

# **Documenting the Impact of Outliers on Decisions about the Number of Factors in Exploratory Factor Analysis**

by

Yan Liu

B.A., Beijing Second Foreign Language University, 2000

M.A., The University of British Columbia, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES  
(Measurement, Evaluation, and Research Methodology)

THE UNIVERSITY OF BRITISH COLUMBIA  
(Vancouver)  
December 2011

© Yan Liu, 2011

## **Abstract**

The overall purpose of this dissertation is to investigate how outliers affect the decisions about the number of factors in exploratory factor analysis (EFA) as determined by four widely used and/or highly recommended methods. Very few studies have looked into this issue in the literature and the conclusions are contradictory— i.e., with studies disagreeing as to whether outliers result in extra factors or a reduced number of factors. For this dissertation I systematically studied the impact of outliers arising from different sources and matched outlier simulation models with different type of outliers.

Chapter 1 provides an overview of the gap between statistical theory regarding outliers and researchers' day-to-day practice and their understanding of the effects of outliers. Chapter 2 presents a review of EFA with an emphasis on the four commonly used or highly recommended decision methods on the number of factors as well as a review of outliers which includes the sources of outliers and problems of outliers in factor analysis. Chapter 3 examines the effects of outliers arising from errors using the deterministic and slippage models. The results revealed that outliers can inflate, deflate, or have no effects on the decisions about the number of factors, which depends on the decision method used and the magnitude and number of outliers. Chapter 4 investigates the effects of outliers arising from an unintended and unknowingly included subpopulation using the mixture contamination model. The general conclusions are similar to chapter 3, but chapter 4 also reveals that symmetric and asymmetric contamination has different effects on different decision methods

and the effects of outliers do not depend on sample size. Chapter 5 provides a general discussion of the findings of this dissertation, describes four novel contributions, and points out the limitations of the present research as well as the future research directions.

This dissertation aims to bridge the gap from day-to-day researchers' practice and understanding of the effects of outliers to current outlier research that emphasizes robust statistics. The findings of this dissertation address the contradictory conclusions made in previous studies.

## Preface

Chapters 3 and 4 can be considered as stand-alone papers. A version of chapters 3 and 4 have been accepted for publication. The citation for the first publication is: Liu, Y., Zumbo, B. D., & Wu, D. W. (in press). A demonstration of the impact of outliers on the decisions about the number of factors in exploratory factor analysis. *Educational and Psychological Measurement*. The citation for the second publication is: Liu, Y., & Zumbo, B. D. (in press). Impact of outliers arising from unintended and unknowingly included subpopulations on the decisions about the number of factors in exploratory factor analysis. *Educational and Psychological Measurement*. My contribution was in the formulation of the research questions, the design of the studies, the computer simulation, the data analyses and interpretation of the results, and the manuscript preparation. Dr. Bruno D. Zumbo contributed to the formulation of the research questions, the design of the studies, interpretation of the results and the manuscript preparation of both chapters 3 and 4. Dr. Amery D. Wu contributed to the formulation of the research questions and the design of the studies of chapter 3.

## Table of Contents

<b>Abstract.....</b>	<b>ii</b>
<b>Preface.....</b>	<b>iv</b>
<b>Table of Contents .....</b>	<b>v</b>
<b>List of Tables.....</b>	<b>viii</b>
<b>List of Figures.....</b>	<b>x</b>
<b>Acknowledgments .....</b>	<b>xii</b>
<b>Chapter 1: Introduction .....</b>	<b>1</b>
<b>Chapter 2: Review of Exploratory Factor Analysis and Outliers .....</b>	<b>9</b>
2.1 Review of Exploratory Factor Analysis (EFA) .....	9
2.1.1 Introduction to EFA.....	9
2.1.2 Four Methods for Determining the Number of Factors in EFA .....	11
2.1.2.1 K-G Rule .....	14
2.1.2.2 Parallel Analysis (PA) .....	15
2.1.2.3 Minimum Average Partial (MAP).....	17
2.1.2.4 Sequential Chi-square Tests .....	19
2.1.3 A Demonstration of the Four Decision Methods.....	21
2.2 Review of Outliers .....	25
2.2.1 Sources of Outliers .....	25
2.2.2 Problems of Outliers in Factor Analysis Models.....	27

2.3	Study Purpose .....	30
<b>Chapter 3: Effects of Outliers Arising from Errors.....</b>		<b>32</b>
3.1	Introduction.....	32
3.2	Study 1: Demonstration of the Impact of Outliers with a Classic Data Set.....	37
3.2.1	Method.....	37
3.2.2	Results .....	41
3.3	Study-2: A Monte Carlo Simulation Documenting the Effects of Outliers .....	45
3.3.1	Methods .....	45
3.3.2	Results .....	47
3.4	Study 3: A Focused Simulation Study to Demonstrate the Effects of Outliers on the Correlation Matrix .....	54
3.5	Discussion .....	59
<b>Chapter 4: Effects of Outliers Arising from an Unintended and Unknowingly Included Subpopulation .....</b>		<b>62</b>
4.1	Introduction.....	62
4.1.1	Sources of Outlier Contamination .....	63
4.1.2	Models Used in Simulation Studies .....	65
4.2	Study 1: Investigating the Effects of Outliers from a Subpopulation Using the Mixture Contamination Model .....	73
4.2.1	Method.....	73
4.2.2	Results .....	77

4.2.3	Demonstration 1 .....	95
4.2.4	Demonstration-2 .....	100
4.3	Discussion .....	102
<b>Chapter 5:</b>	<b>General Discussion .....</b>	<b>105</b>
5.1	Two Types of Outliers and Distinction of Intended and Unintended Subpopulations 105	
5.2	Review of Outlier Simulation & Major Findings .....	109
5.2.1	Outlier Simulation .....	109
5.2.2	Major Findings .....	111
5.3	Implications for Day-to-Day Researchers .....	113
5.4	Novel Contributions.....	118
5.5	Limitations and Suggestions for Future Research .....	121
<b>References</b> .....		<b>124</b>
<b>Appendix</b> .....		<b>142</b>

## List of Tables

Table 2.1	A Demonstration of Four Decision Methods .....	23
Table 3.1	Change in the Number of Factors from Four Approaches Based on the Original Data .....	43
Table 3.2	Change in the Number of Factors Based on the Sequential Chi-square (ML) Test .....	49
Table 3.3	Variable Ordering for a Three-way ANOVA on the Number of Factors Extracted by the K-G Rule .....	50
Table 3.4	Variable Ordering for a Three-way ANOVA on the Number of Factors Extracted by the MAP Approach.....	51
Table 3.5	Variable Ordering for a Three-way ANOVA on the Number of Factors Extracted by the PA Approach.....	53
Table 3.6	Demonstration of Changes in Correlation Coefficients and Eigenvalues of a 4-Variable data in the Presence of Outliers Comparing to the No-Outlier Condition .....	56
Table 4.1	Documenting Simulations Using Mixture of Distributions: Proportion of Outliers in Each Sample across 100 Replications .....	78
Table 4.2	Variable Ordering for a Five-way ANOVA on the Number of Factors Extracted by the K-G Rule .....	82
Table 4.3	Variable Ordering for a Five-way ANOVA on the Number of Factors Extracted by the MAP Approach.....	85



Table 4.4 Variable Ordering for a Five-way ANOVA on the Number of Factors Extracted by the PA Approach.....	88
Table 4.5 Variable Ordering for a Five-way ANOVA on the Number of Factors Decided by the Sequential Chi-square (ML) test .....	91
Table 4.6 Percentage of Non-convergent Replications with the Sequential Chi-square (ML) Tests .....	94
Table 4.7 Demonstration of Changes in Correlation Coefficients, Condition Number and Eigenvalues Using a Four-Variable Data Set in the Presence of Outliers (15%) Compared to the No-Outlier Condition .....	99
Table 4.8 Demonstration of Effects of Outliers on Kurtosis and Skewness .....	101
Table 5.1 Review of Outlier Related Issues in a Sample of Structural Equation Modeling or Factor Analysis Books .....	117

## List of Figures

Figure 2.1 A Demonstration of Parallel Analysis .....	23
Figure 3.1 Graphs for Two-way Interactions of Variables vs. Subjects with outliers and Variables vs. Magnitude of Outliers on the Number of Factors Extracted by the K-G Rule. 50	
Figure 3.2 Graphs for Three-way Interactions of Variables, Subjects with Outliers and the Magnitude of outliers on the Number of Factors Extracted by MAP Approach .....	52
Figure 3.3 Graphs for Main Effect of Variables with Outliers on the Number of Factors Extracted by PA Approach .....	53
Figure 4.1 An Example of Symmetric and Asymmetric Outliers (Proportion of Contamination=0.15) .....	71
Figure 4.2 Graphs for Three-way Interactions of Variables with Outliers vs. Proportion of Contamination by Three Levels of Mean Shift on the Number of Factors Extracted by the K-G Rule .....	84
Figure 4.3 Graphs for Three-way Interactions of Variables with Outliers vs. Proportion of Contamination by Three Levels of Mean Shift on the Number of Factors Extracted by MAP Approach.....	87
Figure 4.4 Three-way Interactions of Variables with Outliers vs. Proportion of Contamination by Three Levels of Mean Shift on the Number of Factors Extracted by PA Approach .....	90

Figure 4.5 Three-way Interactions of Variables with outliers vs. Proportion of Contamination by Three Levels of Standard Deviation Shift on the Number of Factors Decided by the Sequential $\chi^2_{ML}$ Test .....	93
--	----

## **Acknowledgments**

I would like to express my deepest gratitude to my supervisor, Dr. Bruno D. Zumbo, whose expertise and passion for research, consistent mentorship and tremendous support have all been critical factors in my graduate experience at UBC. This dissertation would not have been possible without his ongoing guidance, support and encouragement. I would like to acknowledge and thank Dr. Amery D. Wu for sharing her expertise and time throughout my graduate study. I would also like to thank the other members of my committee, Dr. Anita Hubley and Dr. Sheila Marshall for the mentorship, support and assistance they have provided.

Thank you to the current members of Edgeworth lab for providing support and special thanks to Benjamin Shear who shared his expertise of computer simulation with me. Thank you to my wonderful friends, both those in Canada and China for keeping me motivated throughout my PhD. Most importantly, I would like to thank my family, my parents Shurong Jiang and Xintian Liu, my brother Baohua Liu and my sister in law Hui Xia for their love and support throughout my years of education.

## **Chapter 1: Introduction**

The concern about outliers dates back to at least 1777, when Bernoulli first discussed the practice of discarding outlying observations (Beckman & Cook, 1983). Since then, the field of statistics was formally created and numerous studies of outliers have been published, promoting various methods to deal with outliers. A substantial review can be found in articles or books by Anscombe (1960), Barnett (1978), Barnett and Lewis, (1978, 1994), Beckman and Cook (1983), Grubbs (1969), and Hawkins (1980). A sub-field of statistics is devoted to robust estimation, testing, and data analysis (including data visualization and graphics). Many researchers and statisticians have made considerable contributions to robust estimation methods, such as Huber (1981), Rousseeuw and Leroy (1987), Wilcox (2001, 2005), and Yuan and Bentler (2007).

A variety of definitions of outliers have been provided across different contexts, but generally we can understand outliers using Grubbs' (1969) definition: "an outlying observation, or outlier, is one that appears to deviate markedly from the other members of the sample in which it occurs" (p. 1). Although Grubbs' definition is widely cited, it is not particularly practical or useful and leaves too many ambiguities, such as, "what is the criterion for this?" It is also noteworthy that Grubbs' definition is sample-based and hence a model or population, per se, is not involved in the definition. A single standard definition of outliers cannot be found in the literature because outliers are always relative to some reference such as the sample data or a model that the researcher assumes to be correct.

Therefore, using Grubbs' terminology, a broad definition may be that outliers are data points that deviate markedly from: (i) the sample in which they occur, or (ii) the assumed model, be it a probability distribution or a statistical model such as the least squares regression model. This broad definition allows one to operationalize various strategies for investigating outliers by defining "markedly deviate" and "the model." Furthermore, this broad definition reminds us that an outlier is defined relative to a model or assumed distribution. For example, when studying income, a data point may be regarded as an outlier if income is believed to be normally distributed, but not an outlier if income is believed to be distributed as a skewed random variable such as a member of the exponential, gamma, or chi-square families of distributions (Dragulescu & Yakovenko, 2001). Another commonly seen example is that an observation may not be an outlier in terms of a univariate distribution, but becomes an outlier when put into a multiple regression model. For example, when regressing height onto age, a score of six foot tall is not an outlier but a six foot tall together with 10 year old may well be an outlier.

In the nineteenth century, researchers began to specify the criterion for rejecting outliers. Peirce (1852, 1878) was the first person to propose a specific criterion for the rejection of outliers. In 1863, Chauvenet proposed his criterion for rejecting a single outlier (c.f. Beckman & Cook, 1983). In the 20<sup>th</sup> century, a great number of articles on the methods to identify, reject or accommodate outliers have been published. For example, Thompson (1935) used what would now be called studentized residuals for the rejection of outliers.

Mahalanobis' distance, which was first introduced in 1936 when investigating the dispersion

of multivariate data (Mahalanobis, 1936), has been adapted for identifying outliers and is currently widely used and recommended (Tabachnick & Fidell, 1989). This index measures how far a point is from the center of the sample distribution of points, taking into account the covariance among the variables (Berkane & Bentler, 1988; Rousseeuw & Van Zomeren, 1990). In addition, a variety of robust estimators of the mean have been developed by down-weighting observations in order to accommodate outliers into the data, such as, *L*-estimators, *M*-estimators, and *R*-estimators (Barnett & Lewis, 1994).

Given that outliers have been studied for over 200 years, it might be expected that we have developed a good understanding of effects of outliers and statistical techniques to deal with them. However, there seems to be a large gap between the statistical theory of outliers and the day-to-day research and data analysis practice. Hunter and Schmidt (1990) pointed out that statistical artifacts, such as measurement errors, might result in apparent variability in findings across studies of the same phenomena. Orr, Sackett, and Dubois (1991) also pointed out that researchers' different attitudes towards outliers and their practice in the treatment of outliers might be a source of variation of study findings. Motivated by this idea, they conducted a survey of 100 senior authors who published papers using correlation or regression analysis in the *Journal of Applied Psychology* or *Personnel Psychology* from 1984 to 1987, to investigate attitudes towards detecting and handling outliers. Their results revealed that 67% of these authors advocated removing outliers if there were appropriate reasons, whereas 29% were against removing outliers in any situation, and few of them used multivariate outlier detection techniques. Along the same lines, Bates, Holton, and Burnett

(1999) examined regression studies published from 1993 to 1997 in five leading journals in organizational and human resources research: *Academy of Management Journal*, *Human Resource Development Quarterly*, *Journal of Applied Psychology*, *Journal of Vocational Behavior*, and *Personnel Psychology*. Their results showed that 15% of papers reported that authors checked data and assessed the assumptions of regression analysis, while only 3% reported that authors screened for outliers.

Although statisticians have provided a great number of outlier detection methods and robust estimators for dealing with outliers, many day-to-day researchers have not appreciated statisticians' efforts, rarely using these advanced statistical techniques. This may be because researchers have a limited understanding about how or in what way outliers can affect the model parameter estimates and inferences obtained from some common statistical methods. However, assuming that the effects of outliers are well-known, statisticians have primarily focused on developing and testing statistical methods that are robust to (i.e., not affected by) outliers and have paid less attention to the "effects" of outliers on conventional methods – the idea being that these conventional methods should be superseded by the robust methods. Robust methods have not yet overtaken conventional methods in day-to-day practice, leaving published literature based on conventional methods that may be distorted by outliers; ignoring the effects of outliers on conventional methods does not resolve this issue. Hence it would be helpful to know the effects of outliers when reading and interpreting the extant literature. Clearly, there is a great need to bridge this gap and help day-to-day researchers have a fundamental understanding about the effects of outliers on conventional statistics.



In the literature, a group of researchers have already been making efforts to bridge this gap. Some researchers have demonstrated that even a single outlier can seriously bias descriptive and inferential statistics (Cohen, Cohen, West, & Aiken, 2003; Huber, 1981; Hampel, Ronchetti, Rousseeuw, & Stahel, 1986; Zimmerman & Zumbo, 1993). For instance, Cohen et al. (2003, p. 392) pointed out that, in the presence of outliers, analysis results may reflect a small number of atypical cases rather than a general relationship observed in the data. To illustrate, they showed that in a sample with even one outlier the regression coefficient became not statistically significant and R-square dropped from .43 in the original data (without outliers) to zero in the presence of the outlier. In a correlation study, Devlin, Gnanadesikan, and Kettenring (1981) demonstrated that, using a sample of 29 observations, the correlation coefficient can change from .99 to zero due to a single outlier.

Other researchers pointed out that outliers can result in means shifted further either to the right or left of a distribution, variances can be inflated or deflated, regression coefficients can be biased, estimates of reliability can be inflated, and the inferences made from *t*-tests and *F*-tests may be biased (e.g., Blair & Higgins, 1980; Bradley, 1980; Cook & Weisberg, 1980; Evans, 1999; Lind and Zumbo, 1993; Stevens, 1984; Wilcox, 2005). For example, Zumbo and Jennings (2002) systematically investigated the effects of outliers on one sample *t*-tests using a Monte Carlo simulation, which showed that the type I error rate of *t*-tests was inflated greatly when the proportion of outliers was large. Liu and Zumbo (2007) and Liu, Wu and Zumbo (2010) showed that estimates of Cronbach's coefficient alpha can be severely inflated for continuous as well as ordinal scale item responses, which can be inflated from .40

to .90 in some cases.

However, there has been no systematic study on how outliers affect the decisions on the number of factors in exploratory factor analysis (EFA). EFA is a widely used analytical technique for the purpose of test construction and validation in many fields, such as psychology, education, sociology, marketing, political science, and the social, behavioral and health sciences more generally. Most previous studies focused on identifying and rejecting outliers and developing robust estimators for factor analysis models (e.g., Comrey, 1985; Mavridis & Moustaki, 2008; Pison, Rousseeuw, Filzmoser & Bentler, 2003; Yuan & Bentler, 1998). A few studies have investigated how outliers affect factor analysis, which include EFA and confirmatory factor analysis (CFA), but most of them, using empirical data, only examined effects of outliers on factor loadings, residual variances or chi-square tests (Bollen, 1987; Bollen & Arminger, 1991; Yuan, Marshall, and Bentler, 2002; Yuan & Zhong, 2008). Little is known about how outliers affect the number of factors determined by the commonly used decision methods.

In addition, as Barnett and Lewis (1978, 1994) point out, it has been known that outliers may come from different sources, such as typographical errors, and mistakes in recording data. However, many previous outlier studies used the same mathematical or simulation model to investigate the effects of outliers, neglecting the sources of outliers, which gave researchers the wrong impression that outliers should be treated as errors in their practice no matter what caused the appearance of outliers. Therefore, it is crucial to show day-to-day researchers that outliers can come from different sources and may have different

impacts on the statistical methods currently used by researchers.

The purpose of this dissertation is to provide more information to understand the effects of outliers on decisions about the number of factors in EFA, as determined by four commonly used methods – the Kaiser-Guttman/eigenvalue-greater-than-one rule, parallel analysis, minimum average partial, and  $\chi^2_{ML}$  sequential tests. To systematically investigate the effects of outliers, the present research examined two types of outliers in separate studies (i.e., outliers arising from errors and those from a subpopulation), matched outlier simulation models with the types of outliers, and conducted follow-up studies to provide insight into the potential causes of the findings.

This dissertation is organized as follows. Chapter two provides a brief review of exploratory factor analysis (EFA), which emphasizes four commonly used methods for deciding the number of factors, and a review of outliers, which provides a description of outlier sources and outlier problems in factor analysis, and closes with the study purpose. Chapter 3 examines how outliers arising from errors affect decisions about the number of factors determined by four commonly used methods, with a follow-up study to explain the potential causes. Similar to the purpose of chapter 3, chapter 4 investigates effects of outliers arising from a subpopulation. While chapters 3 and 4 can be considered as stand-alone papers (a version of chapters 3 and 4 are accepted in *Educational and Psychological Measurement*), these chapters also form part of a larger program of my research aimed at providing a systematic investigation of the impact of outliers. The last chapter discusses the results in a more general context, focusing on types of outliers, major findings of the dissertation,

implications for day-to-day researchers, novel contributions, limitations, and future research directions.

## **Chapter 2:      Review of Exploratory Factor Analysis and Outliers**

### **2.1    Review of Exploratory Factor Analysis (EFA)**

#### **2.1.1   Introduction to EFA**

Factor analysis is a statistical method used to describe associations among observed variables with respect to one or more unobserved (latent) variables which are called factors. These factors, such as intelligence, depression, or personality, cannot be measured directly like people's weight or height, but can be generated from a set of observed variables or indicators to account for the covariances among the observed variables. In other words, the generated factor(s) partitions the variance of an observed variable into two parts: *common variance* that is explained by the factor and *residual variance* comprised of random measurement error as well as specific variance not accounted for by the factor(s).

The major purposes of using factor analysis are: (a) identifying and interpreting the underlying constructs, (b) developing operational representatives/observed indicators of the underlying constructs, (c) validating measures/questionnaires using different samples, and (d) providing measurement error free factor scores for subsequent analyses. Since the invention of factor analysis, the first three purposes have been commonly seen in many fields, such as psychology, education, sociology, marketing, political science, and the social sciences, behavioral sciences and health sciences more generally. The last purpose in the list, providing factor scores for subsequent analyses, has recently become a more frequently used strategy

since the development of structural equation modeling (SEM). For example, factor models can be embedded in a path diagram or regression model, hence no longer necessitating factor score estimation. Tests of hypotheses in models with embedded latent variables are generally more statistically powerful because factor scores, unlike the observed variables, are measurement error free.

Generally speaking, there are two classes of factor analysis: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is often used when researchers are investigating the underlying structure, the meaning of factors identified and the relationships between these factors in an exploratory way. In contrast, CFA is used when researchers want to test a model or hypothesis based on substantive theory and/or prior empirical experiences in a confirmatory way. Correspondingly, in order to use CFA researchers must have a clear idea of the number of factors, definition of these factors, the relationships among factors, and also relationships between factors and observed variables. In terms of psychometrics, Gorsuch (1983) pointed out that test development should start with EFA to accumulate researchers' knowledge of the possible factors, the meaning of factors and operational representatives/observed indicators, whereas CFA is used in later stages with different samples to test the quality of scales. Due to the purpose of this dissertation, the present research only focuses on EFA, and more specifically, the four commonly used decision methods to determine the number of factors in EFA studies.

### 2.1.2 Four Methods for Determining the Number of Factors in EFA

EFA investigates the associations among observed variables without any constraints on their structure. Each of  $p$  observed variables ( $Y_j$ ) is modeled as a linear combination of  $m$  factors ( $\eta_m$ ) and a uniqueness component ( $\varepsilon_j$ ). A fundamental equation of the EFA model can be expressed as follows:

$$Y_j = \lambda_{j1}\eta_1 + \lambda_{j2}\eta_2 + \lambda_{j3}\eta_3 + \dots + \lambda_{jm}\eta_m + \varepsilon_j$$

where  $Y_j$  denotes the observed scores for  $j$  items/indicators obtained from a sample of  $N$  independent subjects,  $j = 1, 2, 3, \dots, J$ ;  $\eta_m$  denotes the factor,  $m = 1, 2, 3, \dots, M$ ;

$\lambda_{jm}$  denotes the factor loading relating the  $j$ th indicator to the  $m$ th factor;  $\varepsilon_j$  is the residual variance that is unique to the  $j$ th indicator. The  $\varepsilon_j$  captures both the “uniqueness” and the “random measurement error” components of the observed variables in an undifferentiated manner.

One of the major purposes for using EFA is to identify the factor structure. One or more factors derived in this manner can account for the maximum amount of information among a number of observed variables. The decision about the number of factors in the model is very important as either under- or over-extraction (i.e., too few or too many factors, respectively) can distort the factor solution and lead to misleading interpretations of the resultant factor structure. Under-extraction is a serious problem for EFA because it results in loss of information and introduces substantial errors, and hence the true factor structure cannot be portrayed (Cattell, 1952; Cattell, 1978; Comrey, 1973; Gorsuch, 1983; Thurstone,

1947). Over-extraction is less problematic than under-extraction, but can lead to the “factor splitting” problem, particularly when varimax rotation is used (Comrey, 1973, 1978; Comrey & Lee, 1992). When over-extracting, some trivial factors are created at the price of important factors— the magnitudes of factor loadings of important factors decrease and one or more trivial factors load high on one or more observed variables, which makes the trivial factors artificially important.

Many studies have shown the detrimental effects of both under- and over-extraction. Using real data, Dingman, Miller, and Eyman (1964) found that simple factor structure was obscured when more factors than needed were retained. Levonian and Comery (1966) noted a complete change in factor structure when more factors were retained using real data sets in their study. In two simulation studies, Fava, and Velicer (1992, 1996) found that over-extraction resulted in degradations of factor scores in the cases of low population factor loadings and small sample sizes, and under-extraction led to substantial degradations of factor scores. Also in a simulation study, Wood, Tataryn, and Gorsuch (1996) showed that estimated factor loadings were distorted when over-extraction occurred and the distortion was worse when under-extraction occurred and, as expected, they also found that over-extraction resulted in factor splitting.

In practice, researchers may formulate a construct with an important factor missing when under-extraction occurs or formulate a complicated construct with a lot of noise and nuisance factors when over-extraction occurs. In both situations, the factor structure and relationships between variables and factors can be seriously distorted, and therefore it will



lead to confusion in the later stage when researchers test a theory based on a misspecified factor analysis. In addition, Hoyle and Duvall (2004) suggested that Heywood cases (i.e., negative residual variance) often result from over-extraction (and sometimes under-extraction) when using maximum likelihood (ML) estimation in factor analysis. Thus, the determination of the number of factors to extract is a crucial step in conducting an EFA study.

Over the course of the last 60 years, a variety of methods have been developed to determine the number of factors to retain in the model. The most widely used or highly recommended methods can be categorized into two branches: principal component analysis (PCA) based procedures and Goodness-of-fit statistics. PCA-based procedures, which are eigenvalue related methods, cannot test the adequacy of model-data-fit but rather serve as pointers to the number of factors to retain. The PCA-based procedures include: (i) Kaiser-Guttman (K-G) rule (i.e., eigenvalue-greater-than-one), (ii) parallel analysis (PA), and (iii) minimum average partial (MAP) test. Unlike the PCA-based methods, the goodness-of-fit statistics test the adequacy of model-data-fit. The goodness-of-fit methods are used in a sequential manner such that one starts with a one-factor solution and multiple models are sequentially fit and tested (adding one factor at a time) until the models fit the data adequately, and one can decide the number of factors. The most commonly used goodness-of-fit test, also included in many statistical program packages (e.g., SPSS), is the sequential chi-square tests based on maximum likelihood (ML) estimation (i.e.,  $\chi^2_{ML}$  sequential tests). These four decision methods are briefly reviewed below.

### 2.1.2.1 *K-G Rule*

Guttman (1954) provided procedures for estimating the lower bound for the number of factors, which includes the currently used eigenvalue-greater-than-one rule (Gorsuch, 1983, p.161). Kaiser (1960) further studied the application of Guttman's lower bound methods and developed the rationale for the use of the eigenvalue-greater-than-one rule. To acknowledge the contribution of both of these psychometricians the eigenvalue-greater-than-one rule is called the Kaiser-Guttman or K-G rule. The K-G rule is the most popular and frequently used method to retain factors among researchers. Its popularity is probably because it has been the default option for many years in widely used statistical software packages such as SPSS and SAS (Conway & Huffcutt, 2003; Fabrigar, Wegener, MacCallum & Strahan, 1999; Ford, MacCallum & Tait, 1986).

Eigenvalues represent the amount of standardized total variance of observed variables explained by each factor. Eigenvalues used in the K-G rule are calculated from the principal component analysis (PCA)<sup>1</sup> extraction method of the correlation matrix. The Pearson product-moment correlation matrix is used for PCA extraction with unit "1" on the diagonal. The variance that each observed variable contributes is one, so the total amount of variance explained in PCA is the total number of observed variables. Using the K-G rule, a factor that can explain more variance than 1.0 is regarded a significant factor and should be retained in the model. Kaiser argued that factors which explain variance less than 1.0 should not be

---

<sup>1</sup> EFA divides the total variance into two parts, common variance and residual variance, and works on the common variance whereas PCA explains the total variance.

considered because the internal consistency of component scores approaches zero or becomes negative when eigenvalues are less than 1.0.

However, most psychometric and methodological studies reported that the K-G rule often retains too many factors (e.g., Browne, 1968; Lee & Comrey, 1979; Linn, 1968; Zwick & Velicer, 1980, 1986). This is contrary to the conclusion made by Guttman (1954) that the K-G rule is a lower bound of the number of factors estimated (i.e., tending to underestimate the number of factors). Cliff (1988) explained that the Guttman lower bound rule was developed with respect to the population correlation matrix, but most decisions were based on sample data, which resulted in the overestimation of factor numbers. In general, over-extraction most likely occurs in data with low communalities, a large number of variables, and small sample size. Gorsuch (1983, p. 164) summarized that the K-G rule is more likely to be effective when the sample size is large, the number of variables is less than 40, and the ratio of the number of factors to the number of variables is between 1:3 and 1:5.

#### 2.1.2.2 *Parallel Analysis (PA)*

Parallel analysis (PA) (Horn, 1965) has received substantial attention recently and is a highly recommended statistical technique for determining the number of factors. Though not available directly in commonly used software packages, the programs to implement parallel analysis in SPSS and SAS are made available by O'Connor (2000) online at <http://flash.lakeheadu.ca/~boconno2/nfactors.html>. The PA method has been suggested as the most accurate decision criterion among all currently used eigenvalue-based methods by many

studies. For instance, Humphreys and Montanelli (1975) found parallel analysis to be superior to chi-square tests based on ML estimation. Zwick and Velicer (1986) compared five methods in their simulation study and found the performance of parallel analysis to be the best across all situations.

As a brief description of PA, building on Kaiser and Guttman's idea (i.e., eigenvalue-greater-than-one), Horn (1965) suggested that the meaningful factors from a real data set should account for more variance than the factors derived from random data in which variables are uncorrelated. More specifically, factors or components should be retained as long as the eigenvalues obtained from the real data are larger than the corresponding eigenvalues from random data. The random data are generated from an identity correlation matrix with all zeros on the off-diagonals and the same sample size and number of variables as the original data. The average eigenvalues from random data generated by computer simulation are used in Horn's original procedure. Recently, PA has been further developed and a highly recommended practice is to use eigenvalues that correspond to a desired percentile (e.g. 95<sup>th</sup> percentile) of the distribution of eigenvalues obtained from random data. This practice is regarded as more accurate than the practice of using average eigenvalues (Cota, Longman, Holden, Fekken, & Xinaris, 1993; Glorfeld, 1995; Turner, 1998). Both average and percentile eigenvalues are available in O'Connor's program.

Russell (2002) pointed out that parallel analysis is a variant on the scree plot. Researchers can implement the idea of a scree plot into parallel analysis and make parallel analysis a visualized plot. The scree plot is a commonly used graphical tool to decide on the

number of factors to retain (Cattell, 1966). To use this method, first one needs to plot eigenvalues of factors in a descending order and then one needs to look for a break in the plot, where there is a substantial drop in the eigenvalues. The number of factors to retain is the number of factors prior to this drop. Problems arise when there is no obvious break or there are several breaks, both of which lead to a more subjective judgment and low interrater reliability (e.g., Gorsuch, 1983; Kaiser, 1970; Crawford & Koopman, 1979).

One way to conceptualize PA is that it improves the utilization of the scree plot. To integrate scree plots into parallel analysis, one needs to plot the eigenvalues derived from random data as well as those obtained from a real data set in one plot, and then look for the point where the two lines cross. The factors with eigenvalues from the real data dropping below those from random data are trivial factors, which should not be retained in the model. Unfortunately, the parallel plot is not available yet in either SPSS or SAS. Researchers need to make a plot using other software programs, such as Excel. It should be noted, however, that a numerical comparison of the eigenvalues from real and random data is, of course, functionally equivalent to the plot.

#### 2.1.2.3 *Minimum Average Partial (MAP)*

Velicer (1976) developed the MAP method in the PCA framework. A partial correlation matrix is calculated after each component is partialled out of the correlation matrix and the average of the squared correlations of the off-diagonal partial correlation matrix (i.e., MAP criterion) is computed. The number of factors to retain is determined by the

point where the minimum average of squared partial correlations is obtained. The rationale of the MAP method described by Velicer is that, as common variance is partialled out of the correlation matrix for each successive component, the MAP criterion will continue decreasing until the point where the common variance has been removed and further extraction of additional components will result in partialling out unique variance, which will lead to the increase of the MAP criterion.

Following the description of Velicer, Eaton, and Fava (2000), the specific procedure of MAP is summarized as follows. The first step is to obtain a partial correlation matrix after the first component is partialled out of the original correlation matrix. The partial covariance matrix needs to be computed first,

$$C = R - AA'$$

where  $C$  is the partial covariance matrix,  $R$  is the original correlation matrix, and  $A$  is the pattern matrix. Then the partial correlation matrix is computed

$$R^* = D^{-1/2}CD^{-1/2}$$

where  $R^*$  is the partial correlation matrix and  $D$  is the diagonal of the partial covariance matrix. The second step is to calculate the average of squared coefficients in the off diagonals of the resulting partial correlation matrix (i.e., the MAP criterion) via the following equation.

$$MAP_m = \sum_{i=1}^p \sum_{j=1}^p \frac{(r_{ij}^*)^2}{p(p-1)}$$

where  $r_{ij}^*$  is the value in row  $i$  and column  $j$  of the partial correlation matrix and  $p$  is the total number of observed variables,  $m$  is the  $m$ th component (i.e., the first component here).

These two steps are repeated until the  $p - 1$  components (i.e., the number of variables minus one step) are extracted from the correlation matrix. The reason for not extracting  $p$  components is that partialling out  $p$  components would result in a null partial covariance matrix. Finally, the average values of squared partial correlations from all steps are compared and the number of components is determined by the component analysis which results in the lowest average of squared partial correlations. The average of the squared coefficients in the original correlation matrix (i.e.,  $MAP_0$ ) is also computed, and if this value is lower than the lowest average squared partial correlation, then there is no need to extract any components.

Zwick and Velicer (1982, 1986) showed that the MAP test generally performed better than the K-G rule and scree test across all situations, especially when the average number of variables per component was large, and was not seriously influenced by sample size.

However, they found that the MAP test consistently underestimated the number of components in the cases of low population factor loadings and a low number of variables per component.

#### 2.1.2.4 *Sequential Chi-square Tests*

Currently, the commonly used significance tests to determine the number of factors in statistical software packages are chi-square tests based on Maximum Likelihood (ML). ML estimation is used to maximize the likelihood of the parameters given the data, or, stated differently, to minimize the discrepancy function (i.e., fitting function) of the form:

$$F_{ML} = \log|\Sigma| + tr(S\Sigma^{-1}) - \log|S| - p ,$$

where  $\Sigma$  denotes the model implied covariance matrix,  $S$  denotes the sample covariance matrix, and again  $p$  denotes the number of observed variables. When  $\Sigma = S$ ,  $\text{tr}(S\Sigma^{-1}) = p$  and hence  $F_{ML}$  becomes zero. The value of zero indicates a perfect fit, which leads one to conclude that the model implied covariance matrix perfectly predicts the sample covariance matrix.

The asymptotic distribution of  $(N-1)F_{ML}$  is a chi-square distribution ( $\chi^2_{ML}$ ) with  $\frac{1}{2}p(p+1)-t$  degrees of freedom, where, again,  $N$  is the sample size and  $t$  is the number of free parameters (Browne, 1982; Jöreskog, 1967, 1969). Bollen (1989, pp. 263-269) showed how the chi-square test based on ML is derived, which is also called the likelihood ratio test. Bartlett (1950, 1951) provided an alternative multiplier which is called the Bartlett correction that was claimed to improve the approximation of the chi-square distribution (Hayashi, Bentler, & Yuan, 2007; Lawley & Maxwell, 1971, Equation 4.30). The chi-square distributions with Bartlett correction are written as  $[(N-1)-(2p+4m+5)/6]F_{ML}$ , where  $m$  is the number of factors and all other notations are the same as defined above. The null hypothesis states that the discrepancy between the maintained factor model with  $m$  factors and the saturated model is not statistically significant, that is,  $H_0: \Sigma = S$ . In contrast to most significance tests, we expect that the null hypothesis is not rejected so that a non-significant chi-square test indicates a good model-data fit. Therefore, the chi-square test is applied to a sequential set of analyses in which one more factor is added into the model at a time until the obtained chi-square test is not significant.

However, when applied to EFA, chi-square tests tend to lead to over-extraction



(Gorsuch, 1983; Harris & Harris, 1971; Linn, 1986; Hakstian, Rogers, & Cattell, 1982; Hayashi et al., 2007; Zwick & Velicer, 1986). Many researchers have pointed out that there are several reasons for being cautious in the use of the chi-square test: small sample size, violation of the normal distribution requirements (i.e., kurtosis and skewness), and Heywood cases in which one or more communality estimates exceed 1.0 that results in one or more residual variances with negative values or approaching zero (Bentler & Yuan, 1999; Boomsma, 1983; Browne, 1984; Geweke & Singleton, 1980; Lawley & Maxwell, 1971; Satorra & Bentler, 1988). In a simulation study, Hayashi et al. (2007) demonstrated that over-extraction can distort the chi-square distribution, which results in substantial skewness and kurtosis and lead to biased chi-square tests.

### 2.1.3 A Demonstration of the Four Decision Methods

The four decision methods described above are illustrated with a real data set from Holzinger and Swineford (1939), which is also used for studies in chapters 3 and 4. This data set consists of 24 psychological ability tests of 301 junior high school students. The data have been used throughout the history of factor analysis and are widely studied in the literature (See Appendix for a list of the 24 tests). The data have been shown to consist of four interpretable factors by many researchers, though some researchers found five factors (e.g., Brown, 2001; Gorsuch, 1983; Harman, 1976; Preacher & MacCallum, 2003; Tucker & Lewis, 1973; Wu, 2008).

Table 2.1 presents the results of the number of factors determined by the four methods

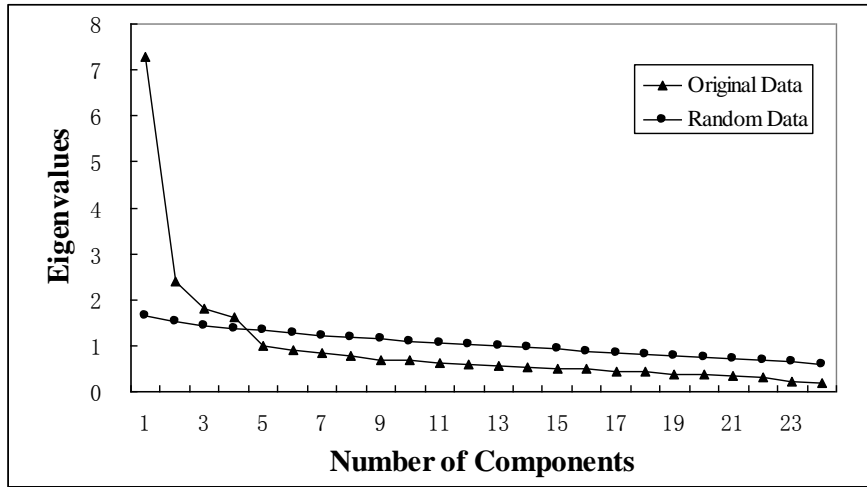
(i.e., K-G, PA, MAP and sequential  $\chi^2_{ML}$  tests). The second column of Table 2.1 shows eigenvalues obtained from the original data, suggesting four factors based on the K-G rule, that is, eigenvalues-greater-than-one. The third column shows the eigenvalues obtained from 5000 random data sets with the same number of variables (24) and sample size (301), which suggested four factors, as the fifth eigenvalue from the original data was smaller than the fifth from the random data. Following the recommendations in the literature, we adopted the practice of 95<sup>th</sup> percentile in this demonstration as well as all studies in this dissertation. Integrating a scree plot into the PA method, Figure 2.1 shows the plot with eigenvalues from the real data as well as the random data with triangles representing eigenvalues from the original data and circles representing those from the random data. The plot shows that the fifth eigenvalue from the original data was smaller than that from the random data, which suggests a solution of four factors.

The fourth column presents the average of squared coefficients in the original correlation matrix (0.086) and a list of MAP criterion values (Velicer, 1976) obtained from extraction of each component. It suggested four factors be retained with the lowest value of 0.014 as the stopping point. Chi-square sequential tests are presented in the last two columns with  $\chi^2_{ML}$  values and the corresponding degrees of freedom as well as  $p$ -values. Using  $p$ -value of .05 as a cut-off criterion, the  $\chi^2_{ML}$  test became non-significant ( $p=0.297$ ) when six factors were retained in the model. Hence, sequential  $\chi^2_{ML}$  tests suggested six factors to retain, which is different from the decision of four factors suggested by all other methods.

**Table 2.1 A Demonstration of Four Decision Methods**

Number of Components	K-G (eigenvalue)	PA (eigenvalue)	MAP	Sequential Chi-square Tests	
				$\chi^2_{ML}$	<i>p</i> -value
0			0.086		
1	7.282	1.636	0.026	1157.10 ( <i>df</i> =252)	< .001
2	2.401	1.524	0.018	644.97 ( <i>df</i> =229)	< .001
3	1.819	1.447	0.015	416.92 ( <i>df</i> =207)	< .001
4	1.606	1.381	0.014	263.06 ( <i>df</i> =186)	< .001
5	0.987	1.326	0.016	203.09 ( <i>df</i> =166)	0.026
6	0.903	1.274	0.019	155.66 ( <i>df</i> =147)	0.297
7	0.826	1.228	0.022		
8	0.786	1.184	0.028		
9	0.700	1.142	0.034		
10	0.682	1.102	0.040		
11	0.634	1.063	0.046		
12	0.599	1.026	0.055		
13	0.573	0.990	0.069		
14	0.535	0.955	0.085		
15	0.498	0.921	0.102		
16	0.485	0.887	0.119		
17	0.448	0.854	0.149		
18	0.423	0.819	0.174		
19	0.382	0.786	0.227		
20	0.361	0.752	0.307		
21	0.337	0.718	0.350		
22	0.319	0.682	0.516		
23	0.215	0.644	1.000		
24	0.202	0.601			

**Figure 2.1 A Demonstration of Parallel Analysis**



## **2.2 Review of Outliers**

### **2.2.1 Sources of Outliers**

In the literature of outlier studies, some researchers have discussed the potential causes of outliers (Barnett & Lewis, 1994; Beckman & Cook, 1983). Liu and Zumbo (2007) summarized the sources of outliers and described three general categories. The first category includes errors that occur during data collection and errors in preparing data for analysis (e.g., data recording or entry errors). The outliers generated from such sources are illegitimate observations and should, where possible, be corrected.

The second category refers to the unpredictable measurement-related errors from participants, such as participant guessing, inattentiveness, which may be caused by fatigue or participants' lack of interest in participation, or misresponding, which happens when, for example, participants misunderstand the instructions or the descriptors on the response scale (Barnette, 1999; Gelin, Beasley, & Zumbo, 2003; Zumbo & Ochieng, 2003).

The third category occurs when researchers unknowingly recruit some individuals who are not members of the target population. For example, a researcher planned to target English as second language (ESL) students and study their language proficiency but unknowingly includes a subgroup of individuals whose first language is English.

Beckman and Cook (1983) pointed out that the appearance of outliers can also be caused by an inappropriate model that fails to provide an adequate model data fit. For

instance, Cohen, Cohen, West, and Aiken (2003) demonstrated that some response variables were treated as outliers when a linear regression model was applied, but became legitimate data points when a non-linear regression model was used. It should be noted that the presence of outliers in a normal distribution may cause skewness, but skewness is not necessarily caused by outliers as the nature of such data is inherently non-normal. Liu and Zumbo (2007) did not include this source of outliers because they defined outliers based on the assumption that the model was correct for the target population. In this case, if replacing the model with a correct model, outlying data points are not outliers anymore.

Given the characteristics of outliers and mechanisms of outlier simulation models, these three categories of outliers were recast into two general types of outliers in this dissertation: outliers arising from errors and those arising from subpopulations. The first type of outliers are errors, including Liu and Zumbo's (2007) first category and part of the second category of outliers, such as data entry errors, data recording errors, or random measurement errors caused by participants' fatigue. Because outliers are errors and are sample specific, it does not make sense to talk about this type of outliers in terms of a population.

The second type of outliers refers to the situation in which one or more subpopulations, which are unplanned and unknowingly recruited, are mixed with the target population. This type of outlier includes Liu and Zumbo's (2007) third category as well as part of the second category of outliers. The observations should be regarded as a subpopulation if outliers in Liu and Zumbo's second category are systematic measurement errors because of characteristics of respondents, such as inherent inattentiveness of

respondents instead of temporary inattentiveness.

In the outlier literature, slippage models and mixture contamination/mixed contamination models have been widely used (Barnett & Lewis, 1994). Both of these models were adopted for simulating outliers from different sources in this dissertation. In addition, deterministic models were also utilized in the present research. The descriptions of these two types of outliers and outlier simulation models are given in details in chapters 3, 4 and 5. This section only gives a simple description of the distinction of these two types of outliers to facilitate the understanding of the material that is forthcoming in this and other chapters.

From a psychometric point of view, one can investigate outliers at either the scale or item level. At the scale level, outliers are defined in terms of scale scores or subscale scores, whereas, at the item level, outliers are defined with respect to item responses. In line with my research interest in continuous variables, outliers in terms of subscale scores or visual analogue scales<sup>2</sup> are considered in the present research because the use of subscales is a common practice in educational, social, behavioral, and psychological studies. In addition, visual analogues scales have gained popularity in computer-based surveys in recent years.

### 2.2.2 Problems of Outliers in Factor Analysis Models

In the literature discussing outliers in factor analysis, researchers have focused on different techniques for detecting and removing outliers (e.g., Comrey, 1985;

---

<sup>2</sup> Visual analogue scale is used widely in assessing human well-being, quality of life, happiness, and psychological and health status; the response format is usually a continuous line between a pair of descriptors that represents opposite ends of a continuum (e.g., “no pain at all” and “worst pain I have ever had”).

Castaño-Tostado & Tanaka, 1991; Kwan & Fung, 1998; Mavridis & Moustaki, 2008; Tanaka & Odaka, 1989) and robust factor analysis models to handle outliers (e.g., Pison, Rousseeuw, Filzmoser, & Bentler, 2003; Yuan & Bentler, 1998). However, few studies have systematically investigated how, and to what extent, outliers affect the performance of factor analysis, especially how outliers affect the decisions about the number of factors in EFA determined by widely used decision methods.

Using the classical dataset of Holzinger and Swineford (1939) with 24 psychological variables collected from 145 subjects, Yuan, Marshall, and Bentler (2002) manipulated one or two subjects' scores, making their values 2 to 5 times larger than the original values, to create outliers in their illustrations. In general, they found that the outliers changed the factor solutions based on ML estimation in the case without robust treatment: (a) the extreme value of outliers, 5 times of the original value, diminished the number of factors from the original five to one, and (b) factor loadings were distorted and lost simple structure. In addition, their study showed that the  $\chi^2$  tests, which include the conventional one, defined as  $(N-1)*F$ , the one with Bartlett's correction, and other rescaled  $\chi^2$  tests, were not affected by one outlier, but were affected by two outliers. In their study, Yuan et al. only examined the K-G rule and  $\chi^2$  tests for determining the number of factors.

Using the same Holzinger and Swineford data, Yuan and Zhong (2008) chose nine observed variables from the data, manipulated different outlier conditions and investigated how outliers affected the CFA model parameter estimates as well as  $\chi^2_{ML}$  tests. In general, their findings showed that all of the parameter estimates (factor loadings, residual variances,



and factor correlations) were distorted to a large magnitude in the presence of outliers and the type I error rate of  $\chi^2_{ML}$  tests were inflated.

A few other studies also examined the effect of outliers in CFA. Bollen (1987) and Bollen and Arminger (1991), using empirical data (Cermak & Bollen, 1983), demonstrated that outliers can cause negative residuals (i.e., Heywood cases), but had relatively less effect on factor loadings. Bollen and Arminger as well as Huber (1981) conjectured, without testing it empirically or via simulation, that a large outlier may create an extra factor. Bollen and Arminger's conjecture was not found to hold up in Yuan, Marshall, and Bentler's (2002) study. Yuan et al. found, using real data, that the number of factors diminished in the presence of outliers.

Selecting four variables from a data set from the National Basketball Association (NBA) descriptive statistics for 105 point guards (a position on the basket team) for the 1992-1993 basketball season (Chatterjee, Handcock, & Simonoff, 1995), Yuan and Bentler (2001) showed that the values of  $\chi^2_{ML}$  tests in a unidimensional factor analysis model were inflated by outliers, that is, the Type I error rate was inflated, and thus pointed out that the presence of outliers could result in misleading results, such as the model failing to fit the data, even if the model is correct.

The results from previous studies showed that outliers could create problems for the decision about the number of factors in EFA, parameter estimates, and model fit indices. However, most of the previous studies are based on empirical data, which are sample dependent, so that we cannot generalize the findings of these studies to other samples or

more general outlier conditions. In addition, there are apparent contradictory conclusions in the literature about how outliers affect the number of factors in EFA, for example Bollen-Arminger-Huber's conjecture of extra factors vs. Yuan-Marshall-Bentler's conclusion of a reduction in factors. Hence, there is a need for more research to systematically investigate how outliers affect factor structure, parameter estimates and fit indices. It should also be noted that, in these studies, investigation of the effects of outliers was not the primary purpose of the studies but rather was used to motivate the need for the mathematical work being presented on outlier detection and/or robust estimation. Therefore, these studies never intended to comprehensively understand the effect of outliers, which is the purpose of this dissertation.

### **2.3 Study Purpose**

Given that there is lack of systematic studies on effects of outliers in EFA in the literature, the purpose of this dissertation was to investigate how outliers affect the decisions about the number of factors determined by four widely used or highly recommended methods (i.e., K-G, PA, MAP and sequential  $\chi^2_{ML}$  tests) for continuous observed variables. The effects of outliers were investigated in chapters 3 and 4 with each chapter focusing on a different type of outliers.

Chapter 3, consisting of three studies, investigates outliers arising from errors. In the first study, the classic data from Holzinger and Swineford (1939) were manipulated to mimic outliers. In the second study, based on the estimates obtained from the data used in

the first study, a Monte Carlo simulation study was conducted with 100 replications for each research design condition. Three outlier conditions were manipulated in both studies: the magnitude of outliers (i.e., 2-5 times of the original values of variables), number of subjects who have outlying responses, and number of variables with outliers. The third study was conducted in order to provide insight into the causes of the findings obtained from the first two studies.

Chapter 4, consisting of two studies, investigates outliers arising from an unintended and unknowingly included subpopulation. A Monte Carlo simulation study was conducted, in which the outlier conditions were manipulated using five factors that resulted in both symmetric and asymmetric outlier conditions (i.e., mean shift, standard deviation shift, proportions of contamination, number of variables having outliers, and sample size). Similar to chapter 3, a follow-up study was also presented after the first study.

## **Chapter 3:      Effects of Outliers Arising from Errors**

### **3.1    Introduction**

Exploratory factor analysis (EFA) is a widely used analytical technique for the purpose of test construction and validation in many fields such as psychology, education, sociology, marketing, and the political, social, behavioral and health sciences. Although several studies have documented the impact of outliers on commonly used statistics and methods, such as the estimation of the center of a distribution (e.g., the mean or median) (e.g., Andrews et al, 1972; Barnett & Lewis, 1994), the estimation of Cronbach's coefficient alpha of reliability (Liu & Zumbo, 2007, Liu, Wu, & Zumbo, 2010), estimated regression coefficients (Cohen, Cohen, Aiken, & West, 2003), and type I and type II error rates of the t-test or analysis of variance (ANOVA) (e.g., Zimmerman & Zumbo, 1993; Zumbo & Jennings, 2002), the impact of outliers on EFA is still largely undocumented.

In discussing the limited use of robust estimators in day-to-day research practice in the social and behavioral sciences, Lind and Zumbo (1993) pointed out the following:

It is often assumed, for example, that data which are observed to deviate only slightly in form from that of the familiar normal curve, will only slightly distort the usual estimates of means, standard deviations, correlations and associated hypothesis tests... Over the past several decades, research in statistics has demonstrated that a continuity principle of the form described above for normal theory based statistics is invalid ... A single outlying observation, for example, can

strongly bias these statistics and thereby provide misleading or invalid results. (p. 407-408)

For example, Cohen et al. (2003, p. 407-409) demonstrated that one outlier distorted the regression parameter estimates greatly. Devlin, Gnanadesikan, and Kettenring (1981) likewise showed that the correlation coefficient dropped from .99 to zero in the presence of a single outlier in a sample of 29 observations.

It should be noted that, to date, a large portion of the multivariate statistical and psychometric research has focused on identifying outliers or developing robust methods for factor analysis models (Castaño-Tostado & Tanaka, 1991; Tanaka & Odaka, 1989; Kwan & Fung, 1998) rather than documenting the impact of outliers, per se. It is as if the impact of outliers is widely known and appreciated, and it most certainly is for those researchers working on the statistical theory of robust estimators, so that they have focused on providing a solution (i.e., robust estimators) to what is a known but, at least for applied researchers, widely undocumented statistical problem. A number of robust methods have been developed such as R-estimators, L-estimators, and M-estimators to down-weight the influence of outliers (e.g., Barnett & Lewis, 1994; Hampel, Ronchetti, Rousseeuw, & Stahel, 1986; Huber, 1981). These terrific statistical advances are largely ignored by day-to-day researchers because of a belief that a small deviation from the familiar normal curve will only slightly distort the usual estimates of means, standard deviations, correlations and associated hypothesis tests (i.e., the continuity principle described above), as pointed out by Lind and Zumbo (1993). A pernicious contributing factor is the limited availability of robust

techniques in widely used statistical software programs such as SPSS and SAS. We believe, however, that the software will come along and respond to market demands when researchers insist on having robust methods at their disposal. The purpose of this paper is to demonstrate to day-to-day researchers the impact of outliers on the decision about the number of factors to extract in EFA and along the way shed some light on apparent contradictory conclusions in the literature to date.

We found very few studies discussing the effect of outliers on EFA performance, particularly the issue of determining the number of factors, and furthermore these studies had somewhat contradictory conclusions. Bollen and Arminger (1991) discussed how outliers affected factor analysis models and conjectured, without testing it empirically or via simulation, that a large outlier may create an extra factor. Likewise, in his description of robust covariance and correlation matrices, Huber (1981, p. 199) also briefly mentioned that sample covariance matrices are excessively sensitive to outliers and that one or two outliers may create an extra factor. The findings from Yuan, Marshall, and Bentler's (2002) study, however, do not support Bollen and Arminger's or Huber's conjecture. Yuan et al. found that the number of factors was reduced in the presence of outliers using the Kaiser-Guttman (K-G) rule (eigenvalues greater than one) – for example, in the extreme case, the number of factors decreased from five factors to one factor. It is worth noting that the purpose of Yuan et al.'s study was focused on the introduction and description of new robust methods so they, in essence, did a very limited study of the impact of outliers to set the stage for their statistical work on robust estimation.

It is important for day-to-day researchers, when reading the extant literature or interpreting their own new findings, to understand whether outliers may have an impact, to what degree, and in what way on the results of exploratory factor analysis. This is a particularly important issue for the essential question of the number of factors to extract during the process of factor analysis. There are many possible sources of outliers (Liu & Zumbo, 2007). The sources of outlier generation in the present study are representative of Liu and Zumbo's first category of outlier sources, that is, errors that mainly occur during data collection or errors in preparing data for analysis (e.g, typographical errors), which are extreme data points. The outliers that resulted from such sources are illegitimate observations and should, when possible, be corrected or removed.

The first two studies reported here demonstrate the impact of outliers on decisions about the number of factors in EFA. In addition, a third study is intended to aid readers in understanding how outliers affect correlation matrices and the eventual decision about the number of factors. The first study is not a simulation, per se, but rather a demonstration of the effects of outliers by systematically inducing outliers and tracing the resultant effects on the decisions about the number of factors. Simulation was used in Studies 2 and 3 because of the need to induce the presence of outliers and document their resultant influence on the decision about the number of factors (in study 2) and the correlation matrix (in study 3). All three studies are influenced by Yuan et al.'s (2002) work. In particular, like Yuan et al, the classic data set of Holzinger and Swineford (1939) is used as to mimic a real data situation.

The effect of outliers was investigated herein using the same framework as Yuan et

al.'s (2002) study, except that in the present research more outlier conditions and more decision methods for deciding the number of factors were included. Alternatively, rather than using real data or simulation, one could investigate the impact of outliers by deriving the mathematical influence function, but the finite sample (empirical) influence function for the statistical test of number of factors such as those based on likelihood theory is a complex mathematical result and not accessible to day-to-day researchers (e.g., Pison, Rousseeuw, Filzmoser, & Croux, 2003). This inaccessibility is further complicated for psychometric tools like the eigenvalues greater than one rule or parallel analysis that are pointers to the number of factors and not founded on a statistical sampling theory; the influence functions for these methods are intractable. And most importantly, our purpose is to demonstrate the impact of outliers to day-to-day researchers in the social and behavioral sciences. Hence, for our target audience, demonstrating the effects of outliers via computer simulation is likely more persuasive and more appropriate than a complex derivation and mathematical analysis.

Therefore, in summary, three studies are reported to demonstrate how outliers affected the decisions about the number of factors to extract using four commonly used or recommended decision methods. The first study is based on real data which are manipulated to mimic outliers. The second study is a more typical Monte Carlo simulation with replications for each design condition; however, the estimates obtained from the real data used in the first study are used as input population parameters. The first study has high fidelity with real data situations; however it lacks replications, which the second study overcomes. The final investigation is a focused simulation of a small correlation matrix that



is a follow-up to the first two studies and is meant to provide insight into how the outliers are altering the elements of the correlation matrix and the corresponding eigenvalues.

### **3.2 Study 1: Demonstration of the Impact of Outliers with a Classic Data Set**

#### **3.2.1 Method**

##### *Study design*

As a follow up to Yuan et al.'s (2002) study, the present study examined how outliers affect four decision methods to determine the number of factors using Holzinger and Swineford's (1939) data. The data set consists of 24 psychological ability test scores from junior high school students. There exist two versions of this classic data set -- a smaller version with a sample of 145 subjects, which was used in Yuan et al.'s (2002) study and a larger version with 301 subjects, which was used for the present study<sup>3</sup>. This classic data set is one of the most widely studied data sets in the factor analysis literature. Previous studies concluded that there are either four or five factors, depending on the decision rule researchers used – with the four factor solution recommended by most researchers (e.g., Gorsuch, 1983; Harman, 1976).

To create outliers for the computer simulation, Yuan et al. (2002) manipulated the raw scores of either the very last subject or the last two subjects in the data set for all 24 variables. Readers should note that Yuan et al.'s purpose was to illustrate their new robust estimation

---

<sup>3</sup> The larger sample size was used for two reasons. First, most of the previous studies of these variables reported the factor structure based on this larger sample. Second, the larger sample size was used because most guidelines in the literature advocate for large sample size of at least 200 to obtain high-quality factor analysis solutions (Jung & Lee, 2011).

methods, and not to systematically study the impact of outliers, per se; therefore, their study is limited by its very nature. The new scores (i.e., outliers) were two to five times that of the original raw scores of the subjects. Therefore Yuan et al.'s study was a  $2 \times 4$  completely crossed factorial design for the outlier conditions—that is, two levels of number of subjects having outlying responses, and four levels of outlier magnitude (2, 3, 4, and 5 times the original scores). Yuan et al. only investigated the K-G rule for deciding the number of factors. Although chi-square tests were used as a reference, they did not conduct chi-square tests in a sequential manner.

The first study mimicked outliers by increasing the magnitude of original raw scores, which is an extension of Yuan et al.'s study design, by varying three factors in a factorial design: (1) the magnitude of outliers, (2) the number of subjects who had outlying responses, and (3) the number of variables with outliers. Like Yuan et al.'s study, new scores (outliers) were 2 to 5 times the original raw scores for some proportion of the last subjects in the data file and a certain number of variables. Adopting the same outlier percentage used by previous studies (Liu & Zumbo, 2007; Liu et al., 2010), the percentage of subjects who had outlying responses was, approximately, 1%, 8% and 15% of the sample size – in this case the last 3, 24 and 45 subjects in the data file. The number of variables with outliers was set at 1, 6, 12 and 24. These numbers of variables with outliers were chosen to represent different degree of outlier contamination, which covered the entire range from the case of only one variable having outliers to the extreme case of all variables having outliers – the latter case was the condition Yuan et al. studied. Therefore, the present study design was a  $3 \times 4 \times 4$  completely

crossed factorial design (48 design conditions) – that is, three levels of the number of outlier cases, four levels of the outlier magnitude, and four levels of the number of variables with outliers, as well as one baseline condition, the original data.

Given that our purpose was to inform day-to-day research practice, the methods for deciding on the number of factors to extract were chosen based on day-to-day practice. A review of decision methods used for deciding the number of factors in EFA was conducted using three journals that routinely report psychometric applications of EFA: *Psychological Assessment*, *International Journal of Educational and Psychological Assessment*, and *European Journal of Psychological Assessment*. Between the years 2000 and 2010, a total of 48 applied factor analysis papers appeared in these journals. Four of these papers did not indicate what method was used for deciding the number of factors. Each of the remaining papers reported using one or more decision rules to determine the number of factors to retain. Among these 44 papers, 28 used the K-G rule (eigenvalues greater than one), 24 used the scree plot, 6 used PA method, 4 used minimum average partial (MAP) method, and 4 used a sequential chi-square test method based on maximum likelihood (ML) estimation. The results of our review are similar to previous reviews (Conway & Huffcutt, 2003; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Ford, MacCallum, & Tait, 1986). Therefore, we included the K-G rule, PA, MAP and sequential chi-square test for our study because they are either commonly used or recommended in the literature (Gorsuch, 2003; Hoyle & Duvall, 2004; Slocum-Gori & Zumbo, in press; Zwick & Velicer, 1986). The use of a scree plot was excluded as it is a subjective tool involving personal judgment.

The first three decision methods (K-G, PA, and MAP) are based on some variation of principal component analysis (PCA), and hence are pointers to help decide on the number of factors whereas the last one is a goodness-of-fit test. Therefore, K-G, MAP, and PA methods are used to help the researcher decide on the region of the possible number of factors wherein the correct number of factors may lie, hence described as ‘pointers’, whereas the chi-square, as a test of fit, is used to help a researcher decide if she/he has specified the correct number of factors in the model. In some cases, a researcher may start with the pointers and then use the fit test to decide on which of the alternative number of factors is correct. It should be noted that the application of factor rotations is not a matter of concern for the decision rules because we only focused on the number of factors suggested by the decision methods, but not on the interpretability of the factor solution.

### *Procedure*

Given that our aim is to inform day-to-day researchers, all the analyses were performed using factor analysis and “Very Simple Structure” packages in the software R2.12.1. Considering that SPSS is a popular software program in many fields, we compared the results of the first simulation study in R with those obtained from SPSS and no difference was found for the K-G approach. The sequential  $\chi^2_{ML}$  tests showed no difference when convergence was reached; however, a 12% increase in non-convergence was found with R due to differing default values compared to SPSS for convergence in the estimation method. The MAP and PA approaches were also conducted using O’Connor’s (2000) SPSS syntax, and the results obtained from the original version of MAP (1976) and the 95<sup>th</sup> percentile of

PA were found to be the same as those in R. It should be noted that the chi-square test was applied in a sequential manner beginning with the extraction of a single factor and an additional factor was added at each step until the obtained chi-square test was not statistically significant (Hoyle & Duvall, 2004, p. 307)

### 3.2.2 Results

Table 3.1 present the results of the first simulation study. The dependent variable is the difference between the number of factors obtained from the data with outliers and the number of factors from the original data – such that zero indicates no effect of outliers and a negative (positive) number means that the outliers decreased (increased) the number of factors. The numbers of factors obtained from the four decision methods based on the original data set, without manipulated outliers, were: four factors based on the K-G rule, PA and MAP approaches; and six factors based on the sequential  $\chi^2_{ML}$  test.

Table 3.1 shows the results of the K-G rule, PA, MAP, and sequential  $\chi^2_{ML}$  tests. Let us take the K-G rule as an example to describe the format of all of the tables. The K-G rule section of Table 1 is divided into four sub-sections for each of the conditions of 1, 6, 12 and 24 variables with outliers. Within each section (e.g., the one outlier variable condition) the subtable was arranged by the magnitude of outliers (rows) and the number of subjects with outliers (columns). For example, for the K-G rule with the magnitude of outliers being twice the original data, the outcome of the K-G rule was the same as the original data (therefore zeros in the table) for the case of three subjects with outliers, but was inflated by resulting in

an additional factor in the cases wherein there were 24 or 45 subjects with outliers.

From Table 3.1, it is evident that, in the presence of outliers, the number of factors sometimes increased and other times decreased depending on the degree of outlier contamination as well as the decision methods used. Some general patterns were found. For the K-G rule, the number of factors increased when the outlier contamination was low and decreased dramatically, from four to one, when more variables had outliers (12 & 24 variables) with 8% and 15% subjects with outliers. For the MAP method, the number of factors was not affected at low outlier contamination (one variable with outliers), but showed strange patterns as the outlier conditions worsened: increasing, decreasing and then returning to the original number of factors. Compared to the other three methods, the PA approach was more resistant to outliers when either one or six variables had outliers. Finally, it is important to note that for the sequential  $\chi^2_{ML}$  tests nonconvergence occurred most often in extreme outlier conditions.

**Table 3.1 Change in the Number of Factors from Four Approaches Based on the Original Data**

K-G				MAP			
Magnitude of Outliers	Number of Outlying Subjects			Magnitude of Outliers	Number of Outlying Subjects		
	3	24	45		3	24	45
	1 Variable with Outliers				1 Variable with Outliers		
2	0	1	1	2	0	0	0
3	0	1	1	3	0	0	0
4	0	1	1	4	0	0	0
5	0	1	1	5	0	0	0
6 Variables with Outliers				6 Variables with Outliers			
2	1	1	1	2	0	1	1
3	1	1	1	3	1	1	1
4	1	1	1	4	1	1	1
5	1	1	1	5	1	1	1
12 Variables with Outliers				12 Variables with Outliers			
2	1	1	1	2	0	1	0
3	1	-1	-1	3	1	-1	-1
4	1	-1	-1	4	1	-1	-1
5	1	-1	-1	5	0	-1	-1
24 Variables with Outliers				24 Variables with Outliers			
2	1	-1	-2	2	1	0	-1
3	0	-2	-3	3	-1	0	0
4	0	-3	-3	4	-1	0	0
5	-1	-3	-3	5	-1	0	0

**Table 3.1 Continued.**

PA				Sequential Chi-square (ML) Tests			
Magnitude of Outliers	Number of Outlying Subjects			Magnitude of Outliers	Number of Outlying Subjects		
	3	24	45		3	24	45
	1 Variable with Outliers				1 Variable with Outliers		
2	0	0	0	2	0	0	0
3	0	0	0	3	0	0	0
4	0	0	0	4	0	0	0
5	0	0	0	5	0	0	0
6 Variables with Outliers				6 Variables with Outliers			
2	0	0	0	2	0	0	0
3	0	0	0	3	0	0	0
4	0	0	0	4	0	1	*
5	0	0	0	5	0	*	*
12 Variables with Outliers				12 Variables with Outliers			
2	0	-1	-1	2	1	0	0
3	0	-1	-1	3	1	1	0
4	-1	-1	-1	4	1	*	0
5	-1	-1	-1	5	1	*	1
24 Variables with Outliers				24 Variables with Outliers			
2	0	-2	-3	2	1	2	*
3	-2	-3	-3	3	1	*	*
4	-2	-3	-3	4	1	*	*
5	-2	-3	-3	5	*	*	*

Note: The \* denotes a cell where there was non-convergence for the sequential  $\chi^2_{ML}$  tests.



### **3.3 Study-2: A Monte Carlo Simulation Documenting the Effects of Outliers**

#### **3.3.1 Methods**

##### *Study Design*

Like Study 1, Study 2 also investigated the effect of outliers on the decisions about the number of factors for the four decision methods. The design of the second simulation study is the same as Study 1 except that the dataset used for outlier manipulation was simulated based on a four-factor model rather than manipulating the elements of a real data set. A four-factor solution based on maximum likelihood exploratory factor analysis was obtained using Holzinger and Swineford's (1939) data. The resulting reproduced correlation matrix (i.e., the implied correlation matrix with "1s" on the diagonal, rather than the reproduced communalities) was used as the population correlation matrix in the simulation to generate multivariate normal datasets with the same number of variables (24) and subjects (301) as the original data. Therefore in Study 2, the numbers of factors obtained from different outlier conditions were compared to a common (known) criterion in the population, four factors.

As in Study 1, the study design was a  $3 \times 4 \times 4$  completely crossed factorial design (48 design conditions) – that is, three levels of the number of outlier cases, four levels of the outlier magnitude and four levels of the number of variables with outliers. One hundred replications were conducted for each cell in the design. It should be noted that, unlike the

first simulation, which, in essence, has only one replication per cell, the average value of the number of factors obtained from 100 replications is reported for each cell in the design. As such, the number of factors reported in each cell need not be a whole number.

### *Data Analysis*

Following the data analysis strategy in Liu and Zumbo (2007) and Liu et al. (2010), three-way ANOVAs (3x4x4) were conducted with the number of factors extracted as the dependent variable and with the different outlier conditions as independent variables separately for the K-G, MAP and PA approaches<sup>4</sup>. An ANOVA was not conducted for sequential  $\chi^2_{ML}$  tests because, due to the non-convergence problem, there were empty cells in the design that substantially complicated the ANOVA. As in Study 1, we presented the change in the number of factors in a tabular manner for the chi-square method.

Following the analytic strategy in our earlier papers, given the large sample size (4800), we utilized eta-squared ( $\eta^2$ ) to orthogonally partition the explained variance from the fixed effect ANOVA models instead of looking at the statistical significance. The proportion of explained variance was used to aid our interpretation of the simulation results. This principle is akin to the strategy in regression analysis involving variable ordering in terms of the most important variables, or practically significant effects, contributing to the

---

<sup>4</sup> We recognize that our dependent variable is a count variable and hence specialized regression techniques such as the Poisson or negative binomial can be used. We used the ANOVA (i.e., a linear regression model) rather than the Poisson or negative binomial because ours is a descriptive purpose to partition the observed variation in the dependent variable rather than specifying a predictive model and/or focusing on significance tests. Furthermore, as Gardner, Mulvey, and Shaw (1995) note, the specialized models are needed when the assumptions of linear regression models are violated – such as normality and homogeneity of variances. However, in our situation those assumptions are not seriously violated, and likewise the variances of our dependent variables are much smaller than the mean which is counter to the expectation for Poisson or negative binomial models.

model  $R^2$  (Thomas, Hughes, & Zumbo, 1998). Like  $R^2$  in regression analysis,  $\eta^2$  is defined as the sum squares of factors under study divided by the total sum squares. Following Ferguson (2009), we used the effect size of 0.04 as the criterion to judge the importance of the main effects and interactions. That is, in Ferguson's terminology, this is the minimum effect size representing a practically significant effect. In addition, if interactions appeared in the model, only higher order interactions were interpreted because main effects and lower order interactions are not interpretable in the presence of higher order interactions.

### 3.3.2 Results

The results obtained in Study 2 were similar to the first study. Table 3.2 presents the results for the sequential chi-square method. Tables 3.3 to 3.5 present the results for K-G, MAP and PA approaches, respectively, and Figures 1 to 3 depict the higher order interactions or the main effects identified as the important factors using  $\eta^2$ .

As in Study 1, Table 3.2 contains the change on the number of factors for the  $\chi^2_{ML}$  method. With only one variable having outliers, there was no change at all compared to the number of factors in the population (i.e., the method resulted in 4 factors) regardless of the magnitude of outliers and number of subjects with outliers. However, the number of factors was inflated from 4 to 10 when the number of variables with outliers increased, and non-convergence problems were found starting with 6 variables with outliers.

Tables 3.3 to 3.5 present the results of sum squares,  $\eta^2$  values, as well as the proportion of variance explained by main effects and interactions for the K-G, MAP and PA

methods. From Table 3.3, one can see that the three important variables for the K-G rule were:

(i) main effect of the number of variables with outliers, (ii) the interaction of number of variables with outliers and number of subjects with outliers, as well as (iii) the interaction of number of variables with outliers and the magnitude of the outliers. Figure 3.1 shows the plots of these two interactions. Both plots showed similar patterns wherein the number of factors was inflated when less than 24 variables had outliers. However, when all 24 variables had outliers the number of factors was deflated. Table 3.4, one can see that the important variables for MAP include a main effect of the number of variables having outliers and a three-way interaction of the number of variables having outliers, the number of subjects having outliers and magnitude of outliers. This three-way interaction is plotted in Figure 3.2. Like Study 1, the number of factors was not affected when only one variable had outliers, but otherwise inflated or deflated based on the various outlier conditions. Table 3.5 lists the ANOVA results for the PA method: only a main effect of the number of variables with outliers was identified as an important factor, as depicted in Figure 3.3. There was no effect of outliers for 1 and 6 variables with outliers but otherwise outliers resulted in a deflation in the number of factors extracted.

**Table 3.2 Change in the Number of Factors Based on the Sequential Chi-square (ML)**

**Test**

Magnitude of Outliers	Number of Outlying Subjects		
	3 (1%)	24 (8%)	45 (15%)
<b>1 Variable with Outliers</b>			
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
<b>6 Variables with Outliers</b>			
2	.8	1.1	1.0
3	1.0	1.2	1.2
4	1.1	1.5	1.5 (2)
5	1.3	1.8 (3)	1.8 (3)
<b>12 Variables with Outliers</b>			
2	.8	1.2	1.3
3	1.2	2.3 (1)	2.4 (11)
4	1.4	3.4 (11)	3.2 (16)
5	1.8	4.1 (32)	3.7 (40)
<b>All 24 Variables with Outliers</b>			
2	.9	2.3 (11)	3.0 (44)
3	1.5	6.0 (99)	*
4	2.0 (6)	*	*
5	2.3 (47)	*	*

Note. The number in brackets for indicates the number of replications, out of 100, with non-convergence. The \* denotes a cell where none of the 100 replications reached convergence.

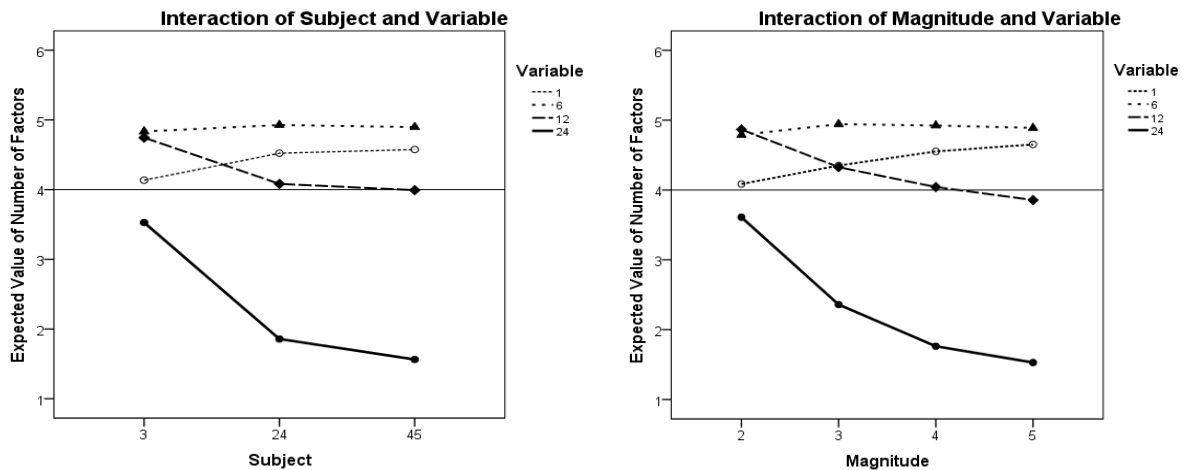
**Table 3.3 Variable Ordering for a Three-way ANOVA on the Number of Factors**

**Extracted by the K-G Rule**

Model	Sum Squares	Eta-square	Percentage of R-Square
Intercept	75715.853		
<b>var#</b>	<b>4635.455</b>	<b>0.610</b>	<b>67.8</b>
<b>var# * sub#</b>	<b>799.059</b>	<b>0.105</b>	<b>11.7</b>
<b>var# * magnitude</b>	<b>756.917</b>	<b>0.100</b>	<b>11.1</b>
sub#	282.313	0.037	4.1
Magnitude	257.928	0.034	3.8
var# * sub# * magnitude	79.180	0.010	1.2
sub# * magnitude	29.255	0.004	0.4
Error	762.040		
Total	83318.000		
Corrected Total	7602.147		

Note. R-Square = .90. In terms of notation, *var#* denotes the number of variables with outliers, *sub#* denotes the number of subjects with outliers, and *magnitude* denotes the magnitude of the outliers.

**Figure 3.1 Graphs for Two-way Interactions of Variables vs. Subjects with outliers and Variables vs. Magnitude of Outliers on the Number of Factors Extracted by the K-G Rule**

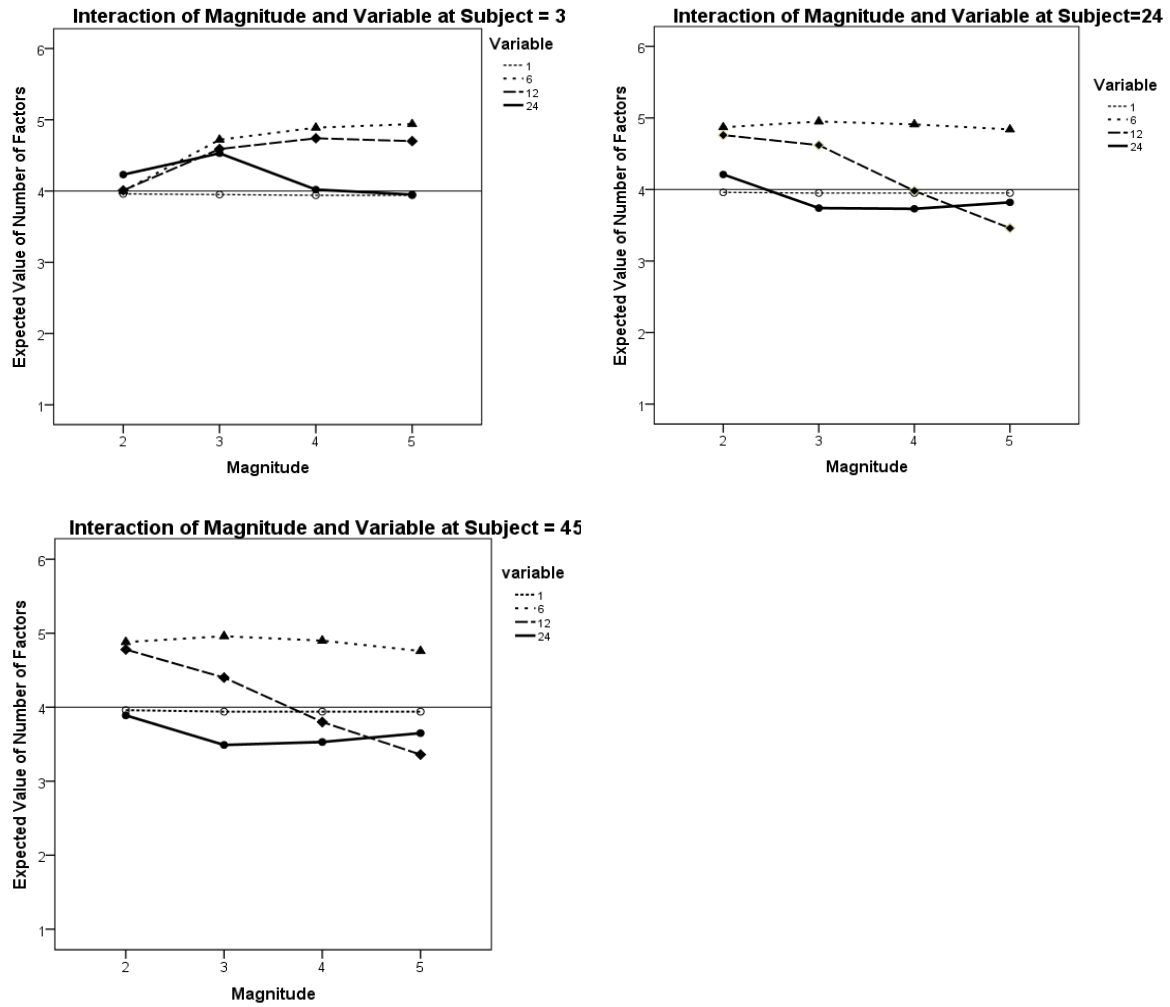


**Table 3.4 Variable Ordering for a Three-way ANOVA on the Number of Factors****Extracted by the MAP Approach**

Model	Sum Squares	Eta-square	Percentage of R-Square
Intercept	85843.625		
<b>var#</b>	<b>620.316</b>	<b>0.231</b>	<b>56.3</b>
<b>var# * sub# * magnitude</b>	<b>132.762</b>	<b>0.049</b>	<b>12.1</b>
var# * magnitude	105.779	0.039	9.6
sub# * magnitude	95.263	0.035	8.6
var# * sub#	87.021	0.032	7.9
Magnitude	33.464	0.012	3.0
sub#	26.840	0.010	2.4
Error	1585.930		
Total	88531.000		
Corrected Total	2687.375		

Note. R-Square = .41. In terms of notation, *var#* denotes the number of variables with outliers, *sub#* denotes the number of subjects with outliers, and *magnitude* denotes the magnitude of the outliers.

**Figure 3.2 Graphs for Three-way Interactions of Variables, Subjects with Outliers and the Magnitude of outliers on the Number of Factors Extracted by MAP Approach**





**Table 3.5 Variable Ordering for a Three-way ANOVA on the Number of Factors**

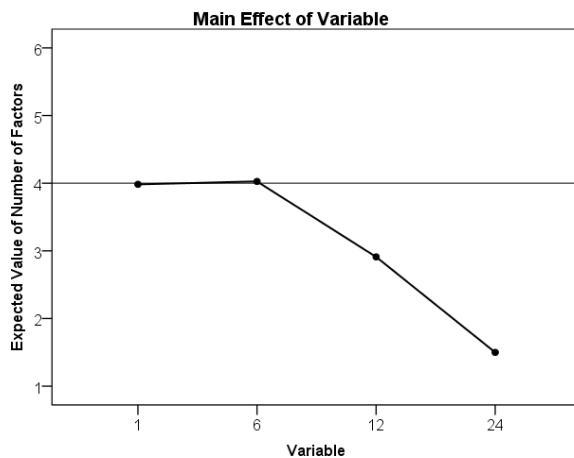
**Extracted by the PA Approach**

Model	Sum Squares	Eta-square	Percentage of R-Square
Intercept	46270.710		
<b>var#</b>	<b>5082.966</b>	<b>0.804</b>	<b>87.2</b>
var# * sub# * magnitude	189.020	0.030	3.2
var# * magnitude	178.246	0.028	3.1
sub# * magnitude	154.734	0.024	2.7
var# * sub#	149.207	0.024	2.6
Magnitude	46.372	0.007	0.8
sub#	31.735	0.005	0.5
Error	488.010		
Total	52591.000		
Corrected Total	6320.290		

Note. R-Square = .92. In terms of notation, *var#* denotes the number of variables with outliers, *sub#* denotes the number of subjects with outliers, and *magnitude* denotes the magnitude of the outliers.

**Figure 3.3 Graphs for Main Effect of Variables with Outliers on the Number of Factors**

**Extracted by PA Approach**



### **3.4 Study 3: A Focused Simulation Study to Demonstrate the Effects of Outliers on the Correlation Matrix**

The simulation methods used herein, as well as by Yuan et al. (2002), were meant to demonstrate the effects of outliers and not necessarily provide an explanation for what causes the effects. However, a small example may help us get an understanding for why the various methods performed differently in the presence of outliers. The statistical engine of factor analysis involves the analysis of a correlation matrix; therefore, a focused study of how outliers affect this matrix may provide some explanatory insights. Table 3.6 is comprised of a table of five rows and two columns of correlation matrices— each cell of the table is a correlation matrix. In addition, the far right column lists the eigenvalues resulting from a PCA of the corresponding correlation matrix on the far left— listed in descending order of magnitude.

Focusing on the far left column in the table, the top left cell is the original correlation matrix of the first four variables of the Holzinger and Swineford (1939) data computed from a simulated data set of 10,000 subjects from a multivariate normal distribution. This is the population analogue of that correlation matrix and hence, by definition, allows us to demonstrate the effect of outliers setting aside any influence of sample size. The remaining four rows of the left column of Table 3.6 are the correlation matrices that result from one to all four variables having 15% of the cases with outliers of a magnitude of 2. In short, we simulated data using our simulation methodology with only the number of variables with

outliers allowed to vary.

**Table 3.6 Demonstration of Changes in Correlation Coefficients and Eigenvalues of a 4-Variable data in the Presence of Outliers Comparing to the No-Outlier Condition**

Correlation Matrix					Change of Correlation Coefficients	Eigenvalues
No Outliers						
	v1	v2	v3	v4		1.976
V1	1					.788
v2	.343	1				.680
v3	.298	.224	1			.556
v4	.443	.341	.280	1		
One Variable with Outliers						
	v1	v2	v3	v4		1.753
v1	1					.829
v2	.210	1				.775
v3	.173	.224	1			.643
v4	.262	.341	.280	1		
Two Variables with Outliers						
	v1	v2	v3	v4		1.965
v1	1					1.097
v2	.770	1				.713
v3	.173	.112	1			.224
v4	.262	.176	.280	1		
Three Variables with Outliers						
	v1	v2	v3	v4		2.595
v1	1					.939
v2	.770	1				.243
v3	.748	.767	1			.223
v4	.262	.176	.146	1		
Four Variables with Outliers						
	v1	v2	v3	v4		3.039
v1	1					.501
v2	.766	1				.232
v3	.762	.768	1			.229
v4	.626	.584	.551	1		

The original correlation matrix, which is on the top of the table, has coefficients

ranging from .224 to .443. The middle column of Table 3.6 are the matrices that result from comparing the corresponding matrix with outliers (to the left of each cell) with the original correlation matrix at the top left cell in the table. Therefore, the matrices in the middle column reflect the change in the correlation matrix due to the outliers. Dashed lines are used to indicate which variables had outliers.

In the case of one variable with outliers, the first variable, the correlations between the first variable and other variables decreased (-0.133, -0.126 and -0.181) and all other coefficients remained the same (zeros). Focusing on the third row of the table, when the first two variables were manipulated with outliers, the correlation between these two variables increased by .427 (increased from .343 to .770), whereas the correlations between these two variables and the others were deflated (ranging from -0.121 to -0.181), and the correlation between the variables without outliers was, as expected, the same as the original. The cell in the third row shows the inflation, deflation, and no effect depending on the number of variables with outliers. When three variables had outliers, the correlation matrix showed either inflation or deflation. When all the variables had outliers, the whole correlation matrix was inflated with coefficient changes ranging from .183 to .544.

It should be noted that, in general, the inflation in the magnitude of correlations of some subset of the variables, when substantial, will, by definition, create an additional factor. Interestingly, if all of the correlations are inflated this creates one dominant factor – in essence enhancing the salience of the dominant factor. One can get a sense of how the number of variables with outliers affects the number of factors extracted by the PCA-based

methods by focusing on the far right column of Table 3.6, the eigenvalues. In short, one can see that an additional factor would be retained, based on the K-G method, when there were two variables with outliers; however, in the case of more than two variables, the first factor was made more salient by the extra-correlation induced by the outliers. This demonstrates the corresponding effects found in Studies 1 and 2 for the K-G and MAP methods. It is difficult to demonstrate the effects on PA because it is sample size dependent; however, with very large sample sizes (i.e., the asymptotic case), PA is the same as K-G.

Unlike the PCA-based approaches, the inflation and deflation of the correlation coefficients when a subset of the variables had outliers did not have a differential effect on the chi-square method. Outliers caused inflation in the number of factors extracted by the sequential chi-square (ML) method and then eventually non-convergence in the extreme outlier condition wherein all variables had outliers. In our studies, as a result of the type of outliers being simulated, both the skewness and kurtosis of the observed variables were greatly inflated by inducing outliers. That is, for our simulated data, the skewness and kurtosis went from nearly zero in the original data to 1.7 and 2.9, respectively, in the extreme case.

As suggested by Boomsma (1983) and Browne (1984), the inflation in skewness and kurtosis would lead to an inflation of the type I error of chi-square tests and correspondingly lead to too many factors being decided by chi-square tests. As for the appearance of non-convergence in the extreme outlier conditions, it happened because of a combination of two sources— Heywood case and failure to find local minimum of the empirical likelihood

solution. A review of the error messages from running ML estimation in our first two studies indicated that, during iterations, one or more communality estimates were greater than one; that is, negative residual variances appeared. Similar to our findings, Bollen (1989, p. 282) showed that outliers can create a Heywood case and result in a negative residual variance. Hoyle and Duvall (2004) also suggested that Heywood cases often result from over-extraction (and sometimes under-extraction) when using maximum likelihood (ML) estimation in factor analysis.

### **3.5 Discussion**

Studies 1 and 2 complement each other in that the first has high fidelity with real data situations because it used real data to mimic errors that mainly occur during data collection or errors in preparing data for analysis (e.g., data entry errors)— that is, Liu and Zumbo's (2007) first category of outlier sources. These types of outliers are illegitimate observations and in practice should, when identified, be corrected or removed. The second study was, in essence, a replication of the first with an eye toward repeated sampling from a population with a known number of factors. This second study also had high fidelity because the population analogue to the number of factors was based on the real data from Study 1.

Our findings, like Yuan et al. (2002), should not be generalized beyond Liu and Zumbo's (2007) first category of outlier sources— that is, errors that occur during data collection and or data entry. Future research should investigate Liu and Zumbo's second and third outlier sources using a probabilistic mixture of probability distributions to reflect a

subgroup of individuals for whom the measure operates differently than for the target population due to unpredictable measurement-related errors, or unknowingly recruiting some individuals who are not members of the target population.

In terms of situating our findings in the theoretical psychometric and statistical literature, our findings supported both (a) the Bollen-Arminger-Huber's conjecture that outliers would result in extra factors, and (b) the Yuan-Marshall-Bentler conclusion that outliers reduced the number of factors. That is, outliers inflated, deflated, or had no effect on the number of factors extracted depending on which decision method was used and the extent of outlier contamination. Applied researchers, however, can be assured that when reading the extant measurement literature or when analyzing their own data that the decisions about the number of factors may be seriously distorted by outliers. For instance, our second study showed that, in some cases, the number of factors was reduced from four to one if using the K-G rule or PA approach, but increased to as many as ten if using the sequential chi-square test. Of course, in some situations outliers had no effect on the decision about the number of factors. However, day-to-day practicing researchers are left with the dis-ease of not knowing, in the context of their own data, which of the situations they may be in: the benign case or the many possible cases of either inflation, deflation of the number of factors, or simply having non-convergence due to outliers.

In terms of day-to-day research practice, like our earlier research on the effect of outliers on the estimation of coefficient alpha (Liu et al., 2010; Liu & Zumbo, 2007), this study is a call for day-to-day researchers to use robust estimators and robust methods, such as



those proposed by Yuan and his colleagues (e.g., Yuan & Bentler, 2007; Yuan & Zhong, 2008) and by Pison et al. (2003) when determining the number of factors to extract in their EFA studies. The number of factors that one extracts for an EFA is sensitive to outliers and can unduly influence theory development and psychometric practice.

## **Chapter 4: Effects of Outliers Arising from an Unintended and Unknowingly Included Subpopulation**

### **4.1 Introduction**

Exploratory factor analysis (EFA) is a widely used statistical technique in the psychosocial, behavioral and health sciences. However, the matter of the impact of outliers on the decisions about the number of factors to retain is largely undocumented in the psychometric and methodological literature. Recently, Liu, Zumbo and Wu (in press) demonstrated the impact of outliers on the decision about the number of factors. Their study focused on outliers that are errors in the data – that is, Liu and Zumbo’s (2007) first category of outlier sources. Liu et al. found that this type of outliers inflated, deflated, or had no effect on the number of factors retained depending on the extent of outlier contamination and which decision method (e.g., parallel analysis, Kaiser-Guttman’s eigenvalues-greater-than-one, minimum average partial, or sequential chi-square tests) was used.

The purpose of the present paper is to continue this line of research and investigate Liu and Zumbo’s (2007) second and third categories of outliers using a probabilistic mixture of distributions. With this purpose in mind, we first discuss the connection between the various sources of outliers and the models used for simulating them in psychometric studies. Next, two studies are reported. The first study demonstrates the impact of these types of outliers on the decision about the number of factors to retain. The second study is a follow-up to the first study, including a focused simulation study of a small correlation matrix and a

report on the skewness and kurtosis of variables that were simulated using the same simulation design as in study 1. The second study provides insight into how the outliers altered the properties of the correlation matrix. Readers interested in a review of the literature as well as a discussion of the need to study outliers in terms of deciding on the number of factors to retain should see Liu et al. (in press).

As Zumbo and Zimmerman (1993) state, computer simulation (including Monte Carlo simulation) is an empirical method of experimental mathematics that is loosely defined as the mimicking of the rules of a model (in our case, a psychological or psychometric phenomenon) via random processes. The key concept in this definition is the correspondence of the psychometric or psychological process to how it is being mimicked in the simulation. For example, in the case of the study of outliers, it is important that the simulation method matches the source of outliers being considered. Below, I briefly review a taxonomy of sources of outliers and then address the simulation methods that can be used to mimic these outliers.

#### 4.1.1 Sources of Outlier Contamination

In terms of psychometric analysis, Liu and Zumbo (2007) described three categories of possible sources of outliers in item responses – that is, a univariate distribution of item responses. As noted above, the first category usually refers to “errors” in the data; instantiations of which include errors that occur during data collection, data recording, or data entry. The outliers generated from such sources are obviously illegitimate observations

and should, when found, be corrected. This first category of outliers arise from mistakes and are hence specific to a particular data set, so that they are a property of a sample but not of a population. For example, typically it does not make sense to talk about the number of typographical data entry errors in a population; however, it does make sense to talk about the number of such typos in a sample. Because of this characteristic, this type of outliers distinguishes itself from the other two categories of outliers in terms of the outlier generation models used in methodological and simulation studies – i.e., deterministic or slippage simulation models.

The second category of outliers refers to the unpredictable measurement-related errors from participants, including guessing and inattentiveness during item responding which may be caused by fatigue or participants' lack of interest in participation. Another example of this category includes item misresponding which happens when, for example, participants misunderstand the instructions or the descriptors on the response scale (e.g., Barnette, 1999). Unlike the first category, which are clearly errors and sample specific, depending on the particular psychological processes in item responding, the second category of outliers may be considered either (sample specific) errors or characteristics and propensities of respondents and hence a population characteristic. For example, misunderstanding the item response instructions may be due to something that reflects momentary inattention or an inherent inattentiveness by the respondent. The former is a sample specific error (and hence akin to an error of the kind in the first category) whereas the latter is, by definition, a characteristic of respondents and hence may reflect a sub-group of

inattentive respondents in the population of possible item respondents.

Liu and Zumbo's (2007) third category of outliers occurs when researchers unknowingly recruit some individuals who are not members of the target population, resulting in a sub-population for whom the measure operates differently than for the target population. Liu and Zumbo described an example of this in the context of self-concept research conducted with a student population wherein some study participants are from Asian countries for which self-concept may be a different construct. There are many examples of this sort of problem as evidenced by the growing number of papers on construct comparability and test adaptation (e.g., Hambleton, Merenda, & Spielberger, 2005).

Outliers from the third category as well as the second category (when they are a characteristic of respondents) reflect an unintended and unknowingly included (here forth referred to as "unintended") subgroup in one's target population, which are usually simulated via probability models. Although they represent different psychological phenomena, these outliers behave the same mathematically and hence can be simulated by the same outlier generation models.

#### 4.1.2 Models Used in Simulation Studies

In the statistical literature, one sees reference to three common models for simulating outliers: deterministic, slippage, and mixture models. Deterministic and slippage models are typically used for the first category and for sample-specific errors in the second category, whereas mixture models are typically used for the second category of outliers that are a

characteristic of respondents (and, hence, a population characteristic) and for the third category of outliers. Whether it is the second or third categories, the mixture model is used to mimic unintended sub-populations. It should be noted, however, that the slippage model can, in particular instances, also be used to model unintended sub-populations except that the number of outliers, in this case, would be fixed from replication to replication.

*Deterministic Model.* The first category of outliers, errors in the data, has been simulated using a deterministic model (Barnett & Lewis, 1994). Because this type of outliers is sample specific, the number of outliers is fixed for a sample and rejection of the null hypothesis of no outliers is deterministically correct as these outliers are obviously different from the majority of observations (Barnett & Lewis, 1994). One way to simulate outliers using a deterministic model is simply to alter the original data, by either multiplying or adding a constant to raw scores. Examples of this type abound in the literature and include EFA studies by Yuan, Marshall, and Bentler (2002) and Study 1 of Liu et al. (in press). In both examples, outliers were created by multiplying raw scores of one or more variables by a constant (2, 3, 4, or 5) for a certain proportion of subjects in a sample.

*Slippage Model.* Another common strategy of simulating outliers as errors in the data (i.e., the first category of outliers) is the slippage model. Like the deterministic model, the number of outliers in a sample is fixed from replication to replication in a simulation study; however, with the slippage model, these outliers arise from some probability distribution.

The slippage model has been widely discussed and used in the literature (e.g., Anscombe, 1960; Barnett & Lewis, 1978; Dixon, 1950; Liu, et al., in press). In its general

form, the null model (without outliers) is

$$H : x_j \in F(j = 1, 2, \dots, n)$$

The alternative model is

$$\bar{H} : x_i \in F(i = 1, 2, \dots, I), \quad x_p \in G(p = 1, 2, \dots, P),$$

where  $I + P = n$ ;  $F$  denotes an target distribution (sometimes called parent distribution);  $G$  denotes a contamination distribution with a different mean and/or variance;  $n$  denotes the total number of observations in a sample;  $I$  is the number of observations from a target distribution; and  $P$  is the number of observations from a contamination distribution. In the null model, all observations are assumed to come from the same population distribution. In the alternative model, a small number of observations are assumed to come from a contamination distribution and the total number of observations in a sample is the sum of observations from a target distribution and from a contamination distribution (Balakrishnan & Childs, 2001).

*Mixture of Distributions.* One can think of this model intuitively as mixing two different population distributions together. A psychological example may help make this concrete: people from Denmark typically rate their life satisfaction as much higher than people from Hungary (OECD, 2005). If we targeted people in Denmark for an investigation, but unknowingly also recruited a small group of people who just immigrated to Denmark from Hungary, the observations recruited are from a mixture of two populations and responses from Hungarian people might appear as outliers. To mimic this kind of outliers,

one would use a *mixture of distributions*<sup>5</sup>, which has been widely used in the research literature and also is utilized in the present research.

A mixture of two distributions is a general model, comprising two weighted probability distributions with positive weights that sum up to one (Blischke, 1978). As the weights represent a probability distribution, the mixture is also a probability distribution. The two distributions thus mixed, depending on the parameter values for the mixing, representing different populations. These components of a mixture of distributions can be normal distributions or non-normal distributions (e.g., Poisson, Negative Binomial distributions). In the statistical and psychometric research literatures, a mixture of two normal distributions has been frequently used for simulating outliers. One of the most well-known and widely used mathematical models is the mixture contamination model -- also referred to as the mixed normal distribution, which was introduced by Tukey (1962) and later extended by Huber (1964), Mosteller and Tukey (1968), and Barnett and Lewis (1994). This mixture contamination model is the one used herein. It is generated by including two normal distributions, a target distribution with mean  $\mu$  and standard deviation  $\sigma$ ,  $N(\mu, \sigma)$ , denoted by  $F$ , and a contamination distribution with some values of mean and/or standard deviation different from  $F$ , denoted by  $G$ .

Given a sample of  $n$  independent observations,  $X_i$  ( $i = 1, 2, \dots, n$ ), the majority of the data points follow the target distribution  $F$  and the proportion of the sample is denoted by  $1-p$ ,

---

<sup>5</sup> Although we do not discuss them in detail herein, “mixture of statistical models” can also be used in this context. It involves situations in which the observed data are a mixture of two statistical models that are different (e.g., one is a one-dimensional and the other a two-dimensional factor analysis model). These sorts of structured populations may be characterized as multi-group factor analysis in the psychometric literature.



while a small fraction,  $p$ , follows the contamination distribution  $G$ . The mixed contamination model is a mixture of  $F$  and  $G$ . The null model is

$$H : x_j \in F(x)$$

The alternative model is

$$\bar{H} : x_j \in (1 - p) * F(x) + p * G(x), 0 < p < 1/2 \quad (1)$$

where the amount of contamination/outliers  $p$  must be less than one half, and often substantially less, which indicates the probability that an observation arises from a contamination distribution  $G$ . If the amount of outliers is as large as near half, any outlier treatment methods, such as robust methods, are not legitimate to apply to these outliers in practice and the outliers should be modeled as another population. It is important to note that, in a simulation study with, for example, 100 replications, the proportion of the sample from the  $G$  distribution is itself a random variable whose average over the 100 replications (i.e., the expected value) is the proportion  $p$  – i.e., the proportion of outliers varies from sample to sample, however, on average it will be  $p$ . In the results section, the varying proportion of outliers from sample to sample was shown in the description of our simulation method (Table 4.1).

Slippage models and the mixture contamination model share some similarities, but have some fundamental differences. Barnett and Lewis (1994) pointed out that the number of outliers is fixed in a slippage model and outliers are regarded as fixed contamination whereas the number of outliers is a random variable in a mixture contamination model and hence outliers are regarded as random contamination. It should be noted that it is not appropriate to

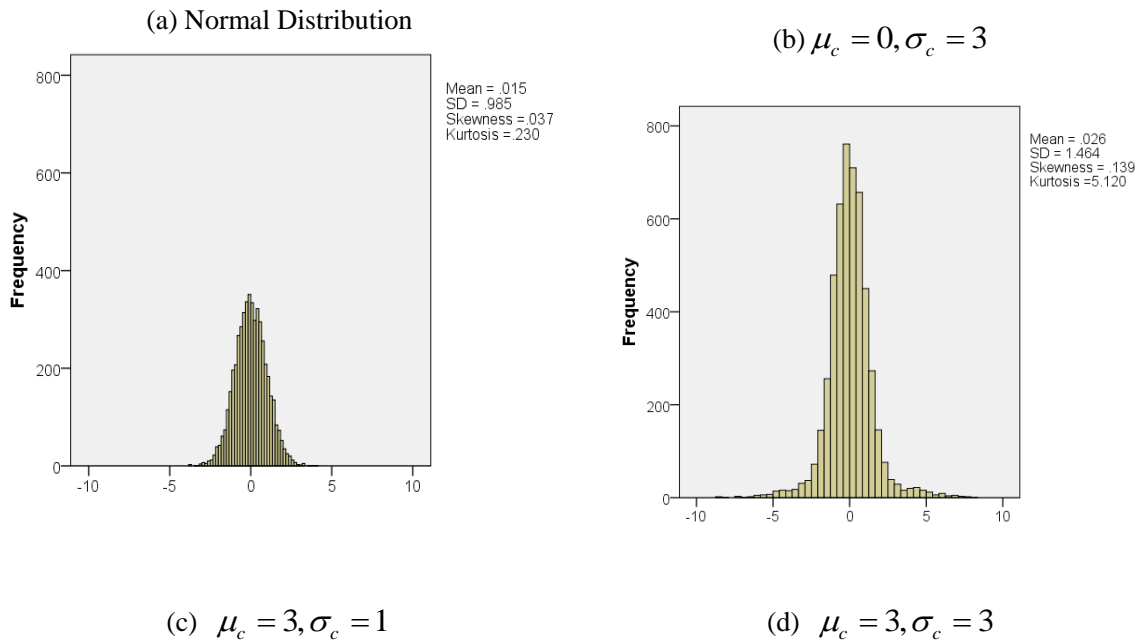
use a mixture of distributions model to simulate outliers from Liu and Zumbo's (2007) first category, but is more appropriate to simulate outliers from their second and third categories. As the first category of outliers is obvious (typographical) errors and sample specific, the randomness of mixture of distributions models does not fit into the fixed property of outliers from the first category. However, slippage models can be used for this kind of outliers because the number of outliers is fixed in each sample.

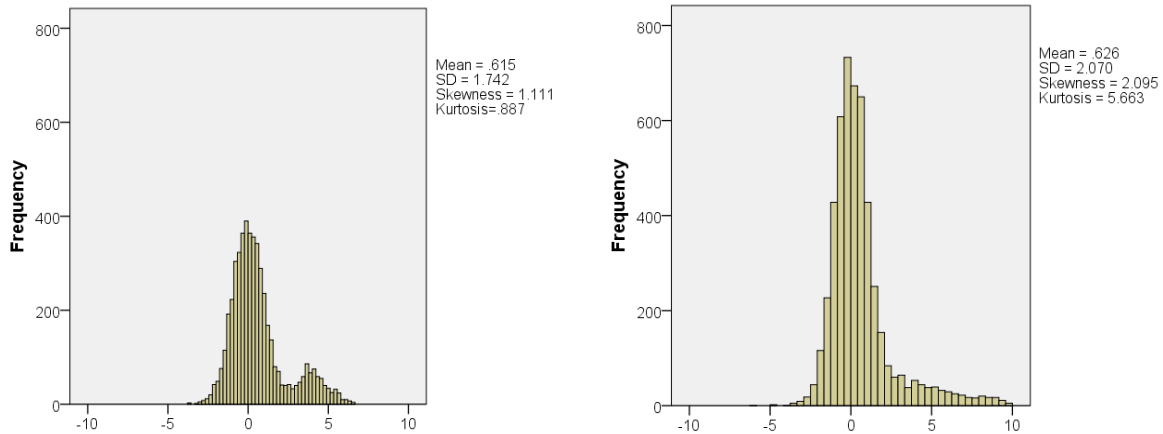
There are two contamination conditions: symmetric and asymmetric contamination. The contamination is symmetric if the population is a mixture of  $N(\mu, \sigma)$  and  $N(\mu, b\sigma)$  where  $b$  is a positive constant greater than one and hence can generate a contamination distribution with a larger standard deviation (SD) than the parent distribution, which is called SD shift in this paper. It is worth noting that if  $b$  is less than one, the condition of inliers should be considered instead of outliers, which is not of interest of the present study. The contamination is asymmetric when the population is a mixture of  $N(\mu, \sigma)$  and  $N(\mu + a, \sigma)$  or  $N(\mu, \sigma)$  and  $N(\mu + a, b\sigma)$  where  $a$  is a constant and  $a \neq 0$ . The mean and standard deviation of  $F$  are usually defined as 0 and 1, respectively, that is  $N(0,1)$ , so adding or subtracting any value to zero will result in the mean shift of a contamination distribution from the center of the population distribution, and hence lead to the asymmetric contamination. Therefore, a variety of contamination conditions can be generated by increasing the three outlier factors, that is, the proportion of contamination, mean shift, and SD shift of the contamination distribution.

An example of outliers in a mixed contamination model is given in Figure 4.1. Figure 4.1a is a normal distribution,  $N(\mu = 0, \sigma = 1)$ . Figure 4.1b presents a case of symmetric

outliers with 15% of outliers, consisting of a parent distribution  $N(\mu = 0, \sigma = 1)$  and a contamination distribution  $N(\mu = 0, \sigma = 3)$ . Outliers are shown as long and heavy tails at each side of the distribution. Figure 4.1c demonstrates a case of asymmetric outliers with 15% of outliers, consisting of a parent distribution  $N(\mu = 0, \sigma = 1)$  and a contamination distribution  $N(\mu = 3, \sigma = 1)$ . Outliers are shown as a heavy tail on one side of the distribution. Figure 4.1d shows another case of asymmetric outliers with a parent distribution  $N(\mu = 0, \sigma = 1)$  and a contamination distribution  $N(\mu = 3, \sigma = 3)$ . Outliers make the distribution have a heavy tail as well as a high peak.

**Figure 4.1 An Example of Symmetric and Asymmetric Outliers (Proportion of Contamination=0.15)**





Building on the findings of Liu et al. (in press), the purpose of the present research was to investigate how outliers, arising from an unintended and unknowingly recruited subpopulation (Liu and Zumbo's second and third categories of outliers), affected the decisions about the number of factors to retain using four commonly used methods, that is, parallel analysis (PA), Kaiser-Guttman's eigenvalues-greater-than-one (K-G), minimum average partial (MAP), or sequential chi-square tests based on maximum likelihood estimation ( $\chi^2_{ML}$ ). The results of a Monte Carlo simulation study were reported first, in which the outlier conditions were manipulated using five factors (i.e., mean shift, SD shift, proportions of contamination of subjects, number of variables with outliers, and sample size). A follow-up study was also presented in order to provide insight into potential causes of our findings.

## **4.2 Study 1: Investigating the Effects of Outliers from a Subpopulation Using the Mixture Contamination Model**

### **4.2.1 Method**

#### *Study Design*

A Monte Carlo simulation study was used to investigate the effects of outliers on decisions about the number of factors by the four decision methods. This study systematically varied five factors with 100 replications for each outlier condition (i.e., simulation condition). These five factors are as follows:

- (a) mean shift of a contamination distribution (0, 1.5, 3)
- (b) standard deviation (SD) shift of a contamination distribution (1, 1.5, 3)
- (c) proportion of contamination (i.e., proportion of the subjects from the contamination distribution) (.01, .08, .15)
- (d) sample size (250, 500, 1000), and
- (e) number of variables with outliers (1, 6, 12, 24).

The study design is therefore a 3x3x3x3x4 completely crossed factorial design with 324 conditions, which also includes the no-outlier conditions (i.e., the comparison condition) that has mean shift of zero and SD shift of one.

To ensure a systematic investigation of outlier effects, the selection of the magnitude of three factors (mean shift, SD shift and proportion of contamination), which are the parameters of a typical mixture contamination model, were guided by previous studies, Blair

and Higgins (1981), Liu and Zumbo (2007), Mosteller and Tukey (1968), and Zumbo and Jennings (2002). Following these studies, the present study adopted similar values of model parameters with some modifications to fit the purpose of the present study. The number of variables with outliers was also included in the present study as it was demonstrated to be an influential factor in determining the number of factors in Liu et al.'s (in press) study. In addition, sample size was found in the literature to affect the performance of the K-G rule as well as chi-square tests (e.g., Gorsuch, 1983, p.164; Hubbard & Allen, 1987; Zwick & Welicer, 1986). Hence, we included samples size as a factor in the present study.

#### *Data Generation*

In line with the earlier work by Liu and Zumbo (2007), Liu et al. (in press) and the psychometric context of our study, the outliers are induced in the item responses, that is, the marginal distributions. The item response format being simulated applies to continuous variables, such as those generated from subscale scores and visual analogue responses.

For the mixture contamination model in equation (1), both the target and contamination data were generated based on the population correlation matrix from Holzinger and Swineford's (1939) classic data set. The original data set consists of 24 psychological ability test scores from 301 junior high school students with a four-factor solution recommended by many researchers (e.g., Gorsuch, 1983; Harman, 1976; Liu, et al., in press). As in Liu et al.'s studies, a four-factor solution based on maximum likelihood exploratory factor analysis was obtained using Holzinger and Swineford's data. The resulting reproduced correlation matrix (i.e., the implied correlation matrix with "1s" on the diagonal, rather than the reproduced

communalities) was used as the population correlation matrix in the simulation to generate multivariate normal datasets with specified marginal means and standard deviations, depending on the experimental condition, that correspond to the target or contamination distribution in equation (1). Multivariate normal data were generated in software R 2.12.1, using a method akin to the Kaiser and Dickman (1962) method wherein, for computational efficiency, we used Cholesky decomposition rather than principal components analysis in the computation. Generating data from a model with a known (pre-specified) number of factors allowed us to compare the number of factors obtained from different outlier conditions to a common criterion in the population: four factors.

### *Outcome Variable*

In each of the 324 experimental conditions, and for each of the 100 replications, the number of factors to retain for the EFA was determined, separately, by the K-G rule, PA, MAP and sequential  $\chi^2_{ML}$  tests. The number of factors retained is the dependent variable for each of these four methods respectively in this simulation study. It should be noted that, as in Liu et al.'s (in press) study 2, for each outlier design condition, an average of the number of factors over 100 replications was obtained and hence the number of factors reported might not be a whole number.

### *Analysis of the Simulation Results*

Following the data analysis strategy used in Liu and Zumbo (2007) and Liu et al. (in press) five-way ANOVAs (3x3x3x3x4) were conducted with the number of factors retained as the dependent variable separately for each of the four decision methods, that is, the K-G rule, PA, MAP and  $\chi^2_{ML}$  sequential test. Given the large sample size (32400-- i.e., 324 cells in design with 100 replications per cell), we used eta-square ( $\eta^2$ ) to orthogonally partition the explained variance obtained from the fixed effect ANOVA models instead of looking at the statistical significance. The proportion of explained variance was used to aid our interpretation of the simulation results, which is like R-square in regression analysis.

Following Liu et al. (in press), we used Ferguson's (2009) minimum effect size of  $\eta^2$  of 0.04 as the criterion to judge the importance of the main effects and interactions. In addition, if interactions appeared in the model, only higher order interactions were interpreted because main effects and lower order interactions are not interpretable in the presence of higher order



interactions.

The sequential  $\chi^2_{ML}$  test can result in no decision about the number of factors to retain because of non-convergence. Therefore, the non-convergence problem can result in unbalanced data for the ANOVA, which can, in turn, distort the orthogonal partition of variance in the outcome variable. We therefore adopted the Type III sum-of-squares method in SPSS, an often-used method for handling unbalanced data with no missing cells (SPSS Inc., 2009), for the data analysis of the simulation results.

#### 4.2.2 Results

##### *Proportion of outliers in a given sample*

As noted earlier, when using the mixture contamination model to simulate outliers, the proportion of outliers in a sample can vary across replications – that is, from sample to sample. To our knowledge, the central tendency and variability in sample-to-sample proportions of outliers has not been documented in simulation studies. To better understand these statistics, we recorded the proportion of outliers across 100 replications for a single variable.

Table 4.1 lists the central tendency (mean, median) and variability (standard deviation, quartiles, minimum and maximum values) for the proportion of outliers for the various conditions in the current simulation study across the 100 replications. Starting from the far left in Table 4.1, one can find the population value of the proportion of contamination, the sample size, and then the seven descriptive statistics computed across the 100 replications.

One can see that, as expected, the mean is equal to the population value of contamination in every case. However, also as expected, there is variability in the proportion of contamination across the samples which depends on the sample size and the population proportion of contamination.

**Table 4.1 Documenting Simulations Using Mixture of Distributions: Proportion of Outliers in Each Sample across 100 Replications**

Pc	n	Mean	SD	Minimum	1st Quartile	Median	3rd Quartile	Maximum
0.010	1000	0.010	0.003	0.003	0.008	0.010	0.012	0.018
	500	0.010	0.004	0.000	0.008	0.010	0.012	0.022
	250	0.010	0.007	0.000	0.004	0.008	0.012	0.028
0.080	1000	0.080	0.010	0.060	0.073	0.080	0.088	0.106
	500	0.080	0.011	0.054	0.072	0.082	0.088	0.106
	250	0.080	0.019	0.040	0.068	0.076	0.088	0.128
0.150	1000	0.150	0.010	0.121	0.143	0.150	0.156	0.178
	500	0.150	0.015	0.108	0.140	0.149	0.162	0.184
	250	0.150	0.023	0.104	0.132	0.150	0.168	0.200

Note. Pc denotes proportion of contamination in the population; n denotes sample size; SD denotes standard deviation.

#### *Results of the simulation study*

Tables 4.2 to 4.5 present the results for the four decision methods (K-G, MAP, PA and sequential  $\chi^2_{ML}$  tests), respectively. Figures 4.2 to 4.5 show the corresponding highest order interactions for these four methods identified as important factors using  $\eta^2$ . For the K-G, MAP and PA methods, the highest order interaction was the same: *mean shift by the number of variables having outliers by the proportion of contamination*. Hence, we only interpreted

this 3-way interaction for the K-G, MAP and PA methods and not any of the lower order interactions and main effects.

Table 4.2 presents the results of the variance decomposition for the K-G rule and Figure 4.2 shows the corresponding plot of the 3-way interaction. With a mean shift of zero (i.e., symmetric outliers), the number of factors was not affected by outliers, which was also the case for the MAP and PA methods. With mean shift of 1.5 and 3 (i.e., asymmetric outliers), the change in the number of factors depended on the number of variables having outliers and the proportion of contamination. When mean shift was 1.5, the number of factors was not affected when one variable and all variables (24) had outliers, but was inflated (from 4 up to 5 factors) when 6 and 12 variables had outliers. With a mean shift of 3, the number of factors was not affected when only one variable had outliers, was inflated (from 4 up to 5 factors) when 6 and 12 variables had outliers, but deflated when all 24 variables had outliers (from 4 to an average of 2.7 factors). There was more deflation with an increase in the proportion of contamination.

Table 4.3 presents the results of the variance decomposition for the MAP method, with the corresponding plots in Figure 4.3. Figure 4.3 showed that the number of factors was not affected when the mean shift was zero and 1.5, but inflated from four to five when the mean shift increased to 3 for the cases of 6 and 12 variables having outliers. The magnitude of the inflation increased with the increase of the proportion of contamination. It is worth noting that the number of factors retained was not affected when all variables had outliers in the MAP method, which was different from the K-G and PA methods.

Table 4.4 presents the results of the variance decomposition for the PA method and Figure 4.4 is the corresponding interaction plot. Similar to the performance of the K-G and MAP methods, PA method was robust to symmetric outliers. In general, the PA method was accurate in retaining the number of factors in the presence of asymmetric outliers, however, it became dysfunctional when all variables had outliers: (a) the number of factors was deflated slightly when the mean shift was 1.5 and deflated dramatically when the mean shift increased to 3; and (b) the magnitude of deflation increased when the proportion of contamination increased.

Unlike the three PCA-based methods, for the sequential  $\chi^2_{ML}$  test, the highest order interaction identified as important using  $\eta^2$  was found to be: *SD shift by the number of variables having outliers by the proportion of contamination*. Table 4.5 presents the results of the variance decomposition for the sequential  $\chi^2_{ML}$  test and Figure 4.5 shows the corresponding interaction plot. The SD shift played an important role in the interaction, which indicated that the symmetric outliers affected the performance of the sequential  $\chi^2_{ML}$  test. When the SD was one (no shift) or 1.5 (mild increase on variations for the contamination distribution), the number of factors retained was either not affected or inflated by a small magnitude. However, the number of factors retained was inflated dramatically when the SD shift increased to 3, and especially when all variables had outliers, the number of factors retained increased from 4 (baseline) to almost 12.

It should be noted that non-convergence was found for the sequential  $\chi^2_{ML}$  tests, ranging from 1 to 19 percent of the replications in a cell of the experimental design. As

shown in Table 4.6, the non-convergence occurred for experimental conditions wherein the SD shift was 3, which we saw in Figure 4.1 involved high kurtosis for the variable with outliers. Within the conditions involving an SD shift of 3, non-convergence happened when the proportion of contamination was either .08 or .15 and either 12 or 24 variables had outliers, and was more likely to happen when all 24 variables had outliers. Sample size also seemed to interact in this finding, wherein the non-convergence occurred predominantly with sample sizes of 1000. Like the findings of Liu et al. (in press), in inspecting the statistical output from the simulation, the non-convergence problem resulted from a combination of a Heywood case and failure to find local minimum of the empirical likelihood solution.

**Table 4.2 Variable Ordering for a Five-way ANOVA on the Number of Factors****Extracted by the K-G Rule**

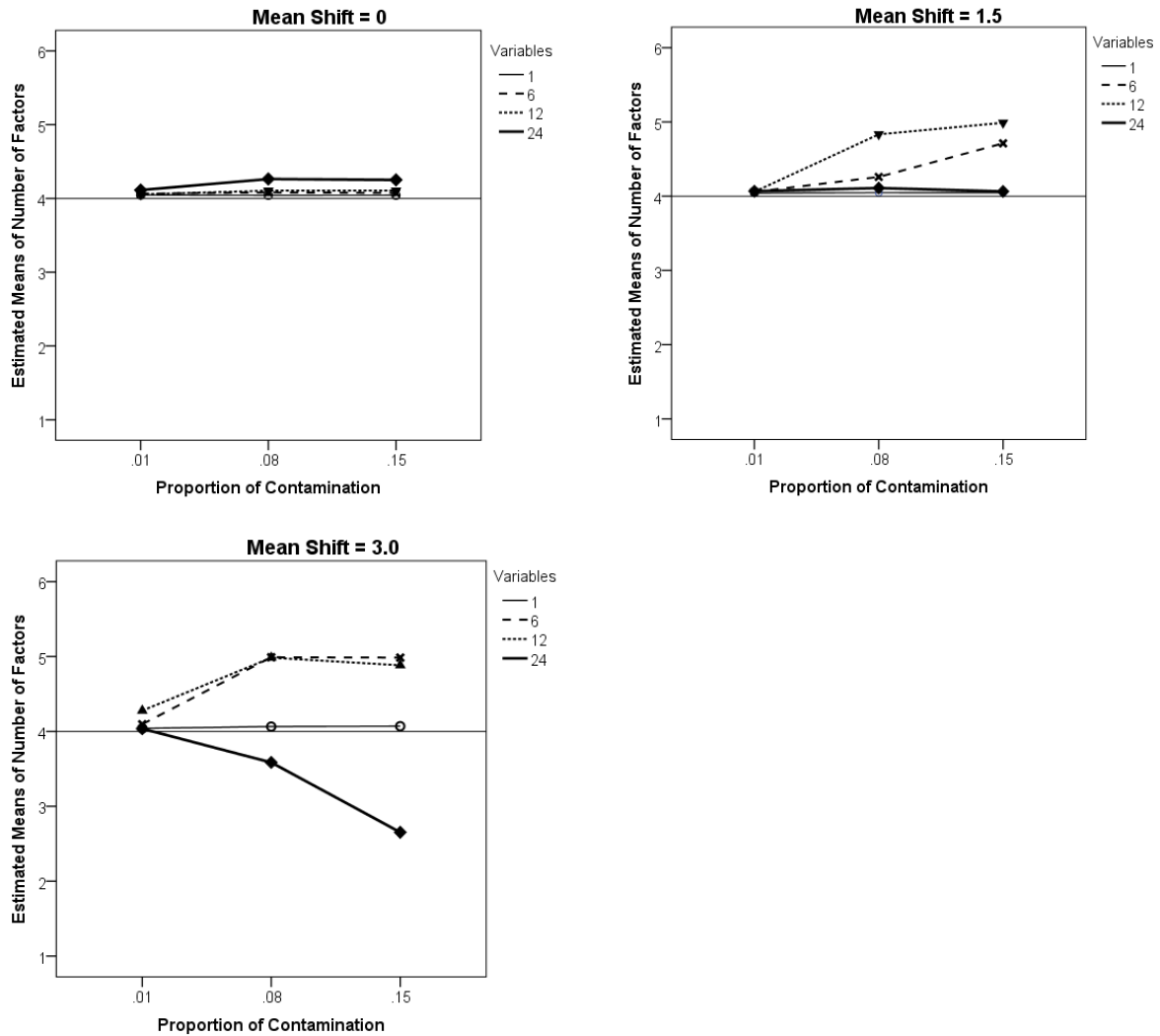
Model	Sum Squares	Eta-square	Percentage of R-Square
<b>vars * mean</b>	<b>1929.155</b>	<b>0.197</b>	<b>25.953</b>
<b>Vars</b>	<b>1742.855</b>	<b>0.178</b>	<b>23.447</b>
<b>pc * vars * mean</b>	<b>930.368</b>	<b>0.095</b>	<b>12.516</b>
<b>pc * vars</b>	<b>887.323</b>	<b>0.090</b>	<b>11.937</b>
Pc	242.723	0.025	3.265
pc * mean	238.286	0.024	3.206
vars * SD	233.350	0.024	3.139
N	232.309	0.024	3.125
SD	169.119	0.017	2.275
Mean	165.230	0.017	2.223
pc * vars * mean * SD	91.428	0.009	1.230
pc * vars * SD	81.851	0.008	1.101
vars * mean * SD	65.865	0.007	0.886
n * SD	56.401	0.006	0.759
n * vars * SD	44.155	0.005	0.594
pc * SD	37.587	0.004	0.506
pc * mean * SD	37.478	0.004	0.504
n * vars * mean	37.212	0.004	0.501
n * vars * mean * SD	30.391	0.003	0.409
n * mean	26.433	0.003	0.356
n * mean * SD	25.753	0.003	0.346
n * pc * vars * mean	24.026	0.002	0.323
n * pc * mean	23.618	0.002	0.318
mean * SD	19.424	0.002	0.261
n * pc * vars	15.457	0.002	0.208
n * vars	14.537	0.001	0.196
n * pc * vars * mean * SD	11.692	0.001	0.157
n * pc * mean * SD	10.674	0.001	0.144
n * pc * vars * SD	3.998	0.000	0.054
n * pc	2.270	0.000	0.031
n * pc * SD	2.220	0.000	0.030
Error	2373.010		
Total	581401.000		
Corrected Total	9806.198		

Note. R-Square = 0.758. ANOVA = analysis of variance; K-G = Kaiser-Guttman rule; mean = mean shift of the contamination distribution; SD = standard deviation shift of the contamination distribution; pc =

proportion of contamination in the population;  $n$  = sample size; vars = number of variables having outliers

**Figure 4.2** Graphs for Three-way Interactions of Variables with Outliers vs. Proportion of Contamination by Three Levels of Mean Shift on the Number of Factors

Extracted by the K-G Rule





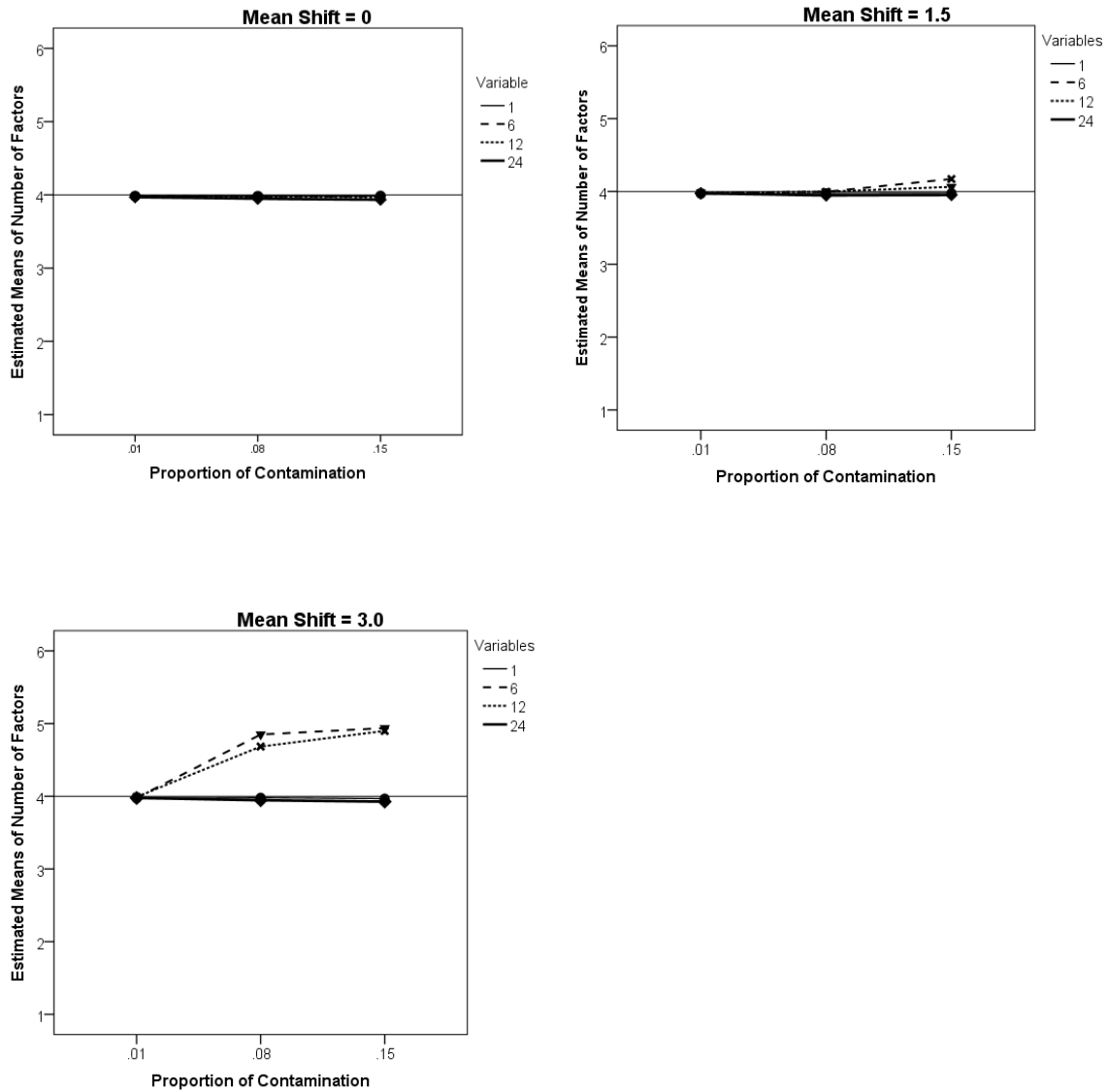
**Table 4.3 Variable Ordering for a Five-way ANOVA on the Number of Factors****Extracted by the MAP Approach**

Model	Sum Squares	Eta-square	Percentage of R-Square
<b>vars * mean</b>	<b>564.060</b>	<b>0.140</b>	<b>21.940</b>
<b>Mean</b>	<b>543.740</b>	<b>0.135</b>	<b>21.149</b>
<b>Vars</b>	<b>414.457</b>	<b>0.103</b>	<b>16.121</b>
<b>pc * vars * mean</b>	<b>292.322</b>	<b>0.072</b>	<b>11.370</b>
<b>pc * mean</b>	<b>273.508</b>	<b>0.068</b>	<b>10.638</b>
<b>pc * vars</b>	<b>204.590</b>	<b>0.051</b>	<b>7.958</b>
<b>Pc</b>	<b>166.922</b>	<b>0.041</b>	<b>6.493</b>
N	28.671	0.007	1.115
SD	14.232	0.004	0.554
pc * vars * mean * SD	6.649	0.002	0.259
vars * SD	5.894	0.001	0.229
pc * SD	5.838	0.001	0.227
vars * mean * SD	4.605	0.001	0.179
n * pc * vars * mean	4.435	0.001	0.173
n * mean	4.411	0.001	0.172
n * vars * mean	4.298	0.001	0.167
pc * mean * SD	4.121	0.001	0.160
n * pc * mean	4.038	0.001	0.157
n * vars * SD	3.618	0.001	0.141
pc * vars * SD	3.613	0.001	0.141
n * pc * vars * SD	2.343	0.001	0.091
n * SD	2.169	0.001	0.084
n * pc * vars * mean * SD	2.156	0.001	0.084
mean * SD	2.028	0.001	0.079
n * vars	1.838	0.000	0.071
n * pc	1.513	0.000	0.059
n * vars * mean * SD	1.390	0.000	0.054
n * pc * vars	1.061	0.000	0.041
n * pc * SD	1.025	0.000	0.040
n * pc * mean * SD	1.020	0.000	0.040
n * mean * SD	0.234	0.000	0.009
Error	1471.610		
Total	542513.000		
Corrected Total	4042.407		

Note. R-Square = 0.636. ANOVA = analysis of variance; MAP = Minimum Average Partial; mean = mean

shift of the contamination distribution; SD = standard deviation shift of contamination distribution; pc = proportion of contamination in the population; n = sample size; vars = number of variables having outliers

**Figure 4.3** Graphs for Three-way Interactions of Variables with Outliers vs. Proportion of Contamination by Three Levels of Mean Shift on the Number of Factors Extracted by MAP Approach



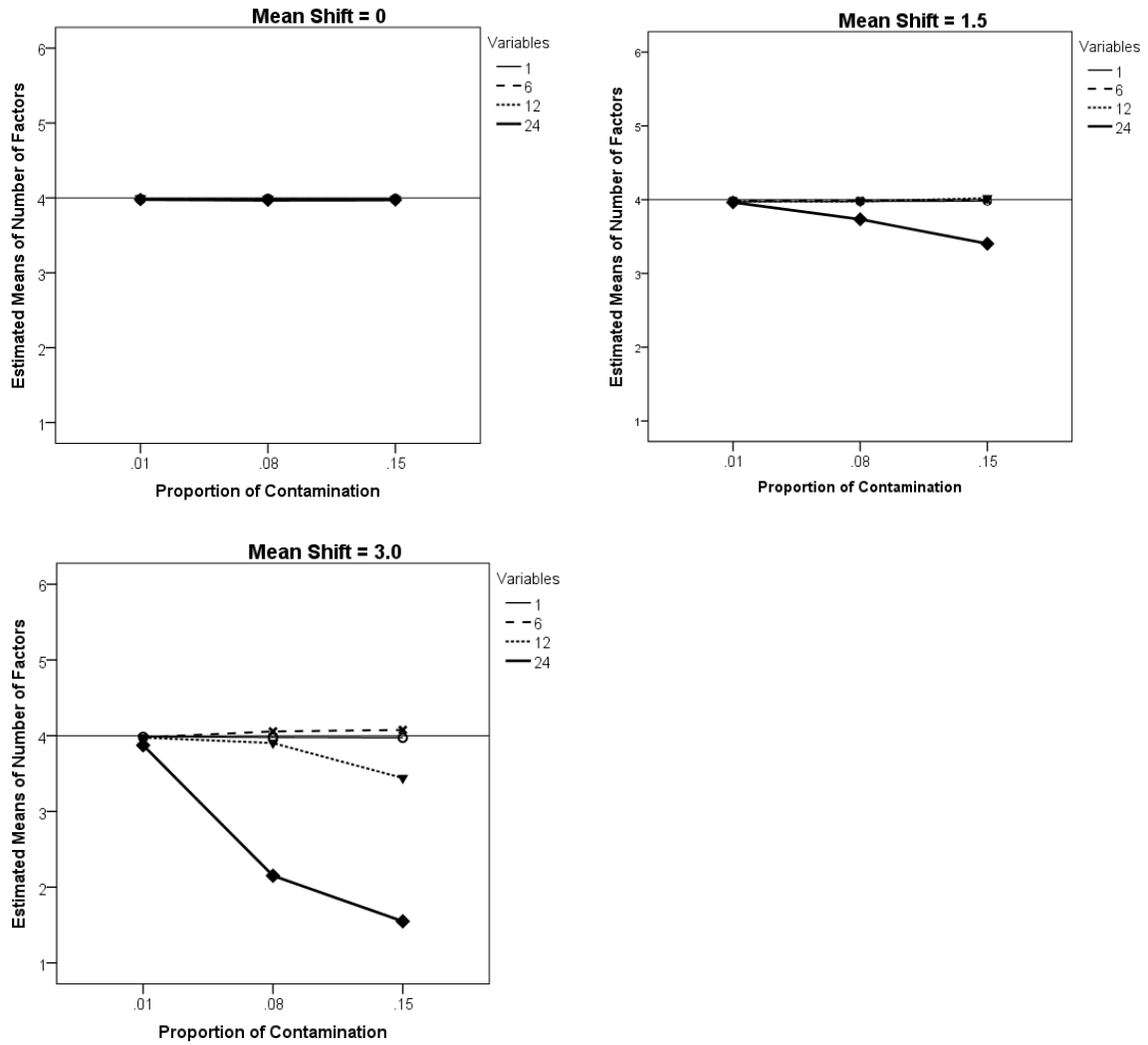
**Table 4.4 Variable Ordering for a Five-way ANOVA on the Number of Factors****Extracted by the PA Approach**

Model	Sum Squares	Eta-square	Percentage of R-Square
<b>vars * mean</b>	<b>2280.285</b>	<b>0.196</b>	<b>23.802</b>
<b>Vars</b>	<b>1985.020</b>	<b>0.171</b>	<b>20.720</b>
<b>pc * vars * mean</b>	<b>1069.485</b>	<b>0.092</b>	<b>11.163</b>
<b>Mean</b>	<b>1006.276</b>	<b>0.087</b>	<b>10.504</b>
<b>pc * vars</b>	<b>940.891</b>	<b>0.081</b>	<b>9.821</b>
<b>pc * mean</b>	<b>498.573</b>	<b>0.043</b>	<b>5.204</b>
Pc	410.573	0.035	4.286
N	234.027	0.020	2.443
n * vars * mean	120.726	0.010	1.260
n * vars	107.233	0.009	1.119
vars * mean * SD	104.751	0.009	1.093
vars * SD	97.287	0.008	1.015
mean * SD	94.225	0.008	0.984
n * pc * vars * mean	84.549	0.007	0.883
SD	80.244	0.007	0.838
n * mean	75.325	0.006	0.786
pc * vars * mean * SD	64.773	0.006	0.676
pc * vars * SD	60.835	0.005	0.635
pc * mean * SD	50.661	0.004	0.529
pc * SD	48.762	0.004	0.509
n * pc * vars	42.428	0.004	0.443
n * pc	41.678	0.004	0.435
n * pc * mean	27.678	0.002	0.289
n * pc * vars * mean * SD	15.749	0.001	0.164
n * pc * mean * SD	10.429	0.001	0.109
n * vars * mean * SD	9.840	0.001	0.103
n * mean * SD	8.414	0.001	0.088
n * pc * vars * SD	7.395	0.001	0.077
n * vars * SD	1.617	0.000	0.017
n * pc * SD	0.393	0.000	0.004
n * SD	0.291	0.000	0.003
Error	2039.620		
Total	485484.000		
Corrected Total	11620.035		

Note. R-Square =0.824. ANOVA=analysis of variance; PA = Parallel Analysis; mean = mean shift of the

contamination distribution; SD=standard deviation shift of contamination distribution; pc =proportion of contamination in the population; n = sample size; vars =number of variables having outliers

**Figure 4.4 Three-way Interactions of Variables with Outliers vs. Proportion of Contamination by Three Levels of Mean Shift on the Number of Factors Extracted by PA Approach**



**Table 4.5 Variable Ordering for a Five-way ANOVA on the Number of Factors****Decided by the Sequential Chi-square (ML) test**

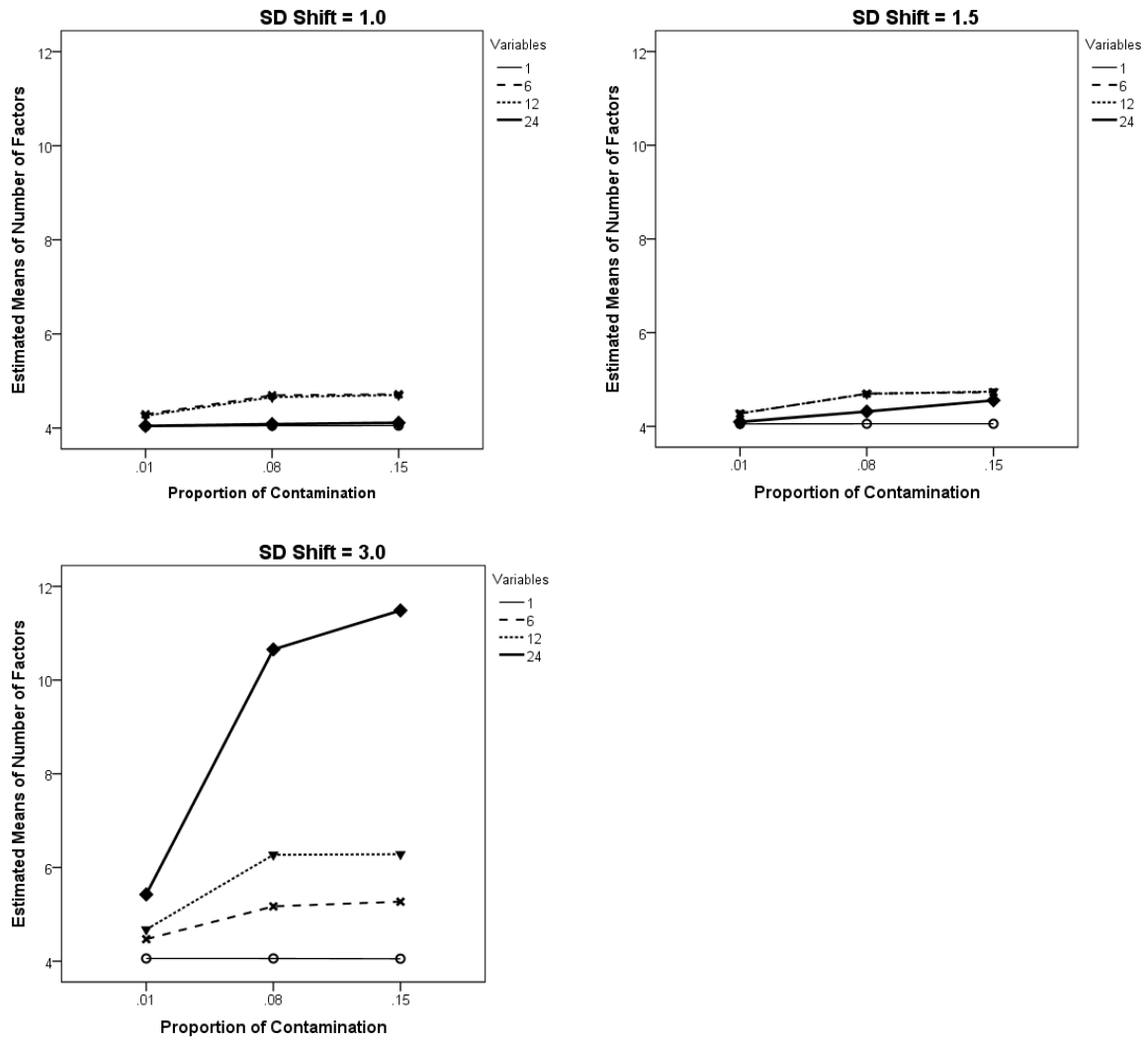
Model	Sum Squares	Eta-square	Percentage of R-Square
<b>vars * SD</b>	<b>27625.757</b>	<b>0.306</b>	<b>32.393</b>
<b>SD</b>	<b>19436.831</b>	<b>0.216</b>	<b>22.791</b>
<b>Vars</b>	<b>13616.464</b>	<b>0.151</b>	<b>15.966</b>
<b>pc * vars * SD</b>	<b>7747.121</b>	<b>0.086</b>	<b>9.084</b>
<b>Pc</b>	<b>5190.159</b>	<b>0.058</b>	<b>6.086</b>
<b>pc * SD</b>	<b>4856.412</b>	<b>0.054</b>	<b>5.694</b>
<b>pc * vars</b>	<b>4018.700</b>	<b>0.045</b>	<b>4.712</b>
Mean	765.987	0.008	0.898
vars * mean	642.238	0.007	0.753
N	214.615	0.002	0.252
n * vars * SD	134.941	0.001	0.158
n * SD	131.215	0.001	0.154
pc * vars * mean	119.643	0.001	0.140
pc * mean	117.121	0.001	0.137
mean * SD	100.804	0.001	0.118
n * vars	99.760	0.001	0.117
n * pc * vars * SD	88.972	0.001	0.104
vars * mean * SD	74.092	0.001	0.087
pc * mean * SD	62.435	0.001	0.073
pc * vars * mean * SD	48.162	0.001	0.056
n * pc * vars * mean	31.810	0.000	0.037
n * pc * mean	29.125	0.000	0.034
n * pc * SD	26.532	0.000	0.031
n * pc	24.936	0.000	0.029
n * pc * vars	22.889	0.000	0.027
n * pc * vars * mean *	18.122	0.000	0.021
SD	10.558	0.000	0.012
n * mean	8.307	0.000	0.010
n * vars * mean * SD	8.049	0.000	0.009
n * mean * SD	7.478	0.000	0.009
n * vars * mean	5.091	0.000	0.006
n * pc * mean * SD	8922.025		
Error	855330.000		
Total	90140.908		
Corrected Total			

Note. R-Square = .946. ANOVA = analysis of variance; mean = mean shift of the contamination distribution; SD = standard deviation shift of the contamination distribution; pc = proportion of contamination in the population; n = sample size; vars = number of variables having outliers



**Figure 4.5 Three-way Interactions of Variables with outliers vs. Proportion of Contamination by Three Levels of Standard Deviation Shift on the Number of Factors Decided by the Sequential  $\chi^2_{ML}$  Test**

Decided by the Sequential  $\chi^2_{ML}$  Test



**Table 4.6 Percentage of Non-convergent Replications with the Sequential Chi-square**

**(ML) Tests**

<b>Pc</b>	<b>Vars</b>	<b>SD Shift of 3</b>		
		<b>Mean Shift</b>		
		<b>.0</b>	<b>1.5</b>	<b>3.0</b>
<b>.08</b>	<b>12</b>	1 (n=250)		
	<b>24</b>	3 (n=500) 8 (n=1000)	4 (n=500) 15 (n=1000)	3 (n=500) 18 (n=1000)
<b>.15</b>	<b>24</b>		1 (n=250)	1 (n=250)
		3 (n=500) 19 (n=1000)	4 (n=500) 17 (n=1000)	10 (n=500) 16 (n=1000)

Note. SD= standard deviation; Pc = proportion of contamination in the population;

Vars = number of variables having outliers

## **Study 2: Demonstrations of Effects of Outliers on Correlation Matrix, and Kurtosis and Skewness of Item Responses**

The purpose of Study 2 was to facilitate our understanding about why these decision methods performed differently in the presence of outliers. As Liu et al. (in press) pointed out; correlation matrices are the engine for the PCA-based methods, and hence are the input data for them. Furthermore, skewness and kurtosis are related to the performance of the  $\chi^2_{ML}$  test in factor analysis (Boomsma, 1983; Browne, 1984). We, therefore, include two demonstrations in Study 2: a small, focused simulation to show how outliers distort the correlation matrix and eigenvalues as well as a demonstration of the change in kurtosis and skewness in the presence of outliers.

### **4.2.3 Demonstration 1**

Researchers usually ignored the effects of outliers on factor analysis partly because they believed that a few outliers should not substantially change the correlation matrix and, as such, a factor analysis should not be affected by outliers. The present small-scale simulation aimed to demonstrate how outliers may distort properties of a correlation matrix. Following Liu et al.'s (in press) study, we also used the original correlation matrix of the first four variables from Holzinger and Swineford's (1939) classic data as the population correlation matrix for simulating multivariate normal data sets. For demonstration purposes, we only included extreme outlier conditions (mean shift=3 and/or SD shift=3) as well as a no-outlier condition with either two or all four variables having outliers. Across all outlier conditions,

the proportion of contamination was 0.15. In Study 1 we did not find sample size effects; therefore, in this demonstration we ruled out this factor and used data sets with 100,000 observations so as to have population analogues.

To examine the change in the correlation matrix under outlier conditions, we used the matrix's condition number to document if it is ill-conditioned and the magnitude of ill-condition -- with larger condition number indicating more ill-conditioned. The advantage of using the condition number is that, when the correlation matrix is close to being singular, we can still obtain a solution, which disguises the problem of being ill-conditioned, but the condition number can reflect if the matrix is ill-conditioned and if the properties of the matrix are distorted. The condition number is a product term,  $cond(A) = \|A\| \times \|A^{-1}\|$ , where  $A$  denotes a correlation matrix,  $\|A\|$  denotes matrix norm,  $A^{-1}$  denotes the inverse of a matrix, and  $\|A^{-1}\|$  denotes the matrix norm for the inverse of a matrix (Watkins, 2010, p. 122).

Given that a correlation matrix is a special case of a square matrix that is symmetric about the major diagonal that contains ones,  $\|A\|$  is the largest eigenvalue of a correlation matrix, and  $\|A^{-1}\|$  is the largest eigenvalue of the inverse of a correlation matrix (Golub & Van Loan, 1993, p. 58).

Table 4.7 presents the results of the simulation with four rows and seven columns. The first column indicates the outlier conditions (i.e., mean shift and SD shift), the second column shows the resulting correlation matrix with only two variables having outliers, the third and fourth columns are the corresponding condition number and eigenvalues, the fifth column shows the resulting correlation matrix with all four variables having outliers, and the sixth

and seventh columns are the corresponding condition number and eigenvalues.

The top row presents the results for the correlation matrix in the no-outlier condition.

The second row shows the results for symmetric outlier condition, with no mean shift and a SD shift of 3. The effects of outliers were not found for either the case of two variables having outliers or that of all variables having outliers. Some of the correlation coefficients were deflated to a small degree when two variables had outliers, and the condition number as well as the magnitude of eigenvalues was not affected by symmetric outliers.

However, there were dramatic changes for the asymmetric outlier condition with mean shift only (mean shift=3, SD shift=1). Echoed in the findings of Liu et al. (in press), we also found that when two variables had outliers, the correlation coefficient for those two variables was inflated, whereas the remaining correlations in the matrix were either deflated when involving combinations of variables with and without outliers or were unchanged when only involving variables without outliers. The complex pattern created an extra factor and resulted in an increase of the condition number from 3.415 (baseline) to 6.481. When all of the variables had outliers, the correlation coefficients were all inflated resulting in the creation of a more dominant (or salient) factor and a large increase in the condition number from 3.415 to 11.363.

For the asymmetric outlier condition with both mean shift and SD shift (mean shift=3, SD shift=3), the effects of outliers were reduced to some degree. Compared to the mean shift only condition (mean shift=3, SD shift=1), the magnitude of inflation in correlation coefficients became smaller; the condition number dropped to some extent, from 6.481 and

11.363 to 4.465 and 7.093; the second eigenvalue for two variables having outliers was not greater than one anymore (i.e., dropped from 1.009 to 0.955); and the magnitude of the largest eigenvalue for all variables having outliers decreased from 3.047 to 2.666.

The interesting findings here were that symmetric outliers did not affect the correlation matrix whereas the asymmetric outliers, especially in the mean shift only condition, distorted the correlation matrix, which either created an extra factor or led to the appearance of a dominant factor that could reduce the number of factors if there were more than one factor. This helps us to understand why mean shift and the number of variables having outliers played important roles for PCA-based methods whereas SD shift did not. Although they are all PCA-based, K-G, MAP and PA methods adopt different procedures and hence one should not be surprised to find some variation among these methods when determining the number of factors, which was shown in our study 1.

**Table 4.7 Demonstration of Changes in Correlation Coefficients, Condition Number and Eigenvalues Using a Four-Variable Data**

**Set in the Presence of Outliers (15%) Compared to the No-Outlier Condition**

Design	Correlation Matrix (Two Variables Having Outliers)	Condition Number	Eigenvalue	Correlation Matrix (All Variables Having Outliers)	Condition Number	Eigenvalue
m=0,SD=1 (No outliers)	1 0.338 1 0.309 0.231 1 0.423 0.328 0.281 1	3.415	1.963 0.776 0.686 0.575			
m=0,SD=3 (SD shift only)	1 .342 1 .272 0.202 1 .374 0.290 0.281 1	3.070	1.888 0.807 0.691 0.615	1 0.342 1 0.315 0.231 1 0.431 0.334 0.287 1	3.489	1.979 0.776 0.679 0.567
m=3,SD=1 (Mean shift only)	1 0.691 1 0.212 0.159 1 0.287 0.222 0.281 1	6.481	<b>1.975</b> <b>1.009</b> 0.712 0.305	1 0.691 1 0.678 0.642 1 0.731 0.686 0.665 1	11.363	<b>3.047</b> 0.362 0.322 0.268
m=3,SD=3 (Mean shift & SD shift)	1 0.566 1 0.222 0.164 1 0.302 0.233 0.281 1	4.465	1.910 0.955 0.708 0.429	1 0.566 1 0.550 0.496 1 0.623 0.561 0.531 1	7.093	2.666 0.509 0.450 0.376

Note. m = mean; SD = standard deviation; dashed lines are used to indicate which variables had outliers.

#### 4.2.4 Demonstration-2

Our findings from study 1 revealed that, for the sequential  $\chi^2_{ML}$  test, the number of factors was inflated with an increase in the magnitude of SD shift, the proportion of contamination, and the number of variables having outliers. Unlike the PCA-based methods, SD shift played an important role, but mean shift did not. In this demonstration, we followed the simulation design in study 1, but did not manipulate the number of variables having outliers. The purpose was to demonstrate the univariate skewness and kurtosis for a single variable in each outlier condition. Like demonstration 1, we did not manipulate sample size and hence reported the population analogues – that is, we reported the values of kurtosis and skewness for a data set with 100,000 observations.

Table 4.8 comprises two parts: the upper part reports kurtosis and the lower part reports skewness<sup>6</sup>. The kurtosis was inflated when the SD shifted from 1 to 1.5, and greatly inflated when SD shift became 3. Mean shift affected kurtosis to some degree for the proportion of contamination of .01 and .08, but not much for a proportion of contamination of .15. It should be noted that the kurtosis was inflated to 8.62 (mean shift=3, SD shift=3) for the proportion of contamination of .08, but was 5.96 for the proportion of contamination of .15. This suggests that a higher level of proportion of contamination (.15) led to less inflation in kurtosis than a lower level (.08). As shown mathematically by Pena and Prieto (2001),

---

<sup>6</sup> Although, due to space limitations, we do not report it herein, our interpretation of the effects of the outlier factors on the kurtosis and skewness are supported by results of ANOVAs of the kurtosis and skewness data in Table 4.8 with an unreplicated design. Please see Liu and Zumbo (2007) for a description of the ANOVA model and the partitioning of the eta-squared with unreplicated simulation designs.



symmetric outliers increase the kurtosis and a small proportion of asymmetric outliers also increase kurtosis, but a large proportion of asymmetric outliers can make kurtosis smaller.

The lower part of Table 4.8 shows that, as expected, skewness was inflated when the mean shifted to 1.5 and 3 and SD shifted to 1.5 and 3. The largest increase of skewness is 1.98 for the mean shift of 3 and SD shift of 3 with .08 proportion of contamination. However, the inflation of skewness was not as large as the inflation of kurtosis. Hence, the inflation in kurtosis likely drove the inflation of type-I error rate of chi-square (ML) sequential tests in our simulation, in which SD shift was an influential factor. This might reflect why SD shift played an important role in explaining the inflation of number of factors in sequential  $\chi^2_{ML}$  tests.

**Table 4.8 Demonstration of Effects of Outliers on Kurtosis and Skewness**

<b>Kurtosis</b>									
		<b>Mean</b>							
		<b>0.0</b>			<b>1.5</b>			<b>3.0</b>	
		<b>SD</b>			<b>SD</b>			<b>SD</b>	
		<b>1.0</b>	<b>1.5</b>	<b>3.0</b>	<b>1.0</b>	<b>1.5</b>	<b>3.0</b>	<b>1.0</b>	<b>1.5</b>
<b>Pc</b>	<b>0.01</b>	-0.02	0.01	1.31	0.01	0.18	2.06	0.61	1.13
	<b>0.08</b>	0.01	0.37	5.89	0.17	1.05	6.86	1.22	2.57
	<b>0.15</b>	0.00	0.48	5.27	0.11	1.09	5.68	0.56	1.72
<b>Skewness</b>									
		<b>Mean</b>							
		<b>0.0</b>			<b>1.5</b>			<b>3.0</b>	
		<b>SD</b>			<b>SD</b>			<b>SD</b>	
		<b>1.0</b>	<b>1.5</b>	<b>3.0</b>	<b>1.0</b>	<b>1.5</b>	<b>3.0</b>	<b>1.0</b>	<b>1.5</b>
<b>Pc</b>	<b>0.01</b>	0.02	0.02	0.01	0.05	0.10	0.33	0.25	0.34
	<b>0.08</b>	0.00	-0.01	-0.08	0.17	0.44	1.13	0.78	1.07
	<b>0.15</b>	0.00	-0.01	-0.03	0.22	0.58	1.24	0.77	1.09

Note. Mean= mean shift; SD = standard deviation shift; Pc = proportion of contamination

### **4.3 Discussion**

The common practices to deal with outliers in data analysis are either to: (a) remove or correct them if they are errors, or (b) use a robust estimator if one is uncertain about the source of the outliers or if one is uncertain that outliers are present (e.g., outliers in high dimensional data are very difficult to detect). However, it should be noted that not all outliers are typographical or data entry (or recording) errors. If outliers arise from unintentionally and unknowingly included subpopulations other than the target population, the outliers are not errors of the first kind described in Liu and Zumbo (2007) but rather, in that sense, legitimate observations that arise from a sub-population different than the target population in a study. The example we provided earlier of the study of life satisfaction in Denmark demonstrates the subtle issues of unintentionally and unknowingly invoking an assumption of measurement universality with heterogeneous populations involved in a multicultural and globalized assessment environment (Hambleton, et al., 2005).

The purpose of the present research was to investigate the effects of outliers, arising from an unintentionally and unknowingly recruited subpopulation, on decisions about the number of factors to retain in an EFA using four decision methods separately. Four important findings are summarized as follows. First, the effects of outliers did not depend on the sample size. This is an important finding because many practitioners believe that having a larger sample size makes them immune to the effects of outliers, which has been shown herein (and

elsewhere) to not be the case. Second, the performance of the three PCA-based methods (K-G, MAP, and PA) was not affected by symmetric contamination, but that of sequential  $\chi^2_{ML}$  tests was affected with inflation of the number of factors retained. Third, for the asymmetric contamination, the number of factors retained was inflated for the MAP method, deflated for the PA method, and either inflated or deflated for the K-G rule, depending on the number of variables having outliers, the proportion of contamination and the level of mean shift, whereas mean shift did not affect sequential  $\chi^2_{ML}$  tests. Finally, the MAP and PA methods are, in general, more resistant to outliers than the K-G rule and sequential  $\chi^2_{ML}$  tests. However, it should be noted that both MAP and PA are still affected by outliers under certain conditions, so they are not fully resistant to outliers.

The present study, along with the earlier study by Liu et al. (in press), provide a broad picture of the effects of outliers on the decisions about the number of factors to retain in an EFA study. When reading extant literature, or conducting an EFA study, readers can be assured that outliers in the item response distributions are likely to have a significant impact on the conclusions, either inflating or deflating the number of factors retained depending on the decision methods used, outlier sources (Liu & Zumbo, 2007), and manifestation of outliers (e.g., asymmetric or symmetric outliers) in the sample. The take-home message in this line of research, however, is still the same: researchers are strongly encouraged to use robust methods in their day-to-day research practice (Wilcox, 2010, in press) and that not doing so may lead to misleading empirical conclusions.

The present research has high fidelity with real data situations, which provides useful

information for applied researchers. However, utilizing real data for simulation also brings some limitations, such as we only mimic one real data situation, so we did not vary the number of variables and number of factors and manipulate different levels of factor loadings and factor correlations. We would encourage future research to investigate these variables when examining the effects of outliers on the decision about the number of factors.

## **Chapter 5:      General Discussion**

Given that the findings reported from chapters 3 and 4 have been discussed in their respective chapters, the purpose of this chapter is to set the findings in a broader context, summarize the common and unique features across these two chapters, describe the implications for day-to-day research practice, present the novel contributions, and discuss the limitations and future directions.

### **5.1 Two Types of Outliers and Distinction of Intended and Unintended Subpopulations**

Liu and Zumbo (2007) described three categories of outlier sources. Based on the characteristics of the outliers and the mathematical and statistical models (i.e., deterministic, slippage, and mixture contamination models) used to simulate the outliers, these three categories were recast into two general categories herein: outliers arising from errors and outliers arising from an unintended and unknowingly included subpopulation. Both types of outliers were investigated in this dissertation.

Outliers are prevalent in day-to-day research practice. For example, human errors often occur during data collection and data preparation. Although researchers have been trying very hard to minimize these error sources, it seems very difficult, if not impossible, to eliminate them. A recent study by Barchard and Pace (2011) compared a single data entry method with more elaborate double entry and visual checking<sup>7</sup> methods, using a group of trained

---

<sup>7</sup>In the single data entry method, one person enters the data once and is done; in a double entry procedure that one person

undergraduate students. Their results revealed that double entry had the fewest errors; visual checking was as bad as single entry, resulting in nearly three times the number of errors as double entry. As Barchard and Pace demonstrated, these data entry errors affected certain statistics tremendously; that is, the coefficient alpha estimate was reduced by as much as .40, correlations turned to zero, and *t*-tests became non-significant.

In addition to data entry errors, outliers can also arise from an unintended subpopulation or subpopulations. These are also likely to appear in day-to-day research practice because the samples that researchers recruit are usually not homogeneous and researchers may not realize that observations from an unintended subpopulation are included in the sample. In social, behavioral, health, and psychological studies, the population under investigation is likely to be heterogeneous because it inherently includes subpopulations, such as gender groups, age groups, or ethnic groups (e.g., Lubke & Muthén, 2005). For example, Orth, Trzesniewski, and Robins (2010) showed that levels of self-esteem and changes in self-esteem differed across age, ethnicity and level of education. In heterogeneous samples, it is essential to distinguish the unintended from the intended subgroups, which affects our decisions regarding the identification of outliers and subsequent analyses.

If a subgroup (or subgroups) in a sample is intended, but is found to behave differently from other observations, the subgroup should not be regarded as outliers. In this case, an alternative model or multi-group model should be adopted. A large amount of related work

---

enters the data twice, and the two entries are compared using a computer program to identify mismatches that can be corrected; in a visual checking procedure, one person enters the data once and then visually compares the entries against the original paper measures, correcting errors if discrepancies are found.

can be found in psychometrics. For example, differential item functioning (DIF)<sup>8</sup> using item response theory (IRT) is checked for gender groups or ethnic groups in high stakes testing, which is used to ensure that the questions do not put different groups at a disadvantage. Similarly, measurement invariance in confirmatory factor analysis (CFA) should be examined for different groups, as researchers want to ensure that observations are measured on the same metric. If group differences are found, either an alternative model is used for modeling the subgroup or multi-group models are adopted, such as partial invariance methods in multi-group confirmatory factor analysis (MG-CFA), to allow parameter estimates to be different for different groups in data analysis.

In addition to observed subpopulations, a heterogeneous population may also include unobserved subpopulations that are unknown to the researcher. Again, these subpopulations should not be regarded as outliers. These subpopulations may be characterized by a complex profile and interaction of variables, not just a single variable such as male/female gender. Conventional statistical methods are not very helpful for modeling unobserved subpopulations. A growing research has shown that latent class analysis or latent mixture models can be a useful tool to model a heterogeneous population/sample when subpopulations or subgroups are unidentified or unobserved (e.g., Lubke & Muthén, 2005; Sawatzky, Ratner, Kopec, Johnson, & Zumbo, 2009). Rather than using one manifest (observed) variable (e.g., ethnicity) in conventional multi-group analysis, a subgroup or

---

<sup>8</sup> DIF occurs when respondents from different populations or subpopulations respond differently to test items or variables in a test or questionnaire after controlling/matching for their true ability or attribute.

subgroups in latent class analysis or latent mixture models can be characterized by a profile comprised of a number of background variables included in the data.

Using factor mixture models, Lubke and Muthén (2005) conducted an investigation of population heterogeneity in American youth taking math and science achievement tests, whereas Sawatzky et al. (2009) examined the sample heterogeneity of adolescents' life satisfaction. Both studies indicated that the latent groups explained the heterogeneity of a population or a sample and background variables were useful for explaining the characteristics of latent populations or groups. Based on the idea that DIF might not be perfectly correlated with manifest groups if a subgroup or subgroups experience, for example, different cognitive processes, mixture IRT models have also been applied to psychometrics to investigate if DIF is caused by latent groups (DeMars & Lau, 2011). Cohen and Bolt (2005) and Samuelsen (2008) showed that mixture IRT models may be a more useful tool than conventional IRT model to identify latent groups and DIF sources.

When a subgroup in a sample is unintended and behaves differently from other observations, these observations should be defined as outliers. If the data include a variable to identify this subgroup, researchers can remove these outliers or model them in a more complex statistical model (e.g., simultaneous multi-group models). Unfortunately, because an unintended subgroup (subgroups) is unknowingly included in a sample, it is, by definition, impossible to identify and incorporate into multi-group models. In these cases, robust statistics are a good option to deal with outliers. Because the idea underlying the use of latent class analysis or latent mixture models is to find unobserved groups, it is possible that latent



mixture models can be used to model a subgroup (subgroups) and distinguish outliers from other observations instead of discarding or down-weighting outliers.

There are two key points to take from this section. First, we have characterized two types of outliers: (1) errors, and (2) unintended and unknowingly included sub-populations. Second, this distinction not only differentiates types of outliers but also implies how one should treat them. During data analysis, if these subpopulations are planned, then observed grouping variables are used in simultaneous multi-group models or used as a design matrix in a single model. If these subpopulations are unplanned then they should be considered outliers and, for example, latent class (factor mixture) analysis may be used to detect and model them. Outliers arising from subpopulations are seldom considered in day-to-day practice and in recommendations in textbooks. Outliers are often only discussed as errors but this is both limiting and problematic.

## **5.2 Review of Outlier Simulation & Major Findings**

### **5.2.1 Outlier Simulation**

Because different types of outliers are a result of different sources, simulation models used and the design of outlier conditions are also different for studying these outliers. This section summarizes the three simulation models used for outlier generation in this dissertation and highlights the distinction between these statistical models.

Chapter 3 investigated outliers arising from errors. Outliers were obtained by simply

multiplying a constant (2, 3, 4, 5) to raw scores. Three factors were examined, that is, magnitude of outliers, number of variables having outliers, and number of subjects having outlying responses. Error type outliers are usually sample specific and thus the number of error type outliers is fixed for a sample. In this chapter, deterministic and slippage models were utilized in study 1 and study 2 separately. Although both models were used for simulating error type outliers and shared some commonality, there are some differences between them.

In study 1 of chapter 3, the original raw data were manipulated for mimicking outliers based on a deterministic model which resulted in only one possible outcome, in this case only one manipulated data set for each outlier condition, and thus the process is deterministic. In study 2, a slippage model was utilized and the data were simulated based on the parameter estimates obtained from the real data with 100 replications for each outlier condition, which is a typical Monte Carlo simulation. In a slippage model, a fixed number of aberrant observations arise independently from a modification of the original distribution; the two distributions (i.e., the original distribution and the modified distribution) are regarded as random variables. Therefore, although the number of outliers was fixed for a sample in both models, the deterministic model deals with one possible outcome whereas the slippage model involves a stochastic process with random outcomes with a central tendency across replications.

Chapter 4 investigated outliers arising from an unintended and unknowingly included subpopulation. In that chapter, symmetric and asymmetric outliers were examined as were

proportion of contamination, number of variables having outliers and sample size. In this case of outliers, two populations are mixed together and the subpopulation (outliers) usually has a different probability distribution from the targeted population. Unlike the error type outliers, these outliers are legitimate observations and should not be treated as errors. In addition, rejection of these outliers is not as easy as error type outliers because it might be hard to identify these outliers in some situations, such as high dimensional data. The mixture contamination model, which is usually used to mix two normal distributions, was used in the simulation study reported in chapter 4. As was shown therein, the number of outliers was a random variable and varied across 100 replications in the simulation, which is different from the slippage model used in Study 2 of chapter 3.

### 5.2.2 Major Findings

The purpose of this dissertation is to investigate how outliers affect the decision about the number of factors by four commonly used decision methods (i.e., K-G, MAP, PA, and sequential  $\chi^2_{ML}$  tests). Chapter 3 focused on error type outliers while chapter 4 focused on outliers arising from an unintended and unknowingly included subpopulation. Some common findings were found across chapters 3 and 4: (a) depending on the decision method and outlier condition, the number of factors retained could be inflated, deflated, or remain the same as the baseline model; (b) MAP and PA methods performed relatively better than the K-G rule and sequential  $\chi^2_{ML}$  tests, although MAP and PA were still affected by outliers; (c) the number of factors was either unchanged or deflated for the PA method; and (d) the

number of factors was either unchanged or inflated for the sequential  $\chi^2_{ML}$  tests.

Given the differences across these studies, some findings were unique to each chapter as well. In chapter 3, the results showed: (a) the number of variables having outliers was an influential factor across all methods and it was the only factor affecting the PA method, (b) for the K-G rule, the number of factors retained was inflated when some variables had outliers, but deflated when all variables had outliers; (c) the performance of MAP method was not affected when only one variable had outliers, but showed strange patterns with the increase in the outlier contamination; and (d) the PA method was the most robust to outliers but was affected by outliers when the number of variables with outliers increased to more than six.

In chapter 4, the results showed: (a) symmetric contamination affected the performance of the sequential  $\chi^2_{ML}$  tests, but did not affect that of other three PCA-based methods (i.e., K-G, MAP and PA); (b) for asymmetric contamination, the number of factors retained was inflated for the MAP method, deflated for the PA method, and either inflated or deflated for the K-G rule, whereas mean shift did not affect sequential  $\chi^2_{ML}$  tests; and (c) effects of outliers did not depend on sample size.

In summary, all decision methods could give researchers misleading results in the presence of outliers. MAP and PA methods were more resistant to outliers than other methods, but they were still not robust to outliers. The studies in this dissertation remind day-to-day researchers to be aware of outliers when conducting factor analysis or reading the extant literature, and to be aware of the implications of the misleading results.

### **5.3 Implications for Day-to-Day Researchers**

This dissertation is a call for day-to-day researchers to be aware of outliers when conducting an EFA study in their own research and to be cautious when they read the extant literature. There are three common pitfalls of researchers' understanding about outliers in day-to-day practice. First, many researchers, as Lind and Zumbo (1993) pointed out, believe that a small deviation from the normal curve will only slightly distort the usual estimates of means, standard deviations, correlations, and associated hypothesis tests. However, even a single outlier can lead to biased results and cause misleading conclusions (e.g., Barchard & Pace, 2011; Cohen, Cohen, West, & Aiken, 2003; Devlin, Gnanadesikan, & Kettenring, 1981; Wilcox, 1998).

The second pitfall is that most researchers believe that having a larger sample size can make them immune to the effects of outliers if only a small proportion of outliers are included in the data. However, chapter 4 showed that the effects of outliers were not eliminated or reduced by larger sample sizes. This has also been shown in other studies. For example, when examining how outliers affected Cronbach's coefficient alpha, Liu and Zumbo (2007) found that outlier effects were not dependent on sample size in their simulation study.

The third common pitfall is that many researchers assume all outliers are errors and hence illegitimate observations that can be easily identified using histograms or box-plots, and feel safe whenever no obvious outlying data points are found in the data. It is worth

noting that not all outliers are errors. If arising from subpopulations, outliers are legitimate observations and should not be treated as errors. Moreover, although the standard outlier identification tools can be used for detecting them in some cases, they may not identify outliers in complex cases such as high dimensional data.

Our findings showed that inclusion of outliers can mislead researchers about the dimensionality of the construct they are studying. This can bring deleterious effects on researchers' work in theory building or theory testing, in validation studies, and in operational testing programs, such as the SAT and GRE. Using an example of the development of social desirability theory, researchers have been trying to identify the number of factors and interpret the meaning of these factors, but the existing studies have shown different numbers of factors obtained for the same scale (e.g., Marlowe-Crowne Social Desirability Scale) using different samples (Helmes, 2000; Leite & Beretvas, 2011). When reading the literature with inconsistent conclusions like this example, researchers should be aware if these were caused by the inclusion of outliers, arising from either errors or subgroups. As our studies showed, the dimensionality of the construct can be seriously distorted by outliers.

Researchers or test developers often want to adapt existing tests to other regions or cultures in order to use them efficiently as well as to save money for developing new tests. In the case of adapting a life satisfaction scale across cultures, for example, validation studies are needed to provide construct validity evidence of the scale based on samples from the range of cultures. If researchers unintentionally include outliers in a new sample, the data

may show a different number of factors, suggesting that the construct under investigation is different across cultures when, in fact, it is not.

Similarly, in operational testing programs, such as the SAT or GRE, outliers may result in artificial differences in factor structures based on different samples even though the number of factors is invariant across samples. Generally, more factors suggest that the test is measuring more skills for a sample of students, whereas fewer factors suggest that a test may be less complex for a sample of students. The inclusion of a few outliers will result in some serious consequences, such as abandoning the items under question and spending more money to develop new items, or scaling test scores for students regarded as disadvantaged.

As the serious consequences herein highlight, there is a need for using modern robust statistical methods whenever one is uncertain about the sources of outliers or whether the data include outliers. Robust statistical methods have been demonstrated to provide better estimates than the standard statistical methods, such as Student's t-tests, ANOVA, and least squares regression analysis when the data have outliers, are non-normal, or do not have equal population variances across groups (e.g., Huber, 1981; Wilcox, 1998, 2005; Wilcox, Charlin, & Thompson, 1986). Some research has shown that robust estimators provided better parameter estimates and less biased model fit indices in EFA as well as CFA (e.g., Yuan, Marshall, & Bentler, 2002; Yuan & Zhong, 2008). Some statistical software programs, such as Mplus, have already made robust estimation available for researchers.

However, as Erceg-Hurn and Mirosevich (2008) pointed out, most day-to-day researchers either do not realize the existence of robust statistical methods or do not know

how to use these robust methods due to lack of exposure to these methods and lack of availability of these methods in commonly used statistical software packages, such as SPSS. Most researchers are taught the standard statistical methods, such as  $t$ -tests,  $F$ -tests, least squares regression, and factor analysis in graduate school, but are rarely exposed to modern robust methods. Most methodology books rarely mention modern robust methods, either paying no attention to outliers or only talking about effects of outliers and how to identify and remove them. Table 5.1 lists how the issue of outliers was addressed in a sample of eight widely used textbooks or methodology books on the topics of structure equation modeling or factor analysis. This sample of books, listed by chronological order, shows that early books did not mention outliers at all, which echoed what Wilcox (1998, p. 300) described in his article, but paid more attention to outliers since the late 1980s and tended to move towards robust estimation methods. However, modern robust methods are still not emphasized enough in the currently used statistics textbooks or methodology books. Therefore, this dissertation is also a call for more teaching on modern robust statistical methods at graduate schools.



**Table 5.1 Review of Outlier Related Issues in a Sample of Structural Equation****Modeling or Factor Analysis Books**

Book by Authors	Effects of Outliers	Detection Methods	Robust Estimators
Fruchter (1954)	-	-	-
Lawley (1971)	-	-	-
McDonald (1985)	-	-	-
Cuttance & Ecob (1987)	Effects of outliers on model fit statistics (demonstrations)	-	-
Bollen (1989)	Heywood case, Effects on Chi-square tests (demonstrations)	Univariate & Multivariate outlier detection	-
Comery & Lee (1992)	-	Described an outlier detection program (Comery, 1985)	-
Hancock & Mueller (2006)	Effects of non-normal data on some fit statistics (not specific to outliers)	-	-
Brown (2006)	Collinearity, non-normality,	-	Robust estimators for non-normal

---

Heywood cases	data, i.e., RML,
(brief description)	WLSM, WLSMV

---

## 5.4 Novel Contributions

In this dissertation, I made four novel contributions to understanding how outliers affect the decisions about the number of factors in an EFA study and how outliers arising from different sources are connected to different outlier simulation models.

The first contribution of this dissertation is that it is the first attempt to systematically document the impact of outliers arising from different sources on the decisions about the number of factors in EFA. In the literature, a few studies had some discussions about how outliers affect the number of factors in EFA or presented small demonstrations, but no systematic study on this issue has been found. As was discussed in chapter 3, statisticians have focused on developing robust estimation methods and often take it for granted that the deleterious effects of outliers are well-known. With an eye toward communicating with social and behavioral researchers, this dissertation is aimed at showing practitioners, as well as statisticians and psychometricians, the impact of outliers on the decision about the number of factors to retain in a factor analysis.

This dissertation systematically studied the impact of outliers in terms of three aspects. First, outliers were classified based on the sources they come from and were studied separately. The categorization of outliers is very important because outliers from different sources have different characteristics and statistical behaviors and hence may have different

impacts on statistical methods. However, this has been paid little attention in the literature.

Moreover, the simulation designs used in the dissertation covered a variety of outlier conditions and different simulation models focused on different outlier conditions. For instance, when investigating error type outliers, magnitude of outliers was examined; when examining outliers arising from a subpopulation, symmetric as well as asymmetric contamination was examined. Hence, this dissertation provided a broad picture for day-to-day researchers as well as researchers in statistical methods.

In addition, this is the first study to consider and systematically investigate the number of variables with outliers. The mathematical statistics literature on robust estimation, as well as the earlier psychometric research such as those by Yuan, Marshall, and Bentler (2002) and Liu and Zumbo (2007), only considered the case in which all variables had outliers. In psychometric theory, however, it is common to consider items as being either sampled or exchangeable with items in a larger domain (or hypothetical set of items). This puts items on a similar footing as respondents (or examinees) in conventional mathematical statistics. This approach to items is widely known as the "item sampling model" of psychometric theory. From this vantage point, the inferences one makes about the items and the examinees (or, more generally, the respondents), and hence the measurement validity, is bounded by this exchangeability of sampled and unsampled items (Zumbo, 2007). The implication of this framework for the study of outliers is that outliers do not need to be equally prevalent for all items, and some items can have no outliers whereas others can have many outliers. In short then, measurement validity itself can be compromised by the presence of outliers in the item

responses.

The second contribution is that the findings from this dissertation resolved the contradictory conclusions in the literature. In the literature, some researchers conjectured, without testing it empirically or via simulation, that a large outlier might create an extra factor (Bollen & Arminger, 1991; Huber, 1981, p. 199), whereas others showed the number of factors was reduced in the presence of outliers (Yuan, Marshall, & Bentler, 2002). The findings from this dissertation showed that outliers could inflate, deflate, or have no effect on the number of factors retained, depending on the decision methods used as well as outlier conditions. Therefore, these contradictory conclusions in the literature were correct under different outlier conditions.

The third contribution is that this dissertation included follow-up studies in order to provide insight into the potential causes of the obtained findings. In practice, day-to-day researchers usually ignore effects of outliers, possibly because the input data in EFA are correlation matrices and researchers believe that a few outliers will not change correlation matrices and hence assume that outliers will not affect factor analysis. These follow-up studies demonstrated how outliers distorted the properties of correlation matrices and inflated skewness and kurtosis, which help readers to understand why outliers affected these decision methods.

Finally, this dissertation made a unique contribution to outlier research. This dissertation articulated the connection between outlier sources and outlier simulation models and demonstrated which simulation models suited different kinds of outliers in three studies.

In chapter 3, deterministic and slippage models were used for simulating error type outliers in which the number of outliers was fixed in the sample. In chapter 4, a mixture contamination model was used for simulating outliers arising from a subpopulation and along the way the central tendency and sample-to-sample variation over 100 replications was documented, which has never been done in the literature on outlier research.

## **5.5 Limitations and Suggestions for Future Research**

The present research has high fidelity with real data situations, which provides useful information for researchers in practice. However, utilizing real data for a simulation also brings some limitations, namely that I only mimic one real data situation, and did not vary the number of variables, number of factors, or levels of factor loadings and factor correlations. Effects of outliers may not depend on the number of variables and factors, but may be moderated by the magnitude of factor loadings and factor correlations. I would encourage future research to include these variables when investigating the effects of outliers on the decision about the number of factors.

This dissertation investigated outliers arising from errors as well as from a subpopulation, which provides a broad picture for researchers, but it did not exhaust all kinds of outlier situations. For example, the mixture contamination model adopted in chapter 4 used two normal distributions. However, it may happen that the subpopulation arises from a non-normal distribution, so effects of outliers should be studied based on a mixture of a normal distribution and a non-normal distribution. Another example is that a mixture of

populations includes a target population and possibly two or more subpopulations; these subpopulations may have different means and standard deviations or have different distribution functions. Therefore, there is a need for more studies to examine how different kinds of outliers affect statistical properties and to what extent.

Building on the findings in this dissertation, and particularly the consideration of outliers as unintended and unknowingly included subpopulations, my program of research has expanded to include a more focused study of statistical methods to investigate outliers. Although outliers are often regarded as either errors or noise in our data, outliers may also come from subpopulations and sometimes provide us valuable information and lead to new discoveries. In his presidential address to the American Statistical Association, Kruskal (1988, p. 929) pointed out that “[i]t is widely argued of outliers that investigation of the mechanism for outlying may be far more important than the original study that led to the outlier; the discovery of penicillium is given as an example”. This is why Kruskal, as a poetic turn of phrase, called outliers 'miracles'.

My program of research has now expanded to include a study of finite mixture models. The development of latent mixture models in recent years has made it possible to model outliers and some researchers have made attempts to identify or model outliers via latent mixture models, such as latent class analysis (e.g., Elliott & Stettler, 2007). As Magidson and Vermunt (2004) pointed out, the basic idea underlying these latent mixture models is that some of the parameters of a postulated statistical model differ across unobserved subgroups. Most importantly, in the spirit of Kruskal's quotation, conditional mixture models can be used

in the context in which one might have either variables that are predictors of class membership or provide a profile description of the classes. It might become a new direction for researchers to use latent mixture models to model outliers, rather than discarding them or downweighting them. However, latent mixture models are not designed for dealing with outliers, per se, so there is a great need to explore how latent mixture models work in the presence of outliers, to determine when they can or cannot be used effectively in this way.

## References

- Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., & Tukey, J. W. (1972). *Robust estimations of location: Survey and advances*. Princeton, NJ: Princeton University Press.
- Anscombe, F. J. (1961). Rejection of outliers. *Technometrics*, 2(123-147).
- Balakrishnan, N., & Childs, A. (2001). Outlier. In M. Hazewinkel (Ed.), *Encyclopaedia of Mathematics (online version)*. New York: Springer-Verlag.
- Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior*, 27(1834-1839).
- Barnett, V. (1999). Nonattending respondent effects on internal consistency of self-administered surveys: A Monte Carlo simulation study. *Educational and Psychological Measurement*, 59, 38-46.
- Barnett, V., & Lewis, T. (Eds.). (1978). *Outliers in statistical data*. New York: John Wiley.
- Barnett, V., & Lewis, T. (1994). *Outliers in statistical data* (3rd ed.). New York: John Wiley.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology (Statistics Section)*, 3, 77-85.
- Bartlett, M. S. (1951). The effect of standardization on a chi-square approximation in factor analysis. *Biometrika*, 38(3-4), 337-344.
- Bates, R. A., Holton III, E. F., & Burnett, M. E. (1999). Assessing the impact of influential



- observations on multiple regression analysis in human resource research. *Human Resources Development Quarterly* 10(4), 343-363.
- Beckman, R. J., & Cook, R. D. (1983). Outlier.....s. *Technometrics*, 25(119-163).
- Bentler, P. M., & Yuan, K. H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34(2), 181-197.
- Blair, R. C., & Higgins, J. J. (1980). The power of t and Wilcoxon statistics: A comparison. *Evaluation Review*, 4, 645-656.
- Blischke, W. S. (1978). Mixtures of distributions. In W. H. Kruskal & J. M. Tanur (Eds.), *International Encyclopedia of Statistics* (Vol. 1). New York: Free Press.
- Bollen, K. A. (1987). Outliers and improper solutions. *Sociological Methods & Research*, 15, 375-384.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley.
- Bollen, K. A., & Aminger, G. (1991). Observational residuals in factor analysis and structural equation models. In P. V. Mardsen (Ed.), *Sociological Methodology* (Vol. 21, pp. 235-262). Cambridge, MA: Blackwell Publishing.
- Boomsma, A. (1983). *On the robustness of LISREL (maximum likelihood estimation) against small sample size and nonnormality*. Unpublished Doctoral dissertation, University of Groningen, Netherlands.
- Bradley, J. V. (1980). Nonrobustness in Z, t, and F tests at large sample sizes. *Bulletin of the Psychonomic Society*, 16, 333-336.
- Browne, M. W. (1968). A comparison of factor analytic techniques. *Psychometrika*, , 33(3),

- 267-334.
- Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis*. Cambridge, UK: Cambridge University Press.
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical & Statistical Psychology*, 37, 62-83.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Beverly Hills, CA: Sage.
- Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, 7(4), 403-421.
- Browne, T. A. (2006). *Confirmatory factor analysis for applied research (methodology in the social sciences)*. New York: The Guilford Press.
- Castaño-Tostado, E., & Tanaka, Y. (1991). Sensitivity measures of influence on the loading matrix in exploratory factor analysis. *Communications in Statistics: Theory and Methods*, 20, 1329-1343.
- Cattell, R. B. (1952). *Factor analysis: an introduction and manual for the psychologist and social scientist*. New York: Harper.
- Cattell, R. B. (1966). Scree test for number of factors. *Multivariate Behavioral Research*, 1(2), 245-276.
- Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*.

- New York: Plenum.
- Cermak, G. W., & Bollen, K. A. (1984). Observer consistency in judging extent of cloud cover. *Atmospheric Environment*, 17, 2109-2110.
- Cliff, N., & Hamburger, C. D. (1967). The study of sampling errors in factor analysis by means of artificial experiments. *Psychological Bulletin*, 68(6), 430-445.
- Cohen, A. S., & Bolt, D. M. (2005). A mixture model analysis of differential item functioning. *Journal of Educational Measurement*, 42(2), 133-148.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). New Jersey: Lawrence Erlbaum Associates.
- Comrey, A. L. (1973). *A first course in factor analysis*. New York: Academic Press.
- Comrey, A. L. (1978). Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology*, 46(4), 648-659.
- Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, N.J: L. Erlbaum Associates.
- Conway, J. M., & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6(2), 147-168.
- Cook, R. D., & Weisberg, S. (1980). Characterizations of an empirical influence function for

- detecting influential cases in regression. *Technometrics*, 22, 495-507.
- 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(3), 585-596.
- Crawford, C. B., & Koopman, P. (1979). Note: Inter-rater reliability of scree test and mean-square ratio test of number of factors. *Perceptual and Motor Skills*, 49(1), 223-226.
- Cudeck, R., & Odell, L. L. (1994). Applications of standard error-estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin*, 115(3), 475-487.
- Cuttance, P., & Ecob, R. (Eds.). (1987). *Structural modeling by example: Applications in educational, sociological, and behavioral research* Cambridge, England: Cambridge University Press.
- DeMars, C. E., & Lau, A. (2011). Differential item functioning detection with latent classes: How accurately can we detect who is responding differentially? *Educational and Psychological Measurement*, 71(4), 597-616.
- Devlin, S. J., Gnanadesikan, R., & Kettenring, J. R. (1981). Robust estimation of dispersion matrices and principal components. *Journal of the American Statistical Association*, 76, 354-362.
- Dingman, H. F., Miller, C. R., & Eyman, R. K. (1964). A comparison between 2 analytic rotational solutions: Where the number of factors is indeterminate. *Behavioral*

- Science*, 9(1), 76-80.
- Dixon, W. J. (1950). Analysis of extreme values. *Annals of Mathematical Statistics*, 21, 488-506.
- Drăgulescu, A., & Yakovenko, V. M. (2001). Evidence for the exponential distribution of income in the USA. *The European Physical Journal B*, 20, 585-589.
- Elliott, M. R., & Stettler, N. (2007). Using a mixture model for multiple imputation in the presence of outliers: The "Healthy for Life" project. *Applied Statistics*, 56(1), 63-78.
- Erceg-Hurn, D. M., & Mirosevich, V. M. (2008). Modern robust statistical methods: An easy way to maximize the accuracy and power of your research. *American Psychologist*, 63(7), 591-601.
- Evans, V. P. A. i. I. E., , (Vol.5, pp. ). . (1999). Strategies for detecting outliers in regression analysis: An introductory primer. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 5, pp. 213-233). Stamford, CT: JAI Press.
- 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.
- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6, 56-83.
- 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 analyses. *Educational and Psychological Measurement*, 56(6), 907-929.
- Ford, J. K., Maccallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical-review and analysis. *Personnel Psychology*, 39(2), 291-314.
- Freund, J. E., & Walpole, R. E. (1980). *Mathematical statistics* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Fruchter, B. (1954). *Introduction to factor analysis*. New York: D. Van Nostrand Company, Inc.
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin*, 118, 392-404.
- Glorfeld, L. W. (1995). An Improvement on Horns parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55(3), 377-393.
- Golub, G. H., & Van Loan, C. F. (1993). *Matrix computation*. London: The Johns Hopkins University Press Ltd.
- Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.
- Gorsuch, R. L. (2003). Factor analysis. In I. B. Weiner, J. A. Schinka & W. F. Velicer (Eds.), *Handbook of psychology* (Vol. 2, pp. 143-164). Hoboken, N J: Wiley.
- Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples.

- Technometrics*, 11(1), 1-21.
- Guttman, L. (1954). Some necessary conditions for common factor analysis. *Psychometrika*, 19(2), 149-161.
- 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.
- Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., & Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions*. New York Wiley.
- Hancock, G. R., & Mueller, R. O. (Eds.). (2006). *Structural equation modeling: A second course*. Greenwich, CT: Information Age Publishing, Inc. .
- Harman, H. H. (1976). *Modern factor analysis*. Chicago: University of Chicago Press.
- Harris, M. L., & Harris, C. W. (1971). A factor analytic interpretation strategy. *Educational and Psychological Measurement*, 31(3), 589-606.
- 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.
- Helmes, E. (2000). The role of social desirability in the assessment of personality constructs. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 21-40). Norwell, Massachusetts:

- Kluwer Academic Publishers.
- Herzog, W., Boomsma, A., & Reinecke, S. (2007). The model size effect on traditional and modified tests of covariance structures. *Structural Equation Modeling, 14*, 361-390.
- Holzinger, K. J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution. *Supplementary Educational Monographs, No.48*. Chicago, IL.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.
- Hoyle, R. H., & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *Handbook of quantitative methodology for the social sciences* (pp. 301-315). Thousand Oaks, CA: Sage Publications.
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria. *Structural Equation Modeling, 6*, 1-55.
- Hubbard, R., & Allen, S. J. (1987). An empirical comparison of alternative methods for principal components extraction. *Journal of Business Research, 15*, 173-190.
- Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics, 35*(73-101).
- Huber, P. J. (1981). *Robust statistics*. New York: Wiley.
- Humphreys, L. G. (1964). Number of cases and number of factors - An example where N is



- very large. *Educational and Psychological Measurement*, 24(3), 457-466.
- Humphreys, L. G., & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, 10, 193-206.
- Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32(4), 443-482.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183-202.
- Jung, S., & Lee, S. (2011). Exploratory factor analysis for small samples. *Behavior Research Methods*, 43, 701-709.
- Kaiser, H. F. (1960). The application of electronic-computers to factor-analysis. *Educational and Psychological Measurement*, 20(1), 141-151.
- Kaiser, H. F. (1970). A second generation little Jiffy. *Psychometrika*, 35(4), 401-415.
- Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika*, 27(2), 179-182.
- Kenny, D. A., & McCoach, D. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10, 333-351.
- Kruskal, W. H. (1988). Miracles and statistics: The casual assumption of independence (ASA presidential address). *Journal of the American Statistical Association* 83 (404), 929-940.

- Kwan, C. W., & Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: Application to factor analysis. *Psychometrika*, 63(1), 35-46.
- Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method*. New York: Elsevier.
- Lee, H. B., & Comrey, A. L. (1979). Distortions in a commonly used factor analytic procedure. *Multivariate Behavioral Research*, 14(3), 301-321.
- Leite, W. L., & Beretvas, S. N. (2005). Validation of scores on the Marlowe-Crowne Social Desirability Scale and the Balanced Inventory of Desirable Responding. *Educational and Psychological Measurement*, 65, 140-154.
- Levonian, E., & Comrey, A. L. (1966). Factorial stability as a function of number of orthogonally-rotated factors. *Behavioral Science*, 11(5), 400-404.
- Lind, J. C., & Zumbo, B. D. (1993). The continuity principle in psychological research: An introduction to robust statistics. *Canadian Psychology*, 34, 407-414.
- Liu, Y., Wu, A. D., & Zumbo, B. D. (2010). The impact of outliers on Cronbach's coefficient alpha estimate of reliability: Ordinal/rating scale item responses. *Educational and Psychological Measurement*, 70, 5-21.
- Liu, Y., & Zumbo, B. D. (2007). The impact of outliers on Cronbach's coefficient alpha estimate of reliability - Visual analogue scales. *Educational and Psychological Measurement*, 67, 620-634.
- Liu, Y., Zumbo, B. D., & Wu, A. D. (in press). A demonstration of the impact of outliers on the decisions about the number of factors in exploratory factor analysis. *Educational*

*and Psychological Measurement.*

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models *Psychological Methods*, 10, 21-39.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 175-198). Thousand Oakes: Sage Publications.

Magidson, J., & Vermunt, J. K. (2004). Latent class models. In D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 175-198). Thousand Oakes: Sage Publications.

Mahalanobis, P. C. (1936). On the generalised distance in statistics. *Proceedings of the National Institute of Sciences of India*, 2, 49–55.

Mavridis, D., & Moustaki, I. (2008). Detecting outliers in factor analysis using the forward search algorithm. *Multivariate Behavioral Research*, 43, 453-475.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NY: Lawrence Erlbaum.

Miles, J., & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and Individual Differences*, 42, 869-874.

Mosteller, F., & Tukey, J. W. (1968). Data analysis, including statistics. In G. Lindzey & E. Aronson (Eds.), *Handbook of Social Psychology (2nd Ed.)* (Vol. 2, pp. 80-203). Reading, MA: Addison-Wesley.

Nevitt, J., & Hancock, G. R. (2000). A Monte Carlo study investigating the impact of item

- parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement*, 63, 729-757.
- Norman, G. R., & Streiner, D. L. (1994). *Biostatistics: The bare essentials*. St. Louis, MO: Mosby.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- 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.
- OECD. (2005). Society at a glance: Organization for Economic Co-operation and Development (OECD) social indicators - 2005 Edition [Electronic Version]. Retrieved November 11, 2006 from <http://www.oecd.org/dataoecd/34/13/34542721.xls>
- Orth, U., Trzesniewski, K. H., & Robins, R. W. (2010). Self-esteem development from young adulthood to old age: A cohort-sequential longitudinal study. *Journal of Personality and Social Psychology*, 98(4), 645-658.
- Pena, D., & Prieto, F. J. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*, 43(3), 286-310.
- Pison, G., Rousseeuw, P. J., Filzmoser, P., & Croux, C. (2003). Robust factor analysis. *Journal of Multivariate Analysis*, 84(1), 145-172.
- Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry of partition variance explained. In T. Pukilla & S. Duntaneu (Eds.), *Proceedings of Second Tampere*

- Conference in Statistics* (pp. 245-260). Finland: University of Tampere.
- Rousseeuw, P. J., & van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*, 85, 633-651.
- Russell, D. W. (2001). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and Social Psychology Bulletin*, 28(12), 1629-1646.
- Samuelsen, K. (2008). Examining differential item function from a latent class perspective. In G. Hancock & K. Samuelsen (Eds.), *Mixture Models in Latent Variable Research*. Greenwich, CT: Information Age Publishing, Inc.
- Saris, W. E., & Satorra, A. (1992). Power evaluations in structural equation models. In K. A. Bollen & S. Long (Eds.), *Testing in structural equation models* (pp. 181-204). London: Sage University Press.
- Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the Business and Economic Statistics Section of the ASA*, 308-313.
- Sawatzky, R. G., Ratner, P. A., Johnson, J. L., Kopec, J., & B.D., Z. (2009). Sample Heterogeneity and the Measurement Structure of the Multidimensional Student's Life Satisfaction Scale. *Social Indicators Research: International Interdisciplinary Journal for Quality of Life Measurement*, 94, 273-296.
- Slocum-Gori, S. L., & Zumbo, B. D. (2011). Assessing the Unidimensionality of Psychological Scales: Using Multiple Criteria from Factor Analysis *Social Indicators*

- Research: An International Interdisciplinary Journal for Quality of Life Measurement*, 102, 443-461.
- SPSS Inc. (2009). PASW STATISTICS 17.0 Command Syntax Reference. Chicago: SPSS Inc.
- Steiger, J. H., & Lind, J. C. (1980). *Statistically-based tests for the number of common factors*. Paper presented at the Annual Meeting of the Psychometric Society, Iowa City, IA.
- Stevens, J. P. (1984). Outliers and influential data points in regression analysis. *Psychological Bulletin*, 95, 334-344.
- Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.
- Tanaka, Y., & Odaka, Y. (1989). Influential observations in principal factor analysis. *Psychometrika*, 54(3), 475-485.
- Thomas, D. R., Hughes, E., & Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research*, 45, 253-275.
- Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. *Biometrika*, 32, 214-219.
- Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: The University of Chicago press.
- Thurstone, L. L. (1947). *Multiple-factor analysis; a development and expansion of the vectors of the mind*. Chicago: The University of Chicago Press.
- Tippett, L. H. C. (1925). On the extreme individuals and the range of samples taken from a

- normal population. *Biometrika*, 17, 364-387.
- Tukey, J. W. (1962). The future of data analysis. *Annals of Mathematical Statistics*, 3, 1-67.
- 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.
- Velicer, W. F. (1976). Determining number of components from matrix of partial correlations. . *Psychometrika*, 41(3), 321-327.
- Watkins, D. S. (2010). *Fundamentals of matrix computations* (3rd ed.). Pullman, MA: A John Wiley & Sons, Inc.
- Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods? *American Psychologist*, 53, 300-314.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Academic Press.
- Wilcox, R. R. (2010). *Fundamentals of modern statistical methods: Substantially improving power and accuracy* (2nd ed.). New York: Springer.
- Wilcox, R. R. (in press). *Modern statistics for the social and behavioral sciences: A practical introduction*. New York: Chapman & Hall/CRC press.
- Wilcox, R. R., Charlin, V. L., & Thompson, K. L. (1986). New Monte Carlo results on the robustness of the ANOVA  $f$ ,  $w$  and  $f$  statistics. *Communications in Statistics: Simulation and Computation*, 15, 933-943.
- 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, 1(4), 354-365
- Yuan, K. H., & Bentler, P. M. (2001). Effects of outliers on estimators and tests in covariance structure analysis. *British Journal of Mathematical & Statistical Psychology*, 54(1), 161-175.
- Yuan, K. H., & Bentler, P. M. (2007). Robust procedures in structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 367-397). Amsterdam, Neitherland: Elsevier.
- Yuan, K. H., Marshall, L. L., & Bentler, P. M. (2002). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. *Psychometrika*, 67(95-122).
- Yuan, K. H., & Zhong, X. L. (2008). Outliers, leverage observations, and influential cases in factor analysis: Using robust procedures to minimize their effect. *Sociological Methodology*, 38, 329-368.
- Zimmerman, D. W., & Zumbo, B. D. (1993). Relative power of parametric and nonparametric statistical methods. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral science* (Vol. 1: Methodological Issues, pp. 481-517). Hillsdale, NJ Lawrence Erlbaum Associates.
- Zumbo, B.D. (2007). Validity: Foundational issues and statistical methodology. In C.R. Rao and S. Sinharay (Eds.) *Handbook of Statistics, Vol. 26: Psychometrics*, (pp. 45-79). Elsevier Science B.V.: The Netherlands.



- Zumbo, B. D., & Jennings, M. (2002). The robustness of validity and efficiency of the related samples t-test in the presence of outliers. *Psicologica*, 23, 415-450.
- Zumbo, B. D., & Zimmerman, D. W. (1993). Is the selection of statistical methods governed by level of measurement? *Canadian Psychology*, 34, 390-400.
- Zwick, W. R., & Velicer, W. F. (1980). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research*, 17, 253-269.
- Zwick, W. R., & Velicer, W. F. (1986). Comparison of 5 rules for determining the number of components to retain. *Psychological Bulletin*, 99, 432-442.

## **Appendix**

### The 24 Psychological Ability Tests in Holzinger and Swineford's (1939) Data

- T1: Visual perception
- T2: Cubes
- T3: Paper form board
- T4: Flags
- T5: General information
- T6: Paragraph comprehension
- T7: Sentence completion
- T8: Word classification
- T9: Word meaning
- T10: Addition
- T11: Code
- T12: Counting dots
- T13: Straight-curved capitals
- T14: Word recognition
- T15: Number - recognition
- T16: Figure recognition
- T17: Object - number
- T18: Number - figure
- T19: Figure - word
- T20: Deduction
- T21: Numerical puzzles
- T22: Problem reasoning
- T23: Series completion

T24: Arithmetic problems