus, for certain manifest variables, factor pattern coefficients on two or more dimensions should be high when complex structures are taken into consideration. Take the scenario of 60 manifest variables with 5 true underlying dimensions as an example. If Type 2 complex structure is specified, 30% of the 60 (i.e., 18) manifest variables (see Table 3.3) need to have high factor pattern coefficients on 3 dimensions ($\text{round}(5/2)=3$).

According to the way data were generated in this dissertation, with the simple structure for the above mentioned example, the first 12 manifest variables have high factor pattern coefficients on only dimension 1, the second 12 manifest variables have high factor pattern coefficients on only dimension 2, ..., the last 12 manifest variables have high factor pattern coefficients on only dimension 5.

To change the simple structure to Type 2 complex structure in the above example, the first 30% of each 12-variable block (i.e., the first 4 manifest variables in the 12-variable block) is identified to make factor pattern coefficient changes. In other words, manifest variables in positions 1, 2, 3, 4 (in the first 12-variable block), 13, 14, 15, 16 (in the second 12-variable block), ..., and 49, 50, 51, 52 (in the last 12-variable block) are identified for making factor pattern coefficient changes. Specifically, the identified-to-change manifest variables in the first block, which have high factor pattern coefficients on only dimension 1, are then changed to have high factor pattern coefficients on dimensions 2 and 3 also. Those identified-to-change manifest variables, which have high factor pattern coefficients on only dimension 2, are then changed to have high factor pattern coefficients on dimensions 3 and 4 also, and so on and so forth. For identified-to-

change manifest variables, which have high factor pattern coefficients on only dimension 5 are then changed to have high factor pattern coefficients on dimensions 1 and 2 also.

In summary, in this dissertation, there are $4 \times 6 \times 9 + 1 = 217$ different condition combinations, with 4 different types of structure complexity, 6 different correlations among underlying dimensions, 9 different levels of true underlying dimensionality (dimension=2 to 10), and an additional special case of dimension=1³⁵. As mentioned above, 25 replications for each condition combination were simulated. Therefore, in total, $217 \times 25 = 5,425$ data sets were generated. The simulation conditions that were fixed and allowed to vary are summarized in Table 3.4. The related program code for data generation can be found in Appendix A.3³⁶.

³⁵ The reason for including dimension=1 as a special case is that complex structure and different correlations among underlying dimensions cannot be simulated when there is only one dimension.

³⁶ The data generation for dimension=1 is included at the end of Appendix A.3.

Table 3.4

Summary of Simulation Conditions that were Fixed and Allowed to Vary

Simulation Conditions	
<u>Fixed</u>	<u>Allowed to Vary</u>
1. Number of observations in each data set (i.e., 10,000)	1. Number of true underlying dimensions
2. Communality (i.e., 0.8)	2. Factor pattern matrix structure
3. Equal correlations among dimensions for each simulation	complexity
4. Equal number of manifest variables per dimension for each simulation	3. Correlations among underlying dimensions across simulations
5. Number of manifest Variables > 50 and also a multiple of the number of underlying dimensions	

Comparison of Factor Retention Rules

In order to make a fair comparison among different factor retention rules, factor analysis with maximum likelihood extraction and principle component analysis (PCA) were both carried out for each generated data set. Different factor retention rules were applied using different analysis results, based on their theoretical development (see Chapter II). Specifically, the following methods were applied to PCA results:

- the original minimum average partial correlation (MAP2)
- the improved minimum average partial correlation (MAP4)
- parallel analysis (PA) using mean eigenvalues with random data generation
- the standard error of scree (SEscree)
- the eigenvalues-greater-than-one (EV1)

Consistent with the literature review, the following methods were applied to results from factor analysis with maximum likelihood extraction:

- the maximum likelihood ratio chi-square test (χ_{ML}^2)
- Akaike's information criterion (AIC)
- Bayesian information criterion (BIC)
- RMK-OEs

As some built-in functions of MATLAB do not provide accurate estimation for maximum likelihood values, and correlations between observed variables and components are not directly given³⁷, additional code was written to produce accurate estimation in order to have a fair comparison among methods. Related program code for each of the first eight methods listed above can be found in Appendix A.4. In order to compare the performance of different decision rules, evaluation criteria are needed.

³⁷ The principle component coefficients from MATLAB cannot be directly used as correlations between observed variables and components. Such correlations can be computed as (coefficients vectors)*sqrt(diagonal matrix with eigenvalues on the diagonal). In addition, after investigation of aberrant results and review of MATLAB documentation and code, I was able to determine that the “maximum likelihood” value resulting from MATLAB factor analysis is actually the minimum value for Equation (2.8) rather than the maximum value for Equation (2.7).

Evaluation Criteria

In this dissertation, two criteria were used for judging the performance of different factor retention rules. The first one is root mean squared error (RMSE) based on the deviation of each estimated number of underlying dimensions from the true number of underlying dimensions. It was computed across iterations as follows for each cell of the 217 condition combinations:

$$RMSE = \sqrt{\frac{\sum (\hat{\theta} - \theta)^2}{n}}. \quad (3.1)$$

In Equation (3.1), θ represents the true number of dimensions, $\hat{\theta}$ denotes the estimated number of dimensions, and n indicates the number of replications for each condition combination. In this dissertation, n equals 25. The related program code can be found in Appendix A.5.

The second criterion is the percent of accurate dimension identification (referred to as Percent Accurate Identification, or PAI hereafter). The PAI is the percentage of estimations where the estimated value equals the true value. For overall performance comparison, I also computed the percentages of the following: (1) under-estimation of at least 2 dimensions, (2) under-estimation of 1 dimension, (3) over-estimation of 1 dimension, and (4) over-estimation of at least 2 dimensions. These percentages were computed in order to observe the occurrence of under- or over-estimation of each factor retention rules when inaccurate identification occurred. The related program code can be found in Appendix A.6.

The mean and standard deviation (SD) of both RMSE and PAI across all condition combinations were used to choose the best performing RMK-OE weights to

define the final RMK-OE rule. Specifically, for RMSE, the desired results are small mean and small SD; whereas for PAI, large mean but small SD is desired. The median and mode of the estimations from several RMK-OEs with the best performing weight combinations were computed, and the larger of the two (with values rounded down if there were decimals) were used as the final RMK-OE value. The choice of the larger value between the two is because under-extraction (i.e., under-estimation of dimensionality) is more severe a problem than over-extraction (i.e., over-estimation of dimensionality) (Fava & Velicer, 1996; Wood et al., 1996). The decision to round down values with decimals is because preliminary analysis results indicated that doing so resulted in greater accuracy, and this was confirmed by the full set of replications as well.

The two evaluation criteria mentioned above were also used to choose the best performing factor retention rules for constructing hybrid methods. The mean and SD of both RMSE and PAI were then examined to define the final hybrid method (in a similar fashion as the final RMK-OE rule was defined). Again, small mean and SD of RMSE, high mean PAI, and small SD of PAI are desired.

Chapter IV

RESULTS

In this chapter, I report results from the simulation study conducted as described in Chapter III. This chapter is organized based on the research questions this dissertation aims to address.

First, the performance of the RMK-OEs based on different weight combinations is described, and the best performing weight combinations are identified. This addresses the first research question listed at the end of Chapter II.

Second, the overall performance (i.e., the performance across all simulation conditions) of the final RMK-OE rule based on the best performing weight combinations is compared to that of other factor retention rules, and hybrid methods based on different groupings of the best performing factor retention rules are evaluated to define the final hybrid method. This addresses the last research question listed at the end of Chapter II.

Third, the overall performance of all ten methods (described in detail in that section) in comparison is presented and the comparison results are reported to target the influence of the following two simulation conditions: (1) different levels of correlation among underlying dimensions, and (2) different levels of structure complexity for factor pattern matrix. This addresses research questions 2 and 3 listed at the end of Chapter II.

Identifying the Best Performing Weight Combinations for RMK-OEs

Table 4.1 presents the results (across all simulation conditions) of different weight combinations. In the first column, the original order of the weight combinations is given.

In the next three columns, the details of these weight combinations are provided. The mean and standard deviation (SD) of root mean squared error (RMSE) are reported in the fifth and sixth columns respectively, and the mean and SD of percent accurate identification (PAI)³⁸ are reported in the seventh and eighth columns respectively.

The last four columns of Table 4.1 present the rank ordering of the weight combinations on the four statistical measures (mean and SD of RMSE and mean and SD of PAI, respectively). Note that for mean PAI, higher values indicate better performance; whereas for other three measures, lower values indicate better performance. The best performance for all four measures corresponds to the rank order of “1” and the worst performance for all four measures corresponds to the rank order of “14.”

When rank orders based on the four different measures were taken into consideration, the focus was given to the best-performing five³⁹ weight combinations in each column (shaded gray in Table 4.1). Because weight combinations in positions 2, 3, and 5 (from the original ordering) were in the top five for all four measures, this set of weight combinations is a strong candidate for constructing the final RMK-OE rule. In addition, the weight combination in position 1 was ranked in the top five on three of the four measures, and was the top measure for both mean and SD of PAI. Therefore, another strong candidate for constructing the final RMK-OE rule is to use the set of weight combinations in positions 1, 2, 3, and 5.

Based on these two candidate sets, the method of defining the final RMK-OE rule described in Chapter III was then carried out. Specifically, for each candidate set of

³⁸ The mean percentage of each RMK-OE weight combination in each estimation error category across all simulation conditions is presented in Appendix B.

³⁹ The decision to focus on the first five values is because 5 is about 1/3 of the total number of weight combinations tried (i.e., 14). To me, this choice can guarantee the focus on the “best performing” weight combinations.

weight combinations, the mode and median (rounded down to the nearest integer if results were with decimal points, to improve accuracy) of the numbers of factors to retain were computed. The larger of the two (median or mode) was considered as the final RMK-OE result based on the corresponding set of weight combinations.

Table 4.1

Values and Rank Orders of Mean and SD of RMSE and PAI across All Simulation

Conditions for Each RMK-OE Weight Combination

Original Order	Weights			Statistic Value				Rank Order			
	Trimmed			RMSE		PAI		RMSE		PAI	
	Median	IQR	Range	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.8	0.1	0.1	0.3773	0.4604	0.9150	0.1358	2	8	1	1
2	0.7	0.2	0.1	0.3785	0.4566	0.9143	0.1360	3	5	2	2
3	0.6	0.3	0.1	0.3796	0.4544	0.9135	0.1368	5	4	4	4
4	0.5	0.4	0.1	0.3811	0.4540	0.9119	0.1377	6	3	8	9
5	0.6	0.2	0.2	0.3750	0.4511	0.9135	0.1368	1	2	3	5
6	0.5	0.3	0.2	0.3793	0.4498	0.9121	0.1374	4	1	7	7
7	0.4	0.4	0.2	0.4106	0.5368	0.9029	0.1618	11	12	12	14
8	0.4	0.3	0.3	0.4106	0.5343	0.9029	0.1608	10	11	11	13
9	1.0	0.0	0.0	0.8898	0.8350	0.8599	0.1598	14	14	14	12
10	0.9	0.1	0.0	0.5450	0.6210	0.8953	0.1421	13	13	13	11
11	0.8	0.2	0.0	0.4294	0.4960	0.9084	0.1377	12	10	10	8
12	0.7	0.3	0.0	0.3871	0.4633	0.9126	0.1363	8	9	6	3
13	0.6	0.4	0.0	0.3835	0.4589	0.9132	0.1370	7	7	5	6
14	0.5	0.5	0.0	0.3918	0.4583	0.9110	0.1380	9	6	9	10

Note. Rank orders were computed based on more decimal points than presented here.

Two sets of RMK-OE results (one for each set of weight combinations mentioned above) were computed for each replication in the simulation. The mean and SD of RMSE and PAI across all simulation conditions were compared, with the best performing set of weight combinations shaded gray for each measure. The results are shown in Table 4.2.

Table 4.2

Mean and SD of RMSE and PAI across All Simulation Conditions for the Two Candidates for the Final RMK Rule

Candidates for the Final RMK	RMSE		PAI	
	Mean	SD	Mean	SD
Based on the four best weight combinations (in positions 1, 2, 3,5)	0.3780	0.4554	0.9143	0.1364
Based on the three best weight combinations (in positions 2, 3, 5)	0.3796	0.4544	0.9135	0.1368

As shown in Table 4.2, the RMK-OE rule defined on the best four weight combinations was found to perform slightly better, as it resulted in a smaller mean of RMSE, a higher mean of PAI, and a smaller SD of PAI. Therefore, the final RMK-OE rule was defined based on these four weight combinations:

- 0.8, 0.1, 0.1 on medians, IQRs, and trimmed ranges respectively,
- 0.7, 0.2, 0.1 on medians, IQRs, and trimmed ranges respectively,
- 0.6, 0.3, 0.1 on medians, IQRs, and trimmed ranges respectively, and
- 0.6, 0.2, 0.2 on medians, IQRs, and trimmed ranges respectively.

Methods Comparison and Development of Hybrid Methods

This section evaluates the performance of the various factor retention rules in comparison (including the newly developed final RMK-OE rule). Multiple hybrid methods based on different groupings of the best performing factor retention rules are described, and their performance results are evaluated to define the final hybrid method. Besides overall performance of each method, I also report the evaluation results focusing on the following two simulation conditions: (1) the different levels of correlation among underlying dimensions, and (2) the different levels of structure complexity for factor pattern matrix.

Overall Performance Comparison and Hybrid Method Construction

The mean and SD of RMSE and PAI for the final RMK-OE rule (referred to as RMK-OE hereafter for simplicity) and other eight rules in comparison were computed. The eight rules in comparison were: the original and the improved minimum average partial correlation (MAP2 and MAP4, respectively), parallel analysis (PA) using mean eigenvalues with random data generation, the standard error of scree (SEscree), the eigenvalues-greater-than-one (EV1) method, the maximum likelihood ratio chi-square test (χ^2_{ML}), Akaike's information criterion (AIC), and Bayesian information criterion (BIC). The values and rank orders for each method on the four measures (mean and SD of RMSE and mean and SD of PAI, respectively) are presented in Table 4.3. The five⁴⁰ best performing methods for each measure is highlighted in gray. Again the mean PAI

⁴⁰ This time, the number five is chosen as there is a large performance gap between the best five and the rest, as can be easily seen from Table 4.3.

was ordered from the highest to the lowest values, while all other three measures were ordered from the smallest to the largest values.

Table 4.3

Values and Rank Orders of Mean and SD of RMSE and PAI across All Simulation Conditions for Different Methods in Comparison

Method	Statistic Value				Rank Order			
	RMSE		PAI		RMSE		PAI	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
RMK-OE ^a	0.3780	0.4554	0.9143	0.1364	4	3	3	3
χ^2_{ML} ^a	0.2449	0.1859	0.9314	0.0837	1	2	2	2
AIC ^a	0.3591	0.1213	0.8653	0.0662	3	1	4	1
BIC ^a	0.2732	1.2148	0.9432	0.2243	2	4	1	4
MAP2 ^b	1.6332	2.9605	0.7425	0.4292	6	7	6	6
MAP4 ^b	1.8769	2.8641	0.6295	0.4738	7	6	7	7
PA ^b	3.1350	3.3235	0.4498	0.4912	9	8	9	8
SEscree ^b	0.6641	1.8858	0.8369	0.3576	5	5	5	5
EV1 ^b	2.9889	3.3473	0.4806	0.4920	8	9	8	9

Note.

^a These rules were applied to results from factor analysis with maximum likelihood extraction.

^b These rules were applied to results from PCA.

It is obvious from Table 4.3 that the four methods based on results from factor analysis with maximum likelihood extraction performed the best. These four methods are maximum likelihood chi-square test (χ_{ML}^2), AIC, BIC, and RMK-OE. The SEscree was found to perform the best among all methods using PCA results. Four possible hybrid methods were thus constructed:

1. RMK-OE, χ_{ML}^2 , AIC, BIC, and SEscree (taking all five best performing rules),
2. RMK-OE, χ_{ML}^2 , AIC, BIC (taking the four best performing rules),
3. χ_{ML}^2 , AIC, BIC, and SEscree (for evaluating the contribution of RMK-OE to rule 1), and
4. χ_{ML}^2 , AIC, and BIC (for evaluating the contribution of RMK-OE to rule 2).

For each hybrid method, the median and mode of the estimated numbers of factors to retain from the participating rules were computed, with the larger value retained as the final result. If the resulting value had decimal points, it was rounded down to the nearest integer to improve accuracy. After each hybrid method was applied to each replication, the mean and SD of RMSE and PAI were computed and compared. The results are presented in Table 4.4, with corresponding rank orders for the four measures (mean and SD of RMSE and PAI) highlighted in gray for the top two performing hybrid methods.

As shown in Table 4.4, the first and second hybrid methods performed the best, as they both resulted in small mean and SD of RMSE, high mean of PAI, and small SD of PAI. Moreover, the values of mean RMSE for the first two hybrid methods are much smaller (about 1/5 in magnitude) than those for hybrid method 3 and 4, indicating much higher overall estimation accuracy. Both of these methods (i.e., the first two hybrid

methods) involved the use of RMK-OE. Therefore, the involvement of RMK-OE is essential in constructing a well-performing final hybrid method.

The first hybrid method also involved the use of the SEscree, while the second did not. These two hybrid methods did not differ much (difference < 0.0005) on mean RMSE, mean PAI, and SD of PAI, whereas the second method resulted in a smaller SD of RMSE (difference > 0.005 but < 0.01) than that of the first hybrid method. The advantage of lower estimation error variability of the second hybrid method is thus obvious. Moreover, the second hybrid method involved one less contributing method—SEscree (which was applied to PCA results rather than factor analysis results), and thus also has the advantage of parsimony. Therefore, I chose the second hybrid method as the final hybrid method, which involved the application of the following four rules: RMK-OE, χ^2_{ML} , AIC, and BIC.

Table 4.4

Mean and SD of RMSE and PAI for Different Hybrid Methods

Hybrid Method	Statistic Value				Rank Order			
	RMSE		PAI		RMSE		PAI	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.0295	0.1369	0.9899	0.0565	1	2	1	1
2*	0.0298	0.1307	0.9897	0.0565	2	1	2	2
3	0.1067	0.4689	0.979	0.0927	3	4	3	4
4	0.1120	0.1545	0.9681	0.0781	4	3	4	3

Note.

* This rule was identified as the final hybrid method.

Overall Performance Comparison of the Ten Methods

In this section, I describe the overall performance comparison results for the following ten methods:

- the final hybrid method (Hybrid)
- the final RMK-OE rule (RMK-OE)
- the maximum likelihood ratio chi-square test (χ_{ML}^2)
- Akaike's information criterion (AIC)
- Bayesian information criterion (BIC)
- the original minimum average partial (MAP2)
- the improved minimum average partial (MAP4)

- parallel analysis (PA) using mean eigenvalues with random data generation
- the standard error of scree (SEscree)
- the eigenvalues-greater-than-one (EV1)

The overall performance (across all simulation conditions) is summarized in two ways. First, the values⁴¹ and rank orders of the mean and SD of RMSE and PAI of these ten methods are presented in Table 4.5. Second, percentages of each of these ten methods in each estimation error categories are reported in Table 4.6. The five estimation error categories are: (1) under-estimation by at least 2 dimensions, (2) under-estimation by 1 dimension, (3) accurate dimension identification (PAI), (4) over-estimation by 1 dimension, and (5) over-estimation by at least 2 dimensions.

⁴¹ These values can all be found in Tables 4.3 and 4.4. For ease in reading, I combined related results and re-rank ordered the ten methods in Table 4.5.

All Methods Comparison Part 1: Mean and SD of RMSE and PAI

Table 4.5

Values and Rank Orders of Mean and SD of RMSE and PAI across All Simulation

Conditions for the Ten Methods in Comparison

Method	Statistic Value				Rank Order			
	RMSE		PAI		RMSE		PAI	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Hybrid ^a	0.0298	0.1307	0.9897	0.0565	1	2	1	1
RMK-OE ^a	0.3780	0.4554	0.9143	0.1364	5	4	4	4
χ^2_{ML} ^a	0.2449	0.1859	0.9314	0.0837	2	3	3	3
AIC ^a	0.3591	0.1213	0.8653	0.0662	4	1	5	2
BIC ^a	0.2732	1.2148	0.9432	0.2243	3	5	2	5
MAP2 ^b	1.6332	2.9605	0.7425	0.4292	7	8	7	7
MAP4 ^b	1.8769	2.8641	0.6295	0.4738	8	7	8	8
PA ^b	3.1350	3.3235	0.4498	0.4912	10	9	10	9
SEscree ^b	0.6641	1.8858	0.8369	0.3576	6	6	6	6
EV1 ^b	2.9889	3.3473	0.4806	0.4920	9	10	9	10

Note.

^a These rules were applied to results from factor analysis with maximum likelihood extraction.

^b These rules were applied to results from PCA.

According to Table 4.5, the final hybrid method (Hybrid) ranked the first on three (i.e., mean RMSE, mean and SD of PAI) out of the four measures, and ranked the second on the fourth measure (i.e., SD of PAI). Specifically, the mean RMSE of Hybrid was found to be about 1/10 in magnitude of that of the second best method (i.e., χ_{ML}^2). The mean PAI of Hybrid was found to be about 99%, with about a 5% mean PAI advantage over that of the second best method (i.e., BIC). Moreover, although the SD of RMSE of Hybrid ranked second, its magnitude (≈ 0.1307) was only slightly smaller (difference < 0.01) than that (≈ 0.1213) of the best method (i.e., AIC).

The reasons for the indispensability of RMK-OE in constructing the well-performing hybrid methods (the first two shown in Table 4.4) are also obvious from Table 4.5. Specifically:

- The mean RMSE of RMK-OE (≈ 0.3780 , ranked fifth) was found to be similar to that (≈ 0.3591) of AIC (the fourth best method), and was about half in magnitude of that (≈ 0.6641) of the sixth best method (i.e., SEscree).
- The SD of RMSE of RMK-OE was ranked fourth and its magnitude was less than 0.5, while that of the fifth method (i.e., BIC) was larger than 1 in magnitude.
- The mean PAI of RMK-OE was about 91%, with little difference from either that (about 93%) of the third best method (i.e., χ_{ML}^2) or that (about 94%) of the second best method (i.e., BIC). The mean PAI of the fifth method (i.e., AIC) was about 86%, much lower in magnitude.
- The SD of PAI of RMK-OE was about 0.1364 and was ranked fourth, while the SD of PAI of the third best method (i.e., χ_{ML}^2) was about 0.0837, and that of the fifth best method (i.e., BIC) was about 0.2243.

Based on these four measures, I observed that the performance of RMK-OE, although not ranked best on any one of them, was only slightly different in magnitude from those ranked above it (excluding the final hybrid method), but was much better (i.e., with large difference advantage in magnitude) from the method ranked just below it. This explains why the hybrid methods that involved the final RMK-OE rule performed much better than those that did not.

All Methods Comparison Part 2: Percentages in Each Estimation Error Category

Besides estimation accuracy, I was also interested in observing the occurrence of under- or over-estimation of each factor retention rule when inaccurate estimation occurred. In order to capture the overall (i.e., across all simulation conditions) pattern of inaccurate estimations for each method, percentages of each of the ten methods in each estimation error category was computed and is presented in Table 4.6. For easy pattern observation, I highlighted some cells with different levels of gray color, with the darker gray indicating worse performance.

As shown in Table 4.6, when only percent accurate identification (PAI, or the middle column) is considered, it is obvious that the final hybrid method (Hybrid) was the most accurate in this simulation (PAI \approx 98.97%). BIC ranked second (PAI \approx 94.32%), followed by χ^2_{ML} (ranked third with PAI \approx 93.14%) and RMK-OE (ranked fourth with PAI \approx 91.43%). PA (PAI \approx 44.98%) was found to be the worst method among all ten methods in comparison. The second worst was EV1 (PAI \approx 48.06%).

When inaccurate estimations occurred, RMK-OE, χ^2_{ML} , and AIC tended toward over-estimation, and were most likely to over-estimate by one factor. Moreover, the over-

estimation by one factor occurrences for AIC were found to be about twice ($\approx 13.07\%$) as many as those for either RMK-OE ($\approx 5.35\%$) or χ_{ML}^2 ($\approx 5.24\%$). RMK-OE demonstrated a much higher percentage ($\approx 2.56\%$) of over-estimation by at least two factors, compared to about 0.61% and 0.29% for χ_{ML}^2 and AIC respectively.

Also, according to Table 4.6, BIC and the methods that utilized results from PCA tended toward under-estimation when estimation errors occurred. Moreover, when under-estimation occurred for these methods, it was more likely for them to severely under-estimate (i.e., under-estimate by at least 2 factors). Specifically, when inaccurate estimation occurred for BIC, about 5 out of 6 cases resulted in severe under-estimation. Moreover, PA and EV1 were found to result in about 50% occurrence of under-estimation by at least 2 factors (about 51.93% and 48.41% for PA and EV1 respectively). MAP4 and MAP2 were found to result in severe under-estimation for about 31.67% and 24.33% of all simulated data sets respectively. SEscree was again found to be the best among all methods using PCA results, resulting in about 11.56% of severe under-estimation of all simulated data.

The most important result from Table 4.6 is that RMK-OE performed generally on a par with AIC, BIC, and χ_{ML}^2 , but when RMK-OE, AIC, BIC, and χ_{ML}^2 were used together to create a hybrid method, both the rate and severity of under- or over-estimation was dramatically reduced.

Table 4.6

Mean Percentage of Each Method in Each Estimation Error Category across All Simulation Conditions

Factor Retention Rules	Percentages in Each Estimation Error Category				
	≤ -2	-1	0	1	≥ 2
Hybrid	0.22	0.39	98.97	0.42	0.00
RMK-OE	0.29	0.37	91.43	5.35	2.56
χ_{ML}^2	0.13	0.88	93.14	5.24	0.61
AIC	0.00	0.11	86.53	13.07	0.29
BIC	4.72	0.96	94.32	0.00	0.00
MAP2	24.33	1.42	74.25	0.00	0.00
MAP4	31.67	5.38	62.95	0.00	0.00
PA	51.93	3.10	44.98	0.00	0.00
SEScree	11.56	4.76	83.69	0.00	0.00
EV1	48.41	3.54	48.06	0.00	0.00

Note.

≤ -2 denotes the category of “under-estimation by at least 2 dimensions”, -1 denotes the category of “under-estimation by 1 dimension”, 0 denotes the category of accurate identification, 1 denotes the category of “over-estimation by 1 dimension”, and ≥ 2 denotes the category of “over-estimation by at least 2 dimensions.”

Performance Comparison for Dimension=1 Only

Because the angle-change based approach can only be employed when the true number of underlying dimensions is larger than or equal to 2, when the true number of underlying dimensions equals 1, the corresponding evaluation component of the RMK-OEs does not involve the evaluation based on angle changes, but rather, on three existing factor analysis (FA) associated methods. Consequently, the final RMK-OE rule and the final hybrid method for evaluating dimension=1 do not involve angle-change based evaluation either, but three existing FA associated methods instead. These three methods are χ_{ML}^2 , AIC, and BIC. RMK-OEs with various weight combinations, the final RMK-OE rule, and the final hybrid method for evaluating dimension=1 resulted in 100% accurate dimensionality identification. When the true number of underlying dimensions is greater than one, however, various RMK-OEs, the final RMK-OE rule, and the final hybrid method never resulted in estimated dimension=1. In contrast, all PCA associated factor retention rules tended to result in dimension=1 when the underlying dimensions were highly correlated, even when the true number of dimensions was at least 2.

When the true number of underlying dimensions equals 1, all methods were found to result in dimension=1. Among the 25 iterations, only 2 iterations were found to result in dimension=2 for χ_{ML}^2 , and 3 iterations were found to result in dimension=2 for AIC. Although all other methods resulted in dimension=1 for all 25 iterations, due to the under-estimation tendency mentioned above for PCA associated rules for dimensions larger than or equal to 2 scenarios, a dimension=1 result from any PCA associated rules should not be considered as the correct result. In other words, further verification for dimension=1 result from any PCA associated rules is needed.

Performance Comparison with Regard to Different Correlations among Factors

The results reported in this section focus solely on cases where the true number of dimensions is ≥ 2 , as different levels of correlation are not possible for dimension=1.

In this section, two groups of figures are presented, with the first group (Figures 4.1-4.6) reporting mean RMSE results, and the second group (Figures 4.7-4.12) reporting mean PAI results. A summary of results are provided for each group of figures. There are six figures in each group, one for each level of correlation among underlying dimensions considered in this dissertation. The six levels of correlations are: 0, 0.3, 0.6, 0.7, 0.8, and 0.9. Note that in all these figures, “nd” indicates the number of true underlying dimensions. Although corresponding tables for these figures are not presented here, I include some numbers from those tables in my explanation of the results shown in these figures.

Figures and Results for Mean RMSE:

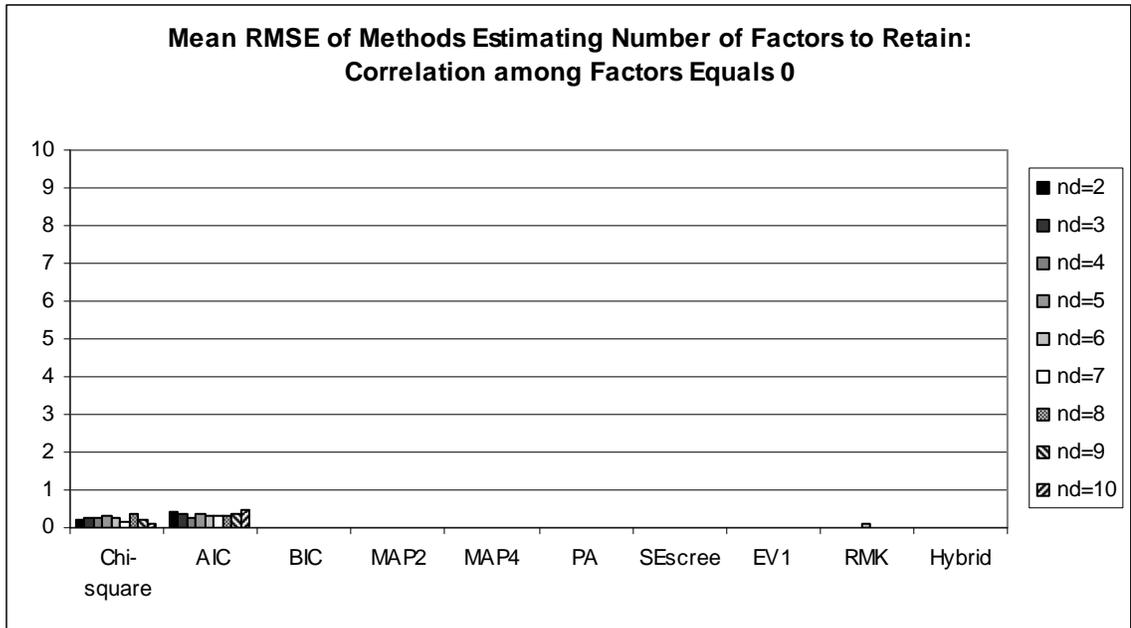


Figure 4.1. Mean RMSE with Correlation among Factors Equals 0

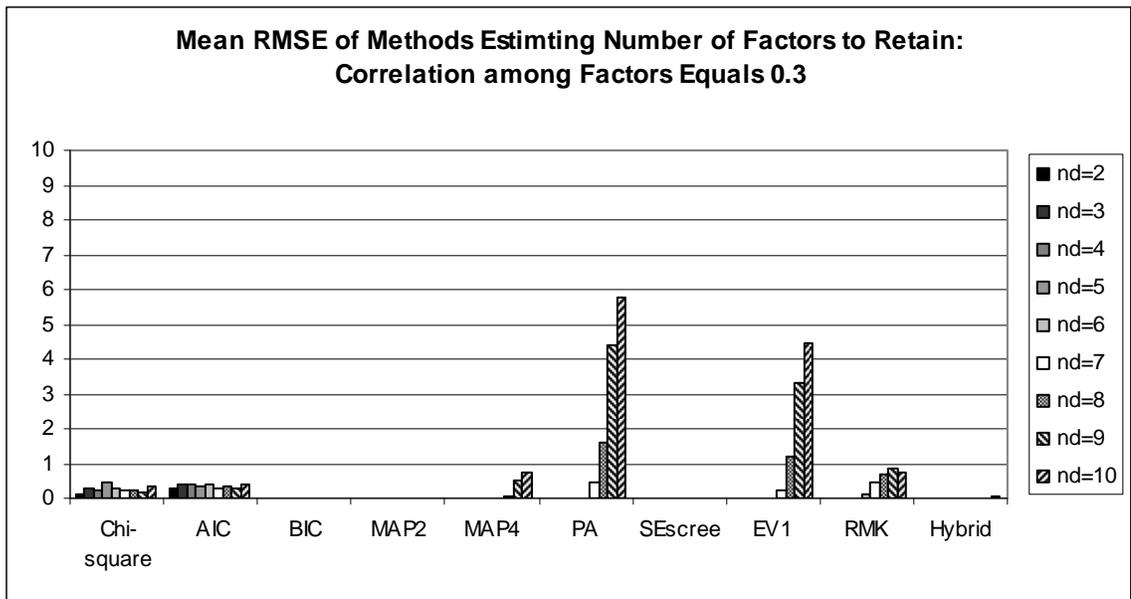


Figure 4.2. Mean RMSE with Correlation among Factors Equals 0.3

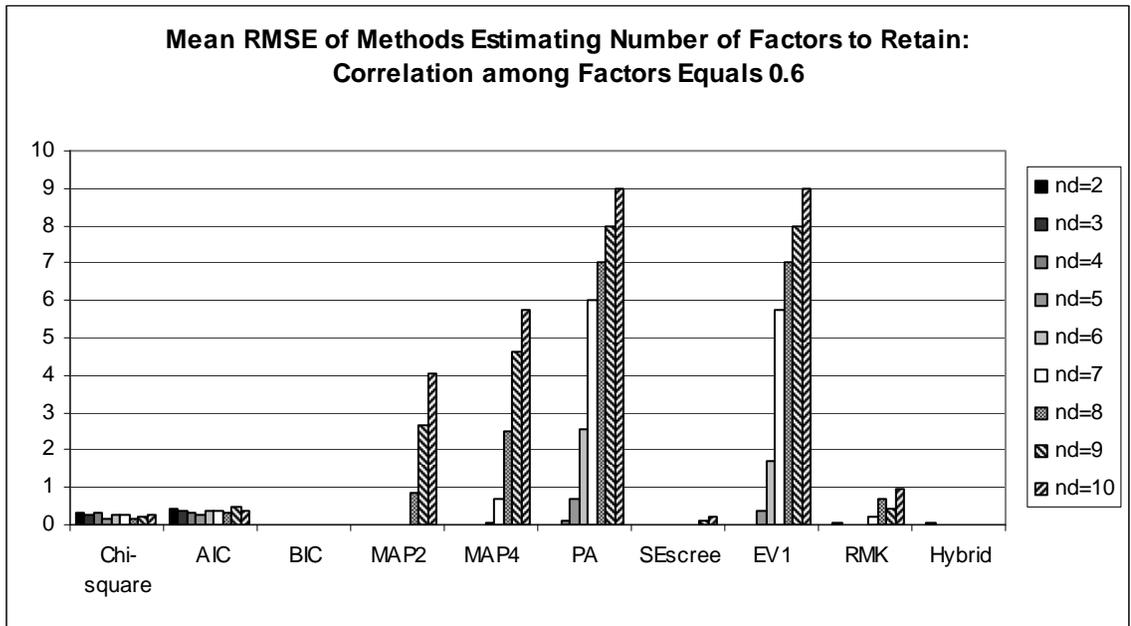


Figure 4.3. Mean RMSE with Correlation among Factors Equals 0.6

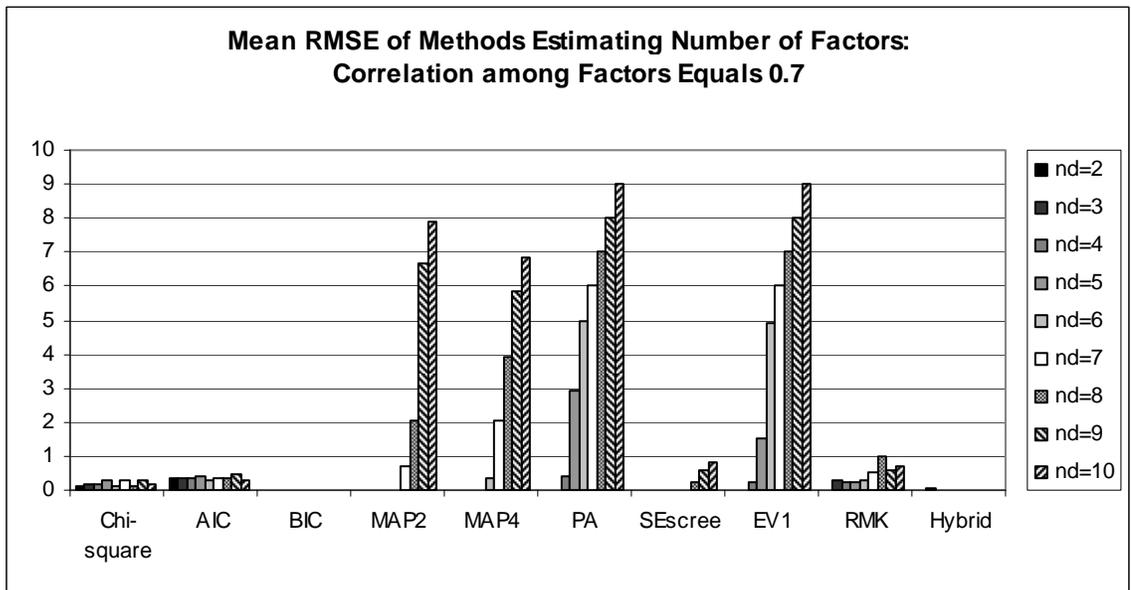


Figure 4.4. Mean RMSE with Correlation among Factors Equals 0.7

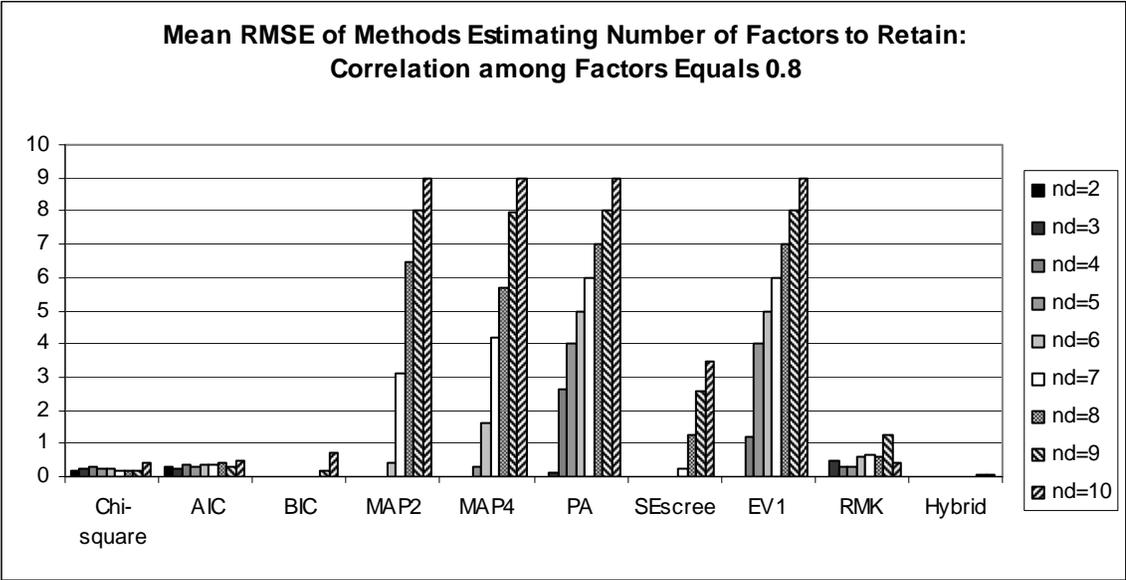


Figure 4.5. Mean RMSE with Correlation among Factors Equals 0.8

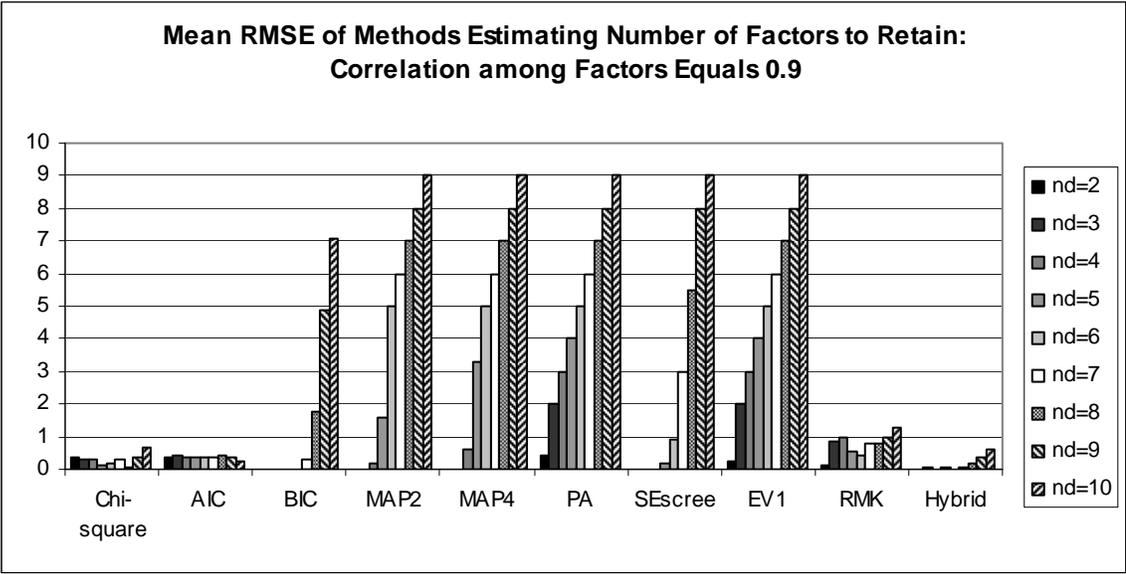


Figure 4.6. Mean RMSE with Correlation among Factors Equals 0.9

I summarize below in bullet points the patterns observed from the above figures on mean RMSE. Note that I group certain methods together in the following summary

(i.e., χ_{ML}^2 and AIC together, MAP2 and MAP4 together, PA and EV1 together) due to their similarity in performance patterns shown in the above figures. From Figures 4.1-4.6, the following conclusions can be made:

- (1) The maximum likelihood ratio chi-square test (χ_{ML}^2) and AIC demonstrated estimation error for each of the 9 different levels of true underlying dimensionality across all six levels of correlation among underlying dimensions. Nevertheless, such estimation error, being represented by mean RMSE, was small all the time (<0.5), with the only exception being dimension=10 for χ_{ML}^2 (mean RMSE ≈ 0.67). Thus, χ_{ML}^2 and AIC generally performed well under all levels of correlations.
- (2) BIC had a mean RMSE=0 for all levels of true underlying dimensionality when correlation=0, 0.3, 0.6, and 0.7 among dimensions. When correlation among underlying dimensions equals 0.8, BIC was found to have estimation error for dimension=9 and 10, although the magnitude of these two mean RMSE were small (<1). When correlation among underlying dimensions equals 0.9, BIC had non-zero estimation error from dimension=7 and up. The mean RMSE was large in magnitude for both dimension=9 and 10 (mean RMSE >4 and >7 for dimension=9 and 10 respectively). Thus, the performance of BIC degraded considerably under the conditions of high correlation among dimensions and high levels of dimensionality.
- (3) MAP2 had a mean RMSE=0 for all levels of true underlying dimensionality when correlation=0 and 0.3 among these dimensions. MAP4 was found to have a mean RMSE=0 across all levels of true underlying dimensionality only when

underlying dimensions were uncorrelated (i.e., correlation=0). Both MAP2 and MAP4 were found to have quite large mean RMSE (>2) when correlation among underlying dimensions was high. Moreover, the higher the correlation among underlying dimensions, the smaller the number of underlying dimensions where mean RMSE started to be larger than 2. Thus, the performance of MAP2 and MAP4 in general degraded in proportion to increasing correlation among underlying dimensions.

- (4) PA and EV1 both had a mean RMSE=0 for all levels of true underlying dimensionality when correlation=0. However, their performance started to deteriorate when correlation=0.3 and the number of underlying dimensions was large (mean RMSE >2 starting from dimension=9 for correlation=0.3 scenario). Moreover, higher correlation among underlying dimensions was found to have mean RMSE larger than 2 with smaller underlying dimensionality. Therefore, the performance of PA and EV1, in general, degraded considerably under conditions of moderate to high correlation among underlying dimensions.
- (5) SEscree was found to have a mean RMSE=0 for all levels of true underlying dimensionality when correlation=0 and 0.3 among these dimensions. When correlation=0.6 and 0.7 among underlying dimensions, SEscree was found to have estimation error for dimension=9 to 10 only (with magnitude of mean RMSE <0.2) and for dimension=8 to 10 only (with magnitude of mean RMSE <1) respectively. When correlation=0.8, the mean RMSE started to be larger than 2 from dimension=9, and when correlation=0.9, the mean RMSE started to be larger than 2 from dimension=7. Thus, the performance of SEscree degraded only

for high correlations among underlying dimensions and high levels of dimensionality.

- (6) RMK-OE was found to have a mean RMSE=0 less frequently than some existing methods. Nevertheless, the magnitude of its mean RMSE was found to be relatively small (i.e., <1) across different conditions except for two scenarios: when correlation=0.8 among underlying dimensions and the number of true underlying dimensions equals 9, and when correlation=0.9 among underlying dimensions and the number of true underlying dimensions equals 10. Even for these two scenarios, the magnitude of its mean RMSE were both found to be less than 1.5. Thus, RMK-OE was found to degrade only slightly with very high correlations among underlying dimensions and high levels of dimensionality.
- (7) The advantage of the final hybrid method (Hybrid) in making accurate factor retention decisions is obvious. Hybrid had a mean RMSE=0 with 43 out of 54 scenarios examined here (i.e., 9 levels of true dimensionality multiplied by 6 levels of correlations among underlying dimensions), and was the second best in this regard (BIC was the best with mean RMSE=0 with 48 out of 54 scenarios). When estimation error occurred (in the remaining 11 scenarios), however, the magnitude of the Hybrid mean RMSE was negligible for 7 out of these 11 scenarios (mean RMSE=0.05). The largest mean RMSE (=0.6) for Hybrid was found when dimension=10 with correlation=0.9 among underlying dimensions, a case in which BIC severely underestimated the true underlying dimensionality.

Figures and Results for Mean PAI:

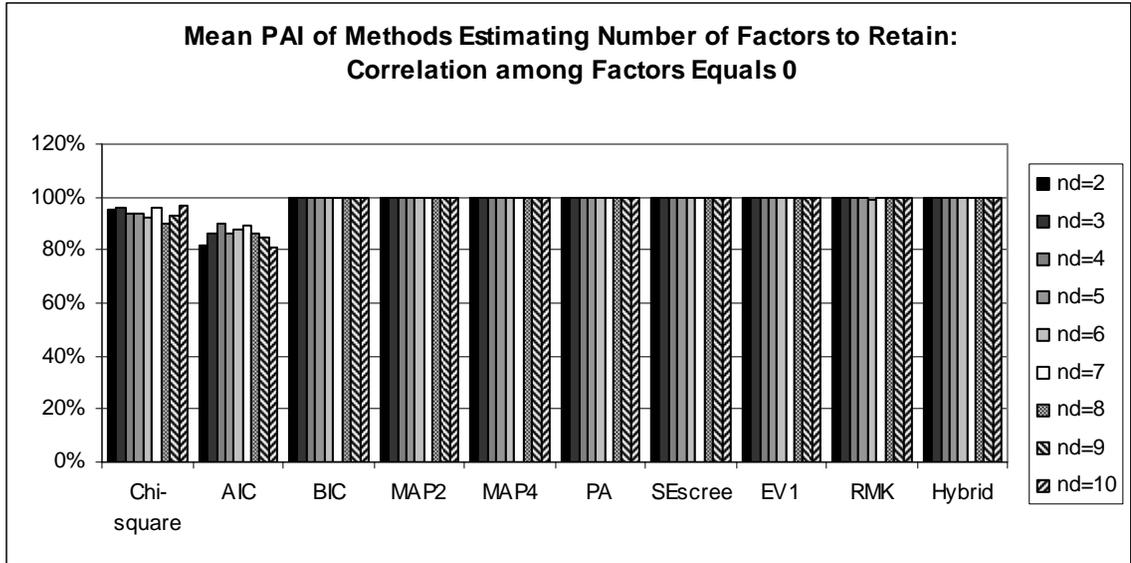


Figure 4.7. Mean PAI with Correlation among Factors Equals 0

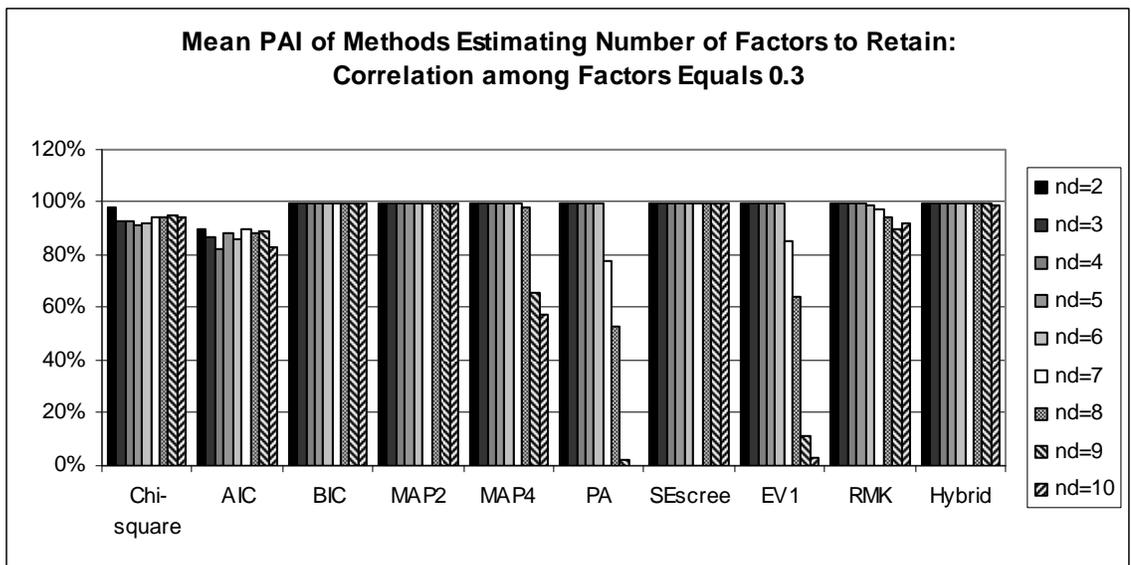


Figure 4.8. Mean PAI with Correlation among Factors Equals 0.3

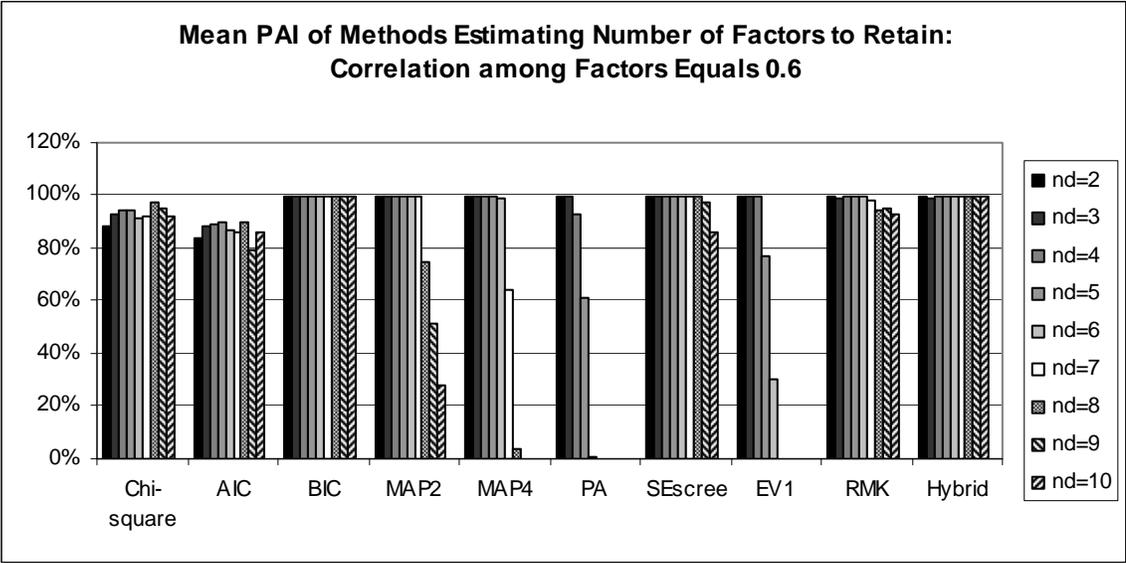


Figure 4.9. Mean PAI with Correlation among Factors Equals 0.6

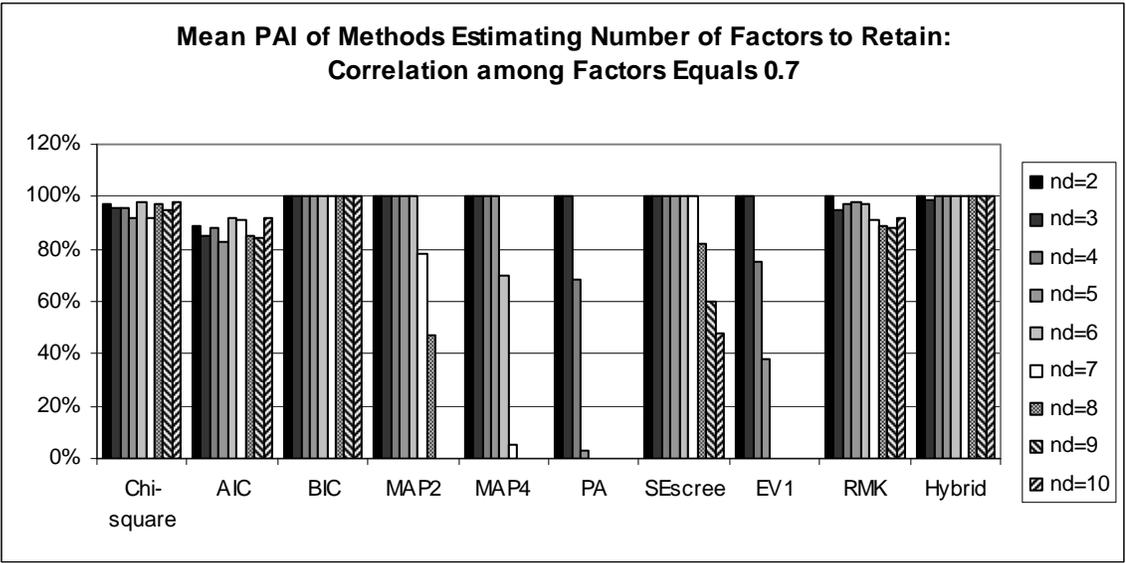


Figure 4.10. Mean PAI with Correlation among Factors Equals 0.7

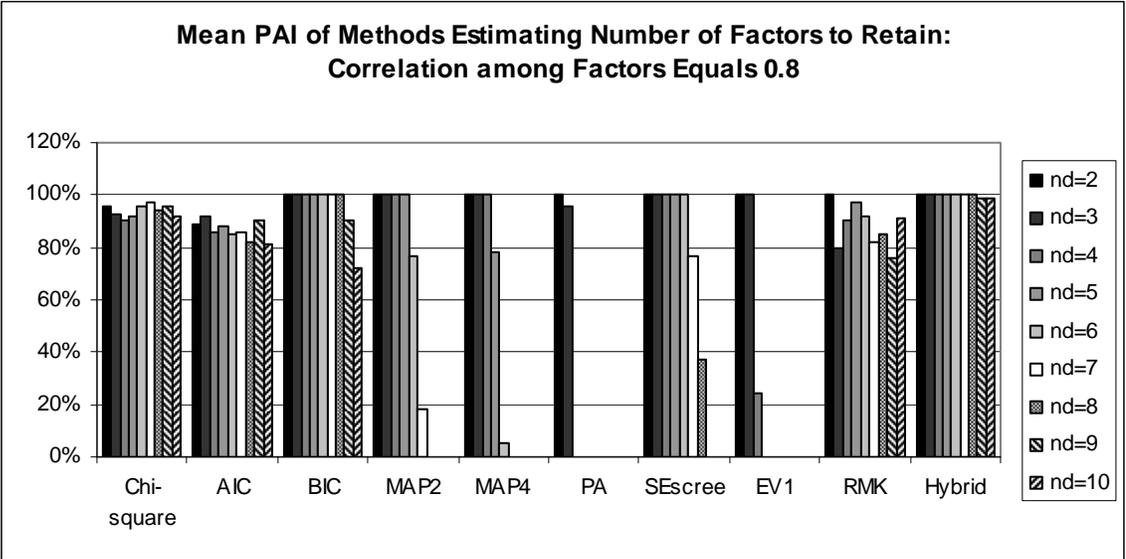


Figure 4.11. Mean PAI with Correlation among Factors Equals 0.8

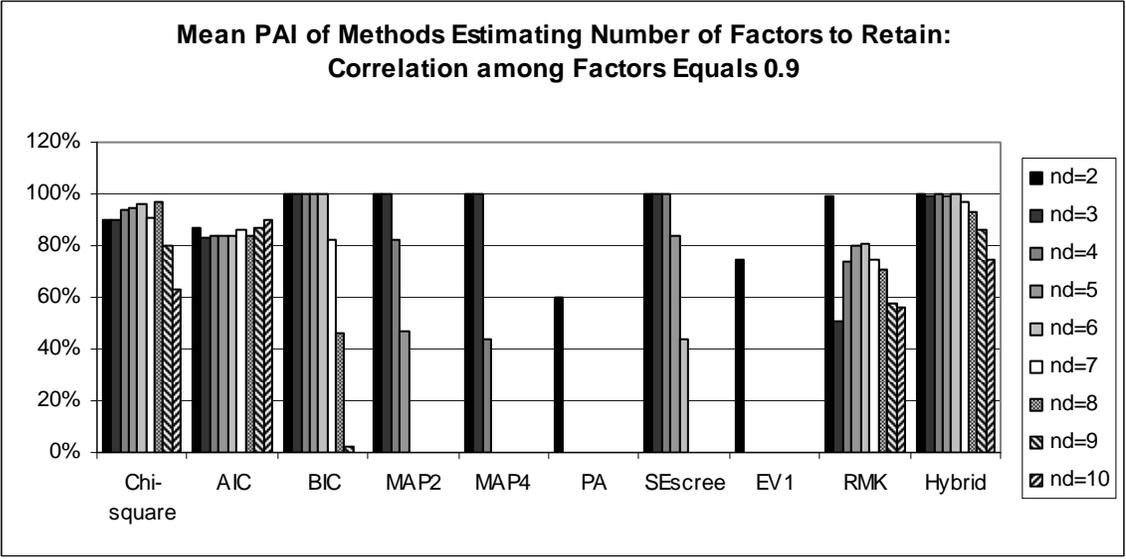


Figure 4.12. Mean PAI with Correlation among Factors Equals 0.9

I summarize below in bullet points the patterns observed from the above figures on mean PAI. The performance patterns shown in Figures 4.7-4.12 correspond to those demonstrated in Figures 4.1-4.6. In other words, in general, when the mean RMSE is

large, the mean PAI is low, and when the mean RMSE is small, the mean PAI is high.

Note that I group certain methods together again in the following summary (i.e., χ_{ML}^2 and AIC together, MAP2 and MAP4 together, PA and EV1 together) due to their similarity in performance patterns shown in above figures. According to Figures 4.7-4.12:

- (1) The maximum likelihood ratio chi-square test (χ_{ML}^2) and AIC never achieved mean PAI=100%. When dimension=10 and correlation among underlying dimensions equals 0.9, the mean PAI of χ_{ML}^2 was 63% and was lower than that of AIC (90%). The mean PAI of χ_{ML}^2 (80%) was also found to be lower than that of AIC (87%) when dimension=9 and correlation among underlying dimensions equals 0.9. For other cases, however, the mean PAI for χ_{ML}^2 was found to be higher than that of AIC. Thus χ_{ML}^2 and AIC performed well, but not perfectly under all conditions.
- (2) BIC was found to have a mean PAI=100% for all levels of true underlying dimensionality when correlation=0, 0.3, 0.6, and 0.7 among these dimensions. The mean PAI was found to be 2% for dimension=9, and 0% for dimension=10 when correlation=0.9 among underlying dimensions. Thus, BIC performed better than χ_{ML}^2 and AIC except for high correlations among underlying dimensions and high levels of dimensionality.
- (3) MAP2 was found to have a mean PAI=100% for all levels of true underlying dimensionality when correlation=0 and 0.3 among these dimensions. MAP4 was found to achieve such accuracy only when underlying dimensions were uncorrelated (i.e., correlation=0). MAP2 was found to always perform better than

MAP4, although both demonstrated the tendency of resulting in the mean PAI=0% when correlation among underlying dimensions was high. Thus both MAP2 and MAP4 performance degraded considerably with moderate to high correlations.

- (4) PA and EV1 were both found to have a mean PAI=100% for all levels of true underlying dimensionality when correlation=0. Both methods were found to result in the mean PAI=0% more frequently than other methods in comparison. And again, higher correlation among underlying dimensions tended to result in mean PAI towards 0% with small number of dimensionality. Thus, the performance of PA and EV1 degraded considerably under conditions of low to high correlations among underlying dimensions.
- (5) SEscree was found to have a mean PAI=100% for all levels of true underlying dimensionality when correlation=0 and 0.3 among these dimensions. SEscree was found to result in the mean PAI=0% only when correlation=0.8 to 0.9 (dimension=9 to 10 and dimension=7 to 10, respectively). These results again demonstrated the advantage of SEscree over other PCA associated factor retention rules examined in this dissertation. Thus, the performance of SEscree degraded only for high correlations among underlying dimensions and high levels of dimensionality.
- (6) The performance of RMK-OE was found to be similar to other existing methods (except for χ_{ML}^2 and AIC) when correlation=0, and was found to be better than or sometimes close to that of χ_{ML}^2 when correlation=0.3 and 0.6. The mean PAI of RMK-OE for large number of true underlying dimensions was also found to

deteriorate when correlation among underlying dimensions was very high (i.e., 0.8 or 0.9). However, its lowest mean PAI was found to be still above 50%. Thus, the influence of correlation among underlying dimensions on estimation accuracy of RMK-OE is much less than that on methods which utilized PCA results.

(7) The worst performance of the final hybrid method (Hybrid) was mean PAI=75% when dimension=10 with correlation=0.9 among underlying dimensions. The second worst performance was the mean PAI=86% when dimension=9 with correlation=0.9 among underlying dimensions. The other mean PAI for Hybrid were all found to be larger than or equal to 93%, with the majority being either 99% or 100%. Thus, except for the two occasions mentioned above, regardless of levels of correlation among underlying dimensions, Hybrid performed as well as or better than all other rules.

Performance Comparison with Regard to Different Structure Complexities

The results reported in this section focus solely on cases where true number of dimensions is ≥ 2 , as different structure complexities for factor pattern matrix cannot be considered for dimension=1.

In this section, again, two groups of figures are presented, with the first group (Figures 4.13-4.16) reporting mean RMSE results, and the second group (Figures 4.17-4.20) reporting mean PAI results. A summary of results are provided for each group of figures. There are four figures in each group, one figure for each type of structure complexity (simple structure and three different types of complex structure) considered in this dissertation. Note that in all these figures, “nd” indicates the number of true underlying dimensions. Although corresponding tables for these figures are not presented

here, I include some numbers from those tables in my explanation of the results shown in these figures.

Figures and Results for Mean RMSE:

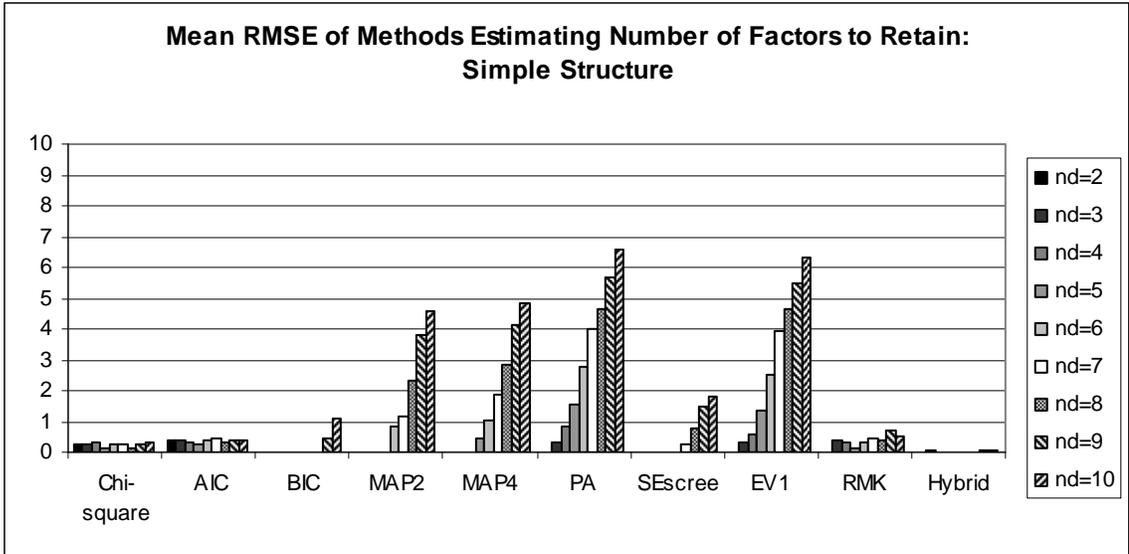


Figure 4.13. Mean RMSE with Simple Structure

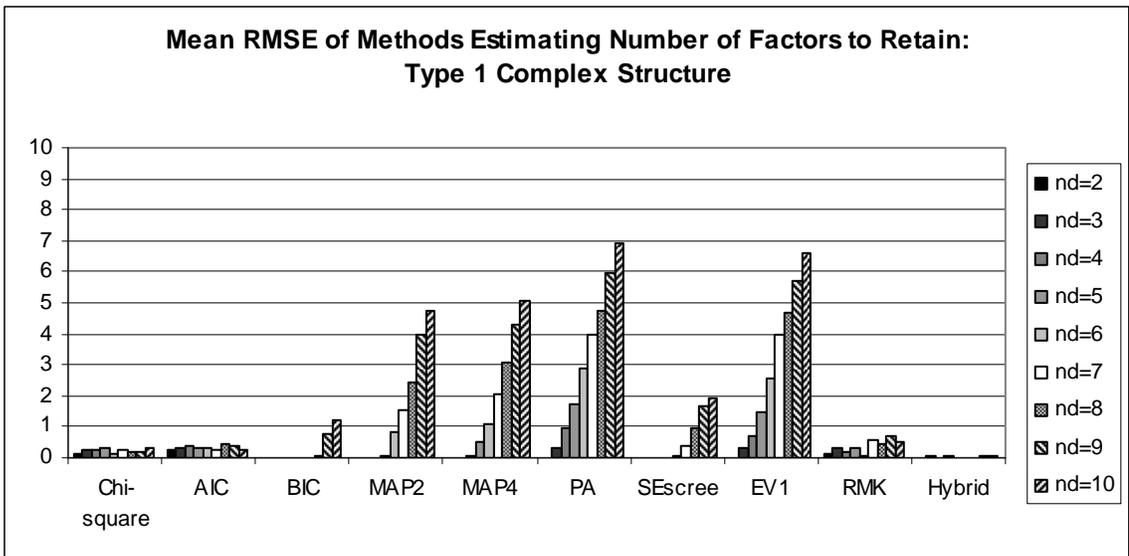


Figure 4.14. Mean RMSE with Type 1 Complex Structure

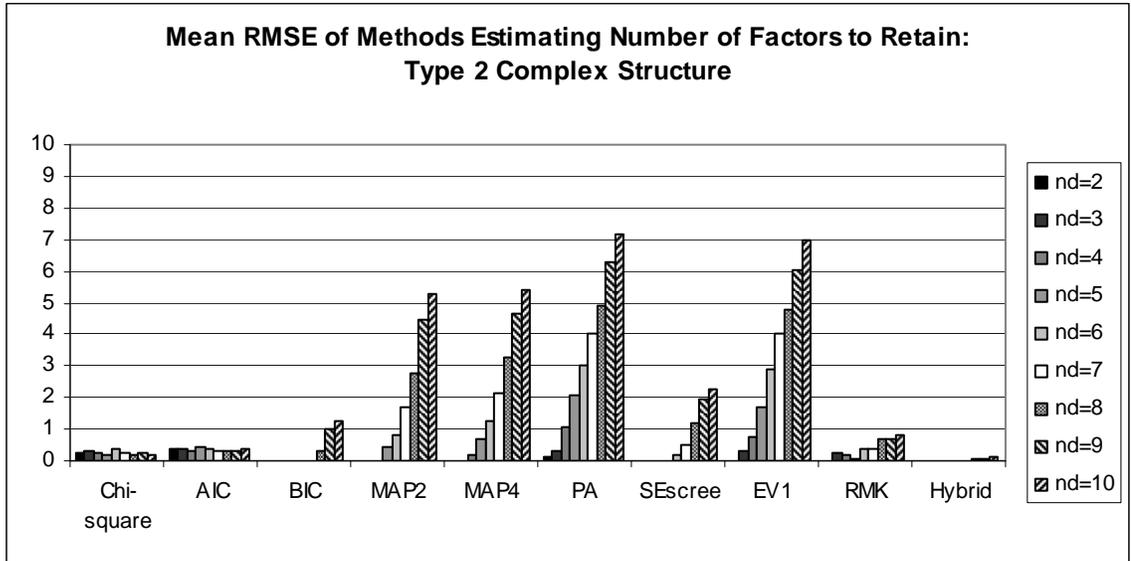


Figure 4.15. Mean RMSE with Type 2 Complex Structure

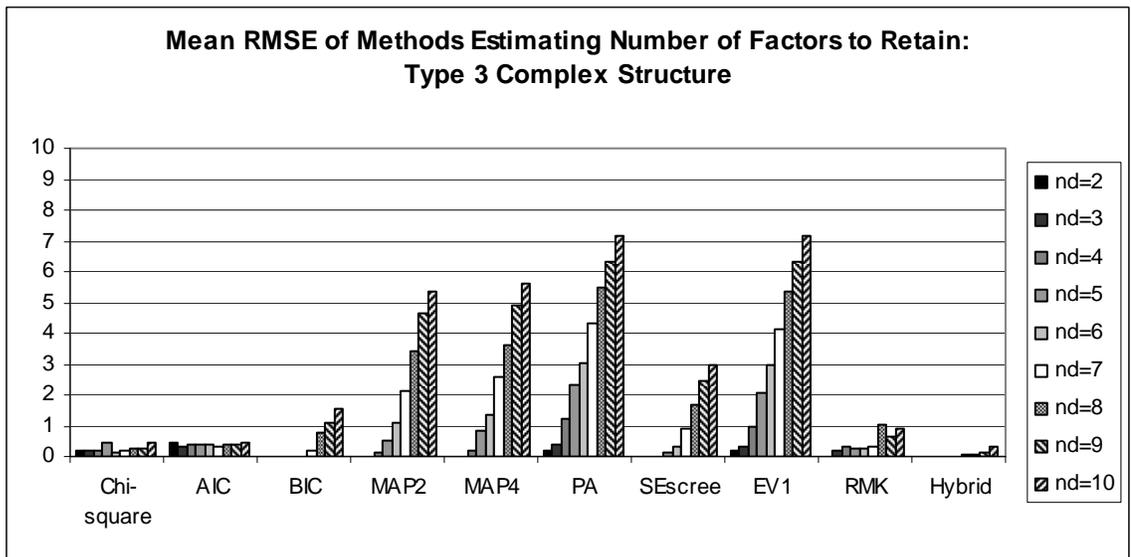


Figure 4.16. Mean RMSE with Type 3 Complex Structure

The following conclusions can be made from the above four figures:

- (1) For each of the four different types of factor pattern matrix structure complexity,

χ_{ML}^2 and AIC were found to have estimation error across all nine different levels

of true underlying dimensionality. Nevertheless, the estimation error, being represented by mean RMSE, was found to be universally small in magnitude (<0.5). Thus, these rules performed well regardless of structure complexity.

- (2) There was a slight influence of factor pattern matrix structure complexity on the accuracy of factor retention decision making for all other rules, including RMK-OE and Hybrid. In general, the more complex the factor pattern matrix structure, the higher the mean RMSE for corresponding levels of true dimensionality.
- (3) Hybrid had an obvious advantage in estimation accuracy over all other rules, as the largest mean RMSE was found to be only 0.3. This occurred when dimension=10 with type 3 complex structure. The next worst performance resulted in mean RMSE ≈ 0.12 when dimension=9 with type 3 complex structure. All other mean RMSE of Hybrid were found to be ≤ 0.01 , and the majority of them were found to have mean RMSE=0. Thus, the final hybrid method performed the best under all levels of structure complexity and for all levels of underlying dimensionality, compared to all other rules.
- (4) RMK-OE was found to have mean RMSE <1 for all cases except for dimension=8 with type 3 complex structure (where mean RMSE ≈ 1.02). Thus, RMK-OE performed relatively well regardless of factor pattern matrix structure complexity.
- (5) PA and EV1 were found to be the worst performing methods, followed by MAP2 and MAP4. SEscree was again found to be the best among methods that utilized PCA results.

Figures and Results for Mean PAI:

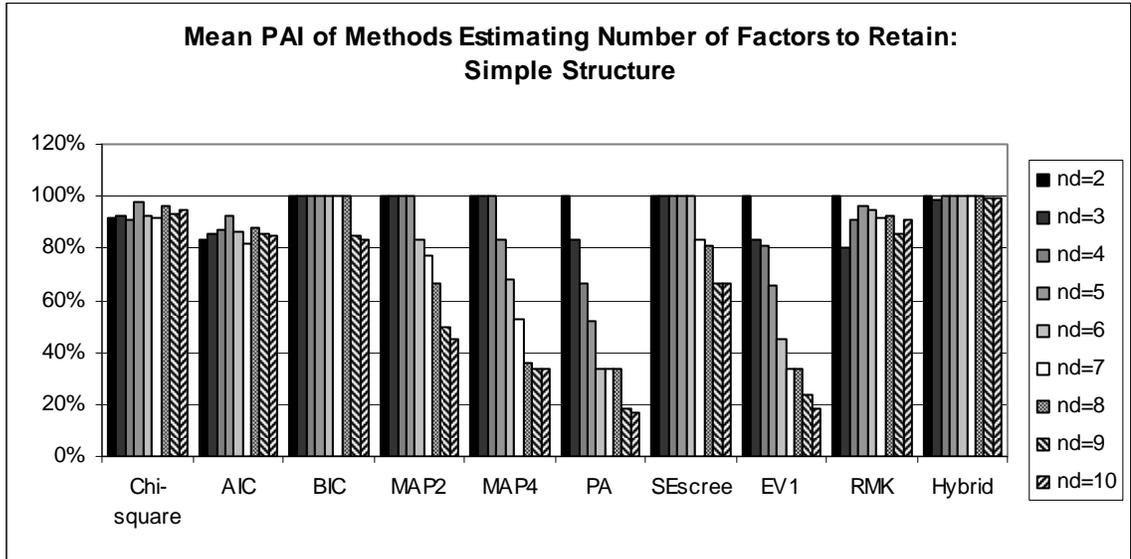


Figure 4.17. Mean PAI with Simple Structure

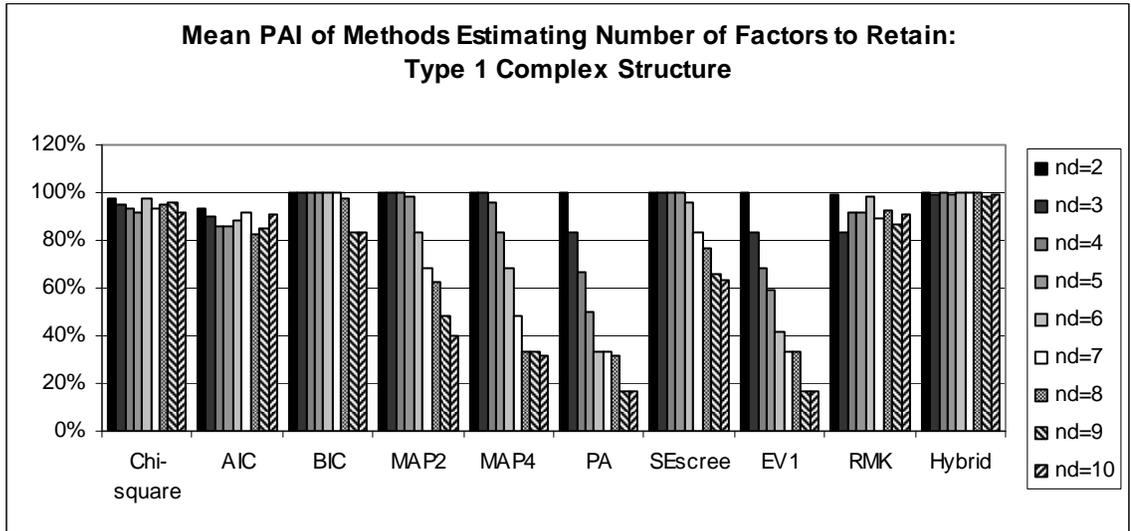


Figure 4.18. Mean PAI with Type 1 Complex Structure

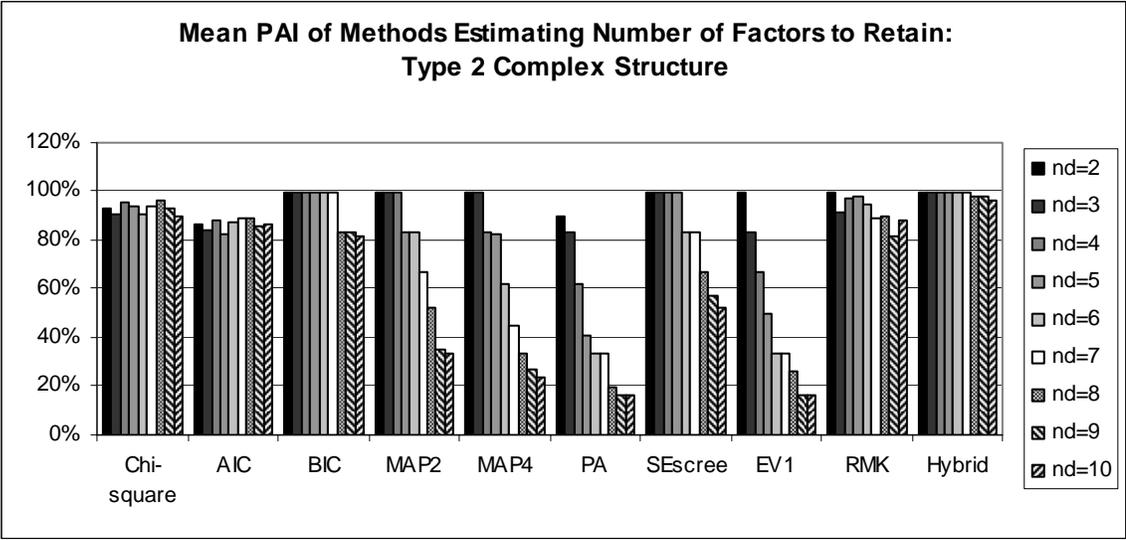


Figure 4.19. Mean PAI with Type 2 Complex Structure

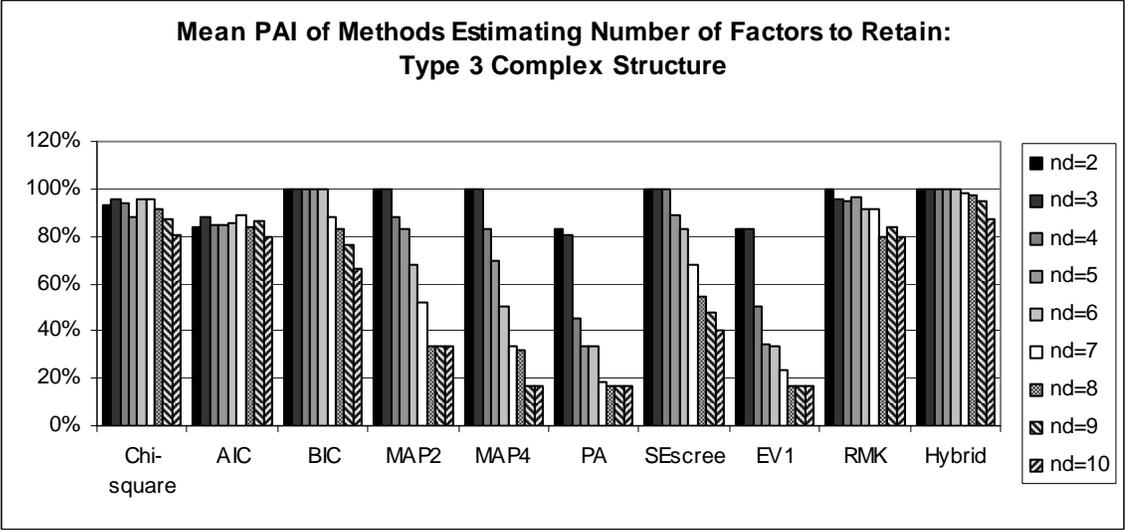


Figure 4.20. Mean PAI with Type 3 Complex Structure

Similar conclusions can be drawn from the above figures on mean PAI, comparing to those on mean RMSE. Specifically, the above mentioned bullet (2) and

bullet (5) for Figures 4.13-4.16 can also be observed from Figures 4.17-4.20. In addition, the following patterns on mean PAI can be observed:

- (1) For each of the four different types of factor pattern matrix structure complexity, χ_{ML}^2 and AIC were found to never have achieved mean PAI=100% (as shown in Figures 4.17-4.20). The mean PAI of χ_{ML}^2 was found to be higher than (or at least the same as) that of AIC across all different types of structure complexity and different levels of true underlying dimensionality. Therefore, χ_{ML}^2 and AIC performed well under various structure complexity and for all levels of underlying dimensionality.
- (2) Except for dimension=10 with type 3 complex structure (where mean PAI \approx 87%), the mean PAI of Hybrid was \geq 95%, with the majority being either 99% or 100%. Thus, the final hybrid method performed the best regardless of structure complexity and levels of underlying dimensionality.
- (3) The lowest mean PAI of RMK-OE was found to be about 79% when dimension=10 with type 3 complexity. The majority of mean PAI of RMK-OE were found to be above 90%. Thus, the final RMK-OE performed well with few exceptions, but not as well as the final hybrid method.

Overall Summary of Results about Correlation and Structure Complexity

In general, the influence of factor pattern matrix structure complexity on the performance of different factor retention rules was not as obvious as that of different correlations among underlying dimensions. Nevertheless, the advantage of the final hybrid method over other factor retention rules in estimation accuracy was obvious, from

both the performance comparison with regard to different correlations among underlying dimensions, and the performance comparison with regard to different factor pattern matrix structure complexity. The advantage of the final hybrid method was also apparent from the overall performance comparison with regard to mean and SD of RMSE and PAI, and the overall performance comparison with estimation error categorization.

Chapter V

DISCUSSION

Accurate dimensionality identification (i.e., the decision about the number of dimensions or factors to retain) is not only important in educational and psychological measurement, but is also important in other areas where factor analysis is used to develop theories. The purpose of this dissertation is to further develop a promising subjective dimension identification rule (i.e., the subjective RMK rule formed through reading the box and whisker plots proposed by Martineau and Reckase [2006]) into a programmable objective method (i.e., the final RMK-OE rule) and to develop a final hybrid method which is based on a compilation of results from a small set of the best performing factor retention rules.

In order to demonstrate the superiority of the final hybrid method and the final RMK-OE rule in making factor retention decisions, their performances were compared to those of other methods. Specifically, four commonly used factor retention rules (plus a variation of the original MAP rule) and three uncommonly used factor retention rules developed for maximum likelihood factor analysis were reviewed and evaluated through simulation studies. For a detailed and complete list of rules included in current comparison, please see the sub-section “Comparison of Factor Retention Rules” in Chapter III.

A simulation study was used for performance comparison of the various factor retention rules. The simulation in this dissertation involved only continuous data, and

took into consideration two conditions that had often been neglected in previous simulation studies: (1) correlation among underlying dimensions or factors especially in the range of 0.6 to 0.9, and (2) various complexities of the relations between observed variables and underlying dimensions (i.e., various structure complexities). Two evaluation criteria were adopted in this dissertation for each simulation condition combination: root mean squared error (RMSE) and percent of accurate identification (PAI).

Results from the simulation study clearly demonstrated the advantage of the final hybrid method over all other existing methods in its dimensionality identification accuracy. In the sections to follow, I first summarize some key findings. I then discuss how these findings contradict or make additions to some common beliefs or previous findings in the field. Additional contributions of these findings are then listed. The chapter ends with a brief description of strengths and limitations of the study, and a list of possible future research.

Summary of Key Findings

Seven key findings are summarized as follows:

- (1) The factor retention rules that applied to principal component analysis (PCA) results (i.e., MAP2, MAP4, PA, EV1, and SEscree, based on their theoretical development) were found to perform perfectly (with mean RMSE=0 and mean PAI=100%) only when underlying dimensions were uncorrelated (i.e., correlation=0). As correlation among underlying dimensions increased, their performance decreased. Moreover, the higher the correlation and the larger the true number of underlying dimensions,

the more the results from these PCA associated rules deviated from factor analysis (FA) associated rules.

- (2) The final hybrid method, which involved the application of four rules (the final RMK-OE rule, maximum likelihood ratio chi-square test (χ^2_{ML}), AIC, and BIC), was found to be very accurate (overall mean PAI $\approx 99\%$ and overall mean RMSE ≈ 0.03). Note, however, this final hybrid method is not simply using the consistent (or the most frequently occurring) number across multiple methods as the final result.
- (3) The final RMK-OE rule was found to be a valuable component resulting in increased accuracy of the final hybrid method.
- (4) In contrast to previous findings, PA was found to be the worst performing method overall, followed by EV1. SEscree was found to be the best performing method among those that utilized PCA results.
- (5) The following three uncommonly used rules: χ^2_{ML} , AIC, and BIC, were found to be the best three existing methods based on overall performance comparisons.
- (6) Correlation among underlying dimensions was a more important observed variable affecting the performance of dimensionality assessment rules than was the level of structure complexity.
- (7) When the true number of underlying dimensions equals 1, all methods performed well (i.e., resulted in dimension=1). Only two cases (out of 25 iterations) of dimension=2 were found with χ^2_{ML} , and three cases of dimension=2 were found with AIC. However, due to the tendency of

severe under-estimation of PCA associated rules (i.e., result in dimension=1 when true number of underlying dimensions is at least two), all PCA associated rules should not be considered as methods for detecting or confirming unidimensionality.

Comparing Key Findings to Common Beliefs or Previous Findings

The above stated key findings either contradict or make additions to some of the common beliefs or previous findings in the field. These common beliefs or previous findings are:

- (1) the number of components obtained from the PCA can be used as a guide in specifying the number of underlying factors;
- (2) the most frequently occurring results across multiple factor retention rules should be used to identify the number of underlying factors; and
- (3) parallel analysis (PA) and minimum average partial correlation (MAP) have been found to be the best performing factor retention rules in previous simulation studies.

The following discussion is divided into three sub-sections, with each of them focusing on one of these common beliefs or previous findings.

Number of Components vs. Number of Factors

Although researchers have realized that the PCA approach is not appropriate for identifying latent traits or underlying dimensions (Fabrigar et al., 1999; Yanai & Ichikawa, 2007), researchers still consider it appropriate to use the number of

components obtained from PCA as a guide to specifying the number of underlying factors to retain (e.g., Velicer et al. (2000)). The reason for this belief is the similar empirical results of PCA to common factor analysis reported in some studies such as Velicer and Jackson (1990) and Ogasawara (2000).

According to the key findings of this dissertation (especially the first key finding), however, it is not appropriate to use the number of components from PCA as a guide for determining the number of factors to retain. This statement is supported by the following two reasons. First, based on the simulation study in this dissertation, the similarity between PCA results and FA results on the number of factors to retain was found only when the underlying dimensions were uncorrelated, or when the number of true underlying (correlated) dimensions was small. Second, many educational and psychological constructs are correlated in the range of 0.6 to 0.9 (Gierl et al., 2006). Based on the findings from this dissertation, this range is the range where PCA associated rules performed the worst when compared to FA associated rules in identifying the number of factors to retain. Therefore, I *do not* recommend using the value of the number of components from PCA associated rules as a guide in specifying the number of factors to retain. My recommendation is thus to use *only* FA associated rules to identify the number of factors to retain.

Most Frequently Occurring Result vs. the Accurate Result

Many researchers have suggested the use of multiple rules to identify appropriate dimensionality, and have recommended adopting the consistent (or most frequently occurring) result as the final result for making factor retention decisions (Henson &

Roberts, 2006; Kim & Mueller, 1978; Netemeyer et al., 2003; Thompson, 2004; Thompson & Daniel, 1996). This suggestion in fact utilizes the “majority wins” rule, which makes sense in general: If various approaches produce the same result, that result should be considered as accurate. According to the key findings of this dissertation (especially the second key finding), however, this recommendation is useful only to some extent and should not be taken literally.

In this dissertation, before the construction of the final hybrid method, I examined three existing FA associated factor retention rules (i.e., χ^2_{ML} , AIC, and BIC), one newly developed rule (i.e., the final RMK-OE, also FA associated), and five PCA associated rules. If the above recommendation were taken literally, then when correlation among underlying dimensions is high (e.g. between 0.6 and 0.9), the final results would always be the number from PCA associated rules, as they were found to be consistent across all five rules. However, the consistent results with the PCA associated rules gave a severe under-estimation when the underlying dimensions were highly correlated.

The final hybrid method proposed in this dissertation (although it is still based on several rules) involved only the rules that were found to perform well overall. Moreover, the final hybrid method was not simply the most frequently occurring number (i.e. mode) across results from the participating rules, but rather a different systematic way to utilize the results from these participating rules.

Specifically, this systematic method utilized both the mode and median, which are resistant to outliers. The larger of the two values was chosen because over-estimation is considered to be less of a problem than under-estimation (Fava & Velicer, 1996; Wood et al., 1996). If results involved decimal points, the values were rounded down to the nearest

integer to improve estimation accuracy (this scenario would only be possible when the median was larger than the mode and was computed as the mean of the two middle-ranked numbers). Rounding down the decimals results in a value closer to the mode, and this produces better estimation accuracy.

The excellent performance of the final hybrid method developed in this dissertation is thus due to the following two reasons. First, participating rules that were used to construct the final hybrid method were carefully chosen based on their overall performance. Second, the method used to combine results from participating rules to produce the final result was carefully designed to be based on a rationale of utilizing two measures of central tendency resistant to outliers. Therefore, I concur with the recommendation of implementing multiple rules to identify appropriate dimensionality, but disagree with the accompanying suggestion of simply picking the most frequently occurring number across results of multiple rules.

PA, MAP, and the Best Performing Factor Retention Rule

Recall that previous simulation studies reported PA as the best performing factor retention rule followed by MAP⁴², and EV1 was found to be the worst performing factor retention rule (e.g., Zwick & Velicer, 1986). The subjective scree test was found to perform better than EV1 but not as well as either PA or MAP (e.g., Velicer et al., 2000). As mentioned in the literature review, these conclusions were drawn from simulation studies in which correlation among underlying dimensions was not considered. This is consistent with the methods involved in those comparison studies as all these factor retention rules are PCA associated, and PCA only considers orthogonal (i.e., uncorrelated) components.

⁴² The MAP method used for comparison in the majority of previous simulations is actually MAP2.

I found in this dissertation (see key findings 1 and 4) that all PCA associated rules performed perfectly when underlying dimensions were uncorrelated. This may be one of the reasons why PA and MAP were found to be well performing in the literature. Those studies, however, also considered some conditions that I did not consider in this dissertation, such as varying communalities, varying sample size, and varying number of observed variables. It is possible that these conditions may be the reason for the claimed advantage of PA and MAP over EV1 in estimation accuracy. Since I did not consider these conditions in this dissertation, an estimation advantage of PA over EV1 was not observed.

Note that the SEscree (a promising objective extension of the subjective scree test) has (until now) not been compared to other factor retention rules in the literature. The results of the simulation study in this dissertation demonstrated an obvious advantage of the SEscree in dimensionality identification accuracy over all other PCA associated rules. This is very different from what other researchers have concluded about the subjective scree test (e.g., Zwick & Velicer, 1986). Despite its unexpected success, the SEscree may be of limited usage when dimensionality identification is the primary purpose.

Additional Contributions of the Key Findings

This section is organized into two parts. First, I reiterate the importance of the final RMK-OE rule in the final hybrid method for dimensionality identification. Second, I

revisit the issue of the most commonly used factor retention rule being the worst or among the worst performing rules in current simulation study.

The Final RMK-OE Rule and the Final Hybrid Method

I evaluated four different hybrid methods based on different groupings of the best overall performing methods. These four types are:

1. The final RMK-OE, the maximum likelihood ratio chi-square test (χ_{ML}^2), AIC, BIC, and SEscree (taking all five best performing rules),
2. The final RMK-OE, χ_{ML}^2 , AIC, BIC (taking the four best performing rules),
3. χ_{ML}^2 , AIC, BIC, and SEscree (for evaluating the contribution of the final RMK-OE to rule 1), and
4. χ_{ML}^2 , AIC, and BIC (for evaluating the contribution of the final RMK-OE to rule 2).

According to Table 4.4, the first two hybrid methods, which involved the final RMK-OE rule, demonstrated a much smaller overall mean RMSE than the other two that did not involve the implementation of the final RMK-OE rule. In addition, these two methods were also found to result in smaller SD of RMSE and PAI, and higher mean PAI. In other words, all four measures showed better performance of those rules that included the final RMK-OE compared to those that did not. This demonstrates the value of the final RMK-OE rule as an indispensable component of the final hybrid method.

Besides its important role in the final hybrid method, the final RMK-OE rule has its own merit due to the fact that it does not require a specific extraction method for implementation. Recall that χ_{ML}^2 , AIC, and BIC are all theoretically more appropriately to be used with maximum likelihood extraction, although they may also be computed

when generalized least squares extraction is employed. The final RMK-OE method does not have this restriction. Therefore, when maximum-likelihood extraction or generalized least squares extraction is not appropriate, desirable, or applicable (i.e., when data do not meet related distributional assumptions), the final RMK-OE rule becomes the only reliable method to use in making factor retention decisions, based on results shown in this dissertation.

Commonly Used Factor Retention Rules and Validity of Developed Theories

As mentioned previously, researchers have used factor analysis in developing theories, and the most frequently used factor retention rules have been found to be either the subjective scree test or EV1 (e.g., Carroll, 1993). According to this dissertation, EV1 was found to be the second worst method in overall performance, and it was found to severely under-estimate when underlying dimensions were highly correlated. Although SEscree was found to be the best performing PCA associated rules, I do not recommend its use in dimensionality identification, because of its performance deterioration when underlying dimensions are highly correlated. The use of the subjective scree test is also a concern, as it is likely to perform similarly to the objective extension (SEscree) but with less reliability due to the subjective nature of graph reading.

Because the vast majority of the many studies that have used factor analysis to identify dimensionality have used the worst-performing rules (i.e., PCA associated rules), I question the validity and reliability of previously developed theories where factor analysis was used. These previously developed theories are likely to have been developed on the basis of severe under-estimation of underlying dimensionality. Such theories may have some unexpected detrimental effects in other areas of research or practice.

For example, if the theories of cognitions have been established with a severe under-estimation of underlying dimensionality, say general intelligence versus at least eight separate intelligences such as verbal, spatial, logical-mathematical intelligence (Woolfolk, 2001), educators cannot accurately pinpoint what to teach (e.g., educators may have a better idea and an easier time to figure out how to enhance students' verbal intelligence than students' general intelligence). In other words, without accurately identifying those (at least) eight separate aspects of intelligence, it will be harder for teachers to make an educational diagnosis and to design appropriate courses to help students strengthen their academic weaknesses.

In educational measurement, the scales constructed on combined dimensions with different weights and on separate dimensions will definitely lead to different rank orderings which in turn will result in different performance level assignment of students. To accurately assign students to performance levels is especially important, as teachers and schools are being evaluated on students' performance on tests (e.g., see Kim & Sunderman, 2005). Likewise, if combined dimensions with different weight combinations or separate dimensions instead of combined dimensions are found in medical studies, different medical approaches or different drug usages may be approved.

Therefore, results of this dissertation not only make a contribution to factor analysis techniques (i.e., a very well-performing hybrid method was proposed in this dissertation), but also call researchers' attention to the possible necessity of re-examining some previously developed theories if factor analysis was used in their development. Specifically, in order to gauge the possible impact of severe under-estimation of underlying dimensionality, it will be useful to apply the proposed final hybrid method or

the final RMK-OE rule to re-evaluate dimensionality in previous studies (especially those that used PCA associated rules in making factor retention decisions). Any of the following investigations may be informative:

- Confirm their theories and findings,
- Determine the effects severe under-estimation of dimensionality would have on the development of theories and interpretations of research findings, and
- Determine what the theories would have looked like had more accurate dimensionality retention rules been applied.

It will be interesting to see how much difference the final hybrid method or the final RMK-OE rule would make in theory development and in the interpretation of scholarly studies in which dimensionality assessment played a significant role.

Strengths and Limitations of the Study, and Possible Future Research

Strengths

There are several strengths of this study.

First, recall that both underlying factors and unique factors were generated from independent multivariate normal distributions. Thus, theoretically speaking, the resulting observed variables were also from multivariate normal distribution (Bartholomew et al., 2008). As stated in Chapter II, the three uncommonly used FA associated methods were developed for maximum likelihood factor analysis, which assumes that all observed variables are a random sample from a multivariate normal distribution. Therefore, the method application for the three FA associated factor retention rules here are consistent with the data generation procedure. Although distributional assumptions are not required for PCA, multivariate normality is still desirable as “if normality is satisfied then the PC

loadings and scores become ML estimators, with the additional desirable property of asymptotic efficiency” (Basilevsky, 1994, p.183). Therefore, generating observed variables from a multivariate normal distribution allows a fair comparison between PCA associated rules and FA associated rules.

Second, the factor retention rule comparison involved in this dissertation is comprehensive. This is because (1) three uncommonly used rules developed for maximum likelihood factor analysis were included for investigation, and (2) the number of true dimensions simulated in this dissertation covered all numbers from 1 to 10. To date, no simulation studies have been performed to cover such a wide range of number of true dimensions. The comprehensive list of factor retention rules and number of true dimensions for comparison help researchers obtain a more complete view of how each method behaves when true dimensionality differs.

Third, the simulation study in this dissertation pays special attention to high correlations among underlying dimensions and various structure complexities of factor pattern matrices. These two conditions have often been neglected by previous simulation studies. Therefore, this dissertation can be considered as the pioneer in this regard. However, results have demonstrated that correlation among underlying dimensions, comparing to factor pattern complexity, demonstrated a larger impact on estimation accuracy of various factor retention rules (especially the PCA associated rules) under examination. My definition of different types of factor pattern complexity in this dissertation differed only in the percentage of manifest variables with high factor pattern coefficients on multiple dimensions. Different definitions of factor pattern complexity

may be used and examined to verify the findings with regard to the impact of factor pattern complexity on estimation accuracy.

Limitations

There are also several limitations of this study.

First, in order to limit the scope of work, I only focused on comparing different factor retention rules using continuous data.

Second, in order to make a fair comparison among factor retention rules in this dissertation, I simulated data based on the formal model (which is mathematically specified in Equation (2.4)) and I only used maximum likelihood as an extraction method (which assumes normality). As the formal model may not represent real data well, the findings may not be directly applicable to the real world. However, real data is likely to add additional complexities rather than doing away with the complexities identified in this dissertation.

Third, I focused most heavily on varying correlations among underlying dimensions in this simulation, and fixed many other factors such as sample size and communality. This may limit the generalizability of the findings in terms of sample size and communalities.

Fourth, due to the computationally intensive nature of this study, I was only able to perform 25 iterations for each condition combination. It may be helpful to confirm and possibly stabilize some of the findings with additional iterations.

Possible Future Research

Future research is needed to address the above mentioned limitations and to further develop the final RMK-OE rule and the final hybrid method to be applicable in a more general framework.

First, adding more iterations (i.e., 100 for each condition combination examined in this dissertation) to examine the stability of two results is desirable. Although the results are unlikely to change significantly, the two findings that may be affected by increasing the number of iterations are:

- (1) The best weight combinations chosen for constructing the final RMK-OE rule, and
- (2) The factor retention rules chosen to construct the final hybrid method.

Second, more simulation conditions and other simulation models may need to be considered for verifying and further generalizing the findings. Specifically, the so called “middle model” which involves an additional term of “minor factors” in Equation (2.4) (e.g., see Tucker, Koopman, & Linn, 1969) may be considered in order to examine the consistency of current findings with data that are closer to the real world. Minor factors are defined as those “real influences, which although contributing somewhat to the covariation among a set of p [observed] variables ... are of little consequences and of a random, unstructured form” (Hakstian & Rogers, 1982). This is especially interesting to researchers in practical applications, as no one would like to retain minor factors as major factors, or in other words, consider unimportant factors as important. Therefore, the ability to distinguish minor and major factors using this methodology will be the topic of

future research. Another related topic would be to determine a threshold (if any) above which accurately identifying the exact number of dimensions is important.

While it is unlikely that more realistic data will reduce the issues identified in this dissertation, other simulation conditions such as the varying communality and the varying sample size may also be considered. Doing so may result in additional findings that may be more useful to educational researchers and practitioners who are dealing with less ideal data.

Third, for identifying dimension=1 case, the final RMK-OE rule utilizes the three factor retention rules that are most appropriate when maximum likelihood extraction method is used. When these three factor retention rules are not appropriate, the component in the final RMK-OE for identifying unidimensionality is not going to produce trustworthy results. Therefore, a better approach for assessing or confirming unidimensionality needs to be developed to be used in conjunction with the final RMK-OE rule.

Fourth, this dissertation is a first step in a thousand mile journey to achieve better equating and scaling results in educational measurement. Therefore, a similar simulation study needs to be carried out with dichotomous data and ordinal data to confirm the application of findings to such test-like data.

Fifth, McDonald (2000) stated that “determining the dimensionality of the model should not be separated from, or treated prior to, determining the fit of an identified structure with a patterned matrix of factor loadings” (p.102). In other words, determining the magnitude of factor pattern coefficients and the number of factors to retain is an iterative process rather than two separate steps. Therefore, the estimation accuracy of

factor pattern coefficients also influences the estimation accuracy of the number of factors to retain. The RMK (and the final RMK-OE) method in particular depends upon the angles among observed variables which, in turn, are calculated from factor pattern coefficients. Consequently, additional research needs to be done to develop a more accurate method of estimating factor pattern matrices, so that a more accurate estimation of the number of dimensions to retain can be achieved.

Finally, a possible next step is to investigate whether highly correlated constructs in real data are impacted differentially by external causes, and if so, the degree of such impact on theory development. Such a study would serve to externally validate the importance of accurately identifying the number of highly correlated factors as discussed in this dissertation.

Appendix A.1
MATLAB Program Code for Bootstrap in RMK-OE

```

function [forplot, lambdaU]=RMKResultstry1(X,nmv,maxDim)
% X = generated data
% nmv = current number of manifest variables.
disp('RMKResultstry1');

% Weights for computing composites
WT=[0.8 0.1 0.1; 0.7 0.2 0.1; 0.6 0.3 0.1; 0.5 0.4 0.1;
    0.6 0.2 0.2; 0.5 0.3 0.2; 0.4 0.4 0.2; 0.4 0.3 0.3;
    1 0 0; 0.9 0.1 0; 0.8 0.2 0; 0.7 0.3 0;
    0.6 0.4 0; 0.5 0.5 0; 0.4 0.6 0; 0.3 0.7 0;
    0 1 0; 0 0 1];

% preallocating empty cells or vectors
forplot=cell(1,numel(WT)/3);
lambdaU=cell(1,maxDim);
lambdaV = cell(1,maxDim);
lambdaP = cell(1,maxDim);
aveAnglesV = zeros(nmv, maxDim);
aveAnglesP = zeros(nmv, maxDim);
anglesV = zeros(nmv*(nmv-1), maxDim);
anglesP = zeros(nmv*(nmv-1), maxDim);

% calculate pattern coefficients with no rotation, varimax rotation, promax
rotation, and calculate angles between manifest variables at different
dimensions based on either varimax rotation or promax rotation.
for d=1:maxDim
    lambdaU{d}=factoran(X,d,'rotate', 'none','maxit',1000000);
    lambdaV{d}=factoran(X,d,'rotate','Varimax','maxit',100000);
    lambdaP{d}=factoran(X,d,'rotate','Promax','maxit',100000);
    [aveAnglesV(:,d),anglesV(:,d)]=calcAngles(lambdaV{d},nmv);
    [aveAnglesP(:,d),anglesP(:,d)]=calcAngles(lambdaP{d},nmv);
end

% compute angle changes or average angle changes.
[diffsV, aveDiffsV]=calcDiffs(anglesV, aveAnglesV, maxDim);
[diffsP, aveDiffsP]=calcDiffs(anglesP, aveAnglesP, maxDim);

% now compute the tr as the whisker of the boxplot;
trV=calcBPTR(diffsV,maxDim);
trP=calcBPTR(diffsP,maxDim);
trAV=calcBPTR(aveDiffsV,maxDim);
trAP=calcBPTR(aveDiffsP,maxDim);

% now compute the bootstrap samples for each column
% first, preallocate empty cells for resulting bootstrap data
btspdiffsV=cell(1,maxDim-1);
btspdiffsP=cell(1,maxDim-1);
btspavediffsV=cell(1,maxDim-1);
btspavediffsP=cell(1,maxDim-1);

for d=2:maxDim
    disp('d=');
    disp(d);
% bootstrap from each column of angle difference matrix based on Varimax
rotation
    btspdiffsV{d-1}=angleboot(diffsV(:,d));

```

```

% bootstrap from each column of angle difference matrix based on Promax
rotation
    btspdifffsP{d-1}=angleboot(difffsP(:,d));
% bootstrap from each column of average angle difference based on Varimax
rotation
    btspavedifffsV{d-1}=angleboot(aveDifffsV(:,d));
% bootstrap from each column of average angle difference based on Promax
rotation
    btspavedifffsP{d-1}=angleboot(aveDifffsP(:,d));
end

for i=1:numel(WT)/3
    wt=WT(i,:); % current weight vector
    wt1=wt(1); % current weight for median
    wt2=wt(2); % current weight for IQR
    wt3=wt(3); % current weight for "trimmed range")
% vector of real composite scores on angle changes or average angle changes
based on current weights and Varimax or Promax rotation
    rlcompV0=wt1.*median(difffsV)'+wt2.*iqr(difffsV)'+wt3.*trV';
    rlcompP0=wt1.*median(difffsP)'+wt2.*iqr(difffsP)'+wt3.*trP';
    rlcompaveV0=wt1.*median(aveDifffsV)'+wt2.*iqr(aveDifffsV)'+wt3.*trAV';
    rlcompaveP0=wt1.*median(aveDifffsP)'+wt2.*iqr(aveDifffsP)'+wt3.*trAP';
% get rid of the first column as it will be all zeros
    rlcompV=rlcompV0(2:maxDim);
    rlcompP=rlcompP0(2:maxDim);
    rlcompaveV=rlcompaveV0(2:maxDim);
    rlcompaveP=rlcompaveP0(2:maxDim);
% preallocating resulting vector
    CIres=zeros(maxDim-1,12);
% composite score on bootstrapped sample of angle changes or average angle
changes for Varimax or Promax rotation
    for j=1:maxDim-1
        compdifffsV=btspdifffsV{j}*wt';
        compdifffsP=btspdifffsP{j}*wt';
        compavedifffsV=btspavedifffsV{j}*wt';
        compavedifffsP=btspavedifffsP{j}*wt';
% rank order composite score
        ncompdifffsV=sort(compdifffsV);
        ncompdifffsP=sort(compdifffsP);
        ncompavedifffsV=sort(compavedifffsV);
        ncompavedifffsP=sort(compavedifffsP);
% Lower bound (2.5%ile) and higher bound (97.5%ile) of the composite score on
angle changes or average angle changes based on Varimax or Promax.
        difffsVLB=ncompdifffsV(25);
        difffsVUB=ncompdifffsV(975);
        difffsPLB=ncompdifffsP(25);
        difffsPUB=ncompdifffsP(975);
        avedifffsVLB=ncompavedifffsV(25);
        avedifffsVUB=ncompavedifffsV(975);
        avedifffsPLB=ncompavedifffsP(25);
        avedifffsPUB=ncompavedifffsP(975);
% store the results of above computation
        CIres(j,:)=[j+1 wt1 wt2 wt3 difffsVLB difffsVUB difffsPLB difffsPUB
avedifffsVLB avedifffsVUB avedifffsPLB avedifffsPUB];
    end

    disp(size(CIres));
    disp(size(rlcompV));
    disp(size(rlcompP));
    disp(size(rlcompaveV));
    disp(size(rlcompaveP));
% store the composite scores based on bootstrapped samples and real values

```

```

        forplot{i}=[CIres rlcompV rlcompP rlcompaveV rlcompaveP];
    end
end

function res=angleboot(angles)
% preallocating results
res=zeros(1000,3);
for i=1:1000
% random sample with replacement (i.e., bootstrap) the vector elements
    sample=randsample(angles,numel(angles),true);
    res(i,1)=median(sample);
    res(i,2)=iqr(sample);
    res(i,3)=calcBPTRvec(sample);
end
end

% compute angle and average angle changes.
function [diffs,aveDiffs]=calcDiffs(angles,aveAngles,maxDim)
angles2=angles;
aveAngles2=aveAngles;
angles2(:,2:end)=angles2(:,1:maxDim-1);
aveAngles2(:,2:end)=aveAngles2(:,1:maxDim-1);
diffs=angles-angles2;
aveDiffs=aveAngles-aveAngles2;
end

% compute angles and average angles
function [aveAngles,angles]=calcAngles(loadings,nVar)
angles=zeros(nVar,nVar);
for i1=1:nVar
    vec1=loadings(i1,:);
    L1=sqrt(vec1*vec1');
    for i2=1:nVar
        vec2=loadings(i2,:);
        if i1~=i2
            L2=sqrt(vec2*vec2');
            cosTheta=(vec1*vec2'/(L1*L2));
            angles(i1,i2)=acosd(cosTheta);
        else
            angles(i1,i2)=NaN;
        end
    end
end
aveAngles=nanmean(angles,2);
angles=reshape(angles,nVar^2,1);
nans=isnan(angles);
angles(nans)=[];
end

% compute the trimmed range
function trSdata=calcBPTR(data,maxDim)
Sdata=sort(data);
% pre-set the upper whisker bar in boxplot
maxw=prctile(Sdata,75)+1.5*iqr(Sdata);
% pre-set the lower whisker bar in boxplot
minw=prctile(Sdata,25)-1.5*iqr(Sdata);
% preallocating
whiSdata = zeros(1, maxDim);
wloSdata = zeros(1, maxDim);
trSdata = zeros(1, maxDim);

```

```

for i=1:maxDim
    B=Sdata(:,i);
    % get the highest value that is less than the pre-set upper bar but closest to
    it
        whiSdata(i)=B(find(Sdata(:,i)<=maxw(i),1,'last'));
    % get the lowest value that is larger than the pre-set lower bar but closest to
    it
        wloSdata(i)=B(find(Sdata(:,i)>=minw(i),1,'first'));
    % define trimmed range as the distance between the two values
        trSdata(i)=whiSdata(i)-wloSdata(i);
end
end

```

```

% This function is similar to the above one, but only computes for one vector
function trSvec=calcBPTRvec(data)
Sdata=sort(data);
maxw=prctile(Sdata,75)+1.5*iqr(Sdata);
minw=prctile(Sdata,25)-1.5*iqr(Sdata);
whiSdata=Sdata(find(Sdata<=maxw,1,'last'));
wloSdata=Sdata(find(Sdata>=minw,1,'first'));
trSvec=whiSdata-wloSdata;
end

```

Appendix A.2
MATLAB Program Code for Decision Process in RMK-OE

```

function res=RMKcompare(forplot,resChisq, resAIC, resBIC)
mxnm=numel(forplot{1}(:,1));
res=zeros(1,18);
slope1=zeros(mxn-2,18);
slope2=zeros(mxn-2,18);
for j=1:18
    mycell = forplot{j}(:,[1,9:10,15]);
    forreg=zeros(3*mxn,2);
    for a=1:mxn-1
        for b=2:4
            forreg((a-1)*3+(b-1),:)=mycell(a,1) mycell(a,b)];
        end
    end
    for i=1:mxn-2
        block1=forreg(1:(i+1)*3,:);
        X1=[ones(size(block1,1),1) block1(:,1)];
        Y1=block1(:,2);
        b1=regress(Y1,X1);
        slope1(i,j)=b1(2);
        block2=forreg((3*i+1):3*mxn,:);
        X2=[ones(size(block2,1),1) block2(:,1)];
        Y2=block2(:,2);
        b2=regress(Y2,X2);
        slope2(i,j)=b2(2);
    end
    for i=1:mxn-4
        if min([resChisq resAIC resBIC])==1 && mode([resChisq resAIC
resBIC])<=2
            res(j)=1;
            elseif (slope1(i,j)<0 && slope2(i,j)>=-0.00075) && ...
                (slope1(1,j)<-5 || (slope2(i+1)-slope2(i)<0.0015 &&
abs((slope2(i+2)-slope2(i+1))-(slope2(i+1)-slope2(i)))<0.0005))
                res(j)=mycell(i,1);
                break;
            end
        end
    end
end
end
end

```

Appendix A.3
MATLAB Program Code for Data Generation

```

function data=comparerulesnd2_10regall(nmv,nD,RD,COMP)
nIter=25;           % number of iteration
count=1;           % current data row
% initialize the results matrix and key matrix.
res=zeros(numel(nD)*numel(RD)*numel(COMP)*nIter,26);
key=zeros(numel(nD)*numel(RD)*numel(COMP)*nIter,5);

for j=1:numel(nD)
    nd=nD(j);
    for k=1:numel(RD);
        rd=RD(k);
        for l=1:numel(COMP)
            comp=COMP(l);
            for m=1:nIter

                if comp==0           % for simple structure
                    X=genSS(nmv,nd,rd); % generate simple structure data
                elseif comp==1       % for complex structure, level 1
                    X=genCS1(nmv,nd,rd); % generate complex structure data
                elseif comp==2       % for complex structure, level 2
                    X=genCS2(nmv,nd,rd); % generate complex structure data
                else                  % for complex structure, level 3
                    X=genCS3(nmv,nd,rd); % generate complex structure data
                end

                % make sure maximum number of dimensions is an integer and is rounding up;
                maxDim=ceil(nmv/2.5);
                R=corrcoef(X);
                S=cov(X);
                [eigenvec,eigenX]=pcacov(R);
                loadX=eigenvec*sqrt(diag(eigenX));
                % compute MAP2, MAP4, PA, Scree, and Eigenvalue>1 rule based on PCA results
                resMAP2=MAP2rule(R, nmv,loadX);
                resMAP4=MAP4rule(R, nmv,loadX);
                resPA=PARule(nmv,eigenX);
                resScree=SEscree(eigenX, nmv);
                resEV1=eigenValDim(eigenX);
                % generate RMK composites and bootstrapped CIs
                [forplot, lambdaU]=RMKResultstry1(X,nmv,maxDim);
                % compute Chisq, AIC, BIC, and RMK based on Factor analysis results
                resChisq=Chisqtest(S,lambdaU, nmv, maxDim);
                resAIC=AICrule(S, lambdaU, nmv, maxDim);
                resBIC=BICrule(S, lambdaU, nmv, maxDim);
                resRMK=RMKcompare(forplot,resChisq, resAIC, resBIC);
                res(count,:)=[resRMK resChisq resAIC resBIC resMAP2 resMAP4
                resPA resScree resEV1]; % store results of this iteration
                key(count,:)=[nmv nd rd comp m]; % store key for this iteration
                count=count+1;           % increase count for next iteration

            end
        end
    end
end

data=[key res]; % final results
end

%For data generation, use (0.5,1.1) range as high discrimination, and
%(0.05,0.2) as low discrimination.

```

```

%generate simple structure
function X=genSS(nmv,nd,rd)
vpd=nmv/nd;
a=(0.20-0.05)*rand(nmv,nd)+0.05;
for j=1:nd
    for i=((j-1)*vpd+1):(j*vpd)
        a(i,j)=(1.1-0.5)*rand(1)+0.5;
    end
end
muuf=zeros(1,nmv);
mucf=zeros(1,nd);
% covariance/correlation matrix of true dimensions;
sigmacf=rd*ones(nd)+(1-rd)*eye(nd);
pc0=zeros(nmv,nd); % initialize the conceptual pattern matrix;
pc=zeros(nmv,nd); % initialize the final population pattern
matrix;
for i=1:nmv
    pc0(i,:)=a(i,:)*((1+a(i,:)*sigmacf*a(i,:)')^(-1/2));
    pc(i,:)=pc0(i,:)*((pc0(i,:)*sigmacf*pc0(i,:)')^(-1/2))*sqrt(0.8);
end
%get diagonal elements on the unique variance matrix;
vuf0=ones(nmv,1)-diag(pc*sigmacf*pc');
sigmauf=diag(vuf0); % covariance matrix of unique factors;

% now generate data using linear combination of cf and uf;
cf=mvnrnd(mucf,sigmacf,10000);
uf=mvnrnd(muuf,sigmauf,10000);
X0=cf*pc'+uf;
X=zscore(X0);
end

function X=genCS1(nmv,nd,rd)
vpd=nmv/nd;
a=(0.20-0.05)*rand(nmv,nd)+0.05;
for j=1:nd
    for i=((j-1)*vpd+1):(j*vpd)
        a(i,j)=(1.1-0.5)*rand(1)+0.5;
    end
end
ixx=indexgen(nmv,nd,0.1); % 10% variable change for complex=1;
for ix=1: numel(ixx);
    i=ixx(ix); % get indexed row number;
    for j=1:nd
        if a(i,j)>0.5
            break;
        end
    end
    for nj=(j+1):(j+round(max(3,nd)/2)-1)
        cj=mod(nj-1,nd)+1; %get the current column for modification;
        a(i,cj)=(1.1-0.5)*rand(1)+0.5;
    end
end
muuf=zeros(1,nmv);
mucf=zeros(1,nd);
sigmacf=rd*ones(nd)+(1-rd)*eye(nd);
pc0=zeros(nmv,nd); % initialize the conceptual pattern matrix;
pc=zeros(nmv,nd); % initialize the final population pattern matrix;
for i=1:nmv
    pc0(i,:)=a(i,:)*((1+a(i,:)*sigmacf*a(i,:)')^(-1/2));
    pc(i,:)=pc0(i,:)*((pc0(i,:)*sigmacf*pc0(i,:)')^(-1/2))*sqrt(0.8);
end
%get diagonal elements on the unique variance matrix;

```

```

vuf0=ones(nmv,1)-diag(pc*sigmacf*pc');
sigmauf=diag(vuf0); % covariance matrix of unique factors;

% now generate data using linear combination of cf and uf;
cf=mvnrnd(muuf,sigmacf,10000);
uf=mvnrnd(muuf,sigmauf,10000);
X0=cf*pc'+uf;
X=zscore(X0);
end

function X=genCS2(nmv,nd,rd)
vpd=nmv/nd;
a=(0.20-0.05)*rand(nmv,nd)+0.05;
for j=1:nd
    for i=((j-1)*vpd+1):(j*vpd)
        a(i,j)=(1.1-0.5)*rand(1)+0.5;
    end
end
ixx=indexgen(nmv,nd,0.3); % 30% variable change for complex=2;
for ix=1:numel(ixx);
    i=ixx(ix); % get indexed row number;
    for j=1:nd
        if a(i,j)>0.5
            break;
        end
    end
    for nj=(j+1):(j+round(max(3,nd)/2)-1)
        cj=mod(nj-1,nd)+1; %get the current column for modification;
        a(i,cj)=(1.1-0.5)*rand(1)+0.5;
    end
end

muuf=zeros(1,nmv);
mucf=zeros(1,nd);
% covariance/correlation matrix of true dimensions;
sigmacf=rd*ones(nd)+(1-rd)*eye(nd);
pc0=zeros(nmv,nd); % initialize the conceptual pattern matrix;
pc=zeros(nmv,nd); % initialize the final population pattern matrix;
for i=1:nmv
    pc0(i,:)=a(i,:)*((1+a(i,:)*sigmacf*a(i,:))^-1/2);
    pc(i,:)=pc0(i,:)*((pc0(i,:)*sigmacf*pc0(i,:))^-1/2)*sqrt(0.8);
end
%get diagonal elements on the unique variance matrix;
vuf0=ones(nmv,1)-diag(pc*sigmacf*pc');
sigmauf=diag(vuf0); % covariance matrix of unique factors;

% now generate data using linear combination of cf and uf;
cf=mvnrnd(muuf,sigmacf,10000);
uf=mvnrnd(muuf,sigmauf,10000);
X0=cf*pc'+uf;
X=zscore(X0);
end

function X=genCS3(nmv,nd,rd)
vpd=nmv/nd;
a=(0.20-0.05)*rand(nmv,nd)+0.05;
for j=1:nd
    for i=((j-1)*vpd+1):(j*vpd)
        a(i,j)=(1.1-0.5)*rand(1)+0.5;
    end
end
end

```

```

ixx=indexgen (nmv, nd, 0.5); % 50% variable change for complex=3;
for ix=1:numel(ixx);
    i=ixx(ix); % get indexed row number;
    for j=1:nd
        if a(i,j)>0.5
            break;
        end
    end
    for nj=(j+1):(j+round(max(3,nd)/2)-1)
        cj=mod(nj-1,nd)+1; %get the current column for modification;
        a(i,cj)=(1.1-0.5)*rand(1)+0.5;
    end
end
muuf=zeros(1,nmv);
mucf=zeros(1,nd);
% covariance/correlation matrix of true dimensions;
sigmacf=rd*ones(nd)+(1-rd)*eye(nd);
pc0=zeros(nmv,nd); % initialize the conceptual pattern matrix;
pc=zeros(nmv,nd); % initialize the final population pattern matrix;
for i=1:nmv
    pc0(i,:)=a(i,:)*((1+a(i,:)*sigmacf*a(i,:))^-1/2);
    pc(i,:)=pc0(i,:)*((pc0(i,:)*sigmacf*pc0(i,:))^-1/2)*sqrt(0.8);
end
%get diagonal elements on the unique variance matrix;
vuf0=ones(nmv,1)-diag(pc*sigmacf*pc');
sigmauf=diag(vuf0); % covariance matrix of unique factors;

% now generate data using linear combination of cf and uf;
cf=mvnrnd(mucf,sigmacf,10000);
uf=mvnrnd(muuf,sigmauf,10000);
X0=cf*pc'+uf;
X=zscore(X0);
end

%%index generation function;
function ixx=indexgen (nmv, nd, complev)
b=zeros(round((nmv/nd)*complev),nd);
a=(1:round((nmv/nd)*complev));
for i=1:numel(a)
    b(i,:)=(a(i):nmv/nd:nmv);
end
ixx=sort(reshape(b,numel(b),1));
end

%% Data generation for dimension=1
function X=genSSnd1(nmv)
h2=0.8*ones(nmv,1); % define communality.
pc=sqrt(h2);
muuf=zeros(1,nmv);
mucf=zeros(1,1);
%get diagonal elements on the unique variance matrix;
vuf0=ones(nmv,1)-diag(pc*pc');
sigmauf=diag(vuf0); % covariance matrix of unique factors;

% now generate data using linear combination of cf and uf;
cf=mvnrnd(mucf,1,10000);
uf=mvnrnd(muuf,sigmauf,10000);
X0=cf*pc'+uf;
X=zscore(X0);
end

```

Appendix A.4

MATLAB Program Code for Existing Factor Retention Rules

```

% Maximum likelihood ratio chi-square test based on factor analysis results.
function res=Chisqtest(S,lambdaU, nmv, maxDim)
h2=cell(1,maxDim);
si=cell(1,maxDim);
sigmahat=cell(1,maxDim);
Chisq=zeros(1,maxDim);
df=zeros(1,maxDim);
n=10000-1; %10000 is the sample size;
for d=1:maxDim
    h2{d}=diag(lambdaU{d}*lambdaU{d}');
    si{d}=eye(nmv)-diag(h2{d});
    sigmahat{d}=lambdaU{d}*lambdaU{d}'+si{d};
    Chisq(d)=n*(log(det(sigmahat{d}))-log(det(S))+trace(S*inv(sigmahat{d}))-
nmv); %Chi-square value;
    df(d)=(1/2)*((nmv-d)^2-(nmv+d)); % df for chi-square test;
end
PI=chi2cdf(Chisq,df);
P=1-PI; % get the p value for goodness-of-fit (chi-square) test;
for d=1:maxDim
    if P(d)>=0.05
        res=d;
        break;
    else res=maxDim;
    end
end
end

% AIC based on factor analysis results.
function res=AICrule(S,lambdaU, nmv, maxDim)
h2=cell(1,maxDim);
si=cell(1,maxDim);
sigmahat=cell(1,maxDim);
n=10000-1; %10000 is the sample size;
ML=zeros(1,maxDim);
AIC=zeros(1,maxDim);
for d=1:maxDim
    h2{d}=diag(lambdaU{d}*lambdaU{d}');
    si{d}=eye(nmv)-diag(h2{d});
    sigmahat{d}=lambdaU{d}*lambdaU{d}'+si{d};
% maximum log-likelihood without the constant part;
    ML(d)=(-1/2)*n*(log(det(sigmahat{d}))+trace(S*inv(sigmahat{d})));
    AIC(d)=(-2)*ML(d)+(2*d*nmv+2*nmv-d^2+d);
end
minAIC=min(AIC);
for d=1:maxDim
    if AIC(d)==minAIC
        res=d;
        break;
    end
end
end

% BIC based on factor analysis results.
function res=BICrule(S, lambdaU, nmv, maxDim)
h2=cell(1,maxDim);
si=cell(1,maxDim);
sigmahat=cell(1,maxDim);
n=10000-1; %10000 is the sample size;

```

```

ML=zeros(1,maxDim);
BIC=zeros(1,maxDim);
for d=1:maxDim
    h2{d}=diag(lambdaU{d}*lambdaU{d}');
    si{d}=eye(nmv)-diag(h2{d});
    sigmahat{d}=lambdaU{d}*lambdaU{d}'+si{d};
    ML(d)=(-1/2)*n*(log(det(sigmahat{d}))+trace(S*inv(sigmahat{d}))); % maximum
log-likelihood without the constant part;
    BIC(d)=(-2)*ML(d)+((log(10000))/2)*(2*d*nmv+2*nmv-d^2+d);
end
minBIC=min(BIC);
for d=1:maxDim
    if BIC(d)==minBIC;
        res=d;
        break;
    end
end
end

```

```

% MAP2 based on PCA results.
function res=MAP2rule(R,nmv,loadX)
fm=zeros(1,nmv-1);
for i=1:(nmv-1)
    A=loadX(:,1:i);
    Rstar0=R-A*A';
    d=diag(1./(sqrt(diag(Rstar0)))); %get D^(-1/2) part where D=diag(Rstar0);
    Rstar=d*Rstar0*d;
    fm(i)=(sum(sum(Rstar.^2))-sum(diag(Rstar.^2)))/(nmv*(nmv-1));
end
f0=(sum(sum(R.^2))-sum(diag(R.^2)))/(nmv*(nmv-1));
if fm(1)>f0
    res=0;
else
    for i=1:(nmv-1)
        if fm(i)==min(fm)
            res=i;
            break;
        end
    end
end
end
end

```

```

% MAP4 based on PCA results.
function res=MAP4rule(R,nmv,loadX)
fm=zeros(1,nmv-1);
for i=1:(nmv-1)
    A=loadX(:,1:i);
    Rstar0=R-A*A';
    d=diag(1./(sqrt(diag(Rstar0))));
    Rstar=d*Rstar0*d;
    fm(i)=(sum(sum(Rstar.^4))-sum(diag(Rstar.^4)))/(nmv*(nmv-1));
end
f0=(sum(sum(R.^4))-sum(diag(R.^4)))/(nmv*(nmv-1));
if fm(1)>f0
    res=0;
else
    for i=1:(nmv-1)
        if fm(i)==min(fm)
            res=i;
            break;
        end
    end
end

```

```

    end
end
end

% PA based on PCA results.
function res=PArule(nmv,eigenX)
D=zeros(1000,nmv);
res0=zeros(1,nmv);
for i=1:1000
    A=randn(10000,nmv);
    [loads,eigen]=pcacov(corrcoef(A));
    D(i,:)=eigen;
end
mneig=mean(D);
for i=1:nmv
    if eigenX(i)>mneig(i)
        res0(i)=1;
    else res0(i)=0;
    end
end
res=sum(res0);
end

% SEscree based on PCA results.
function res=SEscree(eigenX, nmv)
res0=zeros(1,nmv-2);
for d=1:(nmv-2)
    X=zeros(nmv-d+1,2);
    X1=ones(nmv-d+1,1);
    X(:,1)=X1;
    X(:,2)=(d:nmv);
    Y=eigenX(d:nmv);
    [b, bint,r]=regress(Y,X);
    SE=sqrt(sum(r.^2)/(numel(r)-2));
    if SE>(1/nmv)
        res0(d)=1;
    else res0(d)=0;
    end
end
res=sum(res0);
end

% Eigenvalues-greater-than-one based on PCA results.
function n=eigenValDim(list)
n=sum(sum(list>=ones(size(list)))));
end

```

Appendix A.5

MATLAB Program Code for RMSE Comparison of Various Rules

```
%Note that for dimension=1, no correlation among underlying dimensions or
%complex structures can be considered. For dimension=2 to 9, the 6 different
%correlations among underlying dimensions and 4 types of simple or complex
%structures are considered, thus the number of rows in the result matrix is:
%1+4*6*9=217.
```

```
function res=RMSEcompute(data1)
res=zeros(217,size(data1,2)-1);
% the columns in res are the number of columns in data while excluding the
% column of nIter;
for i=1:217
    data=data1(25*(i-1)+1:25*i,:);
    res(i,:)=RMSE(data);
end
end
```

```
function res=RMSE(data)
sumRMK_8_1_1=0;
sumRMK_7_2_1=0;
sumRMK_6_3_1=0;
sumRMK_5_4_1=0;
sumRMK_6_2_2=0;
sumRMK_5_3_2=0;
sumRMK_4_4_2=0;
sumRMK_4_3_3=0;
sumRMK_1_0_0=0;
sumRMK_9_1_0=0;
sumRMK_8_2_0=0;
sumRMK_7_3_0=0;
sumRMK_6_4_0=0;
sumRMK_5_5_0=0;
sumRMK_4_6_0=0;
sumRMK_3_7_0=0;
sumRMK_0_1_0=0;
sumRMK_0_0_1=0;
sumChisq=0;
sumAIC=0;
sumBIC=0;
sumMAP2=0;
sumMAP4=0;
sumPA=0;
sumSEScree=0;
sumEV=0;
for i=1:25
    d2RMK_8_1_1=(data(i,6)-data(i,2))^2;
    sumRMK_8_1_1=sumRMK_8_1_1+d2RMK_8_1_1;
    d2RMK_7_2_1=(data(i,7)-data(i,2))^2;
    sumRMK_7_2_1=sumRMK_7_2_1+d2RMK_7_2_1;
    d2RMK_6_3_1=(data(i,8)-data(i,2))^2;
    sumRMK_6_3_1=sumRMK_6_3_1+d2RMK_6_3_1;
    d2RMK_5_4_1=(data(i,9)-data(i,2))^2;
    sumRMK_5_4_1=sumRMK_5_4_1+d2RMK_5_4_1;
    d2RMK_6_2_2=(data(i,10)-data(i,2))^2;
    sumRMK_6_2_2=sumRMK_6_2_2+d2RMK_6_2_2;
    d2RMK_5_3_2=(data(i,11)-data(i,2))^2;
    sumRMK_5_3_2=sumRMK_5_3_2+d2RMK_5_3_2;
    d2RMK_4_4_2=(data(i,12)-data(i,2))^2;
    sumRMK_4_4_2=sumRMK_4_4_2+d2RMK_4_4_2;
    d2RMK_4_3_3=(data(i,13)-data(i,2))^2;
    sumRMK_4_3_3=sumRMK_4_3_3+d2RMK_4_3_3;
```

```

d2RMK_1_0_0=(data(i,14)-data(i,2))^2;
sumRMK_1_0_0=sumRMK_1_0_0+d2RMK_1_0_0;
d2RMK_9_1_0=(data(i,15)-data(i,2))^2;
sumRMK_9_1_0=sumRMK_9_1_0+d2RMK_9_1_0;
d2RMK_8_2_0=(data(i,16)-data(i,2))^2;
sumRMK_8_2_0=sumRMK_8_2_0+d2RMK_8_2_0;
d2RMK_7_3_0=(data(i,17)-data(i,2))^2;
sumRMK_7_3_0=sumRMK_7_3_0+d2RMK_7_3_0;
d2RMK_6_4_0=(data(i,18)-data(i,2))^2;
sumRMK_6_4_0=sumRMK_6_4_0+d2RMK_6_4_0;
d2RMK_5_5_0=(data(i,19)-data(i,2))^2;
sumRMK_5_5_0=sumRMK_5_5_0+d2RMK_5_5_0;
d2RMK_4_6_0=(data(i,20)-data(i,2))^2;
sumRMK_4_6_0=sumRMK_4_6_0+d2RMK_4_6_0;
d2RMK_3_7_0=(data(i,21)-data(i,2))^2;
sumRMK_3_7_0=sumRMK_3_7_0+d2RMK_3_7_0;
d2RMK_0_1_0=(data(i,22)-data(i,2))^2;
sumRMK_0_1_0=sumRMK_0_1_0+d2RMK_0_1_0;
d2RMK_0_0_1=(data(i,23)-data(i,2))^2;
sumRMK_0_0_1=sumRMK_0_0_1+d2RMK_0_0_1;
d2Chisq=(data(i,24)-data(i,2))^2;
sumChisq=sumChisq+d2Chisq;
d2AIC=(data(i,25)-data(i,2))^2;
sumAIC=sumAIC+d2AIC;
d2BIC=(data(i,26)-data(i,2))^2;
sumBIC=sumBIC+d2BIC;
d2MAP2=(data(i,27)-data(i,2))^2;
sumMAP2=sumMAP2+d2MAP2;
d2MAP4=(data(i,28)-data(i,2))^2;
sumMAP4=sumMAP4+d2MAP4;
d2PA=(data(i,29)-data(i,2))^2;
sumPA=sumPA+d2PA;
d2SEScree=(data(i,30)-data(i,2))^2;
sumSEScree=sumSEScree+d2SEScree;
d2EV=(data(i,31)-data(i,2))^2;
sumEV=sumEV+d2EV;
end
rmseRMK_8_1_1=sqrt(sumRMK_8_1_1/25);
rmseRMK_7_2_1=sqrt(sumRMK_7_2_1/25);
rmseRMK_6_3_1=sqrt(sumRMK_6_3_1/25);
rmseRMK_5_4_1=sqrt(sumRMK_5_4_1/25);
rmseRMK_6_2_2=sqrt(sumRMK_6_2_2/25);
rmseRMK_5_3_2=sqrt(sumRMK_5_3_2/25);
rmseRMK_4_4_2=sqrt(sumRMK_4_4_2/25);
rmseRMK_4_3_3=sqrt(sumRMK_4_3_3/25);
rmseRMK_1_0_0=sqrt(sumRMK_1_0_0/25);
rmseRMK_9_1_0=sqrt(sumRMK_9_1_0/25);
rmseRMK_8_2_0=sqrt(sumRMK_8_2_0/25);
rmseRMK_7_3_0=sqrt(sumRMK_7_3_0/25);
rmseRMK_6_4_0=sqrt(sumRMK_6_4_0/25);
rmseRMK_5_5_0=sqrt(sumRMK_5_5_0/25);
rmseRMK_4_6_0=sqrt(sumRMK_4_6_0/25);
rmseRMK_3_7_0=sqrt(sumRMK_3_7_0/25);
rmseRMK_0_1_0=sqrt(sumRMK_0_1_0/25);
rmseRMK_0_0_1=sqrt(sumRMK_0_0_1/25);
rmseChisq=sqrt(sumChisq/25);
rmseAIC=sqrt(sumAIC/25);
rmseBIC=sqrt(sumBIC/25);
rmseMAP2=sqrt(sumMAP2/25);
rmseMAP4=sqrt(sumMAP4/25);
rmsePA=sqrt(sumPA/25);
rmseSEScree=sqrt(sumSEScree/25);
rmseEV=sqrt(sumEV/25);

```

```
res=[data(1,1:4) rmseRMK_8_1_1 rmseRMK_7_2_1 rmseRMK_6_3_1 rmseRMK_5_4_1
      rmseRMK_6_2_2 rmseRMK_5_3_2 rmseRMK_4_4_2 rmseRMK_4_3_3
      rmseRMK_1_0_0 rmseRMK_9_1_0 rmseRMK_8_2_0 rmseRMK_7_3_0
      rmseRMK_6_4_0 rmseRMK_5_5_0 rmseRMK_4_6_0 rmseRMK_3_7_0
      rmseRMK_0_1_0 rmseRMK_0_0_1 rmseChisq      rmseAIC
      rmseBIC      rmseMAP2      rmseMAP4      rmsePA
      rmseSEScree  rmseEV];
```

end

Appendix A.6

MATLAB Program Code for PAI Comparison of Various Rules

```
%Note that for dimension=1, no correlation among underlying dimensions or
%complex structures can be considered. For dimensions=2 to 9, the 6 different
%correlations among underlying dimensions and 4 types of simple or complex
%structures are considered, thus the number of rows in the result matrix is:
%1+4*6*9=217.
```

```
function res=PCTcompute(data1)
res=zeros(217,26*5+4);
%number of columns equals to number of methods*5+4 (simulation conditions);
for i=1:217
    data=data1(25*(i-1)+1:25*i,:);
    res(i,:)=PCT(data);
end

end
```

```
function res=PCT(data)
difRMK_8_1_1=zeros(1,5);
difRMK_7_2_1=zeros(1,5);
difRMK_6_3_1=zeros(1,5);
difRMK_5_4_1=zeros(1,5);
difRMK_6_2_2=zeros(1,5);
difRMK_5_3_2=zeros(1,5);
difRMK_4_4_2=zeros(1,5);
difRMK_4_3_3=zeros(1,5);
difRMK_1_0_0=zeros(1,5);
difRMK_9_1_0=zeros(1,5);
difRMK_8_2_0=zeros(1,5);
difRMK_7_3_0=zeros(1,5);
difRMK_6_4_0=zeros(1,5);
difRMK_5_5_0=zeros(1,5);
difRMK_4_6_0=zeros(1,5);
difRMK_3_7_0=zeros(1,5);
difRMK_0_1_0=zeros(1,5);
difRMK_0_0_1=zeros(1,5);
difChisq=zeros(1,5);
difAIC=zeros(1,5);
difBIC=zeros(1,5);
difMAP2=zeros(1,5);
difMAP4=zeros(1,5);
difPA=zeros(1,5);
difSEScree=zeros(1,5);
difEV=zeros(1,5);

for i=1:25
    d1RMK_8_1_1=data(i,6)-data(i,2);
    d1RMK_7_2_1=data(i,7)-data(i,2);
    d1RMK_6_3_1=data(i,8)-data(i,2);
    d1RMK_5_4_1=data(i,9)-data(i,2);
    d1RMK_6_2_2=data(i,10)-data(i,2);
    d1RMK_5_3_2=data(i,11)-data(i,2);
    d1RMK_4_4_2=data(i,12)-data(i,2);
    d1RMK_4_3_3=data(i,13)-data(i,2);
    d1RMK_1_0_0=data(i,14)-data(i,2);
    d1RMK_9_1_0=data(i,15)-data(i,2);
    d1RMK_8_2_0=data(i,16)-data(i,2);
    d1RMK_7_3_0=data(i,17)-data(i,2);
    d1RMK_6_4_0=data(i,18)-data(i,2);
    d1RMK_5_5_0=data(i,19)-data(i,2);
    d1RMK_4_6_0=data(i,20)-data(i,2);
```

```

d1RMK_3_7_0=data(i,21)-data(i,2);
d1RMK_0_1_0=data(i,22)-data(i,2);
d1RMK_0_0_1=data(i,23)-data(i,2);
d1Chisq=data(i,24)-data(i,2);
d1AIC=data(i,25)-data(i,2);
d1BIC=data(i,26)-data(i,2);
d1MAP2=data(i,27)-data(i,2);
d1MAP4=data(i,28)-data(i,2);
d1PA=data(i,29)-data(i,2);
d1SEScree=data(i,30)-data(i,2);
d1EV=data(i,31)-data(i,2);

if d1RMK_8_1_1<=-2
    difRMK_8_1_1(1,1)=difRMK_8_1_1(1,1)+0.04;
elseif d1RMK_8_1_1>=2
    difRMK_8_1_1(1,5)=difRMK_8_1_1(1,5)+0.04;
else
    difRMK_8_1_1(1,d1RMK_8_1_1+3)=difRMK_8_1_1(1,d1RMK_8_1_1+3)+0.04;
end

if d1RMK_7_2_1<=-2
    difRMK_7_2_1(1,1)=difRMK_7_2_1(1,1)+0.04;
elseif d1RMK_7_2_1>=2
    difRMK_7_2_1(1,5)=difRMK_7_2_1(1,5)+0.04;
else
    difRMK_7_2_1(1,d1RMK_7_2_1+3)=difRMK_7_2_1(1,d1RMK_7_2_1+3)+0.04;
end

if d1RMK_6_3_1<=-2
    difRMK_6_3_1(1,1)=difRMK_6_3_1(1,1)+0.04;
elseif d1RMK_6_3_1>=2
    difRMK_6_3_1(1,5)=difRMK_6_3_1(1,5)+0.04;
else
    difRMK_6_3_1(1,d1RMK_6_3_1+3)=difRMK_6_3_1(1,d1RMK_6_3_1+3)+0.04;
end

if d1RMK_5_4_1<=-2
    difRMK_5_4_1(1,1)=difRMK_5_4_1(1,1)+0.04;
elseif d1RMK_5_4_1>=2
    difRMK_5_4_1(1,5)=difRMK_5_4_1(1,5)+0.04;
else
    difRMK_5_4_1(1,d1RMK_5_4_1+3)=difRMK_5_4_1(1,d1RMK_5_4_1+3)+0.04;
end

if d1RMK_6_2_2<=-2
    difRMK_6_2_2(1,1)=difRMK_6_2_2(1,1)+0.04;
elseif d1RMK_6_2_2>=2
    difRMK_6_2_2(1,5)=difRMK_6_2_2(1,5)+0.04;
else
    difRMK_6_2_2(1,d1RMK_6_2_2+3)=difRMK_6_2_2(1,d1RMK_6_2_2+3)+0.04;
end

if d1RMK_5_3_2<=-2
    difRMK_5_3_2(1,1)=difRMK_5_3_2(1,1)+0.04;
elseif d1RMK_5_3_2>=2
    difRMK_5_3_2(1,5)=difRMK_5_3_2(1,5)+0.04;
else
    difRMK_5_3_2(1,d1RMK_5_3_2+3)=difRMK_5_3_2(1,d1RMK_5_3_2+3)+0.04;
end

if d1RMK_4_4_2<=-2
    difRMK_4_4_2(1,1)=difRMK_4_4_2(1,1)+0.04;
elseif d1RMK_4_4_2>=2

```

```

        difRMK_4_4_2(1,5)=difRMK_4_4_2(1,5)+0.04;
else
    difRMK_4_4_2(1,d1RMK_4_4_2+3)=difRMK_4_4_2(1,d1RMK_4_4_2+3)+0.04;
end

if d1RMK_4_3_3<=-2
    difRMK_4_3_3(1,1)=difRMK_4_3_3(1,1)+0.04;
elseif d1RMK_4_3_3>=2
    difRMK_4_3_3(1,5)=difRMK_4_3_3(1,5)+0.04;
else
    difRMK_4_3_3(1,d1RMK_4_3_3+3)=difRMK_4_3_3(1,d1RMK_4_3_3+3)+0.04;
end

if d1RMK_1_0_0<=-2
    difRMK_1_0_0(1,1)=difRMK_1_0_0(1,1)+0.04;
elseif d1RMK_1_0_0>=2
    difRMK_1_0_0(1,5)=difRMK_1_0_0(1,5)+0.04;
else
    difRMK_1_0_0(1,d1RMK_1_0_0+3)=difRMK_1_0_0(1,d1RMK_1_0_0+3)+0.04;
end

if d1RMK_9_1_0<=-2
    difRMK_9_1_0(1,1)=difRMK_9_1_0(1,1)+0.04;
elseif d1RMK_9_1_0>=2
    difRMK_9_1_0(1,5)=difRMK_9_1_0(1,5)+0.04;
else
    difRMK_9_1_0(1,d1RMK_9_1_0+3)=difRMK_9_1_0(1,d1RMK_9_1_0+3)+0.04;
end

if d1RMK_8_2_0<=-2
    difRMK_8_2_0(1,1)=difRMK_8_2_0(1,1)+0.04;
elseif d1RMK_8_2_0>=2
    difRMK_8_2_0(1,5)=difRMK_8_2_0(1,5)+0.04;
else
    difRMK_8_2_0(1,d1RMK_8_2_0+3)=difRMK_8_2_0(1,d1RMK_8_2_0+3)+0.04;
end

if d1RMK_7_3_0<=-2
    difRMK_7_3_0(1,1)=difRMK_7_3_0(1,1)+0.04;
elseif d1RMK_7_3_0>=2
    difRMK_7_3_0(1,5)=difRMK_7_3_0(1,5)+0.04;
else
    difRMK_7_3_0(1,d1RMK_7_3_0+3)=difRMK_7_3_0(1,d1RMK_7_3_0+3)+0.04;
end

if d1RMK_6_4_0<=-2
    difRMK_6_4_0(1,1)=difRMK_6_4_0(1,1)+0.04;
elseif d1RMK_6_4_0>=2
    difRMK_6_4_0(1,5)=difRMK_6_4_0(1,5)+0.04;
else
    difRMK_6_4_0(1,d1RMK_6_4_0+3)=difRMK_6_4_0(1,d1RMK_6_4_0+3)+0.04;
end

if d1RMK_5_5_0<=-2
    difRMK_5_5_0(1,1)=difRMK_5_5_0(1,1)+0.04;
elseif d1RMK_5_5_0>=2
    difRMK_5_5_0(1,5)=difRMK_5_5_0(1,5)+0.04;
else
    difRMK_5_5_0(1,d1RMK_5_5_0+3)=difRMK_5_5_0(1,d1RMK_5_5_0+3)+0.04;
end

if d1RMK_4_6_0<=-2
    difRMK_4_6_0(1,1)=difRMK_4_6_0(1,1)+0.04;

```

```

elseif d1RMK_4_6_0>=2
    difRMK_4_6_0(1,5)=difRMK_4_6_0(1,5)+0.04;
else
    difRMK_4_6_0(1,d1RMK_4_6_0+3)=difRMK_4_6_0(1,d1RMK_4_6_0+3)+0.04;
end

if d1RMK_3_7_0<=-2
    difRMK_3_7_0(1,1)=difRMK_3_7_0(1,1)+0.04;
elseif d1RMK_3_7_0>=2
    difRMK_3_7_0(1,5)=difRMK_3_7_0(1,5)+0.04;
else
    difRMK_3_7_0(1,d1RMK_3_7_0+3)=difRMK_3_7_0(1,d1RMK_3_7_0+3)+0.04;
end

if d1RMK_0_1_0<=-2
    difRMK_0_1_0(1,1)=difRMK_0_1_0(1,1)+0.04;
elseif d1RMK_0_1_0>=2
    difRMK_0_1_0(1,5)=difRMK_0_1_0(1,5)+0.04;
else
    difRMK_0_1_0(1,d1RMK_0_1_0+3)=difRMK_0_1_0(1,d1RMK_0_1_0+3)+0.04;
end

if d1RMK_0_0_1<=-2
    difRMK_0_0_1(1,1)=difRMK_0_0_1(1,1)+0.04;
elseif d1RMK_0_0_1>=2
    difRMK_0_0_1(1,5)=difRMK_0_0_1(1,5)+0.04;
else
    difRMK_0_0_1(1,d1RMK_0_0_1+3)=difRMK_0_0_1(1,d1RMK_0_0_1+3)+0.04;
end

if d1Chisq<=-2
    difChisq(1,1)=difChisq(1,1)+0.04;
elseif d1Chisq>=2
    difChisq(1,5)=difChisq(1,5)+0.04;
else
    difChisq(1,d1Chisq+3)=difChisq(1,d1Chisq+3)+0.04;
end

if d1AIC<=-2
    difAIC(1,1)=difAIC(1,1)+0.04;
elseif d1AIC>=2
    difAIC(1,5)=difAIC(1,5)+0.04;
else
    difAIC(1,d1AIC+3)=difAIC(1,d1AIC+3)+0.04;
end

if d1BIC<=-2
    difBIC(1,1)=difBIC(1,1)+0.04;
elseif d1BIC>=2
    difBIC(1,5)=difBIC(1,5)+0.04;
else
    difBIC(1,d1BIC+3)=difBIC(1,d1BIC+3)+0.04;
end

if d1MAP2<=-2
    difMAP2(1,1)=difMAP2(1,1)+0.04;
elseif d1MAP2>=2
    difMAP2(1,5)=difMAP2(1,5)+0.04;
else
    difMAP2(1,d1MAP2+3)=difMAP2(1,d1MAP2+3)+0.04;
end

```

```

if d1MAP4<=-2
    difMAP4(1,1)=difMAP4(1,1)+0.04;
elseif d1MAP4>=2
    difMAP4(1,5)=difMAP4(1,5)+0.04;
else
    difMAP4(1,d1MAP4+3)=difMAP4(1,d1MAP4+3)+0.04;
end

if d1PA<=-2
    difPA(1,1)=difPA(1,1)+0.04;
elseif d1PA>=2
    difPA(1,5)=difPA(1,5)+0.04;
else
    difPA(1,d1PA+3)=difPA(1,d1PA+3)+0.04;
end

if d1SEScree<=-2
    difSEScree(1,1)=difSEScree(1,1)+0.04;
elseif d1SEScree>=2
    difSEScree(1,5)=difSEScree(1,5)+0.04;
else
    difSEScree(1,d1SEScree+3)=difSEScree(1,d1SEScree+3)+0.04;
end

if d1EV<=-2
    difEV(1,1)=difEV(1,1)+0.04;
elseif d1EV>=2
    difEV(1,5)=difEV(1,5)+0.04;
else
    difEV(1,d1EV+3)=difEV(1,d1EV+3)+0.04;
end
end

res=[data(1,1:4) difRMK_8_1_1 difRMK_7_2_1 difRMK_6_3_1 difRMK_5_4_1
      difRMK_6_2_2 difRMK_5_3_2 difRMK_4_4_2 difRMK_4_3_3
      difRMK_1_0_0 difRMK_9_1_0 difRMK_8_2_0 difRMK_7_3_0
      difRMK_6_4_0 difRMK_5_5_0 difRMK_4_6_0 difRMK_3_7_0
      difRMK_0_1_0 difRMK_0_0_1 difChisq    difAIC
      difBIC      difMAP2      difMAP4      difPA
      difSEScree  difEV];
end

```

Appendix B

Mean Percentage of Each RMK-OE Weight Combination in Each Estimation Error

Category across All Simulation Conditions

Original Order	Weights			Percentages in Each Estimation Error				
	Median	IQR	Trimmed Range	Category				
				≤ -2	-1	0	1	≥ 2
1	0.8	0.1	0.1	0.29	0.41	91.50	5.33	2.47
2	0.7	0.2	0.1	0.29	0.39	91.43	5.35	2.54
3	0.6	0.3	0.1	0.29	0.37	91.35	5.40	2.58
4	0.5	0.4	0.1	0.29	0.37	91.19	5.55	2.60
5	0.6	0.2	0.2	0.29	0.37	91.35	5.42	2.56
6	0.5	0.3	0.2	0.29	0.37	91.21	5.55	2.58
7	0.4	0.4	0.2	0.29	0.37	90.29	5.70	3.35
8	0.4	0.3	0.3	0.29	0.37	90.29	5.71	3.34
9	1.0	0.0	0.0	0.48	0.41	85.99	7.52	5.60
10	0.9	0.1	0.0	0.31	0.39	89.53	6.29	3.48
11	0.8	0.2	0.0	0.29	0.37	90.84	5.71	2.78
12	0.7	0.3	0.0	0.29	0.37	91.26	5.53	2.54
13	0.6	0.4	0.0	0.29	0.37	91.32	5.42	2.60
14	0.5	0.5	0.0	0.29	0.37	91.10	5.57	2.67

References

- Afifi, A., Clark, V. A., & May, S. (2004). *Computer-aided multivariate analysis* (4th ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317-332.
- Allen, S. J., & Hubbard, R. (1986). Notes and commentary: Regression equations for the latent roots of random data correlation matrices with unities on the diagonal. *Multivariate Behavioral Research*, 21, 393-398.
- Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neyman (Ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 5*, (pp. 111-150). Berkeley, CA: University of California Press.
- Bartholomew, D. J., & Knott, M. (1999). *Latent variable models and factor analysis* (2nd ed.). New York: Oxford University Press.
- Bartholomew, D. J., Steele, F., Moustaki, I., & Galbraith, J. I. (2008). *Analysis of multivariate social science data* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *The British Journal of Psychology, Statistical Section*, 3, 77-85.
- Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *The British Journal of Psychology, Statistical Section*, 4, 1-2.
- Basilevsky, A. (1994). *Statistical factor analysis and related methods: Theory and applications*. New York: John Wiley & Sons.
- Beauducel, A. (2007). In spite of indeterminacy many common factor score estimates yield an identical reproduced covariance matrix. *Psychometrika*, 72(3), 437-441.

- Bentler, P. M., & Yuan, K-H. (1998). Tests for linear trend in the smallest eigenvalues of the correlation matrix. *Psychometrika*, 63(2), 131-144.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-359.
- Bock, R. D., Gibbons, R., & Muraki, E. (1988). Full-information item factor analysis. *Applied Psychological Measurement*, 12(3), 261-280.
- Bretschler, O. (2001). *Linear algebra with applications* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1(2), 245-276.
- Cattell, R. B., & Jaspers, J. (1967). A general plasmode (No. 30-10-5-2) for factor analytic exercises and research. *Multivariate Behavioral Research Monographs*, No. 67-3.
- Cattell, R. B., & Vogelmann, S. (1977). A comprehensive trial of the scree and KG criteria for determining the number of factors. *Multivariate Behavioral Research*, 12(3), 289-325.
- Cliff, N. (1988). The eigenvalues-greater-than-1 rule and the reliability of components. *Psychological Bulletin*, 103(2), 276-279.
- Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology*, 46(4), 648-659.

- Coover, M. D., & McNelis, K. (1988). Determining the number of common factors in factor analysis: A review and program. *Educational and Psychological Measurement, 48*, 687-692.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7). Available online: <http://pareonline.net/getvn.asp?v=10&n=7>
- Cota, A. A., Longman, R. S., Holden, R. R., & Fekken, G. C. (1991). Anomalies in the Allen and Hubbard parallel analysis procedure. *Applied Psychological Measurement, 15*(1), 95-97.
- Cota, A. A., Longman, R. S., Holden, R. R., Fekken, G. C., & Xinaris, S. (1993). Interpolating 95th percentile eigenvalues from random data: An empirical example. *Educational and Psychological Measurement, 53*, 585-596.
- Crawford, C. B., & Koopman, P. (1973). A note on Horn's test for the number of factors in factor analysis. *Multivariate Behavioral Research, 8*(1), 117-125.
- Doran, H. C. (2003). Adding value to accountability. *Educational Leadership, 16*(3), 55-59.
- Elffers, H., Bethlehem, J., & Gill, R. (1978). Indeterminacy problems and the interpretation of factor analysis results. *Statistica Neerlandica, 32*, 181-199.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.
- Fava, J. L., & Velicer, W. F. (1992). The effects of overextraction on factor and component analysis. *Multivariate Behavioral Research, 27*(3), 387-415.
- Fava, J. L., & Velicer, W. F. (1996). The effects of underextraction in factor and component analysis. *Educational and Psychological Measurement, 56*(6), 907-929.
- Ferré, L. (1995). Selection of components in principal component analysis: A comparison of methods. *Computational Statistics & Data Analysis, 19*, 669-682.
- Finch, H., & Habing, B. (2005). Comparison of NOHARM and DETECT in item cluster recovery: Counting dimensions and allocating items. *Journal of Educational Measurement, 42*(2), 149-169.
- Gessaroli, M. E., & De Champlain A. F. (1996). Using an approximate Chi-square statistic to test the number of dimensions underlying the responses to a set of items. *Journal of Educational Measurement, 33*(2), 157-179.
- Gierl, M. J., Leighton, J. P., & Tan, X. (2006). Evaluating DETECT classification accuracy and consistency when data display complex structure. *Journal of Educational Measurement, 43*(3), 265-289.
- Glorfeld, L. W. (1995). An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement, 55*(3), 377-393.
- Gorsuch, R. L. (1973). Using Bartlett's significance test to determine the number of factors to extract. *Educational and Psychological Measurement, 33*, 361-364.

- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149-161.
- Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hakstian, A. R., & Mueller, V. J. (1973). Some notes on the number of factors problem. *Multivariate Behavioral Research*, *8*(4), 461-475.
- Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, *17*(2), 193-219.
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, *9*(2), 139-164.
- Hayashi, K., Bentler, P. M., & Yuan, K.-H. (2007). On the likelihood ratio test for the number of factors in exploratory factor analysis. *Structural Equation Modeling*, *14*(3), 505-526.
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods*, *7*(2), 191-205.
- Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *The British Journal of Statistical Psychology*, *17*, 65-70.

- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416.
- Hong, S., Mitchell, S. K., & Harshman, R. A. (2006). Bootstrap scree tests: A Monte Carlo simulation and applications to published data. *The British Psychological Society, 59*, 35-57.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.
- Horn, J. L., & Engstrom, R. L. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research, 14*(3), 283-300.
- Hubbard, R., & Allen, S. J. (1987). An empirical comparison of alternative methods for principal component extraction. *Journal of Business Research, 15*, 173-190.
- Humphreys, L. G., & Ilgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement, 29*, 571-578.
- Hurley, J. R., & Cattell, R. B. (1962). The Procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science, 7*(2), 258-262.
- Ichikawa, M. (1988). Empirical assessments of AIC procedure for model selection in factor analysis. *Behaviormetrika, 24*, 33-40.
- Jennrich, R. I. (2007). Rotation methods, algorithms, and standard errors. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp.315-335). Mahwah, NJ: Lawrence Erlbaum Associates.

- Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4), 443-482.
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, 43(4), 443-477.
- Jöreskog, K. G. (2007). Factor analysis and its extensions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp.47-77). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187-200.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141-151.
- Keeling, K. B. (2000). A regression equation for determining the dimensionality of data. *Multivariate Behavioral Research*, 35(4), 457-468.
- Keeling, K. B., & Pavur, R. J. (2004). A comparison of methods for approximating the mean eigenvalues of a random matrix. *Communications in Statistics: Simulation and Computation*, 33(4), 945-961.
- Kim, J. S., & Sunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34(8), 3-13.
- Kim, J.-O., & Mueller, C. W. (1978). *Factor analysis: Statistical methods and practical issues*. Beverly Hills, CA: Sage Publications.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). New York: Springer.

- Koretz, D. M., & Hamilton, L. S. (2006). Testing for Accountability in K-12. In R. L. Brennan (Ed), *Educational Measurement* (4th ed., pp.531-578). Westport, CT: American Council on Education and Praeger.
- Kuha, J. (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods Research*, 33(2), 188-229.
- Lambert, Z. V., Wildt, A. R., & Durand, R. M. (1990). Assessing sampling variation relative to number-of-factors criteria. *Educational and Psychological Measurement*, 50(1), 33-48.
- Lautenschlager, G. J. (1989). A comparison of alternatives to conducting Monte Carlo analyses for determining parallel analysis criteria. *Multivariate Behavioral Research*, 24(3), 365-395.
- Lautenschlager, G. J., Lance, C. E., & Flaherty, V. L. (1989). Parallel analysis criteria: Revised equations for estimating the latent roots of random data correlation matrices. *Educational and Psychological Measurement*, 49, 339-345.
- Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh, Section A*, 60, 64-82.
- Lawley, D.N. (1941). Further investigations in factor estimation. *Proceedings of the Royal Society of Edinburgh, Section A*, 61, 176-185.
- Lee, H. B., & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research*, 14(3), 301-321.
- Linn, R. L. (1968). A Monte Carlo approach to the number of factors problem. *Psychometrika*, 33(1), 37-71.

- Linn, R. L. (2004). *Rethinking of the No Child Left Behind accountability system*. Paper prepared for a forum on No Child Left Behind sponsored by the Center on Education Policy, Washington, DC.
- Loehlin, J. C. (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis* (4th ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Longman, R. S., Cota, A. A., Holden, R. R., & Fekken, G. C. (1989). A regression equation for the parallel analysis criterion in principal components analysis: Mean and 95th percentile eigenvalues. *Multivariate Behavioral Research*, 24(1), 59-69.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Maraun, M. D. (1996). Metaphor taken as math: Indeterminacy in the factor analysis model. *Multivariate Behavioral Research*, 31(4), 517-538.
- Markon, K. E., & Krueger, R. F. (2004). An empirical comparison of information-theoretic selection criteria for multivariate behavior generic models. *Behavior Genetics*, 34(6), 593-610.
- Martineau, J. A. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics*, 31(1), 35-62.
- Martineau, J. A., & Reckase, M. D. (2006, June). *Practical multidimensional vertical equating based on accurate assessment of dimensionality*. Paper presented at the 36th Annual National Conference on Large-Scale Assessment of the Council of Chief State School Officers (CCSSO), San Francisco, CA.
- McDonald, R. P. (1967). *Non-linear factor analysis*. Psychometric Monographs, No. 15.

- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- McDonald, R. P. (1997). Normal-ogive multidimensional model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 257-269). New York: Springer-Verlag.
- McDonald, R. P. (1999). *Test theory: A unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McDonald, R. P. (2000). A Basis for Multidimensional Item Response Theory. *Applied Psychological Measurement*, 24(2), 99-114.
- Montanelli, R. G., Jr., & Humphreys, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika*, 41(3), 341-348.
- Muthén, B. (1978). Contributions to factor analysis of dichotomous variables. *Psychometrika*, 43(4), 551-560.
- Muthén, L. K., & Muthén, B. O. (1998-2007). *Mplus user's guide*. Fifth Edition. Los Angeles, CA: Muthén & Muthén.
- Nasser, F., Benson, J., & Wisenbaker, J. (2002). The performance of regression-based variations of the visual scree for determining the number of common factors. *Educational and Psychological Measurement*, 62(3), 397-419.
- Netemeyer, R.G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage Publications.

- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32(3), 396-402.
- Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika*, 65(2), 167-185.
- Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2004). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics & Data Analysis*, 49, 974-997.
- Raudenbush, S. W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Princeton, NJ: Educational Testing Service.
- Reckase, M. D. (1997). A linear logistic multidimensional model for dichotomous item response data. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 271-286). New York: Springer-Verlag.
- Reckase, M. D. (2004). The real world is more complicated than we would like. *Journal of Educational and Behavioral Statistics*, 29(1), 117-120.
- Reckase, M. D. (2007). Multidimensional item response theory. In C.R. Rao & S. Sinharay (Eds.), *Handbook of statistics 26: Psychometrics* (pp.607-642). The Netherlands: Elsevier B. V.
- Reckase, M. D., Martineau, J. A., & Kim, J.-P. (2000, July). *A vector approach to determining the number of dimensions needed to represent a set of variables*. Paper presented at the annual meeting of the Psychometric Society, Vancouver, Canada.
- Reyment, R., & Jöreskog, K. G. (1993). *Applied factor analysis in the natural sciences*. NY: Cambridge University Press.

- Roussos, L. A. (1995). *A new dimensionality estimation tool for multiple-item tests and a new DIF analysis paradigm based on multidimensionality and construct validity*. Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Illinois.
- Roussos, L. A., Stout, W. F., Marden, J. I. (1998). Using new proximity measures with hierarchical cluster analysis to detect multidimensionality. *Journal of Educational Measurement, 35*(1), 1-30.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.
- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika, 52*(3), 333-343.
- Sharma, S. (1996). *Applied multivariate techniques*. New York: John Wiley & Sons.
- Sijtsma, K., & Meijer, R.R. (2007). Nonparametric item response theory and special topics. In C.R. Rao & S. Sinharay (Eds.), *Handbook of statistics 26: Psychometrics* (pp. 719-746). The Netherlands: Elsevier B. V.
- Song, J., & Belin, T. R. (2008). Choosing an appropriate number of factors in factor analysis with incomplete data. *Computational Statistics and Data Analysis, 52*, 3560-3569.
- Steger, M. F. (2006). An illustration of issues in factor extraction and identification of dimensionality in psychological assessment data. *Journal of Personality Assessment, 86*(3), 263-272.

- Steiger, J. H. , & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye (Ed.), *Theory construction and data analysis in the behavioral sciences*. San Francisco: Jossey-Bass.
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55(2), 293-325.
- Stout, W., Habing, B., Douglas, J. Kim H. R., Roussos, L., & Zhang, J. (1996). Conditional covariance-based nonparametric multidimensionality assessment. *Applied Psychological Measurement*, 20(4), 331-354.
- Swaminathan, H., Hambleton, R. K., & Rogers, H. J. (2007). Assessing the fit of item response theory models. In C.R. Rao & S. Sinharay (Eds.), *Handbook of statistics 26: Psychometrics* (pp. 683-718). The Netherlands: Elsevier B. V.
- Tate, R. (2002). Test dimensionality. In G. Tindal & T.M. Haladyna (Eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation* (p. 181-211). Mahwah, NJ: Erlbaum.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC: American Psychological Association.
- Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement*, 56(2), 197-208.
- Thurstone, L. L. (1947). *Multiple-factor analysis*. Chicago: The University of Chicago Press.
- Timm, N. H. (2002). *Applied multivariate analysis*. New York: Springer-Verlag.

- Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 34(4), 421-459.
- Turner, N. E. (1998). The effect of common variance and structure pattern on random data eigenvalues: Implications for the accuracy of parallel analysis. *Educational and Psychological Measurement*, 58(4), 541-568.
- van Abswoude, A. A., van der Ark, L. A., & Sijtsma, K. (2004). A comparative study of test data dimensionality assessment procedures under nonparametric IRT. *Applied Psychological Measurement*, 28(1), 3-24.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321-327.
- Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research*, 25(1), 1-28.
- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (pp. 41-71). Norwell, MA: Kluwer Academic Publishers.
- Wendler, C. L.W., & Walker, M. E. (2006). Practical issues in designing and maintaining multiple test forms for large-scale programs. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 445-467). Mahwah, NJ: Lawrence Erlbaum Associates.

- Widaman, K. F. (2007). Common factors versus components: Principals and principles, errors and misconceptions. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp.177-203). Mahwah, NJ: Lawrence Erlbaum Associates.
- Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods, 1*(4), 354-365.
- Woolfolk, A. (2001). *Educational psychology* (8th ed.). Needham Heights, MA: Allyn and Bacon.
- Yanai, H., & Ichikawa, M. (2007). Factor analysis. In C.R. Rao & S. Sinharay (Eds.), *Handbook of statistics 26: Psychometrics* (pp.257-296). The Netherlands: Elsevier B. V.
- Yang, J. (2008). *Statistical methods for estimating the dimension of multivariate data*. Unpublished doctoral dissertation, The George Washington University, Washington DC.
- Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 111-153). Westport, CT: American Council on Education and Praeger Publishers.
- Zeng, J., & Martineau, J. A. (2008, March). *Recovering dimensionality with dimensions correlated: A comparison of a new vector approach to determining dimensionality with established exploratory factor analysis procedures*. Paper presented at the annual conference of the American Educational Research Association, New York, NY.

- Zhang, J., & Stout, W. (1999). The theoretical DETECT index of dimensionality and its application to approximate simple structure. *Psychometrika*, 64(2), 213-249.
- Zoski, K., & Jurs, S. (1990). Priority determination in surveys: An application of the scree test. *Evaluation Review*, 14(2), 214-219.
- Zoski, K., & Jurs, S. (1993). Using multiple regression to determine the number of factors to retain in factor analysis. *Multiple Linear Regression Viewpoints*, 20, 5-9.
- Zoski, K., & Jurs, S. (1996). An objective counterpart to the visual scree test for factor analysis: The standard error scree. *Educational and Psychological Measurement*, 56(3), 443-451.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99(3), 432-442.