

https://dspace.org/

THE VALIDITY OF EXPLORATORY

Maddern, Kevin Richard

2011-05-13
http://hdl.handle.net/10539/9756

Downloaded from a DSpace repository. https://dspace.org/

THE VALIDITY OF EXPLORATORY FACTOR ANALYSIS WITH ORDINAL DATA USING THE NORMAL DISTRIBUTION FITTING ALGORITHM

Kevin Richard Maddern

A research report submitted to the faculty of Commerce, Law and Management, University of Witwatersrand, in partial fulfilment of the requirements for the degree of Master of Business Administration

Johannesburg, 2008

ABSTRACT

The purpose of this research was to determine the impact that ordinal data (Likert type responses) has on principle axis method exploratory factor analysis, and whether these results could be improved through first rescaling the ordinal data using the distribution fitting method proposed by Stacey (2005), or using analysis that ignores the inter scale distances by using Spearman-rank correlation analysis, prior to performing exploratory factor analysis.

This research was a simulation study. It simulated the continuous attitudes/beliefs of respondents given a known (unobserved) factor structure, and then used these to create responses on an ordinal scale. The study compared the robustness of exploratory factor analysis using varimax rotation when applied; directly to the ordinal data, to the Spearman-rank correlation matrix of the same ordinal data, and to the transformed ordinal data using a distribution fitting algorithmic approach. Comparison was done between these methods and the known factors of the contrived 'latent' continuous population from which the ordinal data was derived.

The research found that discretisation creates both a negative bias and an increase in the variance in the factor loading results. The amount of negative bias and variance increase both as the number of categories used in the response scale decrease, and as the underlying communalities in the factor structure decrease. Given various forms of scale distortion introduced during the discretisation process (skew, kurtosis, and bimodality), it was found that both the negative bias and the variance of the results will increase even further as the scale distortion increases. These scale distortions are akin to the various biases that a group of respondents may exhibit when completing an ordinal scale questionnaire. Skewed distortion was found to be the most destructive, followed by bimodal distortion. Although no improvement was found in the results when using normal distribution fitting algorithm prior to exploratory factor analysis, it was found that, in cases of skewed or bimodal distortion, performing exploratory factor analysis on the Spearman-rank correlation matrix of the

ordinal data was superior to performing the factor analysis directly on the ordinal data.

These results are relevant to many fields of research, including business, given the prolific use of ordinal scales to gather data due to their simplicity and ease of use, and given the need by researchers to simplify or further analyse the ordinal data using multivariate techniques intended for interval level data (such as exploratory factor analysis). It was found that researchers need to take care in the selection of the response scale for their analysis, since it is shown that the correct choice in category size is the most effective way to ensure that data loss due to discretisation error is minimised. Data loss was shown to constitute both a strengthening and a weakening in the observed factor loading when compared to the unobserved (intended) factor structure. The risk is therefore shown that a researcher may unknowingly discard good data, or include spurious data, through the incorrect use of the response scale. These results begin to guide the researcher to avoid these pitfalls.

DECLARATION

I, Kevin Richard Maddern, declare that this research report is my own work. It is submitted in partial fulfilment of the requirements for the degree of Master of Business Administration in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in this or any other university.

Kevin Richard Maddern

9 August 2008

DEDICATION

Nine tenths of education is encouragement – Anatole France.

This is dedicated to those that encourage.

ACKNOWLEDGEMENTS

I would like to thank the following people for their support, without whom none of this would have been possible:

- My wife, for all the support and understanding throughout my studies;
- My supervisor, Dr Anthony Stacey, for all his time and guidance;
- My friends and family, for remaining in touch when I was distracted.

TABLE OF CONTENTS

ABSTRACT II

DECLARATION IV

DEDICATION V

ACKNOWLEDGEMENTS VI

LIST OF TABLES X

LIST OF FIGURES XII

1 INTRODUCTION 1

1.1 PURPOSE OF STUDY 1

1.2 CONTEXT OF STUDY 1

1.3 PROBLEM STATEMENT 2

1.3.1 PROBLEM STATEMENT 2

1.3.2 SUB-PROBLEMS 2

1.4 SIGNIFICANCE OF STUDY 2

1.5 DELIMITATIONS AND LIMITATIONS 3

1.5.1 DELIMITATIONS 3

1.6 DEFINITION OF TERMS 4

2 LITERATURE REVIEW 5

2.1 INTRODUCTION 5

2.2 THE USE OF ORDINAL DATA IN MULTIVARIATE ANALYSIS 5

2.2.1 ISSUES WITH FACTOR ANALYSIS ON ORDINAL DATA 7

2.2.2 VALIDITY AND RELIABILITY 10

2.2.3 PARAMETERS TO CONSIDER FOR THE METHODOLOGY 13

2.3 DISTRIBUTION FITTING PRIOR TO FACTOR ANALYSIS ON ORDINAL DATA 16

2.4 CONCLUSION OF LITERATURE REVIEW 21

2.4.1 RESEARCH QUESTION 1: 21

2.4.2 RESEARCH QUESTION 2: 21

2.4.3 RESEARCH QUESTION 3: 21

3 RESEARCH METHODOLOGY 22

3.1 RESEARCH METHODOLOGY /PARADIGM 22

3.2 RESEARCH DESIGN 22

3.2.1 METHODOLOGY OVERVIEW 23

3.2.2 METHODOLOGY - INPUTS TO THE STUDY (STEPS 1, 2, 3 & 8) 25

3.2.3 METHODOLOGY - CREATING THE LATENT CONTINUOUS DATA (STEPS 5 TO 8) 27
3.2.4 METHODOLOGY - CATEGORISATION OF THE LATENT DATA (STEP 9) 30
3.2.5 METHODOLOGY - ANALYSIS OF OUTPUT (STEP 10 TO 16) 30
3.2.6 METHODOLOGY – COMPARISON OF RESULTS (STEP 17) 32
3.2.7 MODEL INPUTS AND CONSTRAINTS 32
3.2.8 MODEL OUTPUTS 39
3.3 POPULATION AND SAMPLE 40
3.3.1 POPULATION 40
3.3.2 SAMPLE 41
3.4 CREATION OF THE INPUT DATA 41
3.4.1 MODEL SUMMARY 43
3.5 DATA ANALYSIS AND INTERPRETATION 44
3.5.1 COMPARATIVE MEASURES USED 44
3.5.2 TABULAR PRESENTATION OF RB AND CV 46
3.5.3 PLOTS USED 47
3.6 VALIDITY AND RELIABILITY 49
3.6.1 EXTERNAL VALIDITY 49
3.6.2 INTERNAL VALIDITY 49
3.6.3 RELIABILITY 50
4 INTERNAL VALIDITY ACHIEVED IN THE STUDY 51
4.1 INTRODUCTION 51
4.2 SAMPLING ERROR AND EFA STABILITY IN THE CONTINUOUS DATA 51
4.2.1 DESCRIPTIVE STATISTICS OF THE UNDERLYING SAMPLES 52
4.2.2 CORRELATION ERROR BETWEEN INTENDED (R) AND SAMPLE (R*) 53
4.2.3 FACTOR ANALYSIS ERROR 54
4.3 MANIFEST CATEGORICAL DATA DESCRIPTIVE STATISTICS 57
4.3.1 LIMITED SCALE DISTORTION MANIFEST CATEGORICAL DATA 57
4.3.2 SKEWED DISTORTION MANIFEST CATEGORICAL DATA 60
4.3.3 KURTOTIC DISTORTION MANIFEST CATEGORICAL DATA 62
4.3.4 BIMODAL DISTORTION MANIFEST CATEGORICAL DATA 64
5 PRESENTATION OF RESULTS 66
5.1 INTRODUCTION 66
5.2 THE EFFECTS OF CATEGORISATION ON FA WITH LIMITED DISTORTION 67
5.3 THE EFFECTS OF CATEGORISATION ON FA WITH SKEWED DISTORTION 72
5.4 THE EFFECTS OF CATEGORISATION ON FA WITH KURTOTIC DISTORTION 74
5.5 THE EFFECTS OF CATEGORISATION ON FA WITH BIMODAL DISTORTION 76
5.6 COMPARISON OF METHODS IN LIMITED DISTORTION SITUATIONS 78
5.7 COMPARISON OF METHODS IN SKEWED DISTORTION SITUATIONS 79
5.8 COMPARISON OF METHODS IN KURTOTIC DISTORTION SITUATIONS 80
5.9 COMPARISON OF METHODS IN BIMODAL DISTORTION SITUATIONS 80
6 CONCLUSION 82
6.1 INTRODUCTION 82

6.2 SUMMARY OF THE MAIN FINDINGS 82
6.2.1 RECOMMENDATIONS FOR RESEARCH DESIGN 83
6.2.2 RECOMMENDATIONS FOR DATA ANALYSIS 86
6.3 RECOMMENDATIONS FOR FUTURE RESEARCH 89

REFERENCES 91

7 APPENDICES 95

7.1 CATEGORISATION – A PRACTICAL EXAMPLE 96
7.2 SCHEMATICS OF MODELS 97
7.3 FACTOR LOADING DESCRIPTIVE STATISTICS – LITTLE SCALE DISTORTION 105
7.4 FACTOR LOADING DESCRIPTIVE STATISTICS – SKEW DISTORTION 126
7.5 FACTOR LOADING DESCRIPTIVE STATISTICS – KURTOTIC DISTORTION 147
7.6 FACTOR LOADING DESCRIPTIVE STATISTICS – BIMODAL DISTORTION 168

LIST OF TABLES

Table 1 - Model Summary 44

Table 2 - Continuous Data Descriptive Statistics 53

Table 3 - Correlation Error of Samples 54

Table 4 - Control Results on positively loaded variables 56

Table 5 – Group 1 Thresholds 57

Table 6 - Manifest Categorical Data Descriptive Statistics (Limited Distortion) () . 58

Table 7 - Histogram for Discrete Manifest Data (Limited Distortion) 59

Table 8 - Group 2 Thresholds - Skewed Distortion 60

Table 9 - Manifest Categorical Data Descriptive Statistics (Skewed Distortion)... 60

Table 10 - Histogram for Discrete Manifest Data (Skewed Distortion) 61

Table 11 - Group 2 Thresholds - Kurtotic Distortion 62

Table 12 - Manifest Categorical Data Descriptive Statistics (Kurtotic Distortion) . 62

Table 13 - Histogram for Discrete Manifest Data (Kurtotic Distortion)..... 63

Table 14 - Group 2 Thresholds - Bimodal Distortion 64

Table 15 - Manifest Categorical Data Descriptive Statistics (Bimodal Distortion) 64

Table 16 - Histogram for Discrete Manifest Data (Bimodal Distortion)..... 65

Table 17 - RB of Factor Loadings with Little Scale Distortion 67

Table 18 - CV of Factor Loadings with Little Scale Distortion 69

Table 19 - RB of Factor Loadings with Skew Scale Distortion 73

Table 20 - CV of Factor Loadings with Skew Scale Distortion 73

Table 21 - RB of Factor Loadings with Kurtotic Scale Distortion 75

Table 22 - CV of Factor Loadings with Kurtotic Scale Distortion 75

Table 23 - RB of Factor Loadings with Bimodal Scale Distortion 76

Table 24 - CV of Factor Loadings with Bimodal Scale Distortion 77

Table 25 - Category size recommendation 85

Table 26 - Recommended Methods 88

Table 27 - FL Descriptive Statistics - Little Scale Distortion on 2 Categories 105

Table 28 - FL Descriptive Statistics - Little Scale Distortion on 3 Categories 106

Table 29 - FL Descriptive Statistics - Little Scale Distortion on 4 Categories 107

Table 30 - FL Descriptive Statistics - Little Scale Distortion on 5 Categories 108

Table 31 - FL Descriptive Statistics - Little Scale Distortion on 6 Categories 109

Table 32 - FL Descriptive Statistics - Little Scale Distortion on 7 Categories 110

Table 33 - FL Descriptive Statistics - Little Scale Distortion on 8 Categories 111

Table 34 - FL Descriptive Statistics - Skew Scale Distortion on 2 Categories 126

Table 35 - FL Descriptive Statistics - Skew Scale Distortion on 3 Categories 127

Table 36 - FL Descriptive Statistics - Skew Scale Distortion on 4 Categories 128

Table 37 - FL Descriptive Statistics - Skew Scale Distortion on 5 Categories 129

Table 38 - FL Descriptive Statistics - Skew Scale Distortion on 6 Categories 130

Table 39 - FL Descriptive Statistics - Skew Scale Distortion on 7 Categories 131

Table 40 - FL Descriptive Statistics - Skew Scale Distortion on 7 Categories 132

Table 41 - FL Descriptive Statistics - Kurtotic Scale Distortion on 2 Categories 147

Table 42 - FL Descriptive Statistics - Kurtotic Scale Distortion on 3 Categories 148

Table 43 - FL Descriptive Statistics - Kurtotic Scale Distortion on 4 Categories 149

Table 44 - FL Descriptive Statistics - Kurtotic Scale Distortion on 5 Categories 150

Table 45 - FL Descriptive Statistics - Kurtotic Scale Distortion on 6 Categories 151

Table 46 - FL Descriptive Statistics - Kurtotic Scale Distortion on 7 Categories 152

Table 47 - FL Descriptive Statistics - Kurtotic Scale Distortion on 8 Categories 153

Table 48 - FL Descriptive Statistics - Bimodal Scale Distortion on 2 Categories 168

Table 49 - FL Descriptive Statistics - Bimodal Scale Distortion on 3 Categories 169

Table 50 - FL Descriptive Statistics - Bimodal Scale Distortion on 4 Categories 170

Table 51 - FL Descriptive Statistics - Bimodal Scale Distortion on 5 Categories 171

Table 52 - FL Descriptive Statistics - Bimodal Scale Distortion on 6 Categories 172

Table 53 - FL Descriptive Statistics - Bimodal Scale Distortion on 7 Categories 173

Table 54 - FL Descriptive Statistics - Bimodal Scale Distortion on 8 Categories 174

LIST OF FIGURES

Figure 1: Distortions in Applying Scale - redrawn from Kim (1975)..... 7

Figure 2 - NDF Method - Conceptual Diagram..... 17

Figure 3 - NDF Initial setup 18

Figure 4 - NDF after Solving (right), & after Standardisation (left)..... 19

Figure 5: Graphic Representation of the Simulation Methodology 24

Figure 6 - Obtaining Correlation Matrix from Factor Loading Matrix 26

Figure 7 - Choleski Matrix 28

Figure 8 - Model Tree 39

Figure 9 - Tabular presentation of RB or CV 46

Figure 10 - Example 'singular' modified box plot 48

Figure 11 - Factor Loading as % of Intended Factor Loading 68

Figure 12 - 6 Category Plot - Little Scale Distortion 70

Figure 13 - Percentiles - CAT Method - Limited Distortion 112

Figure 14 - Modified Box Plots Category 2, 4, 5, 6, 7 and 8 (Limited Distortion) 118

Figure 15 – Percentiles – SPR Method - Limited Distortion 119

Figure 16 – Percentiles – NDF^1 Method - Limited Distortion 120

Figure 17 – Percentiles – NDF^2 Method - Limited Distortion 121

Figure 18 – Method RB Comparison - Limited Distortion 122

Figure 19 – Method CV Comparison - Limited Distortion 123

Figure 20 – Method RB % Underperformance - Limited Distortion 124

Figure 21 – Method CV % Underperformance - Limited Distortion 125

Figure 22 - Percentiles – CAT Method - Skewed Distortion 133

Figure 23 - Modified Box Plots Category 2, 4, 5, 6, 7 and 8 (Skewed Distortion) 139

Figure 24 – Percentiles – SPR Method – Skewed Distortion 140

Figure 25 – Percentiles – NDF^1 Method - Skewed Distortion 141

Figure 26 – Percentiles – NDF^2 Method - Skewed Distortion..... 142

Figure 27 – Method RB Comparison - Skewed Distortion..... 143

Figure 28 – Method CV Comparison - Skewed Distortion..... 144

Figure 29 – Method RB % Underperformance - Skewed Distortion..... 145

Figure 30 – Method CV % Underperformance - Skewed Distortion..... 146

Figure 31 - Percentiles – CAT Method - Kurtotic Distortion..... 154

Figure 32 - Modified Box Plots Category 2, 4, 5, 6, 7 and 8 (Kurtotic Distortion) 160

Figure 33 – Percentiles – SPR Method - Kurtotic Distortion 161

Figure 34 – Percentiles – NDF^1 Method - Kurtotic Distortion 162

Figure 35 – Percentiles – NDF^2 Method - Kurtotic Distortion 163

Figure 36 – Method RB Comparison - Kurtotic Distortion 164

Figure 37 – Method CV Comparison - Kurtotic Distortion 165

Figure 38 – Method RB % Underperformance - Kurtotic Distortion..... 166

Figure 39 – Method CV % Underperformance - Kurtotic Distortion..... 167

Figure 40 - Percentiles – CAT Method - Bimodal Distortion..... 175

Figure 41 - Modified Box Plots Category 2, 4, 5, 6, 7 and 8 (Bimodal Distortion) 181

Figure 42 - Percentiles – SPR Method - Bimodal Distortion 182

Figure 43 - Percentiles – NDF^1 Method - Bimodal Distortion 183

Figure 44 - Percentiles – NDF^2 Method - Bimodal Distortion 184

Figure 45 – Method RB Comparison - Bimodal Distortion 185

Figure 46 – Method CV Comparison - Bimodal Distortion 186

Figure 47 – Method RB % Underperformance - Bimodal Distortion..... 187

Figure 48 – Method CV % Underperformance - Bimodal Distortion..... 188

1 INTRODUCTION

1.1 Purpose of study

The purpose of the research was to determine the impact that ordinal data has on principle axis method exploratory factor analysis (EFA), and whether these results can be improved through first transforming the ordinal data into interval data using the distribution fitting method proposed by Stacey (2005).

1.2 Context of study

A literature review was conducted in 1984 on various fields of study to determine the extent of use of multivariate techniques (Gnanadesikan and Kettenring, 1984). It reported that in five of eight surveyed fields, factor analysis was the most widely used multivariate technique. In education, psychology, and sociology it was applied six to eight times as much as discriminant analysis, the next most popular multivariate method. In business, the use of factor analysis has certainly increased as the number of variables for analysis has increased (Hair et al., 2005) .

No doubt many of these studies used ordinal verbal scales (e.g. Likert scale) to gather the data due to its simplicity for gathering responses (Srinivasan and Basu, 1989), yet this simplicity introduces potential problems when these ordinal response categories are simply assigned numeric values (e.g. 1 – 5 for a Likert scale) for the purposes of any parametric statistical analysis. Many of these statistical analysis techniques assume continuous data, yet the observed responses have been categorised into just a few discrete measures and in so doing cast a shadow on the validity or accuracy of the results (Labovitz, 1967, Somers, 1968, Hawkes, 1971, Wilson, 1971, Olsson, 1979, Vierra and Carlson, 1981, Young, 1981, Babakus et al., 1987, Potthast, 1993, Velleman and Wilkinson, 1993, Dolan, 1994, Hutchinson and Olmos, 1998, Lubke and Muthén, 2004, Stacey, 2005, Temme, 2006).

It is the purpose of this study to attempt to highlight the problems that may arise in using EFA on ordinal data, so that researchers may understand the implications this could have on the results they obtain and the conclusions they reach. It will attempt to caution for those

conditions under which the results of EFA on ordinal data may be invalid. It will also test the application of distribution fitting to transform the data from ordinal to interval prior to the EFA being performed, and hopefully make meaningful recommendations for its use.

1.3 Problem statement

1.3.1 *Problem Statement*

What could be some of the conditions under which the accuracy of EFA is adversely affected through the simple allocation of a discrete interval scale to ordinal data, and would the accuracy of the results improve through the use of distribution fitting (Stacey, 2005) as a method to transform ordinal data to interval data prior to performing EFA?

1.3.2 *Sub-problems*

The first sub-problem is: to identify conditions under which the accuracy of EFA is adversely affected through the simple allocation of a discrete interval scale to ordinal data?

The second sub-problem is: to determine the accuracy of the results obtained through the use of distribution fitting (Stacey, 2005) as a method to analyse ordinal data prior to performing EFA?

1.4 Significance of study

The study adds to the literature by giving guidance to researchers in many fields, including business, who use EFA as a method for analysing or simplifying ordinal data. Ordinal data is commonly found in much research and includes Likert-type response scales, age observations, income observations etc.

Much of the existing research is done on confirmatory factor analysis which in effect works in reverse to exploratory factor analysis. The principle behind confirmatory factor analysis is that it tests a set of observations (sample) against a previously developed or known model to confirm whether the expected factors exist within the newly observed sample. EFA on the other hand is used by researchers who do not have any known model against which to

compare their data, but are trying to find underlying factors that have never before been observed, or are trying to reduce the complexity of their observed data.

The study attempts to provide guidance on when not to use EFA on ordinal data, and also attempts to provide a more robust method.

1.5 Delimitations and limitations

1.5.1 *Delimitations*

The study was performed using EFA using the principle components method only, and it did not attempt to compare the results against any other factor analysis methods.

This study was based purely on simulated (contrived) data and did not attempt to test on empirical data - with simulated data the results can be better analysed for specific changes in the input data (skewness, kurtosis, sample size, number of categories etc.).

The study is restricted to ordered categories where the latent variables are continuous.

1.6 Definition of terms

Latent population is used to refer to the unobserved continuous population. In this study the latent population is the simulated attitudes and beliefs that are categorised, and the categorised population is referred to as the observed or manifest population.

Scaling or categorization or discretisation are terms used to refer to a process where continuous data is collapsed into categories through either applying thresholds, or in the field through applying a questionnaire with Likert type responses (since the underlying attitude may be continuous, observations are practically implemented using a few discrete categories).

The terms ‘interval’ and ‘ordinal’ refer only to the scales used or the data produced, and does not refer to the underlying (unobserved or latent) data.

EFA is always meant to refer to principle component method exploratory factor analysis.

Communalities referred to with regards to factor analysis refer to the correlations amongst variables and factors.

The following labels will be used to denote Matrices in factor analysis (Tabachnick and Fidell, 2007):

- R – correlation matrix (Matrix of correlation between variables)
- A – factor loading matrix (Matrix of regression-like weights used to estimate the unique contribution of each factor to the variance in a variable. In this study the factors are created to be orthogonal, and therefore this is also the correlations between the variables and the factors).

2 LITERATURE REVIEW

2.1 Introduction

The literature review attempted to uncover any known weaknesses of performing EFA on ordinal data, and in so doing hoped to guide the design of the methodology for this simulation study.

2.2 The use of Ordinal Data in Multivariate Analysis

Treating ordinal data as continuous data in multivariate analysis potentially violates the assumptions of multivariate normality as discussed by many authors (Labovitz, 1967, Somers, 1968, Hawkes, 1971, Wilson, 1971, Olsson, 1979, Bollen and Barb, 1981, Vierra and Carlson, 1981, Young, 1981, Babakus et al., 1987, Potthast, 1993, Velleman and Wilkinson, 1993, Dolan, 1994, Hutchinson and Olmos, 1998, Lubke and Muthén, 2004, Stacey, 2005, Temme, 2006, Tabachnick and Fidell, 2007). The subject is not new and has had much methodological debate beginning with the work of Stevens (1946). One of the main strategies of dealing with ordinal data is to assume it to be interval in nature, and continue with the various parametric and multivariate statistical methods. Labovitz (1967, 1971) seems to be quoted the most as being at the forefront of this strategy since part of his research showed that “certain assumptions of both descriptive and inference statistics *can* be violated without unduly altering the conclusions” (Labovitz, 1967 ,pg. 151). Yet Labovitz does not blindly suggest converting all ordinal to interval scales regardless of the underlying distributions, and makes it clear to his critics in his 1971 paper that “the procedure is not risky if care is taken to avoid extreme exponential distributions” (Labovitz, 1971, pg. 521). In response to a critic, Mayer (1970), who criticised this strategy stating that not all statistics had been tested, Labovitz replied that he too welcomed further evidence. It is therefore in line with this statement that this study attempts to add greater clarity to a statistical method that has not been fully analysed under ordinal data conditions.

Harwell and Gatti (2001) attempted to estimate the prevalence of ordinal data being used as interval data. They limited their analysis to three publications in 1997 only; American Educational Research Journal, Sociology of Education, and Journal of Education

Psychology. They found that 73% of the dependant variables used in studies published in these three articles in that year alone appeared to be measured using an ordinal scale, yet the statistical tests that were used required assumptions of normality.

The purist's view to this alternate strategy of using the parametric approach would be to use the non-parametric measures exclusively to analyse ordinal data (Kim, 1975). Yet Kim found that there was no ordinal multivariate-analysis method that compared to those like multiple regression, and that the practical advantages of using such multivariate techniques outweighed the compromises made in violating the measurement assumptions. Kim goes on to advocate the use of the parametric strategy, not because it produces minor errors, but because of the power of the techniques.

Both Nunnally (1975) and Hensler and Stipak (1979) acknowledge the power that interval level analysis brings, even though interval level measurement has not been achieved. Hensler and Stipak (1979) goes on to state that because of the clear advantages of interval-level statistics, data analysts will continue to use them.

Hensler and Stipak (1979) provides a cautionary note to researchers though who unthinkingly use the pre-existing code values that may already exist on datasets since they may have been nonlinearly distorted from the underlying variables through the process of categorization. The authors do however consider the most prevalent practice, of assigning equal interval values as a coding scheme, to be preferable as a general strategy, but continue by advising that the researcher needs to understand the loss in information in doing so.

Following on from this, care should also obviously be taken by a researcher in the construction of the scale used to observe the data (Kim, 1975). Kim states that the entire purpose of analysing the association between two observed (measured) variables is to understand the nature of the underlying unobserved variables. The extent to which the relationship between these unobserved variables can be understood depends however on the relationship between the underlying values and the observed values, and this can potentially be distorted through the improper use of scale (Kim, 1975). This is directly relevant to this study in that all exploratory factor analysis methods attempt to uncover and simplify underlying relationships between variables, yet if those relationships are altered

through the incorrect use of scale, then the technique may in fact be unable to show the underlying relationship or may yield erroneous results.

Kim continues by explaining how scale may distort the understanding of the underlying distribution. With interval data it can be assumed that the underlying distribution is the same as the observed distribution, but with ordinal data it is not possible to make this assumption due to the potential distortions created in applying the scale, as can be seen in Figure 1 (Kim, 1975).

Scale in Use	Underlying Pattern	Resultant Pattern between Measured Variables
Interval	Linear	Linear
	Monotonic	Monotonic
	Nonmonotonic	Nonmonotonic
Ordinal	Linear	Linear or monotonic
	Monotonic	Linear or monotonic
	Nonmonotonic	Nonmonotonic

Figure 1: Distortions in Applying Scale - redrawn from Kim (1975)

2.2.1 *Issues with Factor Analysis on Ordinal Data*

Some simulation studies using confirmatory factor analysis have shown that as the skewness or positive kurtosis of the observed ordinal variables moved away from zero, the estimation of the factor structure deteriorated from the known underlying unobserved variables (Babakus et al, 1987, Potthast, 1993, Dolan, 1994, Hutchinson and Olmos, 1998).

In one study using maximum likelihood factor analysis, the classification of data caused through assigning integer values to ordinal data created a goodness of fit problem with the model, and lead to an erroneous indication that more factors were needed than were actually required (Olsson, 1979). Although some other studies had shown the method to be

robust against the effects of ordinal data, these were based on methods that took variables from the same distribution (Fuller and Hemmerle, 1966), and were later proven to be incorrect when variables were used that were skewed in opposite directions and had high true loadings (Olsson, 1979).

Bartholomew et al. (2002) did not recommend performing factor analysis on categorical data since this was likely to result in biased estimates of the factor loadings.

Since exploratory factor analysis calculates a matrix of association between variables as a first step, and indeed this is why it is possible to provide a correlation matrix as the input to an exploratory factor analysis process, further evidence could be found of the potential impact that ordinal data could have on EFA by looking at the effects that ordinal data has on correlation. Bollen and Barb (1981) analysed the effects of categorization on correlation. They simulated a normally distributed continuous data set, consisting of two variables which were correlated with one another, from which they drew samples that they then categorized (collapsed) into a simple ordinal scale. The means and standard deviations of the correlation coefficients of the collapsed variables were then compared to the corresponding metrics of the original continuous data set. The simulation varied the magnitude of the correlation (0.2, 0.4, 0.6, 0.8, and 0.9) and the size of the categories (2 through 10). What they found was that the collapsed correlation coefficients would be smaller than the actual correlation coefficients, and that this difference would be greatest for low correlation and low number of categories, and smallest for high number of categories with high underlying correlation. It appeared there was a threshold at 5 categories, above which the results for the correlation were found to have an ‘acceptable’ difference (the collapsed correlation coefficient is 89% of the value of the underlying correlation – 11% under). Bollen and Barb (1981) also found that the standard deviation of the collapsed correlation coefficients would be worst when the underlying correlation is high and the number of categories is low, but came to the same conclusion about the threshold at 5 categories where the error was ‘acceptable’.

This would therefore lead to the belief that EFA, under these circumstances, could be expected to have very little error, but it does draw attention to the fact that the number of categories should not be below 5. As an example, this would impact on researchers that categorise Age or Salary into less than five categories and then attempt to run multivariate

analysis. Gorsuch (1983) cautions that low correlations (communalities) may allow for the result of factor analysis to capitalize on chance, and this may very well influence the result (this is discussed in more detail in section 2.2.2).

Yet Bollen and Barb (1981) goes on to highlight that their simulation used normally distributed continuous variables, and highlights results found in another study (Wylie, 1976) that showed that correlation errors (differences) occurred most when continuous variables were dichotomized and had extreme opposite skews. If it is accepted that this would then affect factor analysis, then this supports Olsson's (1979) findings sited earlier.

Bollen and Barb (1981) does go on to analyse a skewed set of real-life data containing two variables, in which they find the results to be more satisfactory than in their simulation, but they do not mention the amount of skew present in the data except to mention that it was positively skewed.

Perhaps a method of correlation worth investigation as an input to EFA would be the Spearman Rank Correlation method (Spearman, 1904) which analyses the rank and not the distance between scale points. It makes sense that a scale that has distorted the inter-category distance should use a method of analysis that ignores these distances (i.e. uses rank only), although this would constitute a loss of data (Hensler and Stipak, 1979), and does not account for the scale creating distortions in the relationship with the underlying distribution (Kim, 1975).

It can therefore be stated that, based on the literature review, there is an expected error or bias in the EFA results as a result of the discretisation of continuous data into ordinal data, but it is unclear from the literature review what magnitude of error could be expected.

It must be noted that a separate body of research exists that proposes alternative approaches to the factor analysis of ordinal data. Ordinal factor analysis, or more specifically latent class analysis and latent trait analysis, are approaches designed for the factor analysis of binary or categorical data. For polytomous categorical variables (variables with more than two categories) there are two main approaches, namely; item response function approach, and underlying variable approach. An explanation and comparison of these methods is provided by Moustaki (2000), Jöreskog and Moustaki (2001), or Bartholomew et al. (2002). These approaches are considered superior to performing EFA

directly on ordinal data, yet are not as readily available or as simple to implement, and in some cases may not allow for the analysis of models with more than two factors (Bartholomew et al., 2002). These approaches are not considered here though, since this research paper is concerned with highlighting the effects of EFA on ordinal data, and determining whether normal distribution fitting is able to improve the results of EFA on ordinal data. While it is highly recommended that these superior approaches be used by researchers, as cited previously, researchers continue to use EFA with ordinal level data. This research is aimed at providing guidance under these circumstances and therefore remains relevant.

Research Question 1:

What will the effects of ordinal level data be on the output of EFA?

Research Question 2:

Will EFA on the Spearman-rank correlation matrix of the ordinal level data yield better results than performing EFA on the ordinal level data directly?

2.2.2 *Validity and Reliability*

A researcher is interested in factor analysis as a method to better understand and perhaps simplify his data correctly to reveal the underlying constructs. Neuman (2006) says that "validity suggests truthfulness" (Neuman, 2006, pp188), and that this indicates how well the analysis represents or 'fits' the underlying construct, and therefore reliability and validity are central issues in the use of factor analysis as with any statistical instrument. Various references highlight the different types of validity that a researcher must take into account (for example, Bailey, 1987, Lewis-Beck, 1993, Neuman, 2006), but it is specifically the impact that factor analysis could have on the validity of an experiment using ordinal data that are of concern here. Does the analysis correctly represent the underlying construct?

The invalidity of an instrument is discussed at length in many references as a factor in both the internal and external validity of a study (for example, Bailey, 1987, Lewis-Beck, 1993, Neuman, 2006). Part of this instrument would be the scale used to gather the observations,

which in the context of this study are those specific scales that at best produce ordinal data. As an example, these could be a Likert scale, a summated rating, or a Guttman scalogram analysis technique. These scales, if used correctly, may certainly have measurement validity as an instrument since they measure what they are supposed to measure (Bailey, 1987, Lewis-Beck, 1994), since the measures are ordinal rather than interval level data (Bailey, 1987), statistical validity is threatened when one performs factor analysis since the data does not meet the statistical methods' assumptions – that factor analysis assumes an interval scale at worst (Neuman, 2006, Gorsuch, 1983). There are scales that produce interval level data as a way of avoiding such validity threats, such as Thurstone's method of equal-appearing intervals (Bailey, 1987), but this report is specifically interested in determining the extent of the threat of ordinal data on the statistical validity of studies that use EFA. This part of the instrument is therefore not of concern here, and it is assumed that at best ordinal level data is the input to this statistical analysis.

Another part of the instrument that potentially affects internal and external validity of a study is the method used for data analysis, and since reliability is necessary for validity (Neuman, 2006), it is also of interest in this study to determine the reliability of EFA. It would be impossible to evaluate the entire validity of the EFA method under all conditions since these would be specific to each researcher's study, and would depend in that specific case on whether the entire instrument was designed correctly by the researcher to produce valid results, i.e. representative of the underlying constructs in that specific instance. It is however possible to attempt to evaluate the reliability of EFA in isolation within the context of its use on ordinal data, and therefore the impact of this reliability of the method on the validity of the instrument, and as a result, the study.

Maximizing the generalizability of a factor analysis study is important since there are rarely significance tests, and there are many possibilities for capitalizing on chance (Gorsuch, 1983). Gorsuch (1983) discusses the issues that are more likely to make factor analysis more reliable between studies of different random samples from the same population (higher replicability):

- **Method of Factoring** - Gorsuch (1983) concluded that the component model will yield different results to the common factor model when a low number of variables (10 or 20) are used, and where some do not correlate well with other variables. The

author also found evidence that the common factor model may lead to more replicable factors than the component model when communalities were low. Gorsuch states also that the number of factors extracted interacts with the rotation, and that adding an additional factor to a model of only a few extracted factors will cause the factors to shift considerably when rotated, but adding to a moderate number of extracted factors will not. He concludes that it appears that extracting too few factors decreases replicability. No differences between the replicability of oblique and orthogonal factor rotations can be given due to lack of evidence, yet some evidence does suggest that rotated factors are more stable than unrotated factors. Factor analysis does however capitalize on chance in the extraction procedure and in the rotation procedure, and this may result in variability of results.

- **Importance of communalities** – Gorsuch (1983) states that when the communalities are high, the unique factor weights are low in the common factor model. In this model the chance correlations are multiplied with the unique weights, so therefore if the communalities are high then these result in a lower chance correlation and therefore greater replicability of the factors. The author notes that it also appears from cited research that, adding variables with low communalities may decrease the replicability through increasing the capitalization on chance, and that high communalities increase stability and replicability. “Therefore, variables without a prior history of good reliability estimates and good correlations with other variables in the analysis are not desired in a factor analysis” (Gorsuch, 1983, pg. 331). In exploratory factor analysis the luxury of any ‘prior history’ of variables may not exist, and other precautions would need to be employed as mentioned in this section to ensure replicability.

- **Variables per Factor** – Gorsuch (1983) states that the larger the number of variables in the study (he considers 30 to be moderately large) the less dependent the study should be on the choice of method and the more replicable the results should be. As the number of variables per factor increases it will increase the unique rotational position, reduce the capitalization on chance, and increase the replicability. In the common factor model the greater the number of factors the greater will be the

factor scores that are determined, which leads to greater replicability. The goal would be to have a minimum of four to six variables per factor (Gorsuch, 1983).

- **Number of individuals** – The number of variables has a similar increase in the expected replicability, and should not be below 5 individuals per variable, or below 100 individuals for an analysis (Gorsuch, 1983). These minimums only apply when the expected communalities are high and the number of variables per factor is sufficient, otherwise they should be increased.

Gorsuch (1983) suggests that for analysing data that may have issues with replicability, to split the data in half and run factor analysis on these two halves. The study should then only proceed with those factors that matched across the two studies (sample sizes would need to be sufficient to ensure significance). Any further statistics required for another study could now be taken by running factor analysis on the entire data set, but only extracting and analysing those factors identified in the first analysis.

2.2.3 *Parameters to consider for the methodology*

Certain parameters were found in the literature worth consideration for the methodology of this study. Although no study was found that identically matched a methodology for obtaining the outcomes this study was targeting, various smaller 'components' were taken from the several different studies referenced up to this point. The 'components' for consideration were;

- **Sample size** – Flora and Curran (2004) used sample sizes of 100, 200, 500, and 1000, in a one and two factor design. Muthen and Kaplan (1992) used a 500 and 1000 sample size for a two and three factor design, and reported that Harlow (1985) used a 200 and 400 sample size in a two factor design.

  Gorsuch (1983) advised against any study with less than 100 respondents.

- **Number of Categories** – Bollen and Barb (1981) highlighted the potential issues with the number of categories chosen and the effects on the correlation coefficient,

and it was postulated earlier that although that study did not specifically analyse the effects on factor analysis, this could have an effect on the factor analysis output since correlation is the underlying input for this multivariate technique. Their study analysed 2, 3, 4, 5, 6, 7, 8, 9, and 10 categories, and they found an apparent threshold was reached at 5 categories.

- **Correlation strength** – Gorsuch (1983) stated that variables with low correlation could influence the replicability of a factor analysis output (this is discuss in more detail under the section on validity 2.2.2). Bollen and Barb (1981) also showed the effects that ordinal data had on the strength of correlation, and therefore the potential effect this could have on factor analysis was postulated.

  Hair et al (2005) discussed the importance of statistical significance of factor loadings and provided guidelines of which factor loadings to consider given a specific sample size. For example, given a sample size of 200, 0.4 factor loading should be considered as the smallest factor loading to be considered to ensure significance. As a rule of thumb, the authors consider ±.30 and ±.40 factor loadings as minimally acceptable, with values exceeding ±.50 generally necessary for practical significance.

- **Scale distortion** – Flora and Curran (2004) manipulated the unobserved continuous distribution, and then applied thresholds, to generate varying observed ordinal variables with varying levels of skewness and kurtosis. In this way it directly tests the robustness of the underlying unobserved variables being non-normally distributed. There were five different continuous distributions used as input; multivariate normal, and four with different levels of non-normality (low skewness vs. low kurtosis, moderate skewness vs. low kurtosis, low skewness vs. moderate kurtosis, moderate skewness vs. moderate kurtosis). These were chosen to test the separate effects of skewness and kurtosis, and were indicative of those distributions encountered in applied psychology. Thresholds were used to transform the continuous 'unobserved' distributions into ordinal data of five categories (Flora and Curran, 2004).

Muthén and Kaplan (1985, 1992) used a continuous normal distribution and varied the inter-category distances to produce ordinal data of varying levels of skewness and kurtosis.

As an example, scale distortion would be the effect that the questionnaire design would have on not correctly representing the underlying attitudes of the respondents. When the respondents 'fit' their attitudes to the ordinal scale provided for each response item the scale may distort the true underlying attitudes' distribution. Muthen and Kaplan (1992) refers to an unpublished study by Harlow (1985) that chose skewness values ranging from -2.0 to +2.0 and kurtosis values ranging from -1.0 to +8.0. Muthen and Kaplan (1992) considered these settings to be 'moderately normal', and used 6 cases of scale distortion in their study with skewness and kurtosis respectively of; 0.0 and 0.0, -0.74 and -0.33, -1.22 and 0.85, -2.03 and 2.9, 0.0 and 2.79, 0.0 and -1.3. Both these studies varied the inter-scale distances of the categories applied to the normal distribution, and did not vary the underlying distribution itself, to effectively vary the input distribution or scale distortion. Flora and Curran (2004) created underlying distributions with varying skewness and kurtosis, then applied standard equal distance category thresholds to generate the ordinal data. They utilized input skewness and kurtosis respectively of; 0.0 and 0.0, 0.75 and 1.75, 0.75 and 3.75, 1.25 and 1.75, 1.25 and 3.75. This was not what was actually tested though, since the categorisation changed the skewness and kurtosis values from what was originally intended – a categorised result will have a different skewness and kurtosis to the original continuous distribution because it has been categorised and the continuous distribution has been collapsed into an ordinal distribution of only a few numbers.

Since it was the intention of this study to provide a control which isolated the affects of categorisation, it was thought more prudent to follow the methods of Muthen and Kaplan (1992) of varying categories rather than varying the underlying distribution, since this allowed the control to be conducted on a constant normal distribution without the effects of scale distortion, and mimicked the in-field practice of incorrectly applying scale to normally distributed latent variables (underlying attitudes). It also provided a meaningful control input to factor analysis (continuous normal distribution) that did not violate any of the multivariate assumptions of the method.

The literature review also uncovered another form of scale distortion which was expected to have an effect on factor analysis (Olsson, 1979, Wylie, 1976) and distribution fitting, that of an oppositely skewed set of item responses, which follows the real-life possibility of respondent's inconsistently applying the scale from question to question to represent their underlying attitudes or beliefs.

2.3 Distribution Fitting Prior to Factor Analysis on Ordinal Data

It has been previously mentioned that researchers facing ordinal data have a choice of using either parametric or non-parametric approaches to analyse their data. One potential option is to attempt to transform the ordinal data into interval data (Gautam et al., 1996, Hensler and Stipak, 1979, Stacey, 2005) or to rescale the ordinal data (Stacey, 2005) to allow for the correct use of more powerful statistical analysis techniques. Hensler and Stipak (1979) stated that using only rank-order statistics on an ordinal scale discards information about inter-category distances. Scaling attempts to correct this while minimizing nonlinear distortion. Hensler and Stipak (1979) refered to this as attempting to preserve the inter-category distances, while trying to construct observed variables that are linear transformations of their latent (underlying) variable (but which have an error term that represents the loss of intra-category information). Therefore to attempt to increase the external validity of this study, it would be important to analyse as many of the options available to a researcher, and it is therefore important to test for the robustness of EFA following some sort of data rescaling.

Various methods exist in the literature for rescaling ordinal data. Hensler and Stipak (1979) demonstrated a method using the frequency of the observed distribution to find category thresholds on a table for area under a standard normal curve, and Bendixen and Sandler (1995) used correspondence analysis as a method for rescaling. Stacey (2005) used an algorithmic approach, which he compared against the correspondence analysis method for transformation of ordinal variables (Bendixen and Sandler, 1995) and found it to be superior.

Stacey (2005) did not test the distribution fitting method as an input for EFA, nor did he test the model with multivariate data.

Stacey's (2005) normal distribution fitting algorithm (NDF) is based on using a linear solving algorithm (such as Solver for Excel®) to find the best distribution fit between a normal distribution and the observed categorical data. Conceptually this is shown in Figure 2.

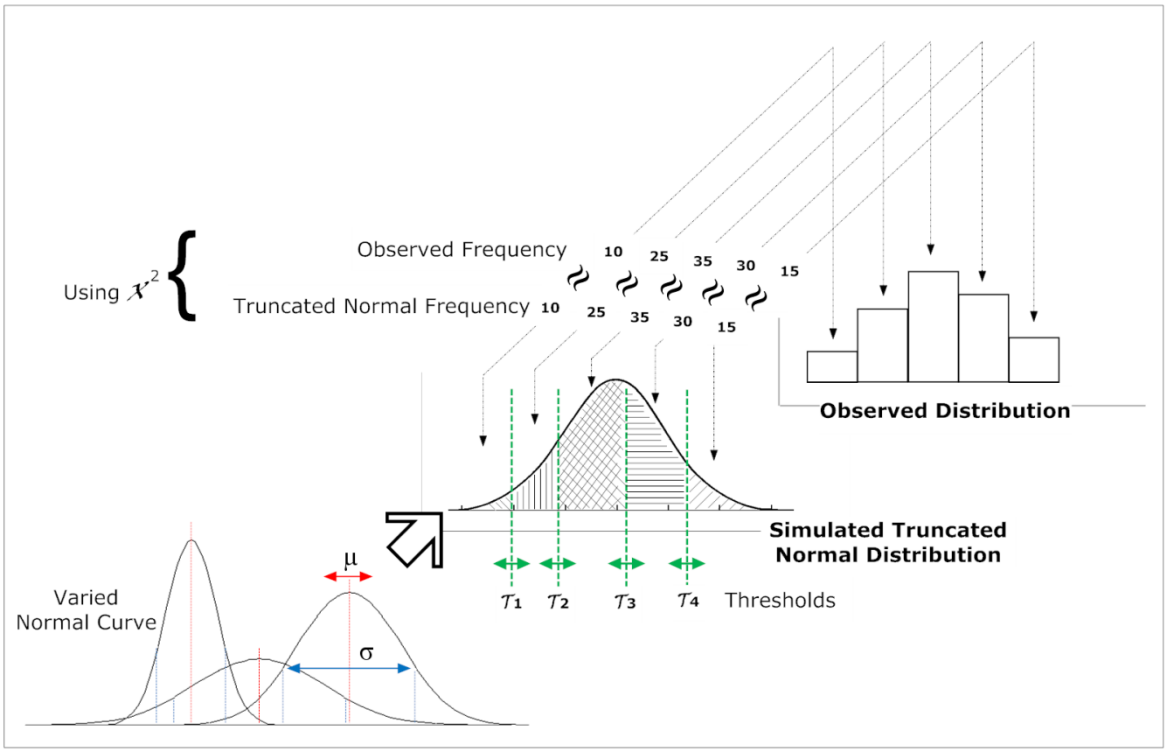


Figure 2 - NDF Method - Conceptual Diagram

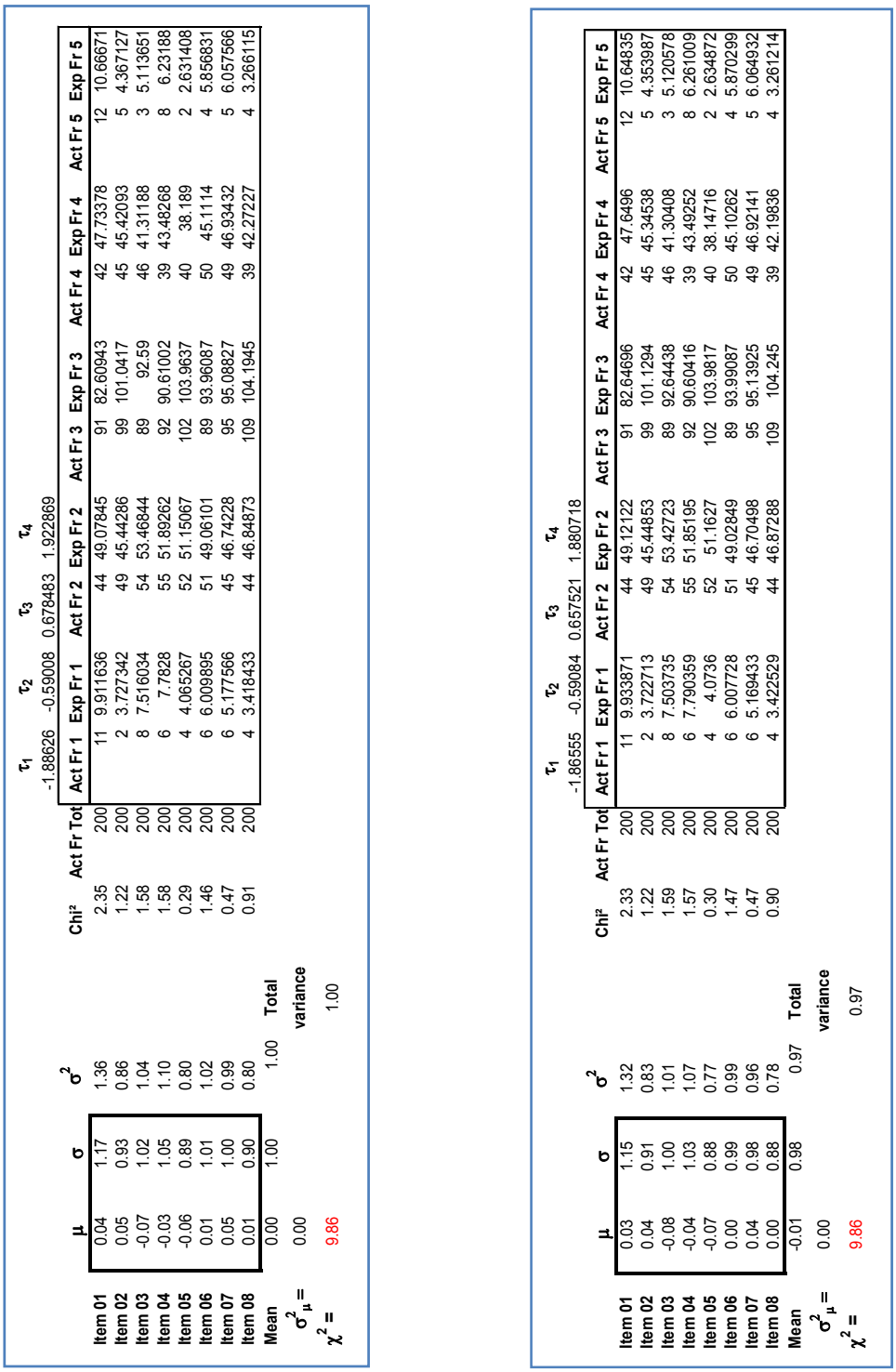
It begins by calculating the observed frequencies for every ordinal response (say 1-5 in a 5 category questionnaire) for every item (question). It then tabulates this against the expected frequency for the same responses and items, given a standard normal distribution of a given mean, standard deviation, and set of thresholds. At the start of running the algorithm these input parameters for the normal distribution (mean, standard deviation, and thresholds) are guessed.

Figure 3 shows a typical layout of an NDF for a 5 Category solution with 8 questions. The values show the initial 'guesses' prior to solving.



Figure 3 - NDF Initial setup

The two frequencies (expected and observed) are compared using the χ^2 statistic (Chi Squared) which is a 'goodness of fit' measurement.



						τ1	τ2	τ3	τ4										
						-1.88626	-0.59008	0.678483	1.922869										
	μ	σ	σ²		Chi²	Act Fr Tot	Act Fr 1	Exp Fr 1	Act Fr 2	Exp Fr 2	Act Fr 3	Exp Fr 3	Act Fr 4	Exp Fr 4	Act Fr 5	Exp Fr 5
Item 01	0.04	1.17	1.36		2.35	200	11	9.911636	44	49.07845	91	82.60943	42	47.73378	12	10.66671
Item 02	0.05	0.93	0.86		1.22	200	2	3.727342	49	45.44286	99	101.0417	45	45.42093	5	4.367127
Item 03	-0.07	1.02	1.04		1.58	200	8	7.516034	54	53.46844	89	92.59	46	41.31188	3	5.113651
Item 04	-0.03	1.05	1.10		1.58	200	6	7.7828	55	51.89262	92	90.61002	39	43.48268	8	6.23188
Item 05	-0.06	0.89	0.80		0.29	200	4	4.065267	52	51.15067	102	103.9637	40	38.189	2	2.631408
Item 06	0.01	1.01	1.02		1.46	200	6	6.009895	51	49.06101	89	93.96087	50	45.1114	4	5.856831
Item 07	0.05	1.00	0.99		0.47	200	6	5.177566	45	46.74228	95	95.08827	49	46.93432	5	6.057566
Item 08	0.01	0.90	0.80		0.91	200	4	3.418433	44	46.84873	109	104.1945	39	42.27227	4	3.266115
Mean	0.00	1.00	1.00	Total												
				variance												
σ^2_μ =	0.00															
χ^2 =	9.86			1.00												

						τ1	τ2	τ3	τ4										
						-1.86555	-0.59084	0.657521	1.880718										
	μ	σ	σ²		Chi²	Act Fr Tot	Act Fr 1	Exp Fr 1	Act Fr 2	Exp Fr 2	Act Fr 3	Exp Fr 3	Act Fr 4	Exp Fr 4	Act Fr 5	Exp Fr 5
Item 01	0.03	1.15	1.32		2.33	200	11	9.933871	44	49.12122	91	82.64696	42	47.6496	12	10.64835
Item 02	0.04	0.91	0.83		1.22	200	2	3.722713	49	45.44853	99	101.1294	45	45.34538	5	4.353987
Item 03	-0.08	1.00	1.01		1.59	200	8	7.503735	54	53.42723	89	92.64438	46	41.30408	3	5.120578
Item 04	-0.04	1.03	1.07		1.57	200	6	7.790359	55	51.85195	92	90.60416	39	43.49252	8	6.261009
Item 05	-0.07	0.88	0.77		0.30	200	4	4.0736	52	51.1627	102	103.9817	40	38.14716	2	2.634872
Item 06	0.00	0.99	0.99		1.47	200	6	6.007728	51	49.02849	89	93.99087	50	45.10262	4	5.870299
Item 07	0.04	0.98	0.96		0.47	200	6	5.169433	45	46.70498	95	95.13925	49	46.92141	5	6.064932
Item 08	0.00	0.88	0.78		0.90	200	4	3.422529	44	46.87288	109	104.245	39	42.19836	4	3.261214
Mean	-0.01	0.98	0.97	Total												
				variance												
σ^2_μ =	0.00															
χ^2 =	9.86			0.97												

Figure 4 - NDF after Solving (right), & after Standardisation (left)

The linear solving algorithm now solves to reduce the total χ^2 number by manipulating the mean and standard deviation of the normal distributions for each item, and the thresholds for all items. Once solved (Figure 4), 8 different means and standard deviations are left (one for each item), with one set of thresholds, that best match the observed data. As the last step of the NDF the distributions are standardized by making changes to the;

- Thresholds … $New\ \tau_n = {(Old\ \tau_n - \mu_T)}/{\sqrt{\sigma_T^2}}$
    - Where τ_n is the nth Threshold ('Old' being the value prior to standardisation, and 'New' being the new standardised value), μ_T is the total mean across all 8 normal distributions prior to standardisation, σ_T^2 is the total variance across all 8 normal distributions prior to standardisation.
- Means… $New\ \mu_n = {(Old\ \mu_n - \mu_T)}/{\sqrt{\sigma_T^2}}$
    - Where μ_n is the nth mean ('Old' being the value prior to standardisation, and 'New' being the new standardised value), μ_T is the total mean across all 8 normal distributions prior to standardisation, σ_T^2 is the total variance across all 8 normal distributions prior to standardisation.
- Standard Deviation… $New\ \sigma_n = {Old\ \sigma_n}/{\sqrt{\sigma_T^2}}$
    - Where σ_n is the nth standard deviation ('Old' being the value prior to standardisation, and 'New' being the new standardised value), σ_T^2 is the total variance across all 8 normal distributions prior to standardisation.

Once standardised, a threshold set for a normal distribution with an overall mean of 0 and variance of 1 remains (Figure 4). Using the threshold, scaled responses can be 'created' from the old responses.

Research Question 3:

Will the normal distribution fitting algorithm proposed by Stacey (2005) improve the results of the Factor Analysis over the other two methods (No transformation, and Spearman-rank)?

2.4 Conclusion of Literature Review

The literature review revealed little detailed information about the effects of ordinal data on EFA, but it did allude to expected difficulties that could be uncovered in the study.

The study is certainly worthwhile, having shown that factor analysis is widely used, and that ordinal data is routinely treated as interval level data.

2.4.1 *Research Question 1:*

What will the effects of ordinal level data be on the output of EFA?

(Bollen and Barb, 1981, Gorsuch, 1983, Muthén and Kaplan, 1985, Babakus et al., 1987, Muthén and Kaplan, 1992, Flora and Curran, 2004)

2.4.2 *Research Question 2:*

Will EFA performed on the Spearman-rank correlation matrix of the ordinal level data yield better results than performing EFA on the ordinal level data directly?

(Spearman, 1904, Bollen and Barb, 1981, Gorsuch, 1983)

2.4.3 *Research Question 3:*

Will the normal distribution fitting algorithm proposed by Stacey (2005) improve the results of the factor analysis over the other two methods (No transformation, and Spearman-rank)?

(Hensler and Stipak, 1979, Bendixen and Sandler, 1995, Gautam et al., 1996, Stacey, 2005)

3 RESEARCH METHODOLOGY

The study compared the robustness of EFA using varimax rotation when applied; directly to ordinal data, to the Spearman-rank correlation matrix of the same ordinal data, and to the transformed ordinal data using a distribution fitting algorithmic approach. Comparison was done between these methods and the known factors of the contrived 'latent' continuous population from which the ordinal data was derived. The ordinal data was derived through a discretisation process - allocating a simple interval scale (applying thresholds) to the contrived 'latent' continuous data.

3.1 Research methodology /paradigm

This study was run as a simulation using contrived data. The synthesis of data in favour of using empirical data was important to allow for the isolated test of certain parameters' effects on the EFA output. External validity of this study depended on it being able to isolate the parameter influence from other issues of measurement or sampling error that may have been difficult to isolate if empirical data were used.

3.2 Research Design

The methodology is shown in graphical form in Figure 5. The aim of the study was to test the effects that ordinal data has on Exploratory Factor Analysis, which in effect is actually testing the effects of incorrect categorization or scaling (which produce ordinal data) on Factor Analysis. A continuous normal latent (underlying) population is assumed, which was then discretised or categorised to produce the ordinal responses which was then analysed.

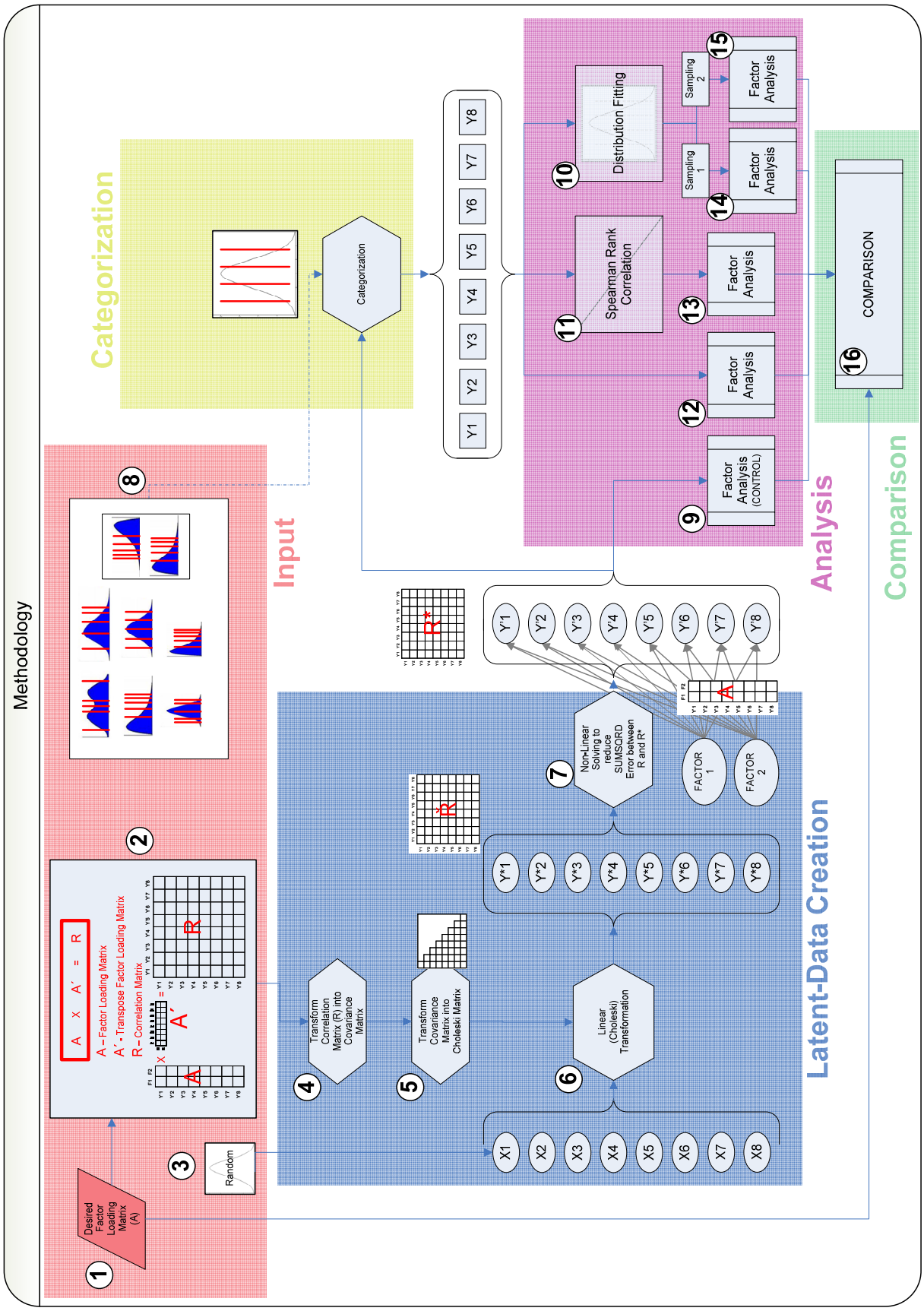
Varying the inter-threshold distances applied to obtain the categorical data is what produced the scale distortion (Muthén and Kaplan, 1992, Stacey, 2005).

The simulation was similar to those done before on confirmatory factor analysis (Muthén and Kaplan, 1985, Muthén and Kaplan, 1992, Flora and Curran, 2004), except that this study used EFA.

3.2.1 *Methodology Overview*

The methodology was implemented by building an automated model programmed in Visual Basic for Applications (VBA) that operated out of Microsoft® Excel® 2007, and once given a set of factor loadings and category thresholds, would automatically generate the necessary data. The factor analysis was performed using NCSS 2007, and the results were re-imported into Excel® files for storage and further analysis. All linear and non-linear solving was performed using Premium Solver Platform for Excel® version 7.1 (Frontline Systems).

Figure 5: Graphic Representation of the Simulation Methodology



3.2.2 *Methodology - Inputs to the study (Steps 1, 2, 3 & 8)*

The study compared the results of EFA on categorical data against the results of the same technique on the underlying continuous data from which the ordinal data was produced. The method started with a continuous population with a known factor loading structure (A).

Step 1

Item 1 in the schematic of the methodology (Figure 5) shows this factor structure input as the start to the entire process, and the comparison point for evaluating the results.

Step 2

From factor analysis theory[1] it is known that multiplying the factor loading matrix (A) by its transpose (A') yields the correlation matrix (R) of the data set if the factors are orthogonal (Tabachnick and Fidell, 2007). This second step was done in the various study's iterations (discussed later) whenever a new factor loading structure (A) and correlation structure (R) was analysed.

[1] For a simple explanation of the Factor Analysis method see GORSUCH, R. L. (1983) *Factor Analysis,* New Jersey, Lawrence Erlbaum Associated, Inc. , or TABACHNICK, B. G. & FIDELL, L. S. (2007) *Using Multivariate Statistics,* Boston, Pearson Education, Inc.



Figure 6 - Obtaining Correlation Matrix from Factor Loading Matrix

Step 3

The second input to the study was normally-distributed random continuous data with a mean of 0 (μ) and a standard deviation of 1 (σ), represented in Figure 5 by Step 3, and was obtained by using the Microsoft® Excel® Function "=NORMSINV(RAND())". It is known that criticism may exist for this function on its accuracy to generate normal data. Since the accuracy of the method has improved in later versions of Excel® to 14 or 15 decimal places (Microsoft, 2007), and since this study does not utilize this function in the extreme tails of the distribution, it is considered practically sufficient for its use here (West, 2005).

Step 4

The other main inputs to the study were the thresholds (Step 4) used to distort the continuous data in the categorisation step. These will be discussed later in this section.

3.2.3 *Methodology - Creating the latent continuous data (Steps 5 to 8)*

The creation of the latent data was done by taking the desired factor loading structure's correlation matrix (R) gained in Step 2 and using this to synthesise the perfect samples for the study by applying Steps 5 – 8. These steps are crucial to removing sampling error from each sample in an attempt to isolate only those effects or errors that are to be observed (namely categorization error). In effect, each sample is exactly representative of the underlying population since sampling error is entirely removed, and therefore follows exactly the intended factor structure introduced into each model. Each sample then can be considered as a population, with each population used in each iteration of every model being identical in their underlying parameters.

Step 5

The correlation matrix that formed the input from Step 2 consists only of the common components and so the diagonal of the correlation matrix will not be 1's after the matrix multiplication has created it, but will be less than 1 since they exclude the unique components (Figure 6, shown earlier, is an example of this).

For this reason the diagonals are forced to 1's in Step 5 for the purposes of generating the covariance matrix. Generating the covariance matrix from the correlation matrix is necessary since it is required as the input for the next step, the Choleski transformation[2]. The covariance matrix (variance-covariance) was obtained through multiplying the correlation (r) between two variables with the standard deviation (σ) of the two variables. Therefore the diagonals where r = 1 are the variance of each variable (σ^2), and off diagonal elements are $(r_{AB} \times \sigma_A \times \sigma_B)$ which is the covariance of each pair of variables.

[2] A Choleski Transformation is a linear transformation of random data (X) performed to obtain a new set of data with a known covariance matrix (Y*) – i.e. multivariate data.

Step 6

Once the covariance matrix was obtained it is used to produce a Choleski matrix (Step 6 in Figure 5), which is a lower half matrix only (Figure 7). Here the bottom left corner of the example Choleski matrix shown in Figure 7 are zero's because of the method of 'pure' loading utilized in this study, i.e. variables do not cross load onto both factors (see Section **Error! Reference source not found.**).

Choleski Decomposition

1.003551604	0	0	0	0	0	0	0
0.613539402	0.73660669	0	0	0	0	0	0
0.643320199	0.301409207	0.711121752	0	0	0	0	0
0.652696959	0.305802419	0.202522599	0.692479415	0	0	0	0
0	0	0	0	0.940591653	0	0	0
0	0	0	0	0.483018335	0.859301873	0	0
0	0	0	0	0.456981326	0.267356291	0.767763	0
0	0	0	0	0.445692126	0.26075156	0.185308	0.725504

Figure 7 - Choleski Matrix

Three methods for obtaining the Choleski matrix were found (Hunt, 2001, Lynch, 1994, Moore, 2001) but the simplest to incorporate into an automated model was the code shown by Moore (2001) which could be incorporated into VBA to run automatically when called by the model.

Step 7

Once the matrix is completed, the random data was then multiplied with the elements of the Choleski matrix, according to the Choleski method, to produce the final data set (Step 7).

Y*1 = $(CH_{11} \times X1)$

Y*2 = $(CH_{21} \times X1) + (CH_{22} \times X2)$

Y*3 = $(CH_{31} \times X1) + (CH_{32} \times X2) + (CH_{33} \times X2)$

…etc

Where Y^*n is the nth variable output of the Choleski transformation, and where $CH_{r,c}$ is the rth row and c^{th} column element of the Choleski matrix (CH), and Xn is the nth random variable generated in Step 3.

Step 8

The final step for creating each sample set for each model was to remove all remaining correlation deviation using non-linear solving in Step 8 (Premium Solver Platform). This is necessary since steps 5-7 do not produce a sample with a perfectly matched correlation matrix (shown as Ř in Figure 5) to the initial intended correlation structure (R); there is still an error component present (since the covariance matrix was obtained using standard deviations prior to linear transformation). Therefore the model was made to step through each sample set of every model and perform a non-linear solving routine to reduce the SUM² error between the actual correlation matrix (Ř) of each sample set and the intended correlation matrix (R) by manipulating the Choleski matrix. The result still contained a small error component due to the finite sample size, and hence the correlation matrix of this sample is referred to as R* and not R. This error will be shown later when the results are presented, but it is considered practically insignificant to the results of the study.

To skip Steps 5-7, although possible, would make the non-linear solving of each model extremely time consuming. It is simpler, and far quicker, to use these steps to start the non-linear algorithm much closer to the final solution (R*) and therefore reduce the required processing time to reach the final sample with the desired correlation matrix (R*).

The result is a set of continuous data that follow a multivariate normal distribution consisting of 8 variables (Y'1 – Y'8), with a known correlation matrix (R*), given an intended factor loading structure (A). This is considered as the latent or unobserved data, and would represent a respondent's underlying attitudes or beliefs which a researcher is trying to uncover through a Likert-type questionnaire (ordinal response scale). This data is the input to the categorisation process, just as a respondent's beliefs or attitudes would be the input when completing a questionnaire. For the purposes of this study it also formed the input to a EFA 'control process' which enables the measurement of both the discretisation error and the replicability of the EFA method.

3.2.4 *Methodology - Categorisation of the latent data (Step 9)*

The threshold inputs from Step 4 were applied to the latent continuous data from Step 8 to produce the ordinal or categorical data (marked Y1 – Y8 in Figure 5) as the output of the categorisation step (Step 9 – shown in yellow).

$$Y_{ij} = C_n \text{ if } \tau_{n-1} < Y'_{ij} \leq \tau_n$$

Where Y_{ij} is the simulated response generated in Step 9 of the ith respondent to the jth survey item, C_n is the nth ordinal response, τ_n is the threshold that differentiates C_k from C_{k+1} , Y'_{ij} is the simulated underlying attitude or belief generated in Step 8 of the ith respondent to the jth survey item. A practical example is included in the Appendix for ease of understanding (Section 7.1).

This data represents a respondent's answers to a set of categorical questions, and by varying and then applying the thresholds in step 4, the model produced a set of ordinal data that contained a discretisation error which represents:

- the loss of data through collapsing a continuous set of beliefs into a much smaller set of discrete ordinal responses, and
- The distortions in the respondents underlying attitudes and beliefs created through their interpretation of the questions posed or perhaps their cultural predisposition to how they answer questionnaires with rigid categorical responses.

3.2.5 *Methodology - Analysis of output (Step 10 to 16)*

The categorical data (Y) were then all analysed using EFA, using four different approaches:

- The first approach was to simply apply EFA to the ordinal data (Step 11).
- The second method attempted to compensate for the incorrect use of Pearson's correlation analysis on ordinal data by using Spearman's rank method which ignores inter category distance (inter category distance is potentially distorted through the categorisation process). In this approach the data was used to create a Spearman rank correlation matrix (Step 12) which then became the input to the factor analysis process (Step 13).

- The third and fourth methods both used Normal Distribution Fitting (NDF) to attempt to 'rescale' the ordinal data (Step 14) prior to factor analysis. The NDF was applied using the linear solving algorithm of Premium Solver in the manner set out by Stacey (2005), and is described in the literature review of this study. However, the NDF process produces thresholds, not data points, and so further sampling needed to take place before factor analysis could be done.
    - The third method sampled randomly within the truncated normal distribution formed by applying the thresholds obtained from the NDF process in Step 15. This produced true interval data, yet it was unclear prior to the analysis what this random sampling would do to the correlation structures and therefore the factor analysis output, and therefore a fourth method was introduced.
    - The fourth method (Step 16) assumed that random sampling, even within the confines of the thresholds produced by the NDF process, would further degrade the correlation structure by adding a random error. It also assumed that the rescaling exercise of the NDF process was necessary to bring the ordinal level data closer to interval level data before the factor analysis could take place. Therefore if it was accepted that the thresholds were accurate (Stacey, 2005), then the mean could be taken of the truncated normal distributions formed by applying the rescaled thresholds, and these would represent rescaled values for the original categorical data. This would return the correct inter-scale distance since instead of 1, 2, 3, 4, and 5, it would now have a more correctly scaled data set, perhaps 0.8, 1.1, 2.6, 3.7, and 4.8. Even though this would still represent pseudo categorical type data, it would be assumed to have retained the correlation structure of the original categorical data. Finding the mean of a truncated normal distribution is possible through calculating the integral formula[3]

      $$Y_{k,j} = \int_{\tau_{k-1}}^{\tau_k} x.e^{\frac{-(x-\mu_j)^2}{2\sigma_j^2}} dx \Big/ x.e^{\frac{-(x-\mu_j)^2}{2\sigma_j^2}} dx$$ (Stacey, 2005).

[3] This study estimated the mean for practical simplicity by obtaining 1000 random samples from each truncated normal distribution between each threshold and then obtaining the mean of each of these sample sets.

3.2.6 *Methodology – Comparison of Results (Step 17)*

Step 17

Step 17 then formed the point of comparison and analysis by which to compare; firstly the effects that ordinal data had on factor analysis, and secondly the improvement (if any) that the various methods studied here had on the results. The methods for analysis and comparison are discussed in more detail in Section 3.5 on Data analysis and interpretation.

It must be noted that factor analysis does not necessarily return the two factors in the order that was intended, i.e. what was used as input for factor 1 may be returned as factor 2. It may also not return the signs intended, i.e. two positively loaded variables may be returned as two negatively loaded variables. The model was designed to automatically switch signs and factor names depending on the relative strength of variable 1 and variable 8 (since these were never intended to be loaded on the same factor).

3.2.7 *Model Inputs and Constraints*

Given the methodology approach in the previous section, there were three inputs that could be varied to produce different results to analyse and compare, namely; different factor loading structures, different number of categories to collapse the data into, and differing amounts of scale distortion to apply by manipulating the thresholds. There are also various other constraints that remained fixed (sample size, number of variables, etc.) and these will be discussed further in this section.

Fixed constraints

Although sample size would have an impact on the stability of the results, just as it is a factor in any other statistical analysis, it was fixed here to 200 samples per analysis (iteration) purely to reduce the amount of data analysed. Gorsuch (1983) advised against any study with less than 100 respondents, and therefore a number of 200 was chosen which has been used in various other similar studies (Harlow, 1985, Flora and Curran, 2004).

Hair et al (2005) stated that factor loadings of ±.40 could be considered as a minimum for practical significance.

The number of variables (akin to the number of survey items in real-life) were fixed at 8, this allowed for a 2 factor design to have split loading with 4 variables loading uniquely onto each factor while still adhering to the minimums advised by Gorsuch (1983).

The number of simulations (iterations) performed per model was set to 100. This provided 100 different iteration results on every metric to allow the study to analyse the distribution/variability of each model, and to provide some idea of confidence intervals. This provided a measure of replicability based on the combination of input variables, and will be discussed further in Section 3.5. Only the random data input changed for each of the 100 iterations, the input factor structure, thresholds, and categories remained identical for all iterations (100 in total) of a specific model.

Model inputs

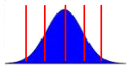
As discussed at the beginning of this section, there were three variable inputs to this study that could be varied for each model;

Factor Loading Structure – the number of variables loading onto each factor remained the same (4), as did the number of factors (2), but the individual loadings of each variable on each factor were varied for each model (remaining the same for each iteration). Factor loadings for this study were set to either 0.4, 0.5, 0.6, 0.7, 0.8 or 0.9 depending on the model being analysed (the models are discussed in more detail in the next section). As a rule, if variables 1 to 4 were loaded onto Factor 1, then variables 5 to 8 would be given factor loadings of 0 for Factor 1 to ensure that the design did not intend[4] to cross-load any factors.

[4] The term 'intend' is used here since although the model may have used 0 (zero) as a loading input for a variable on a particular Factor when the data was generated, the results of any of the EFA outputs may have revealed a positive or negative loading since the categorization degraded the data and therefore could provide the opportunity for the factor analysis process to capitalize on chance, and perhaps create a spurious loading.

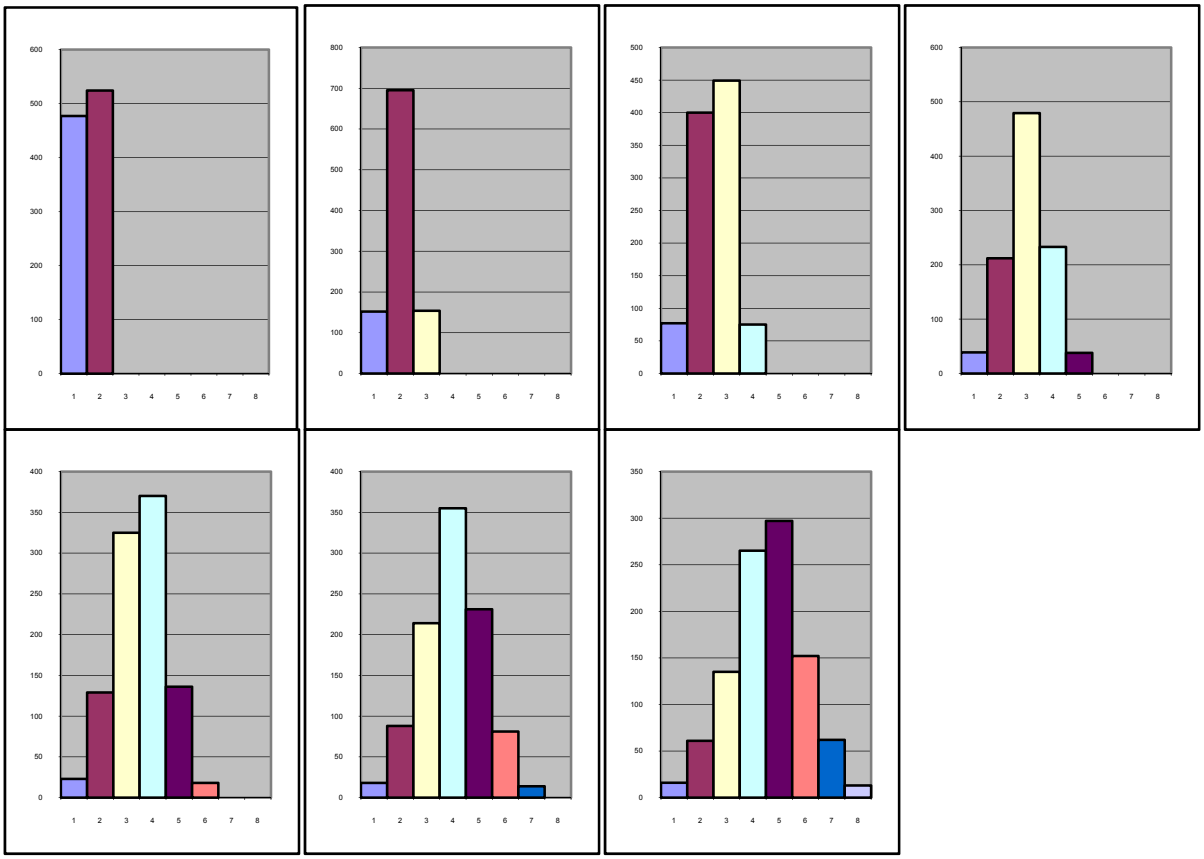
Numbers of Categories – the number of categories were varied in each iteration to produce a range of outputs for comparison. The categories chosen for the study were 2, 3, 4, 5, 6, 7, and 8.

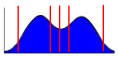
Scale Distortion – was achieved by manipulating the thresholds applied to the continuous data during the categorization process. As discussed in the literature review, various different studies have applied different scale distortions in two different methods; firstly varying the underlying continuous distribution, and secondly by manipulating the thresholds applied to a fixed normal distribution. This study adopted the later for reasons already discussed. It should be noted that both the variation in random data and the categorization process affected the final skewness and kurtosis from what was intended, and so the numbers provided below are merely a guideline to the intent (actual resultant skewness and kurtosis figures are provided later). The various different methods of manipulation were;

- Little /No Scale Distortion 

The following thresholds were applied for the different desired categories to create as little distortion on the distribution as possible (as close to skewness 0, kurtosis 0 as possible):

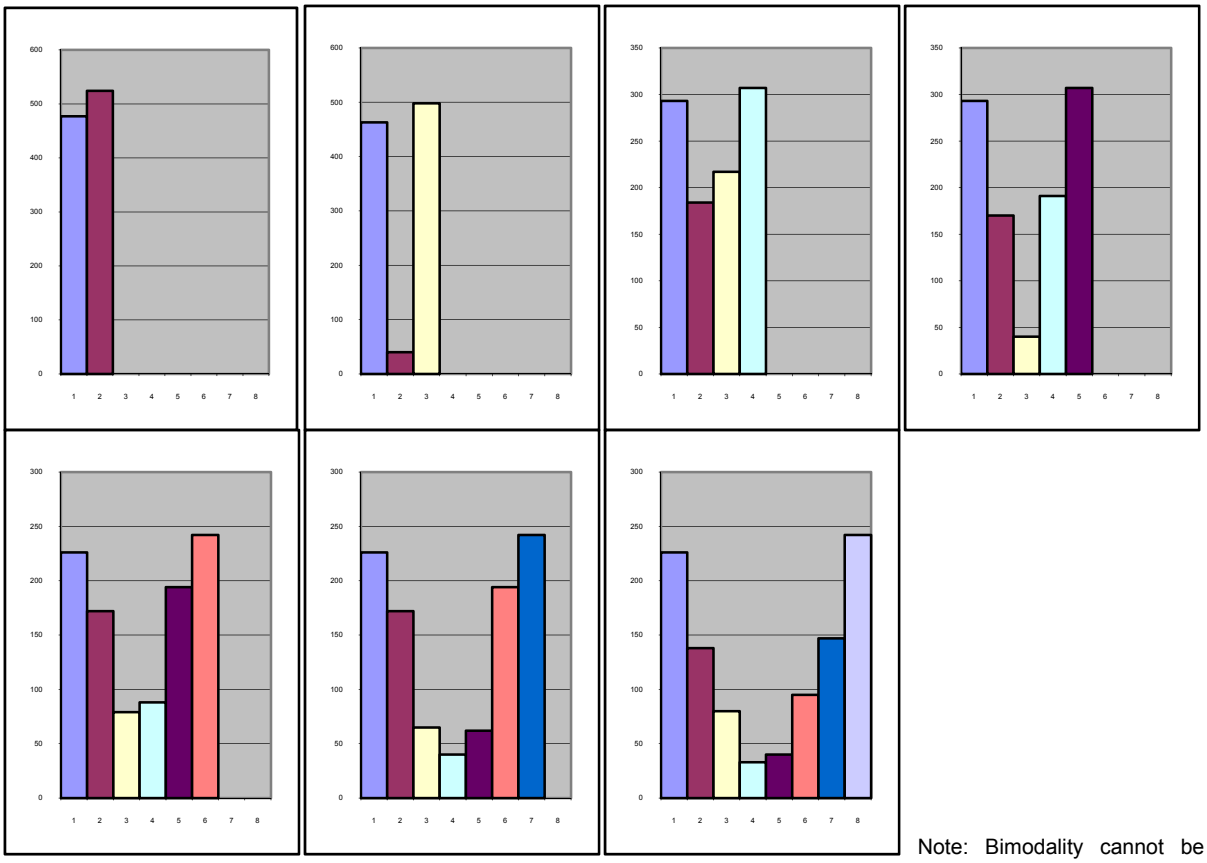
	2 Category	3 Category	4 Category	5 Category	6 Category	7 Category	8 Category
Threshold 1	0	-1	-1.5	-1.8	-2	-2.14286	-2.25
Threshold 2		1	0	-0.6	-1	-1.28571	-1.5
Threshold 3			1.5	0.6	0	-0.42857	-0.75
Threshold 4				1.8	1	0.42857	0
Threshold 5					2	1.28571	0.75
Threshold 6						2.14286	1.5
Threshold 7							2.25



- Bimodal Distribution 

  The following thresholds were applied for the different desired categories to achieve a bimodal distribution:

	2 Category	3 Category	4 Category	5 Category	6 Category	7 Category	8 Category
Threshold 1	0	-0.05	-0.5	-0.5	-0.7	-0.7	-0.7
Threshold 2		0.05	0	-0.05	-0.2	-0.2	-0.3
Threshold 3			0.5	0.05	0	-0.05	-0.08
Threshold 4				0.5	0.2	0.05	0
Threshold 5					0.7	0.2	0.08
Threshold 6						0.7	0.3
Threshold 7							0.7

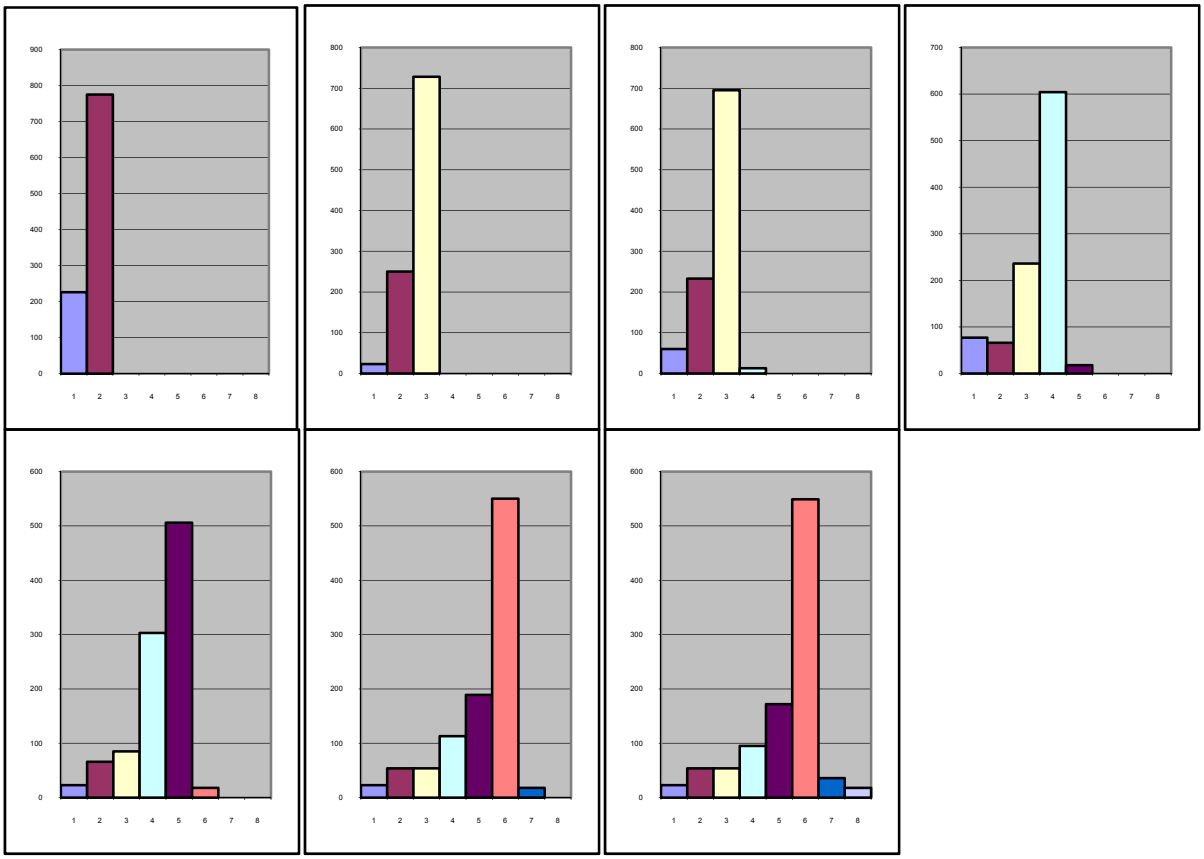


Note: Bimodality cannot be achieved on a 2 Category Distribution.

- Skewed Distortion 

The following thresholds were applied for the different desired categories to achieve a desired skewness of -1.22 and a kurtosis of 0.85:

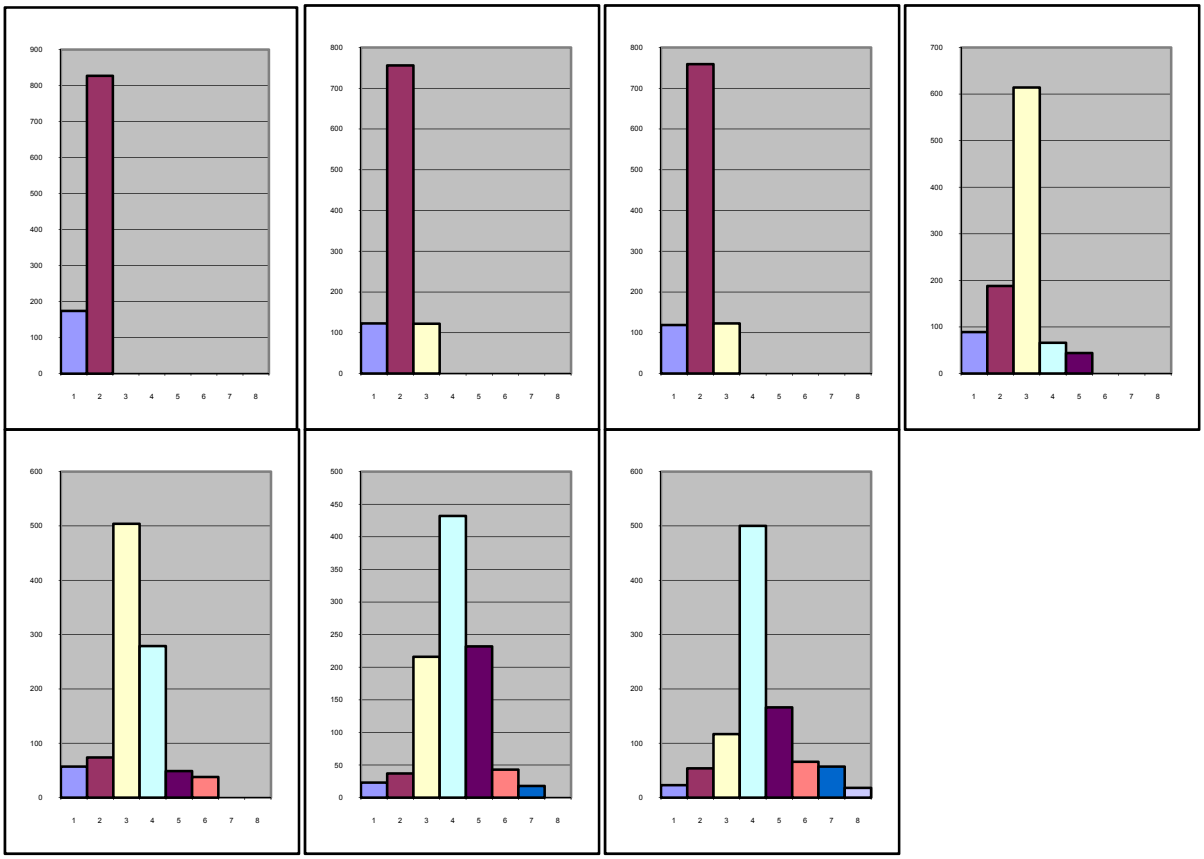
	2 Category	3 Category	4 Category	5 Category	6 Category	7 Category	8 Category
Threshold 1	-0.7	-2	-1.6	-1.5	-2	-2	-2
Threshold 2		-0.55	-0.5	-1.05	-1.4	-1.5	-1.5
Threshold 3			2.2	-0.25	-0.89	-1.1	-1.1
Threshold 4				2	0	-0.62	-0.7
Threshold 5					2	-0.1	-0.2
Threshold 6						2	1.65
Threshold 7							2



- Kurtotic Distortion 

  The following thresholds were applied for the different desired categories to achieve a kurtosis of 0.85 and a skewness of 0:

	2 Category	3 Category	4 Category	5 Category	6 Category	7 Category	8 Category
Threshold 1	-0.89	-1.15	-1.19	-1.40	-1.65	-2	-2
Threshold 2		1.15	1.13	-0.54	-1.1	-1.6	-1.5
Threshold 3			3.91	1.20	0.35	-0.54	-0.8
Threshold 4				1.72	1.36	0.54	0.5
Threshold 5					1.8	1.6	1.05
Threshold 6						2	1.5
Threshold 7							2



The actual results for kurtosis and skewness naturally varied between the various category sizes from the intended numbers. The actual resulting skewness and kurtosis numbers can be seen in Section 4.3.

3.2.8 *Model Outputs*

Each model produced all the information necessary to analyse its part of all three questions, i.e. it produced the outputs as shown in the tree below (Figure 8).

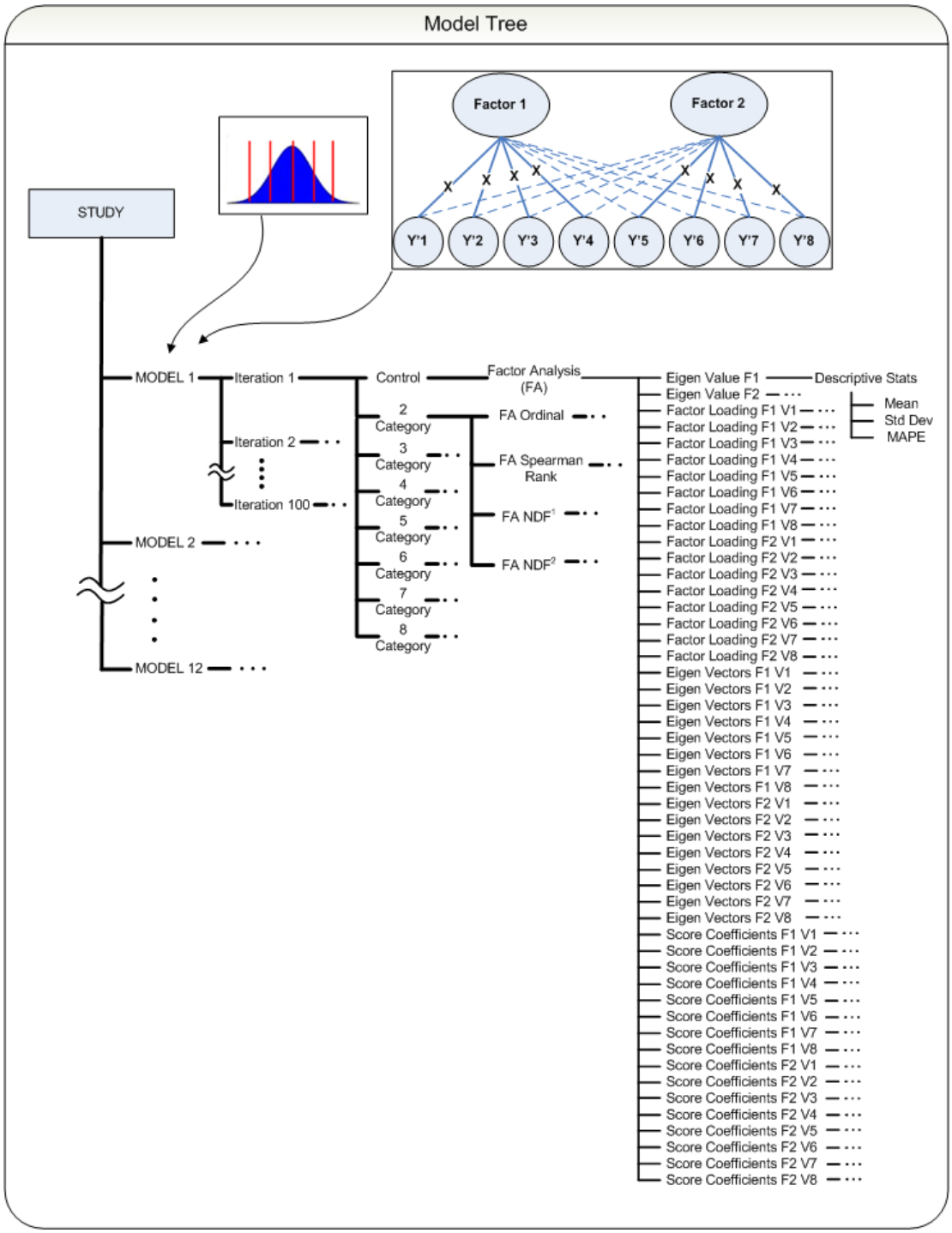


Figure 8 - Model Tree

Therefore, each model produced (reading from left to right in Figure 8); 100 iteration outputs, each consisting of 7 category outputs, each of these analysed by the 4 methods, and each containing a control, each of those producing 50 EFA metrics (of which the factor loadings are the metric most utilized – 16 in total).

Therefore one model produced 100 X 7 X 5 X 200(samples) X 8(variables) = 4,480,000 data points for factor analysis, and required 3,500 separate EFA's to be performed per model. Once passed through EFA it produced a control plus 4 sets of method results (for each analysis method; ordinal, Spearman, NDF^1, and NDF^2), for every category size (2-8), each set containing 100 sets of eigen values, factor loadings, factor coefficients and eigen vectors. This study was only concerned with analysing the factor loading's output as a means of evaluation and comparison, therefore the eventual output of each model was 100 X 16 (factor loadings) X 7 (categories) X 4 (methods) + 100 control points = 44,900 analysed points split into the various distributions and descriptive statistics as required to answer or analyse each question (factor loading's means and standard deviation mainly - discussed later in Section 3.5 Data analysis and interpretation).

The large size of data created is the reason that sample size, pertinent to any statistical investigation, was not treated as another variable to be analysed here but was fixed for the entire study to reduce the amount of data being analysed (200 sample size).

3.3 Population and sample

3.3.1 *Population*

The data for this study was created to simulate sets of underlying attitudes of respondents, yet it is possible to think of the study as having 12 separate populations (1 for each model), since in this case a population would be any set of responses that follow the same underlying characteristics of; factor structure, and category distortion (designed underlying attitudes). The population for each model in this study was therefore different to other models.

Each population was a contrived set of continuous data, made up of 100 sample sets created with known relationships between 8 variables and two unobserved factors (factor structure). Since the methodology aimed to remove as much sampling error as was practical, the population is almost exactly represented by each of the 100 samples created for it. The samples were not randomly selected from a larger sample frame as would be done in a bootstrap resampling. This study would more closely resemble a Monte Carlo simulation in the way each analysis (sample) was created based on a set of governing rules (population).

3.3.2 *Sample*

Sets of samples were created repeatedly (100 times) according to the process outlined in Section 3.2.3, and according to the set input parameters for each model (those of the population). The total sum of these 100 samples could be considered as the population for each model, each exactly representative of their population (although different from each other due to their random origins), and each consisting of 200 individual responses to 8 items (variables). Repeating the analysis with 100 sample sets provided a better understanding of the accuracy and precision of each method under each set of test criteria (please see comments of external validity versus precision in section 3.6).

Each time a sample set was created (Step 7 of Figure 5) it was analysed in its continuous state (prior to categorisation) using EFA to provide a control for internal validity (Muthén and Kaplan, 1992), after which it was passed through a categorisation phase (Step 9 of Figure 5) to produce the ordinal data for the analysis (akin to observed data obtained in the field in empirical studies).

3.4 Creation of the Input Data

The method for the creation of the data has been discussed in Section 3.2 Research Design, yet it is important at this stage to explain the various inputs to each of the 12 models, and how each of these individually contributed to analysing or answering the various research questions.

To recap, there were 3 research questions. They were:

1. What will the effects of ordinal level data be on the output of EFA?
2. Will performing EFA on the Spearman-rank correlation matrix of the ordinal level data yield better results than performing EFA on the ordinal level data directly?
3. Will the normal distribution fitting algorithm proposed by Stacey (2005) improve the results of the EFA over the other two methods (No transformation, and Spearman-rank)?

Data for Research Question 1

Various models were run to generate the large amount of data required to analyse and answer the research questions posed. Question 1, "what is the effect of ordinal data on EFA", is a very broad question, and as such has required that this study approach it from two main angles (groups) to make analysis and presentation simpler:

- **Group 1**. The effects on factor analysis of pure categorization with as little distortion of the underlying distribution as possible – the goal was to ascertain what the effects on EFA are through the loss of data when a continuous distribution is collapsed into discrete categories, but provide as little distortion to the underlying distribution as possible during this categorisation process.
- **Group 2**. The effects on factor analysis of the most common forms of distribution distortion – the goal was to ascertain not only the effects of discretisation error, but also what is the effect when the categorization process misrepresents (distorts) the underlying distribution.

Data for Research Questions 2 & 3

Every model run in the two groups presented above to answer Question 1 produced an output in all four analysis methods (ordinal, Spearman, NDF^1, and NDF^2) and therefore

while running separate models to analyse question 1, the models automatically produced the data necessary for analysis of research questions 2 and 3.

3.4.1 *Model Summary*

In summary each model with its inputs can be tabulated. A schematic representation of each model can be found in Section 7.2 of the appendix.

Model	Input Factor Loading								Input Distribution Distortion									Input to Research Question Number		
	F1 V1	F1 V2	F1 V3	F1 V4	F1 V5	F1 V6	F1 V7	F1 V8	F2 V1	F2 V2	F2 V3	F2 V4	F2 V5	F2 V6	F2 V7	F2 V8		1	2	3
1	0.4	0.4	0.4	0.4	0	0	0	0	0	0	0	0	0.5	0.5	0.5	0.5	None	✓	✓	✓
2	0.6	0.6	0.6	0.6	0	0	0	0	0	0	0	0	0.7	0.7	0.7	0.7	None	✓	✓	✓
3	0.8	0.8	0.8	0.8	0	0	0	0	0	0	0	0	0.9	0.9	0.9	0.9	None	✓	✓	✓
4	0.4	0.4	0.4	0.4	0	0	0	0	0	0	0	0	0.5	0.5	0.5	0.5	Skew	✓	✓	✓
5	0.6	0.6	0.6	0.6	0	0	0	0	0	0	0	0	0.7	0.7	0.7	0.7	Skew	✓	✓	✓
6	0.8	0.8	0.8	0.8	0	0	0	0	0	0	0	0	0.9	0.9	0.9	0.9	Skew	✓	✓	✓
7	0.4	0.4	0.4	0.4	0	0	0	0	0	0	0	0	0.5	0.5	0.5	0.5	Kurtotic	✓	✓	✓
8	0.6	0.6	0.6	0.6	0	0	0	0	0	0	0	0	0.7	0.7	0.7	0.7	Kurtotic	✓	✓	✓
9	0.8	0.8	0.8	0.8	0	0	0	0	0	0	0	0	0.9	0.9	0.9	0.9	Kurtotic	✓	✓	✓
10	0.4	0.4	0.4	0.4	0	0	0	0	0	0	0	0	0.5	0.5	0.5	0.5	Bimodal	✓	✓	✓

11	0.6	0.6	0.6	0.6	0	0	0	0	0	0	0	0	0.7	0.7	0.7	0.7	Bimodal	✓	✓	✓
12	0.8	0.8	0.8	0.8	0	0	0	0	0	0	0	0	0.9	0.9	0.9	0.9	Bimodal	✓	✓	✓

Table 1 - Model Summary

3.5 Data analysis and interpretation

The main statistics used for comparison, as already mentioned, were the factor analysis output of factor loading because these could be directly compared back to the 'intended' factor structure (A), and therefore there is a clear indication of the amount of data lost (discretisation error) through the various categorisation and scale distortion methods.

Since 100 separate iterations for every model were performed, and 4 variables were loaded homogenously onto every factor, the output was 400 (100 X 4) separate factor loading data points. This provided a distribution that could be analysed to better compare accuracy and bias of each model and provide confidence intervals to further guide researchers. It is important to note that in many respects it is more important to look at the range of values received from the 100 iterations than it is to look at the average value (bias) of those 100 iterations. This is because a researcher faced with only one empirical sample, would not have the luxury of multiple iterations from which to draw an average bias, but would in fact have only one factor loading value that could fall anywhere within the distributions shown here. Therefore although bias is shown in these results, it is the confidence intervals and CV results that are of more value to any researchers using this study to guide the interpretation of results.

3.5.1 *Comparative Measures Used*

The main measures for the average value and the spread of the results are the mean and standard deviation respectively. Relative measures of mean and standard deviation were chosen however that would make results more comparable across models, intended factor loading strengths, and category sizes. Relative Bias (RB) was chosen as a relative

measure to represent the mean value, and is a percentage measure obtained by applying the formula:

$$RB = \left(\frac{\hat{\theta} - \theta}{\theta}\right) \times 100$$

Where θ is the population parameter (or in this case the intended factor loading), and $\hat{\theta}$ is the estimated statistic (or in this case the observed factor loading).

The Coefficient of Variation (CV) was chosen as a relative measure to represent the standard deviation, and is a percentage measure obtained by applying the formula:

$$CV = \left(\frac{\hat{\sigma}}{\hat{\mu}}\right) \times 100$$

Where $\hat{\sigma}$ is the observed standard deviation (in this case the standard deviation of the 100 iterations X 4 variables for the given model – the standard deviation of 400 points), and $\hat{\mu}$ is the observed mean (in this case the mean of the 100 iterations X 4 variables for the given model - the mean of 400 points). If the 'raw' statistics are of interest for mean or standard deviation, then these can be found in the appendices.

In both groups the factor loadings in each factor are homogenous (for example all 0.7), therefore the factor loading mean's presented for a particular model are in fact the mean of all iterations in all homogenous variables under the same input conditions, i.e. values presented are based on 400 data points taken from the 4 homogenous variables in each of the 100 iterations. It is felt that since these variables are in fact homogenous, and that the 4 variables were chosen to load equally and purely (i.e. no cross loading) onto each factor to increase the replicability of the study, analysis can be done on these as though they are one sample and thereby not reduce any information. Likewise, Standard Deviation, RB, and CV, are calculated in this same manner.

3.5.2 *Tabular Presentation of RB and CV*

Figure 9 shows a typical table used to present a set of RB or CV values for a given group of models.

Relative Bias

RB		**Category**						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	-25.3%	-13.8%	-9.9%	-5.0%	-3.9%	-2.7%	-2.9%
0.5	**CAT**	-21.9%	-13.2%	-8.6%	-5.9%	-3.8%	-3.1%	-2.4%
0.6	**CAT**	-18.4%	-14.6%	-8.1%	-5.7%	-4.1%	-3.3%	-2.2%
0.7	**CAT**	-18.1%	-14.3%	-8.4%	-5.6%	-4.0%	-3.0%	-2.2%
0.8	**CAT**	-16.6%	-14.0%	-8.5%	-5.6%	-4.1%	-3.0%	-2.3%
0.9	**CAT**	-13.7%	-12.8%	-8.5%	-6.0%	-4.2%	-3.2%	-2.4%

Figure 9 - Tabular presentation of RB or CV

Firstly, all groups of models are presented with two varying axes; intended factor loading, and category size. They are always presented using intended factor loadings. These represent the factor loadings represented in the underlying continuous (unobserved) population, and never represent the observed factor loadings because these are represented in the data themselves as either an RB or CV measure.

Secondly, in Figure 9 there is a 'method' column, in this case all are marked as 'CAT'. This column will be used to denote the method used in obtaining the results being presented, namely:

CAT – represents factor loading results obtained directly from the ordinal data with absolutely no rescaling, and it uses the standard Pearson's method to obtain the correlation matrix from the categorical data as the first step of the factor analysis process. It represents the status quo method used by most researchers.

SPR – represents factor loading results obtained directly from the categorical data with absolutely no rescaling, BUT it uses the Spearman Rank method to obtain the correlation matrix from the categorical data as the first step of the factor analysis process.

NDF[1] – represents factor loading results obtained after first using the Normal Distribution Fitting Algorithm discussed in Section 3.2.5, and then sampling once randomly within the thresholds obtained.

NDF[2] – represents factor loading results obtained after first using the Normal Distribution Fitting Algorithm discussed in Section 3.2.5, then drawing 1000 random samples from each truncated normal distribution found between thresholds, and using the mean of these values as the input to the factor analysis.

Thirdly, the relative measure presented (RB or CV) in the table is shown in the top left hand corner of the table, in this case the RB values are shown for the group of models in question.

Lastly, the shading used in the tabular presentation, as can be seen in the example (Figure 9), is colour coded for quick visual reference indicating the amount of variance or bias presented. Values below 5% have been shaded green and represent a trivial bias/variance, values between 5% and 10% are shaded yellow and represent a moderate bias/variance, and values at or exceeding 10% are shaded red and constitute a substantial bias/variance. The chosen bands and the naming (trivial, moderate, substantial) follow Bollen and Barb (1981).

3.5.3 *Plots Used*

All plots represent actual nominal values, and do not represent relative measures unless specifically stated.

When comparability is desired between more results, a modified box plot is used as can be seen in Figure 10. It is felt that these give a simple visual comparison in less space than would be possible with traditional box plots.

Figure 10 is used to easily compare a single measure at various different factor loading values. The vertical axis represents the intended factor loading, while the horizontal axis represents the observed factor loadings in each of 100 iterations. Each plot then represents a separate set of input parameters which is highlighted in the legend, and comprises; a

'central' dot representing the mean value for the 100 iterations, two open spaces left and right of this dot representing the inter-quartile range, and a line to the left of the left hand gap and to the right of the right hand gap representing the lower and upper quartiles respectively. The diagonal line provides a visual reference for the intended loading, and therefore is used to easily see the amount of error and spread represented in the data.

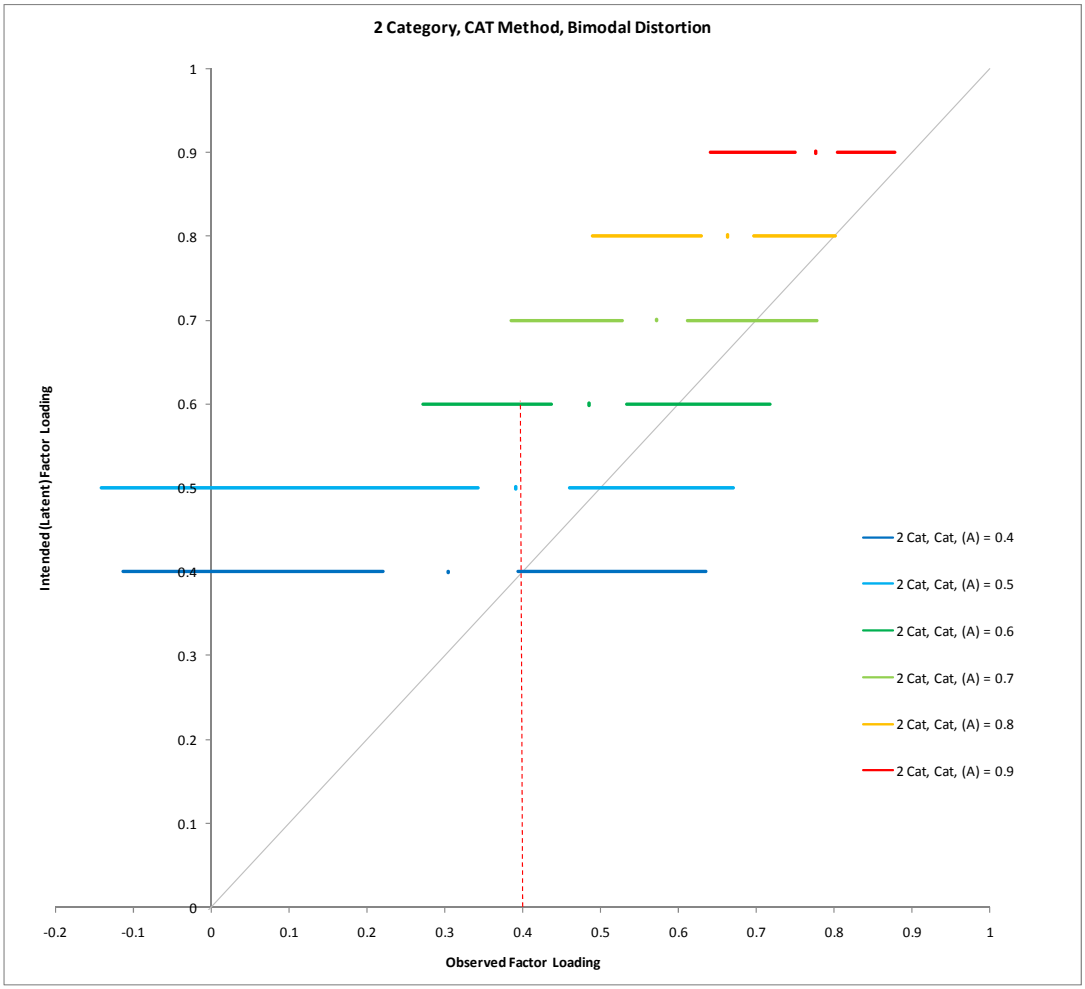


Figure 10 - Example 'singular' modified box plot

3.6 Validity and reliability

3.6.1 *External validity*

External validity of this study is important since its purpose is to provide guidelines to other researchers, and therefore in this context external validity would mean the applicability of these results to other studies of Exploratory Factor Analysis on ordinal data. The study ran only 100 iterations per model (Muthen et al (1992) ran 1000 replications per analysis set). This freed up processor resources that were better used on a greater variation in the input parameters, and in so doing perhaps traded off accuracy for greater external validity (greater applicability to a larger range of external studies), and provided a better indication of where further research could be focused.

3.6.2 *Internal validity*

Internal validity would have meant that the study measured what it intended to measure within the confines of this study – i.e. the effect of ordinal data on factor analysis. This would mean that the study would have wanted to exclude any variability in the factor output that may be as a result of some unintended causes. The study provided a 'control' to monitor this internal validity. It did this by running Exploratory Factor Analysis on the interval data prior to categorization for each analysis (Muthén and Kaplan, 1992) and by 'engineering' samples to exclude sampling error. The control provided a measure of the success of factor analysis on the continuous data for the input criteria of every model. If variability was observed in the control, then it would indicate that the input parameters were creating instability in factor analysis even prior to categorization (issues eluded to in the literature review namely; too few variables, variables with low communalities, or too few individuals etc. (Gorsuch, 1983)). If the control showed very little (or no) instability, then any instability observed post categorization must have been introduced at this stage - i.e. the effects of categorisation distortion only was successfully isolated, which *was* the effect being studied.

3.6.3 *Reliability*

Reliability of Factor Analysis has been discussed at length under the Literature Review section, and was controlled by excluding the undesirable affects that may create reliability issues in Factor Analysis as discussed in 3.6.2 above.

4 INTERNAL VALIDITY ACHIEVED IN THE STUDY

4.1 Introduction

This section presents the actual descriptive statistics of the input data, both continuous and categorical, as well as the factor analysis output of the control (being the factor analysis output applied to the continuous or un-categorized input data). The purpose of this section is to prove the internal validity of this study. It is possible to show that;

1. Sampling error is almost entirely removed from the input data – each sample almost perfectly represents the intended underlying parameters. This shows that the categorization error was successfully isolated to within practical limits.
2. There was virtually no variability in the factor analysis output performed on the underlying continuous data, which means that the input parameters of the continuous data did not create any variability in the accuracy or reliability of the factor analysis method, and therefore any variations shown in the results of the factor analysis performed on the categorized data could be attributed to the categorization error introduced.

4.2 Sampling Error and EFA Stability in the Continuous Data

The underlying continuous data represents the unobserved or latent data that a researcher is attempting to analyse and understand. Perhaps these represent the attitudes of some population under research. In this study it is assumed that the researcher attempts to analyse this population using a simple questionnaire method that results in categorical (ordinal) data. However, as discussed before, this research does not sample from a population, but rather constructs a set of continuous samples with near perfect factor structures. This is necessary to attempt to eliminate the effects of measurement error and sampling error that would normally occur in real life since it is the intention to isolate and analyse the categorisation error only. Therefore to show the reduction in all errors except that of categorisation error it is important to present:

a) The descriptive statistics of this constructed underlying set of continuous sample data. This shows the normality of the underlying distribution, void of any distortion

introduced in the engineering of the continuous samples that may influence the results.

b) The error between the desired correlation structures and the actual correlation structures of the continuous samples. This shows the reduction of the effect of sampling error on the correlation structures and proves that each sample is representative of the intended parameters. This proves that the FA result obtained from the categorised data can be compared to the intended parameters as a measure of categorisation error.

c) The resultant factor analysis on the continuous sample as the control. This shows any error created by the factor analysis method itself, and speaks to the reliability of the factor analysis instrument (as discussed in Section 2.2.2 Validity and Reliability).

4.2.1 *Descriptive Statistics of the Underlying Samples*

In the Methodology section it was discussed that the underlying continuous distribution was created by; first taking random data from a normal distribution which had a mean of 0 and a standard deviation of 1, and then applying a Choleski linear transformation to obtain a data set with the desired correlation structure, that was then further perfected using a Non-linear Solving Algorithm.

Analysis of the constructed/contrived continuous data showed no significant errors introduced through the Choleski process and were all considered normal, showing little to no distortion in skewness or kurtosis, and showed no significant shift in mean or significant variance in the amount of standard deviation. The descriptive statistics for 400 of the samples analysed for these errors (each measured on 200 data points X 8 variables combined) are shown in Table 2 below. This is shown as a representative example of all the studies samples, purely to show the robustness of the method used to create the data sets. In the interests of space, there is no attempt to present all the data for the entire study.

Sample's	*Mean*	*Median*	*Std Dev*	*UCL (95%)*	*LCL (95%)*
Mean (intended = 0)	0.0003	0.0012	0.0394	0.0041	−0.0036
Std Dev (intended = 1)	0.9972	0.9967	0.0394	1.0012	0.9934
Kurtosis	0.0628	0.0550	0.1565	0.0782	0.0474
Skewness	0.0059	0.0051	0.0769	0.0135	−0.0016

Table 2 - Continuous Data Descriptive Statistics

4.2.2 *Correlation Error between Intended (R) and Sample (R*)*

The correlation error is the error between; the intended[5] correlation matrices (R), and the actual correlation matrices of the continuous 'unobserved' or contrived data (R*). The error is measured using the sum of squared differences between the two matrices. The descriptive statistics for 400 of the samples are shown in Table 3.

Great care was taken to construct the data with a correlation structure as close to the desired matrices as possible[6]. The correlation is the single most important factor in this study since it is the one input that could introduce the greatest unintended error into the analysis – sampling error. If a Choleski Matrix is used to derive a set of data from random inputs a SUM^2 error in the order of greater than 0.1 would be observed, and this error would be expected to vary greatly from sample to sample. This error can be likened to sampling

[5] The desired correlation matrices are derived through multiplying the desired factor loadings with their transverse matrices as discussed in the Methodology section.
[6] Given the time and resource constraints.

error since it would effectively distort the view of the underlying/intended population and it would introduce variability from sample to sample. This would combine with, and therefore inflate, discretisation error and therefore deserves particular attention and great care to avoid/reduce it. With non-linear solving in addition to the Choleski transformation (as discussed in the Methodology section), the error is reduced to a mean of 0.0000005, and is maintained within a narrow band to prevent huge variability in samples (standard deviation of 0.0000005). It should therefore be noted that these errors are extremely small, and therefore sampling error was considered to have been practically eliminated.

Correlation	*Mean*	*Median*	*Std Dev*	*UCL (95%)*	*LCL (95%)*
Sum Squared Error	0.0000005	0.0000003	0.0000005	0.0000005	0.0000004

Table 3 - Correlation Error of Samples

4.2.3 *Factor Analysis Error*

The error produced by the instrument of factor analysis directly affects the reliability of the results and the study took care not to subject factor analysis to input parameters in the continuous data that would affect reliability. Therefore the 'control' factor analysis was introduced (as discussed in Section 3.6 Validity and reliability) on all the latent continuous data iterations to test the reliability of the instrument across iterations (samples) as well as across models, and it was done to test validity of the method itself.

Validity ensures that no input model is selected that factor analysis is unable to solve using the continuous latent data (issues affecting factor analysis validity and reliability were discussed in section 2.2.2). Once an input is known to yield valid factor results, then

reliability can be shown by analysing whether factor analysis performs accurately in a reliable manner across all latent samples.

Any introduction of error by factor analysis would reduce the ability to isolate the errors introduced through categorisation (discretisation error). Conversely, if it can be shown that factor analysis performs reliably and accurately on the latent data, then any difference between the intended results and the results of factor analysis on the categorised data can be attributed to discretisation error or the affects of ordinal data on factor analysis. Table 4 shows the descriptive statistics of the control factor loading results. It can be seen from the mean and the relative bias (RB) of the factor loadings that the output of the factor analysis performed on the continuous data (control) is extremely accurate, while the low standard deviation and coefficient of variation (CV) shows the high precision. As would be expected though, these small errors do increase as the factor loadings decrease, but even at the worst are only contributing fractions of a percentage to the final outcome.

Controls' Intended Factor Loading	*Actual Factor Loading Mean*	*Actual Factor Loading Std Deviation*	*RB*	*CV*
0.4 [7]	0.400002	0.0001	0.0003%	0.03%
0.5 [8]	0.500008	0.0001	0.0007%	0.02%
0.6 [9]	0.600002	0.0001	0.0003%	0.01%
0.7 [10]	0.700002	0.0001	0.0003%	0.01%
0.8 [11]	0.799999	0.0001	-0.0002%	0.01%
0.9 [12]	0.900001	0.0001	0.0001%	0.01%

Table 4 - Control Results on positively loaded variables

Presented in Table 4 are only the controls from Group 1. This is done to provide an example of the level of stability obtained throughout the study for varying factor strengths. In the interest of space the detailed control outputs for all models is not shown, but were all in line with the results presented above.

[7] Taken from the factor loading results of variables 1-4 for Factor 1 of Model 2.
[8] Taken from the factor loading results of variables 5-8 for Factor 2 of Model 2.
[9] Taken from the factor loading results of variables 1-4 for Factor 1 of Model 3.
[10] Taken from the factor loading results of variables 5-8 for Factor 2 of Model 3.
[11] Taken from the factor loading results of variables 1-4 for Factor 1 of Model 4.
[12] Taken from the factor loadings results of variables 5-8 for Factor 1 of Model 4.

4.3 Manifest Categorical Data Descriptive Statistics

4.3.1 *Limited Scale Distortion Manifest Categorical Data*

The first group aimed to reduce the amount of scale distortion of the underlying (unobserved) continuous data during the categorisation process. By attempting to set the thresholds used for categorisation to values that best retain a 'pseudo' normal distribution in the resultant categorised data, the study introduced as little skewness, kurtosis, or distribution distortion as possible, and in so doing the results represent the loss of data created in simply collapsing continuous data (attitudes) into categorical data (Likert-type responses) without any distortion of the underlying attitudes. The loss of data creates a bias and instability into the factor analysis output.

The thresholds selected for this model are shown below in Table 5.

	2 Category	3 Category	4 Category	5 Category	6 Category	7 Category	8 Category
Threshold 1	0	-1	-1.5	-1.8	-2	-2.14286	-2.25
Threshold 2		1	0	-0.6	-1	-1.28571	-1.5
Threshold 3			1.5	0.6	0	-0.42857	-0.75
Threshold 4				1.8	1	0.42857	0
Threshold 5					2	1.28571	0.75
Threshold 6						2.14286	1.5
Threshold 7							2.25

Table 5 – Group 1 Thresholds

It is impossible to not distort the scale and distribution in the lower category sizes, even though normality tests like Martinez-Iglewicz could not reject normality on even the 2 category distribution, it is hard to appreciate that a 2 category histogram 'appears' normal. It is therefore important to present the actual descriptive statistics of the manifest categorical samples with their histograms to gain a better understanding of what constitutes 'minimal scale distortion' in this study.

Plots of the resultant (manifest) categorical data presented in this section are shown per category size, across the 8 variables for the first iteration of each Model (1-3). This provides a sufficient view of the resultant categorical distribution (6,400 data points per category size), rather than presenting the distribution across all iterations for all Models in Group 1

(640,000 data points per category size). The descriptive stats are gathered from all iterations.

	μ	σ	μ_{sk}	σ_{sk}	μ_{ku}	σ_{ku}
2 Category	1.4992	0.5000	0.0032	0.1459	-1.9988	0.0309
3 Category	1.9986	0.5602	0.0008	0.0487	0.2908	0.5933
4 Category	2.4989	0.7174	0.0025	0.1438	-0.2824	0.1033
5 Category	2.9983	0.8709	0.0038	0.1398	-0.2064	0.2311
6 Category	3.4979	1.0293	0.0047	0.1427	-0.1918	0.2318
7 Category	3.9972	1.1903	0.0012	0.1436	-0.1753	0.2366
8 Category	4.4974	1.3517	0.0035	0.1452	-0.1608	0.2395

Table 6 - Manifest Categorical Data Descriptive Statistics (Limited Distortion) [13]

Table 6 presents the average value for each category size for this group (see footnote on how these were derived), and as would be expect, the value is roughly half the difference of the two outer category values i.e. for an 8 category sample with little distortion to left or right an average value of (8 – 1)/2 + 1 = 4.5 is expected, and is observed. The average skewness (μ_{sk}) confirms that there is little side to side distortion of the distributions during the categorisation process, and little variation or non-uniformity in this skewness represented across all category sizes (shown by σ_{sk}). Likewise, the effect of categorisation in Group 1 on kurtosis (μ_{ku} and σ_{ku}) is limited and uniform, except for the 2 category samples, but it is impossible to not affect Kurtosis when collapsing a continuous distribution into just two ordinal values, and it is therefore accepted to be unavoidable.

Table 7 presents density plots of each category without any significant distortion, and are shown along with a superimposed normal distribution curve for comparison.

[13] In this table, μ is the average across variables of iterations of the average value within variables of iterations, i.e. the average of variable 1 for iteration 1 is taken along with all other variable of all other iterations, and these are then averaged to get μ. Likewise, σ is the average of all the standard deviations obtained from every variable of every iteration. The next four measures are the average skew and its standard deviation, and average kurtosis and its standard deviation, obtained in a similar method to the first two measures.

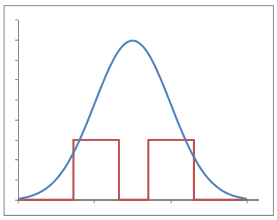
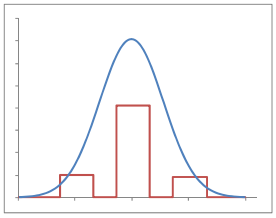
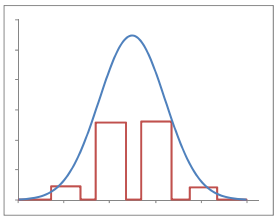
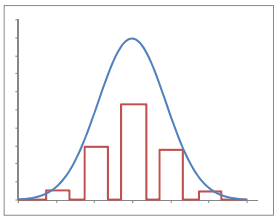
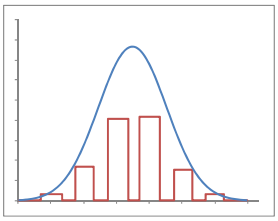
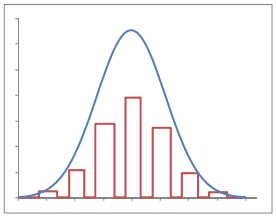
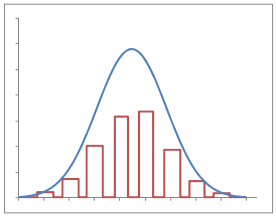
 2 Category	 3 Category
 4 Category	 5 Category
 6 Category	 7 Category
 8 Category	

Table 7 - Histogram for Discrete Manifest Data (Limited Distortion)

4.3.2 *Skewed Distortion Manifest Categorical Data*

As part of Group 2 the study investigated the effects on factor analysis of a skewed distortion applied during categorisation on the underlying continuous normal distribution. The statistics and plots shown here were obtained, and are presented, in the same manner as the previous section.

The thresholds used in this Sub-Group are shown in Table 8 and were designed to attempt to obtain a similar skewness (-1.22) and kurtosis (0.85) across all category samples.

	2 Category	3 Category	4 Category	5 Category	6 Category	7 Category	8 Category
Threshold 1	-0.7	-2	-1.6	-1.5	-2	-2	-2
Threshold 2		-0.55	-0.5	-1.05	-1.4	-1.5	-1.5
Threshold 3			2.2	-0.25	-0.89	-1.1	-1.1
Threshold 4				2	0	-0.62	-0.7
Threshold 5					2	-0.1	-0.2
Threshold 6						2	1.65
Threshold 7							2

Table 8 - Group 2 Thresholds - Skewed Distortion

	μ	σ	μ_{sk}	σ_{sk}	μ_{ku}	σ_{ku}
2 Category	1.7596	0.4264	-1.2421	0.2352	-0.4060	0.6184
3 Category	2.6878	0.5081	-1.3013	0.2026	0.6183	0.6810
4 Category	2.6527	0.5994	-1.1620	0.2152	0.8599	0.5200
5 Category	3.4103	0.9128	-1.2413	0.1776	0.9843	0.6313
6 Category	4.2343	1.0191	-1.1628	0.1569	1.1110	0.5850
7 Category	5.0735	1.3195	-1.2732	0.1643	0.9776	0.6080
8 Category	5.1874	1.3832	-1.1268	0.2002	1.0962	0.6423

Table 9 - Manifest Categorical Data Descriptive Statistics (Skewed Distortion)

Table 10 presents density plots of each category for the skew distortion sample set, and are shown along with a superimposed normal distribution curve for comparison.

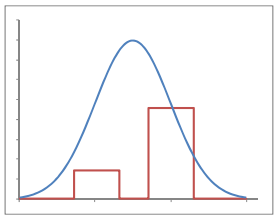
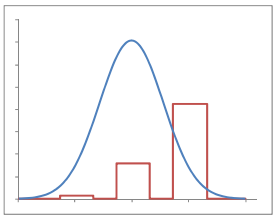
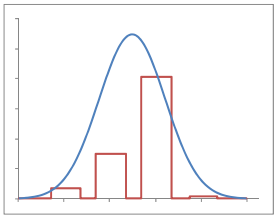
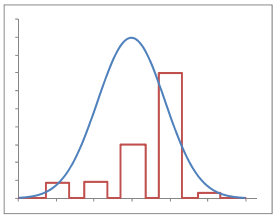
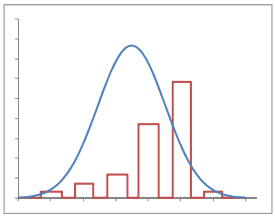
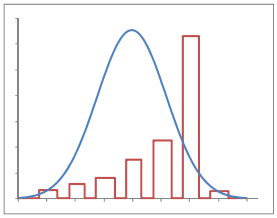
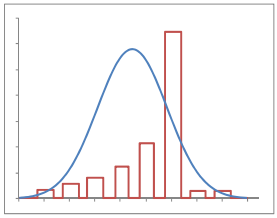
 2 Category	 3 Category
 4 Category	 5 Category
 6 Category	 7 Category
 8 Category	

Table 10 - Histogram for Discrete Manifest Data (Skewed Distortion)

4.3.3 *Kurtotic Distortion Manifest Categorical Data*

As part of Group 2 the study investigated the effects on factor analysis of a kurtotic distortion applied during categorisation on the underlying continuous normal distribution. The statistics and plots shown here were obtained, and are presented, in the same manner as the previous section.

The thresholds used in this Sub-Group are shown in Table 11 and were designed to attempt to obtain a similar skewness (0) and kurtosis (0.85) across all category samples.

	2 Category	3 Category	4 Category	5 Category	6 Category	7 Category	8 Category
Threshold 1	-0.89	-1.15	-1.19	-1.40	-1.65	-2.00	-2.00
Threshold 2		1.15	1.13	-0.54	-1.10	-1.60	-1.50
Threshold 3			3.91	1.20	0.35	-0.54	-0.80
Threshold 4				1.72	1.36	0.54	0.50
Threshold 5					1.80	1.60	1.05
Threshold 6						2.00	1.50
Threshold 7							2.00

Table 11 - Group 2 Thresholds - Kurtotic Distortion

The actual descriptive statistics for the categorical samples created with kurtotic distortion are shown in Table 12. With the exception of the 2 Category samples, all samples were created with low levels of skewness, thereby effectively isolating the effects of kurtosis as far as possible. Kurtosis values averaged close to or above 0.85 as can be seen in Table 12, although a large degree of standard deviation in kurtosis was observed if compared to the much lower amounts observed for kurtosis and skewness in the previous sections.

	μ	σ	μ_{sk}	σ_{sk}	μ_{ku}	σ_{ku}
2 Category	1.8158	0.3862	-1.6744	0.3077	0.9072	1.1067
3 Category	1.9993	0.4946	0.0005	0.1124	1.2637	0.9197
4 Category	2.0124	0.4897	0.0395	0.1263	1.3572	0.9480
5 Category	2.7816	0.8478	0.0316	0.2026	0.9740	0.5251
6 Category	3.2981	0.9868	0.2445	0.2060	1.1383	0.4997
7 Category	3.9981	1.0596	0.0005	0.2338	0.9389	0.3954
8 Category	4.2392	1.2915	0.4898	0.2098	1.1618	0.5650

Table 12 - Manifest Categorical Data Descriptive Statistics (Kurtotic Distortion)

Table 13 presents density plots of each category for the kurtotic distortion, and are shown along with a superimposed normal distribution curve for comparison.

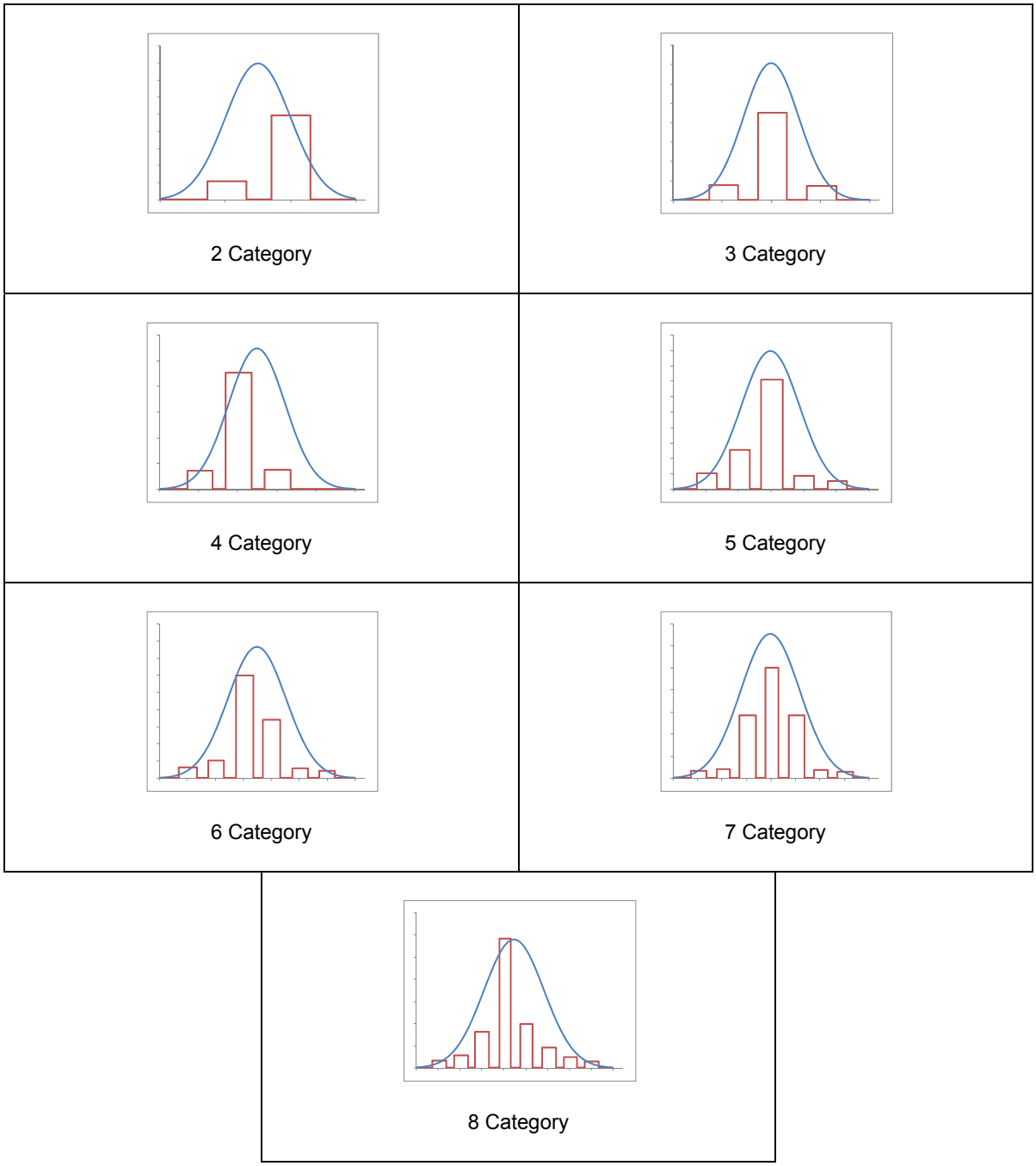


Table 13 - Histogram for Discrete Manifest Data (Kurtotic Distortion)

4.3.4 *Bimodal Distortion Manifest Categorical Data*

As part of Group 2 the study investigated the effects on factor analysis of a bimodal distortion applied during categorisation on the underlying continuous normal distribution. The statistics and plots shown here were obtained, and are presented, in the same manner as the previous sections.

The thresholds used in this Sub-Group are shown in Table 14 and were designed to produce some form of bimodality across all category samples.

	2 Category	3 Category	4 Category	5 Category	6 Category	7 Category	8 Category
Threshold 1	0	-0.05	-0.5	-0.5	-0.7	-0.7	-0.7
Threshold 2		0.05	0	-0.05	-0.2	-0.2	-0.3
Threshold 3			0.5	0.05	0	-0.05	-0.08
Threshold 4				0.5	0.2	0.05	0
Threshold 5					0.7	0.2	0.08
Threshold 6						0.7	0.3
Threshold 7							0.7

Table 14 - Group 2 Thresholds - Bimodal Distortion

Although the descriptive statistics are shown for this sample set, there is no firm measure of bimodality that could be included for comparative purposes.

	μ	σ	μ_{sk}	σ_{sk}	μ_{ku}	σ_{ku}
2 Category	1.4992	0.5000	0.0034	0.1454	-1.9990	0.0300
3 Category	1.9981	0.9795	0.0038	0.1427	-1.9560	0.0332
4 Category	2.4979	1.2140	0.0029	0.1160	-1.5531	0.0777
5 Category	2.9969	1.6720	0.0034	0.1266	-1.6967	0.0524
6 Category	3.4952	1.9593	0.0033	0.1213	-1.6017	0.0618
7 Category	3.9942	2.4219	0.0035	0.1273	-1.6983	0.0495
8 Category	4.4937	2.8338	0.0032	0.1272	-1.6975	0.0505

Table 15 - Manifest Categorical Data Descriptive Statistics (Bimodal Distortion)

Table 16 presents density plots of each category, and are shown along with a superimposed normal distribution curve for comparison.

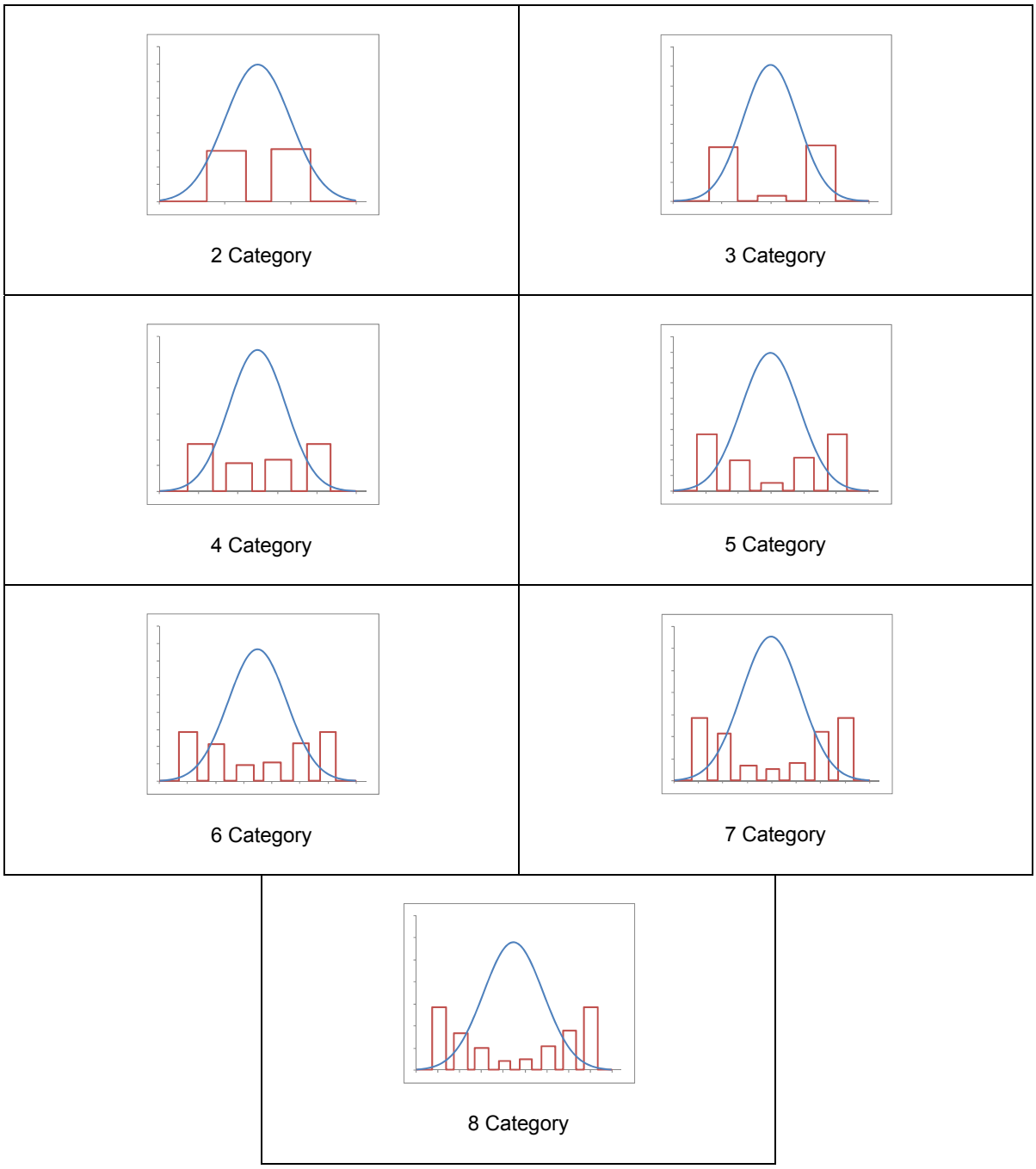


Table 16 - Histogram for Discrete Manifest Data (Bimodal Distortion)

5 PRESENTATION OF RESULTS

5.1 Introduction

The results are presented and discussed below. Simplified tables and figures are included with the discussions where they can be referenced as they are discussed. More detailed views of the data are available in the Appendices.

Section 5.2 represents the results from Group 1, and answers part of the first research question by attempting to show the affects of the categorisation process (discretisation error) on factor analysis results. Sections 5.3, 5.4, and 5.5 represent the results from Group 2. These sections complete the answer for the first research question by including the effects of scale distortion.

The second and third research questions – understanding the benefits of Spearman Rank and Normal Distribution Fitting Algorithm in these situations – are covered together in sections 5.6, 5.7, 5.8, and 5.9, as it is believed that this gives a better comparison when methods are seen side by side.

As discussed before, the study chose two relative measures to do the main comparisons on, one that represents the bias introduced (RB), and one that shows the relative amount of variance in the factor loading (CV). In all sections below these values are tabulated and discussed, or included in the Appendices where more detail is required. In all tables, shading and discussion on the amount of bias and variance (trivial, moderate, substantial) has been done as per the guidelines in Section 3.5.2.

5.2 The Effects of Categorisation on FA with Limited Distortion

The full descriptive statistics for factor loadings are presented in the Appendices Section 7.3. Shaded RB and CV results for limited scale distortion are shown below in Table 17 and Table 18 respectively.

Relative Bias

RB		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	-25.3%	-13.8%	-9.9%	-5.0%	-3.9%	-2.7%	-2.9%
0.5	**CAT**	-21.9%	-13.2%	-8.6%	-5.9%	-3.8%	-3.1%	-2.4%
0.6	**CAT**	-18.4%	-14.6%	-8.1%	-5.7%	-4.1%	-3.3%	-2.2%
0.7	**CAT**	-18.1%	-14.3%	-8.4%	-5.6%	-4.0%	-3.0%	-2.2%
0.8	**CAT**	-16.6%	-14.0%	-8.5%	-5.6%	-4.1%	-3.0%	-2.3%
0.9	**CAT**	-13.7%	-12.8%	-8.5%	-6.0%	-4.2%	-3.2%	-2.4%

Table 17 - RB of Factor Loadings with Little Scale Distortion

The RB under limited scale distortion conditions shows substantial bias below 4 categories, and moderate bias below 6 categories. In general, bias improves as category size increases, yet the increase in intended (underlying) factor loading under these conditions appears to have less of an impact on bias above 4 categories. Going back to the results presented earlier on the work on categorisation's effect on correlation strength (Bollen and Barb, 1981) the authors pointed to an apparent threshold at 5 Categories where results at varying intended correlations converged with moderate bias. Here the relative bias of factor loadings becomes moderate at 4 categories, and trivial at 6 categories. Convergence of the bias results, similarly as with Bollen and Barb (1981), occurs at 5 or perhaps 6 categories as can be seen in Figure 11.

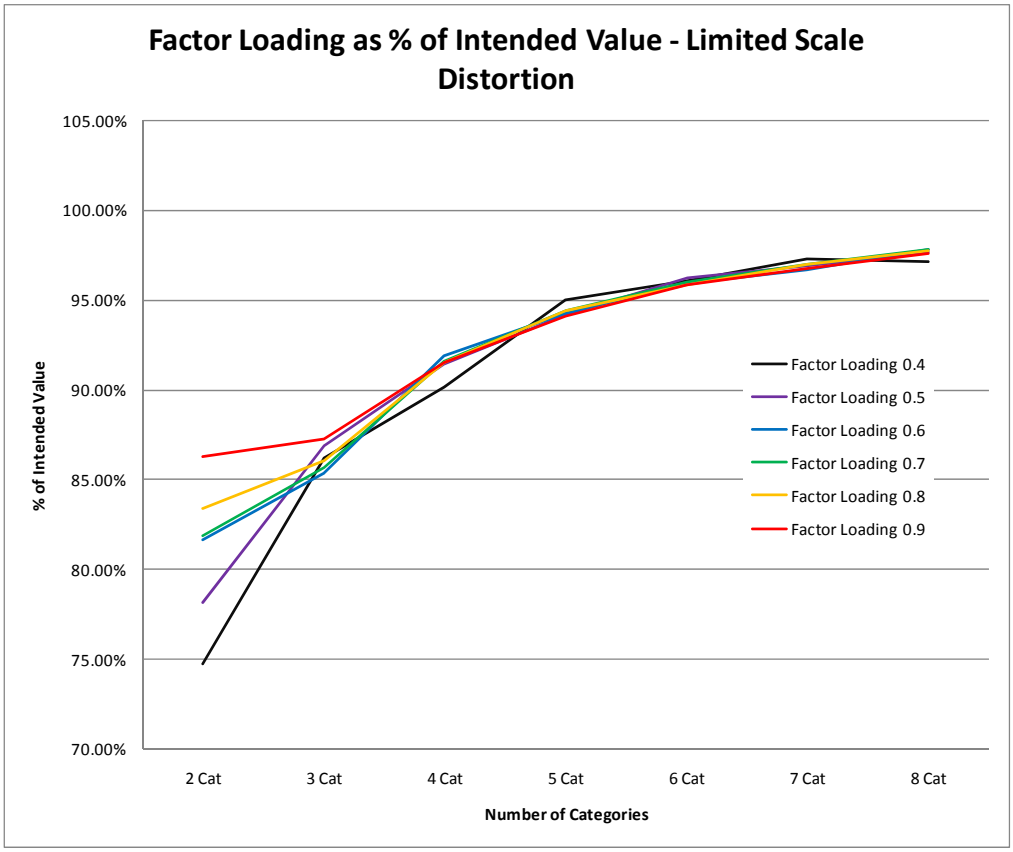


Figure 11 - Factor Loading as % of Intended Factor Loading

Looking at the coefficient of variance (CV), in general it is found that the CV improves both as the category size increases and as the factor strength increases. This can be seen by the diagonal pattern in Table 18.

At 5, 6, and 7 category results, for low underlying loadings of 0.4, the CV is substantial. Therefore even though the bias may be considered borderline trivial at this factor loading, the CV shows a wild fluctuation of the results around this bias. This shows the 'destruction' of factor structure caused by categorization, even at high category values of 6 and 7.

Coefficient of Variation

CV		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	48.1%	26.8%	19.1%	14.8%	12.7%	10.5%	9.3%
0.5	CAT	29.7%	16.9%	12.3%	9.5%	7.6%	6.7%	5.9%
0.6	CAT	13.7%	11.7%	8.0%	6.3%	5.2%	4.6%	4.1%
0.7	CAT	11.6%	8.4%	5.9%	4.7%	3.8%	3.2%	2.7%
0.8	CAT	8.4%	6.3%	4.3%	3.1%	2.5%	2.2%	1.9%
0.9	CAT	5.2%	5.0%	2.9%	2.3%	1.8%	1.4%	1.2%

Table 18 - CV of Factor Loadings with Little Scale Distortion

Suppose a researcher was analysing a factor analysis result obtained from a 6 Category response scale. As part of wanting to retain only significant results the researcher follows the guidelines of Hair et al (2005) and sets a threshold at 0.4 factor loading (observed) for a sample size of N=200 (as with this study). It can be seen from this study that the CV shows a fluctuation of over 12% given a 6 category response scale and an underlying factor loading of 0.4, and that this fluctuation occurs not around 0.4 but around a value that is 3.9% less than 0.4 (negative bias). It could be expected then that if the researcher sets their threshold at 0.4 factor loading (observed), then there is a high probability that a true underlying attitude of 0.4 factor loading will load below the 0.4 threshold due to discretisation error, and would be excluded from the researcher's results. Figure 13 in the Appendices shows that according to these results, there would be a 64% probability that the data would load below a factor loading of 0.4, and would be excluded as a result.

Figure 12 visually shows this 'loss' of data below the 0.4 threshold of both the 0.4 and 0.5 true underlying factor loading results. Yet it also shows the amount of spurious loading that occurs by the 0.4 intended data over the 0.5 factor loading mark. This represents the 'creation' of spurious loadings by the discretisation error.

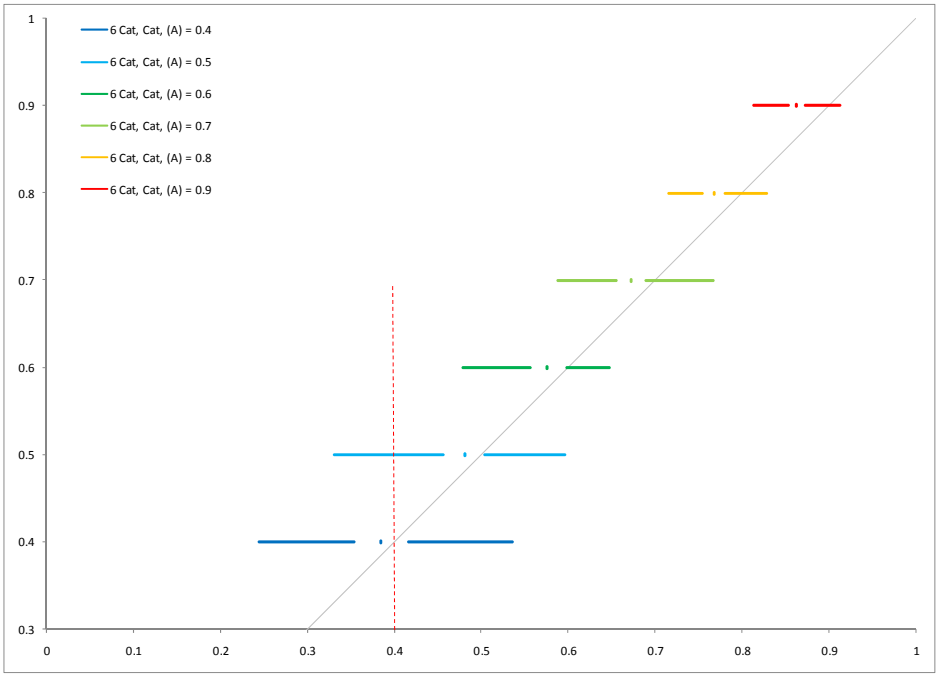


Figure 12 - 6 Category Plot - Little Scale Distortion

As with any statistic then, the choice of where to 'draw the line' is a balance of what the researcher finds more acceptable between the typical Type I and Type II errors. The gut reaction of an inexperienced researcher may be to just lower the threshold in Figure 12 to attempt to capture more of this missed data. They may believe that lowering the threshold would be justified on the basis that this study shows that categorisation creates a negative bias in general, and lowering this threshold then captures more of the negatively biased results. However, it is noticed that CV increases as underlying intended factor loading strength decreases, that is to say that the spread of data becomes wider as factor loadings become weaker. Also, it can be seen from Figure 12 that many observed values end up above their true underlying values. These spurious loadings, which can be seen where the plotted lines extend to the right beyond the diagonal line (which represents the true underlying values), shows that this increase in CV also extends above the true underlying value. In these cases the categorisation process actually incorrectly strengthened the underlying relationships. Since CV increases as factor strength decreases, and that some of this spread would be manifest as spurious loading above both the bias and the actual underlying loading, it could potentially be expected that true factor loadings below 0.4 may

very well overlap with those values from a true underlying factor loading of 0.4 and above. Therefore a variable with a true underlying factor loading of below 0.4, which should rightfully be excluded from the analysis due to lack of significance based on Hair et al (2005), could now be unknowingly included by the researcher if they lowered the threshold in an attempt to capture more of the missed data. The lower the threshold is made, the more conceivable the likelihood that this could occur. Since this study only analysed loadings at a minimum of 0.4 true underlying factor loading, it is unclear what level of spurious loadings could be experienced by smaller factor loadings. More research would be required to understand at which point this would become critical.

Besides the immediate information these results present around the loss of data caused by categorisation, and also the potential spurious loadings created by the same process, it can clearly be seen how any ordinal response scale with less than 5 or even 6 categories should be avoided under these conditions of limited scale distortion. It also becomes clear how advantageous it would be to rather employ a 7 Category response scale that would allow the inclusion of much more of the smaller significant communalities without the greater chance of including spurious ones.

When presented visually, a greater comparison of the vast differences between high and low categories, and between high and low communality strengths, can be observed. Figure 14 in the Appendices shows modified box plots of categories 2, 4, 5, 6, 7, and 8, across all factor strengths. What are immediately evident are the vast differences in the magnitude of the spread of outliers, and the spread of the interquartile ranges, as the underlying factor strength increases/decreases. What is also evident is the difference in outliers and interquartile ranges as the number of categories increase/decrease, with a 2 Category response showing dismal results, and 7 and 8 Category responses showing much improved results. Added to this, Figure 14 also shows the increase in bias in nominal terms as factor strength increases. This is the reason that a percentage bias measure (RB) was used in this study instead of a nominal bias measurement, because in nominal terms it appears at first glance that the bias is increasing as intended factor loading increases, where in reality it is improving, but this is only visible when compared on a percentage (relative) measure.

For a more detailed view of the spread of the results compared to the 0.4 chosen threshold, refer to Figure 13 that shows the percentage of each tail above a chosen observed factor

loading. When using a 2 Category response scale, there is approximately 30% probability that a data point with a true underlying factor loading of 0.4 would be included if the threshold is set at 0.4 observed. This means that there is about a 70% probability that this data would be lost! This is not much better for the true underlying loadings of 0.5, with approximately 51% probability of the data being missed (only 49% included in the tail above 0.4), and 7% being missed in the case of 0.6 true underlying. Therefore choosing a response scale of only 2 Categories vastly increases the probability that lower factor loading values would be lost due to categorisation error, with only the very largest of factor loadings being preserved (although negatively biased), and this is without the influences of sampling error experienced in real life. Changing the response scale to 7 Categories would immediately improve all data retention, decreasing the probability of exclusion to approximately 58% for 0.4 true underlying factor loadings.

5.3 The Effects of Categorisation on FA with Skewed Distortion

The full descriptive statistic results for the factor loadings for this section can be found in the appendices in Section 7.4. The results here will be presented using the relative measures (RB, CV) used before, and these are presented in Table 19 and Table 20 respectively, utilising the same shading method as before.

In the previous section the data was analysed for a situation where the response scale added little distortion to the underlying continuous data's normal distribution. In this section the results are presented where the response scale skewed the underlying normal distribution.

Table 19 shows the greatly reduced area with trivial bias when compared to the results in Section 5.2 (no intended scale distortion). Firstly it can be seen that moderate bias can only be expected on a 5 Category response above a factor loadings of 0.7 (true underlying), but that in essence the clear cut threshold has perhaps shifted to the 6 category mark, but only for moderate biased results!

Relative Bias

RB		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	-26.5%	-24.7%	-16.9%	-12.2%	-9.1%	-10.2%	-8.9%
0.5	CAT	-27.1%	-25.0%	-16.0%	-11.2%	-8.6%	-8.9%	-8.1%
0.6	CAT	-23.5%	-19.0%	-15.7%	-11.4%	-8.4%	-9.1%	-8.5%
0.7	CAT	-21.7%	-17.1%	-14.3%	-10.0%	-7.4%	-7.6%	-7.5%
0.8	CAT	-19.2%	-15.2%	-12.8%	-8.5%	-6.3%	-6.1%	-6.0%
0.9	CAT	-15.5%	-12.4%	-10.3%	-6.6%	-5.3%	-4.3%	-4.4%

Table 19 - RB of Factor Loadings with Skew Scale Distortion

If trivial bias is the desired result then the scale is confined to 6 categories, and only for very high underlying factor loadings of 0.9 and perhaps 0.8. Hair et al (2005) notes how very unlikely high loadings of 0.9 are in empirical data, and therefore it would appear that based on these results the expectation would be that skewing of the underlying distribution would create at best moderate biases in EFA results. This highlights the data destruction present if a response scale distorts the underlying attitudes/beliefs through skewing them, even unfortunately at the desired larger category sizes.

Coefficient of Variance

CV		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	50.5%	47.5%	37.0%	28.8%	23.4%	25.6%	23.6%
0.5	CAT	38.8%	38.5%	22.1%	17.7%	14.1%	15.0%	14.1%
0.6	CAT	20.8%	17.1%	14.7%	11.0%	9.2%	10.4%	9.7%
0.7	CAT	15.3%	11.8%	10.1%	8.0%	6.8%	6.9%	6.9%
0.8	CAT	10.2%	8.1%	6.9%	5.8%	4.8%	4.8%	4.6%
0.9	CAT	7.5%	5.9%	4.4%	3.2%	2.5%	2.5%	2.6%

Table 20 - CV of Factor Loadings with Skew Scale Distortion

Table 20 shows the CV for the skewed results, with a slight change (only 1% or 2% weakening) in the trivial variance area (green) when compared to the previous results in Section 5.2., yet with a ‘sharper’ and more severe ‘drop off’ to substantial variances which reach much higher values at the extremes (top left corner).

As before, analysing the variances is important to understand the impact that lower-loading variables could have for researchers when they are incorrectly observed as higher loadings due to the discretisation and scale distortion errors. Figure 23 shows a modified box plot of 2, 4, 5, 6, 7, and 8 Category responses with skewed scale distortion. The increase in bias and the increase in spread under these distortion conditions can be seen clearly in this visual representation. In the previous section it was shown how at 6 categories the 0.4 intended loading results spuriously loaded above the 0.5 mark. With skewed distortion however, given a 6 category response scale, the results show spurious loadings of the intended 0.4 factor loadings to over the 0.6 factor loading mark. Not only is this factor loading spuriously loading higher as a result of the skewed distortion, but so too are the 0.5 and 0.6 intended factor loadings.

Again the simplified percentile plots in the Appendices (Figure 22) give a better understanding of the probability of spurious loadings and lost data. Looking at Figure 22 it is immediately evident the probability of the loss in data when keeping to the 0.4 threshold discussed earlier. Given a 6 category response scale, probable data loss would extend to the 0.5 true underlying factor loading, with as much as a 70% and 17% probability of missing the data on the 0.4 and 0.5 true underlying factor loadings respectively. Moving to a 7 category response scale would decrease the probability of losing data to 67% and 21% for the 0.4 and 0.5 true underlying factor loadings respectively, yet still far worse than the levels of around 58% and 0% respectively that were observed when there was little scale distortion.

5.4 The Effects of Categorisation on FA with Kurtotic Distortion

As with the previous sections, full descriptive statistics for the factor loading results on the kurtotic distortion can be found in the appendices, Section 7.5. Here again the relative measures (RB and CV) are presented, shown below in Table 21 and Table 22 respectively, shaded in the same manner as before.

Relative Bias

RB		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	-35.2%	-18.1%	-17.8%	-8.7%	-6.2%	-4.7%	-4.7%
0.5	CAT	-34.5%	-20.1%	-20.3%	-9.4%	-6.5%	-4.9%	-5.2%
0.6	CAT	-26.7%	-18.1%	-18.4%	-8.8%	-6.6%	-4.8%	-4.7%
0.7	CAT	-23.8%	-18.2%	-18.5%	-9.1%	-6.5%	-4.8%	-5.2%
0.8	CAT	-20.8%	-16.1%	-16.5%	-8.3%	-6.1%	-4.8%	-4.6%
0.9	CAT	-16.9%	-14.6%	-14.9%	-7.4%	-5.9%	-4.6%	-4.0%

Table 21 - RB of Factor Loadings with Kurtotic Scale Distortion

The RB for the kurtotic-distorted data seems an improvement on the skewed results at higher category sizes, but slightly worse at lower category sizes. Here a clear threshold of borderline trivial bias is apparent at the 7 Category mark. The 5 Category level is clearly moderate at all true underlying factor loadings levels. As with limited distortion results, kurtotic RB results tend to be more affected by the category scale chosen than by the underlying factor loading (with the obvious exception of the 2 category scale).

Coefficient of Variance

CV		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	67.6%	34.9%	33.4%	20.2%	16.2%	13.9%	14.6%
0.5	CAT	54.0%	25.6%	25.9%	13.0%	11.2%	8.9%	9.2%
0.6	CAT	23.7%	13.1%	14.0%	8.5%	7.7%	5.8%	6.5%
0.7	CAT	17.9%	10.9%	10.9%	6.4%	5.3%	4.2%	4.8%
0.8	CAT	12.5%	7.4%	7.6%	4.5%	3.5%	2.9%	3.2%
0.9	CAT	7.9%	5.6%	5.9%	3.1%	2.4%	2.0%	2.1%

Table 22 - CV of Factor Loadings with Kurtotic Scale Distortion

The CVs are very similar in pattern to those in Section 5.2 and Section 5.3 in that they are almost split into an upper and lower diagonal. With kurtotic distortion however, it appears that CV results for the lower categories seem to be affected even more than with skewed

distortion, yet the higher category values resemble more closely those results for limited scale distortion. Perhaps this gives even more weight to the argument to avoid lower category sizes, where it appears in the case of kurtotic distortion that both RB and CV results become more negatively pronounced as the category size decreases.

Figure 32 gives a visual representation of the results under these conditions. Figure 31 shows the percentile plots under kurtotic distortion conditions. Given a 6 category design the expected probability of losing data would be 63% and 9% for 0.4 and 0.5 underlying factor loading respectively, whereas for a 7 category scale it would be 62% and 3% respectively. The higher category sizes perform much better than in the skewed distortion situation.

5.5 The Effects of Categorisation on FA with Bimodal Distortion

As with the previous sections, full descriptive statistics for the factor loading results on the kurtotic distortion can be found in the appendices, Section 7.6. Here again the relative measures (RB and CV) are presented, shown below in Table 23 and Table 24 respectively, shaded in the same manner as before.

Relative Bias

RB		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	-23.7%	-20.6%	-9.8%	-11.1%	-8.7%	-10.0%	-9.6%
0.5	**CAT**	-21.6%	-17.6%	-8.4%	-9.2%	-7.0%	-8.0%	-7.9%
0.6	**CAT**	-19.0%	-17.6%	-8.3%	-9.4%	-7.3%	-8.4%	-8.1%
0.7	**CAT**	-18.3%	-16.6%	-7.9%	-8.7%	-6.6%	-7.6%	-7.4%
0.8	**CAT**	-17.0%	-15.5%	-7.3%	-8.1%	-6.0%	-7.0%	-6.8%
0.9	**CAT**	-13.8%	-12.2%	-5.0%	-5.5%	-4.0%	-4.7%	-4.5%

Table 23 - RB of Factor Loadings with Bimodal Scale Distortion

The RB results for bimodal distortion conditions show an apparent shift in the moderate threshold down to the 4 Category level. However, if the objective is again to get trivial bias results only, then this is now a threshold for greater than 0.8 factor loadings and 6 categories or higher only. The results show a very narrow band of trivial bias, in a region of factor loading strength that is rarely encountered in practice. As with skewed results it appears that at best moderate RB results are obtained, and the bimodal bias results appear to be affected more by factor loading strength, becoming more and more negatively biased as the factor loading has decreased, even in the higher category sizes. The much larger band of moderate bias is now very similar in size and shape to the trivial bias band found in Section 5.2 (for results with limited scale distortion).

Coefficient of Variance

CV		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	43.7%	37.8%	20.2%	22.8%	19.0%	21.4%	20.2%
0.5	**CAT**	29.0%	21.8%	12.9%	13.4%	11.6%	12.4%	12.7%
0.6	**CAT**	15.0%	14.6%	8.3%	9.4%	8.0%	8.9%	8.7%
0.7	**CAT**	10.9%	10.2%	6.0%	6.5%	5.4%	6.0%	5.9%
0.8	**CAT**	7.9%	7.3%	4.1%	4.5%	3.8%	4.1%	4.2%
0.9	**CAT**	5.4%	5.1%	2.4%	2.7%	2.2%	2.5%	2.5%

Table 24 - CV of Factor Loadings with Bimodal Scale Distortion

The CV result show the same aversion to factor loading strength as RB did, even in the higher category sizes. The previous ‘diagonal’ pattern observed in the CV results of the previous sections has almost flattened.

Modified box plots are provided in Figure 41 in the Appendices, and immediately one can see the uniformity in the spread of the results for the higher category sizes as discussed above. Figure 40 shows the simplified percentiles for bimodal distortion. Given a 6 category design, the expected probability of losing data would be 70% and 9% for 0.4 and 0.5

underlying factor loading respectively, whereas for a 7 category scale it would be 71% and 12% respectively.

5.6 Comparison of Methods in Limited Distortion Situations

Figure 18 and Figure 19 show the comparative statistics of RB and CV, respectively, for the limited distortion models tested. Figure 20 and Figure 21 show the differences of each method against the best performing method (amount of underperformance), for the given intended factor loading and category size, of RB and CV respectively, and includes a "total best performer" metric which is the average underperformance by category size for the method.

The CAT method (performing FA directly on ordinal data) and the NDF^2 method (NDF with the average of truncated distribution) perform very well, with the difference in performance between the two methods rarely exceeding 0.2% in RB or 0.5% in CV measures[14]. For categories 4 to 8 the CAT method is the best performer in both RB and CV. SPR method (Spearman Rank) fairs relatively the same as CAT and NDF^2 at 2 and 3 categories for both RB and CV, then slightly poorer for the rest; performing less than 1.5% worse on CV and RB than the best performer (in this case CAT) for 4 to 6 categories, and just over 1.5% worse for CV for 7 and 8 categories and remaining below 1.5% underperformance on RB for these two categories.

NDF^1 method (NDF with random sampling) is outperformed by the other methods, and as logic would suggest, random sampling, even within the bounds of a truncated normal distribution, further degrades the correlation strength during the re-scaling process.

The question is whether to discard NDF^2, the next best performer, as a method for applying factor analysis to categorical data? Certainly if the researcher is planning to obtain statistical measures other than factor analysis from empirical data, then there is a strong case made by Stacey (2006) for first 'standardising' the data before further analysis. It can

[14] One notable exception being an anomaly for NDF^2 for 4 Category at 0.5 intended factor loading.

be seen from these results that applying EFA on this basis would yield satisfactory results. However, if it is not the intention or the desire to first scale the data, then it can be seen that EFA applied directly to the categorical data under these situations yields the best results, and with much less time and effort considering the lengthy processing time required for the NDF algorithm (especially with large numbers of categories and items).

The study found that the NDF algorithm struggled in fitting a single threshold set to standard normal distributions for 8 items of 200 responses at higher category sizes. Perhaps future studies should investigate better ways to solve for this (different solving algorithms), or ways to perform NDF with a unique threshold per item instead of one shared threshold set across all items. The potential increase in fit may very well yield superior performance to that of CAT.

The RB and CV results give a fair idea of the performance of each method under these circumstances, but showing percentiles gives a better understanding of the probability of lost data in the various methods when a particular threshold is chosen. Percentile plots are provided in the Appendices (Figure 13, Figure 15, Figure 16, and Figure 17).

5.7 Comparison of Methods in Skewed Distortion Situations

For skewed distortion it is found that NDF^2 is now relegated to third best performer, with CAT performing best at RB in general, and SPR performing best at CV in general (Figure 27, Figure 28, Figure 29, and Figure 30). The differences in percentage between CAT and SPR are very slight; generally less than 1% for CV and approximately 1% difference for RB (for categories greater than 5). Section 5.3 showed that skew distortion creates on average the most spread in the results out of all the other distortion types. Increased spread leads to greater probability of data loss below the 0.4 acceptance criteria and greater spurious loadings, and therefore choosing a method that limits this is greatly advantages. Therefore SPR would be a preferred method when data is skewed, however the difference in performance is marginal based on the CV results. Looking at the percentile plots (Figure 22, Figure 24, Figure 25, and Figure 26) shows this slight improvement. As an example, on a 7 category response, using CAT method would give a probability of loss of data of 67%

and 21% for a 0.4 and 0.5 underlying factor loading respectively, while the SPR method shows 70% and 19% respectively, but with almost 1% in lost data for the 0.6 underlying factor loading.

5.8 Comparison of Methods in Kurtotic Distortion Situations

Under kurtotic distortion conditions, the basic CAT method appears again from the CV and RB results to be the overall preferred method, except on low category sizes on CV where SPR is preferred (Figure 36, Figure 37, Figure 38, and Figure 39).

NDF^2 method appears not to outperform on any of the parameters. Perhaps again this may be proved to be a valid method if future research is able to reduce the fitting process within the algorithm as discussed in Section 5.6.

It would appear from the face value results of RB and CV that kurtotic distortion is best analysed using the standard CAT method, and that the effort of rescaling prior to factor analysis certainly holds no advantage at all. Looking at the percentile plots (Figure 31, Figure 33, Figure 34, and Figure 35) further confirms this. As an example, given a 7 category response scale, using CAT method would give a probability of loss of data of 62% and 3% for a 0.4 and 0.5 underlying factor loading respectively, while the SPR method shows 67% and 7% respectively.

5.9 Comparison of Methods in Bimodal Distortion Situations

As with skewed results, the results show that when the responses have bimodal distortion, SPR becomes a more universally reliable method (Figure 18, Figure 19, Figure 20, and Figure 21). In this case, unlike the skewed distortion results, even on the RB measure it yields superior results. In many cases this outperformance is larger than 1%. Looking at the percentile plots (Figure 40, Figure 42, Figure 43, and Figure 44) further confirms this. As an example, given a 7 category response scale, using CAT method would give a probability of

loss of data of 71% and 12% for a 0.4 and 0.5 underlying factor loading respectively, while the SPR method shows 66% and 6% respectively.

NDF^2 here shows promising results, in most cases outperforming the CAT method on RB, and fairing well on CV against CAT (even bettering it on higher categories of 7 and 8).

6 CONCLUSION

6.1 Introduction

This study set out to begin to understand the isolated effect that ordinal data has on the reliability of exploratory factor analysis. It was considered as an important goal given the prolific use of this multivariate technique in all research fields, including business, and considering how often ordinal scales are used to gather data due to their simplicity and ease of use. The conclusions of this study, and suggestions for future research, are set out below.

6.2 Summary of the main findings

Similar to those studies cited earlier from the literature, on the effects of ordinal data on Confirmatory Factor Analysis, Maximum Likelihood Factor Analysis, and correlation strength, this study also finds a general deterioration of the factor structure due to categorization, and has referred to this as discretisation error.

Key findings:

- In general it has been seen that discretisation error's effect on factor loadings is worse as category size decreases, as lower underlying factor loading strengths are encountered, or as the distortion of the underlying normal distribution increases.
- The most destructive forms of distortion in the study were skewness, followed by bimodality.
- The benefits of selecting a higher category size far outweigh any attempts to 'recover' data after the fact. The recommendation is to choose a minimum category size of 7 when performing such data analysis, and the benefits in reducing

discretisation error's effects on EFA in doing this are far greater than using normal distribution fitting algorithm or inter scale distance dependant methods like Spearman-rank after the fact.

- Normal distribution fitting prior to EFA is not consistently superior enough to be considered favourable over performing EFA directly on ordinal data, yet it shows some promise and can perhaps be refined further.
- Performing EFA on Spearman rank correlation matrices is consistently superior under conditions of skewness or bimodality to be considered favourable over performing EFA directly on ordinal data.
- Discretisation error's affect on factor loading results are manifest as both a weakening in observed factor loading below the unobserved factor loading construct, as well as an incorrect strengthening of the observed factor loading above the unobserved factor loading construct (a spurious loading). This potentially leads to a researcher's incorrect inclusion of spurious data, as well as the unfortunate exclusion of valid data.

Based on the findings of this research there are two sets of recommendations to be made that could reduce the effects observed in this study, namely; those measures that a researcher can take during the design of their research, and those that can be applied after data has been gathered in the data analysis phase. These recommendations are summarised and discussed in Sections 6.2.1 and 6.2.2 below.

6.2.1 *Recommendations for Research Design*

This study has highlighted that a researcher should ensure that the minimal amount of distortion of the underlying attitudes or beliefs are introduced by the chosen response scale, and has provided a general guide to suitable methods when faced with such distortions. Yet it is the results that point to the choice in the number of categories to use in a study that is the most critical guideline offered here. The findings of this research show that the precautions a researcher can take in the design of their research, when using EFA on

ordinal data, will be far more effective at preventing data loss and distortion than any measures could be at attempting to rescale distorted data once it has already been gathered.

It is highly advisable to avoid the lower category sizes at all costs for two main reasons:

- The results show that the lower the number of categories, and the lower the communalities, the greater the variance of the observed factor loading results and the greater the mean negative bias. This higher variance and negative bias increase the risk of data being lost as significant findings load below the chosen factor loading threshold, as was observed in this research. It also raises a concern that lower communalities spuriously load higher than their actual underlying values. If in this study the factor loadings of a 0.4 true underlying factor loading were observed to spuriously load above 0.5, then one would need to ask how many non-significant underlying factor loadings of below 0.4 would spuriously load above this threshold, and as a result would be incorrectly interpreted as significant results? This study did not investigate the spurious loadings of non-significant underlying factor loadings, and so it can only be speculated as to whether or not this could be a real concern. Yet it has been observed that the variance of the factor loadings increase as the communalities decrease in strength, and therefore this risk of spurious loading would be highly probable.
- Since the results show that lower category sizes are affected even greater when underlying attitudes are distorted, and it may be unknown in advance by a researcher what sort of distortions may be introduced in their study or what weakness of communalities may be uncovered, it would be more prudent of the researcher to select higher category sizes when designing the research. Selecting higher category sizes would be more robust to sometimes unavoidable distortions introduced through the way respondents capture their underlying attitudes or beliefs.

This study suggests the use of category sizes of 7 or more if at all possible, given that if communalities are closer to the 0.4 level and high levels of distortion are experienced, then at least the researcher is assured the highest practical probability of not missing this data,

and is given some comfort in the reduced possibility of spurious loadings contaminating the results, as can be seen in Table 25.

Factor Loading Threshold[15]	Recommended Minimum Category Size	Expectation[16]
0.4 FL[17]	7 Category	Under worst conditions of distortion observed (Skewness), 67% probability of 0.4 true underlying factor loading data, and 21% probability of 0.5 true underlying factor loading data would be lost, with an unknown amount of spurious loading from non-significant factor loadings below 0.4.
0.5 FL[18]	7 Category	Under worst conditions of distortion observed (Skewness), 93% probability of 0.4 true underlying factor loading data successfully excluded from the practical significant findings. 75% probability of 0.5 true underlying factor loading data, and 22% probability of 0.6 true underlying factor loading data, lost. 7% probability of spurious loading from the 0.4 true underlying factor loading data, with an unknown amount of spurious loading from non-significant factor loadings below 0.4.

Table 25 - Category size recommendation

[15] The threshold chosen by the researcher given a sample size of 200 (as with this study). Factor loadings observed below this threshold would not be interpreted as significant to their study.
[16] Is the worst expectation based on the constraints of this study, namely; n=200 and skew distortion as tested.
[17] Considered to be the minimum factor loading for significance at n=200, and the minimum to interpret structure (Hair et al, 2005).
[18] Considered to be the minimum factor loading for practical significance at n=200 (Hair et al, 2005).

6.2.2 *Recommendations for Data Analysis*

Once the data has been gathered, it is recommended that the following steps are followed for data analysis.

Step 1 – Analyse the observed data and determine if any distortion is present.

Step 2 – Refer to Table 26 below, and determine the suggested method to apply (if any) prior to EFA.

Step 3 – Perform EFA.

Step 5 – Choose a factor loading threshold using the table below as a guide; below which the factor will be excluded from the analysis and deemed not to be practically significant. Setting it higher than normal would reduce the chance of spurious loadings but would also exclude much of the wanted data, and setting it lower would include the wanted data but increase the chances of erroneous data being included in the results. Care should still be taken in the selection of this threshold, and it is strongly recommended that the percentile plots in the appendix are consulted, rather than just blindly applying the thresholds included below. It must be remembered that these results are based on a sample size of 200.

Step 6 – Proceed with caution when interpreting the results. Where possible seek further validation of the results, and be very wary of unexpected or unusual constructs. This cautious approach is particularly prudent; when using a category size below 7, when confronted with a study that deviates from the constructs of this study, when the most devastating forms of scale distortion analysed here are encountered (skewness and bimodality), or when a type of distortion is encountered that was not studied here (such as opposite skew).

Distortion Observed	Recommended Method prior to EFA	Recommended Factor Loading Threshold[19]	Expectation
Limited Distortion	None	0.46 FL	96.8% probability of 0.4 true underlying factor loading data being successfully excluded. 77.4% probability of 0.5 true underlying factor loading data being observed on or above the threshold (successfully included). An unknown amount of spurious loading from non-significant factor loadings below 0.4.
Skew Distortion	Spearman Rank	0.49 FL	95.7% probability of 0.4 true underlying factor loading data being successfully excluded. 28.3% probability of 0.5 true underlying factor loading data, and 88.4% probability of 0.6 true underlying factor loading data, being observed on or above the threshold (successfully included). An unknown amount of spurious loading from non-significant factor loadings below 0.4.
Kurtotic Distortion	None	0.47 FL	95.1% probability of 0.4 true underlying factor loading data being successfully excluded. 56.8% probability of 0.5 true underlying factor loading data being observed on or above the threshold (successfully included). An unknown amount of spurious loading from non-significant factor loadings below 0.4.

[19] Considered to be the estimated factor loading threshold to maximise the probability of including the true underlying factor loadings of 0.5 or more, and minimize the probability (5% or less) of including the 0.4 true underlying factor loading, for practical significance at n=200 (Hair et al, 2005).

Distortion Observed	Recommended Method prior to EFA	Recommended Factor Loading Threshold[19]	Expectation
Bimodal Distortion	Spearman Rank	0.48 FL	96% probability of 0.4 true underlying factor loading data being successfully excluded. 43.2% probability of 0.5 true underlying factor loading data, and 97.7% probability of 0.6 true underlying factor loading data, being observed on or above the threshold (successfully included). An unknown amount of spurious loading from non-significant factor loadings below 0.4.

Table 26 - Recommended Methods

6.3 Recommendations for future research

Many questions were highlighted in this research and remain unanswered.

1. ***What could be some of the conditions under which discretisation error creates spurious significant factor loadings from EFA performed on factor structures that have underlying non-significant factor loadings?*** Given that this study observed variances in factor loading results of EFA performed on ordinal data, and it can be seen that much of this variance appears as a spurious loading above the true underlying factor loading, then what is the probability of smaller underlying factor loadings that are not significant to begin with, spuriously loading high enough (due to discretisation error) for them to be incorrectly accepted as significant loadings? This is probable given that it can be observed from this study that the variance increases as communalities decrease. The significance of a study of this nature would be that if spurious loadings are not found to be as problematic as are expected, then the recommended thresholds given in this study can be revised to increase the probability of capturing more relevant factor loadings. However, if the spurious loading is indeed an issue, then further guidelines can be provided for the researcher to ensure that these spurious loadings are excluded from any analysis.
2. ***What is the effect on the validity of EFA performed on ordinal data when factors are composed of different sized ordinal scales?*** Given that large category sizes perform better than smaller category sizes in EFA performed on ordinal data, yet research questionnaires are typically made up of at least some mixed ordinal scales, do large category sizes increase the reliability of the results of smaller category sizes within the same factor, performing a stabilising effect, or is there a weakening effect on the larger category by the smaller category?
3. ***What is the effect of sample size on the validity of EFA performed on ordinal data?*** This study focused only on one sample size given the physical constraints, yet this is extremely pertinent to any researcher for any statistical analysis. Given that this study has narrowed the focus of the most relevant parameters affecting EFA

performed on ordinal data, further research with sample size as central research parameter would add greater external validity to these results providing greater guidelines to future researchers.

4. ***What is the effect of opposite skew in ordinal variables within a factor on the validity of EFA performed on these variables?*** It was discussed in the literature review that both Olsson (1979) and Wylie (1979) found opposite skew in variables to be the most destructive distortion for analyses like these, yet this was not within the scope of this research. It is conceivable that distortions of attitudes and beliefs created through completing ordinal scales would not all follow the same form of distortion from item to item, that is; the same level or direction of skewness. Therefore an analysis like this would add greater insight into the effects of scale distortion prior to EFA.

REFERENCES

BABAKUS, E., FERGUSON, C. E., JR. & JÖRESKOG, K. G. (1987) The Sensitivity of Confirmatory Maximum Likelihood Factor Analysis to Violations of Measurement Scale and Distributional Assumptions. *Journal of Marketing Research,* Vol. 24, No. 2, pp. 222-228.

BARTHOLOMEW, D. J., STEELE, F., MOUSTAKI, I. & GALBRAITH, J. I. (2002) *The analysis and interpretation of multivariate data for social scientists*, Boca Raton: Chapman & Hall.

BAILEY, K. D. (1987) *Methods of Social Research,* New York, The Free Press.

BENDIXEN, M. & SANDLER, M. (1995) Converting verbal scales to interval scales using correspondence analysis. *Management Dynamics: Contemporary Research,* Vol. 4, No. 1, pp. 32-50.

BOLLEN, K. A. & BARB, K. H. (1981) Pearson's R and Coarsely Categorized Measures. *American Sociological Review,* vol. 46, No. 2, pp. 232-239.

DOLAN, C. V. (1994) Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology,* Vol. 47, pp. 309-326.

FLORA, D. B. & CURRAN, P. J. (2004) An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods,* Vol. 9, No. 4, pp. 466-491.

FULLER, E. L. & HEMMERLE, W. J. (1966) Robustness of the maximum likelihood estimation procedure in factor analysis. *Psychometrika,* Vol. 31, pp. 255-266.

GAUTAM, S., KIMELDORF, G. & SAMPSON, A. R. (1996) Optimized scorings for ordinal data for the general linear model. *Statistics and Probability Letters,* Vol. 27, pp. 231-239.

GNANADESIKAN, R. & KETTENRING, J. R. (1984) A pragmatic review of multivariate methods in applications. IN DAVID, H. A. & DAVID, H. T. (Eds.) *Statistics: An Appraisal.* Ames, Iowa State University Press.

GORSUCH, R. L. (1983) *Factor Analysis,* New Jersey, Lawrence Erlbaum Associated, Inc.

HAIR, J. F. J., BLACK, W. C., BABIN, B. J., ANDERSON, R. E. & TATHAM, R. L. (2005) *Multivariate Data Analysis*, Prentice Hall.

HARLOW, L. L. (1985) Behaviour of some elliptical theory estimators with nonnormal data in a covariance structure framework: A Monte Carlo study. *Unpublished PhD dissertation.* Los Angeles, University of California.

HARWELL, M. R. & GATTI, G. G. (2001) Rescaling Ordinal Data to Interval Data in Educational Research. *Review of Educational Research,* Vol. 71, No. 1, pp. 105-131.

HAWKES, R. K. (1971) The Multivariate Analysis of Ordinal Measures. *The American Journal of Sociology,* Vol. 76, No. 5, pp. 908-926.

HENSLER, C. & STIPAK, B. (1979) Estimating Interval Scale Values for Survey Item Response Categories. *American Journal of Political Science,* Vol. 23, No. 3, pp. 627-649.

HUNT, N. (2001) Generating Multivariate Normal Data in Excel. *Teaching Statistics,* 23, 2.

HUTCHINSON, S. R. & OLMOS, A. (1998) Behaviour of descriptive fit indexes in confirmatory factor analysis using ordered categorical data. *Structural Equation Modeling,* Vol. 5, pp. 344-364.

JÖRESKOG, K. G. & MOUSTAKI, I. (2001) Factor analysis or ordinal variables: a comparison of three approaches. *Multivariate Behavioural Research*, Vol. 36, pp. 347-387.

KIM, J.-O. (1975) Multivariate Analysis of Ordinal Variables. *The American Journal of Sociology,* Vol. 81, No. 2, pp. 261-298.

LABOVITZ, S. (1967) Some Observations on Measurement and Statistics. *Social Forces,* Vol. 46, No. 2, pp. 151-160.

LABOVITZ, S. (1971) In Defence of Assuming Numbers to Ranks. *American Sociological Review,* Vol. 36, No. 3, pp. 521-522.

LEWIS-BECK, M. S. (Ed.) (1993) *International Handbooks of Quantitative Applications in the Social Sciences,* United Kingdom, Toppan Company, with the cooperation of Sage Publications, Inc.

LEWIS-BECK, M. S. (Ed.) (1994) *International Handbooks of Quantitative Applications in the Social Sciences,* United Kingdom, Toppan Company, with the cooperation of Sage Publications, Inc.

LUBKE, G. H. & MUTHÉN, B. O. (2004) Applying Multigroup Confirmatory Factor Models for Continuous Outcomes to Likert Scale Data Complicates Meaningful Group Comparisons. *Structural Equation Modeling,* Vol. 11, No. 4, pp. 514-534.

LYNCH, R. M. (1994) Common Elements Correlation. *Teaching Statistics,* 16, 1.

MAYER, L. S. (1970) Comments on 'The assignment of numbers to rank order categories'. *American Sociological Review,* Vol. 35, pp. 916-917.

MICROSOFT (2007) Description of NORMDIST function in Excel.

MOORE, T. (2001) Generating Multivariate Normal Pseudo Random Data. *Teaching Statistics,* 23, 1.

MOUSTAKI, I. (2000) A review of exploratory factor analysis for ordinal categorical data. IN CUDECK, R., DU TOIT, S. & SÖRBOM, D. (Eds.) *Structural Equation Modelling: present and future*. Scientific Software International.

MUTHÉN, B. O. & KAPLAN, D. (1985) A comparison of some methodologies for the factor-analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology,* Vol. 38, pp. 171-180.

MUTHÉN, B. O. & KAPLAN, D. (1992) A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology,* Vol. 45, No. 1, pp. 19-30.

NEUMAN, L. W. (2006) *Social Research Methods: Qualitative and Quantitative Approaches,* Boston, Pearson International, Inc.

NUNNALLY, J. C. (1975) Psychometric Theory. 25 Years Ago and Now. *Educational Researcher,* Vol. 4, No. 10, pp. 7-14+19-21.

OLSSON, U. (1979) On the Robustness of Factor Analysis Against Crude Classification of the Observations. *Multivariate Behavioural Research,* Vol. 14, pp. 485-500.

POTTHAST, M. J. (1993) Confirmatory factor analysis of ordered categorical variables with large models. *British Journal of Mathematical and Statistical Psychology,* Vol. 46, pp. 273-286.

SOMERS, R. H. (1968) An Approach to the Multivariate Analysis of Ordinal Data. *American Sociological Review,* Vol. 33, No. 6, pp. 971-977.

SPEARMAN, C. (1904) The Proof and Measurement of the Association between Two Things. *American Journal of Psychology,* Vol. 15.

SRINIVASAN, V. & BASU, A. K. (1989) The Metric Quality of Ordered Categorical Data. *Marketing Science,* Vol. 8, No. 3, pp. 205-230.

STACEY, A. G. (2005) The reliability and validity of the item means and standard deviations of ordinal level response data. *Management Dynamics,* Vol. 14, No. 3, pp. 2-25.

STEVENS, S. S. (1946) On the theory of measurement. *Science,* 103, pp677-680.

TABACHNICK, B. G. & FIDELL, L. S. (2007) *Using Multivariate Statistics,* Boston, Pearson Education, Inc.

TEMME, D. (2006) Assessing measurement invariance of ordinal indicators in cross-national research. IN TERLUTTER, R. (Ed.) *International Advertising and Communication.* Wiesbaden, DUV.

VELLEMAN, P. F. & WILKINSON, L. (1993) Nominal, Ordinal, Interval, and Ratio Typologies Are Misleading. *The American Statistician,* Vol. 47, No. 1, pp. 65-72.

VIERRA, R. K. & CARLSON, D. L. (1981) Factor Analysis, Random Data, and Patterned Results. *American Antiquity,* Vol. 46, No. 2, pp. 272-283.

WEST, G. (2005) Better Approximations to Cumulative Normal Functions. *Wilmott Magazine.*

WILSON, T. P. (1971) Critique of Ordinal Variables. *Social Forces,* Vol. 49, No. 3, pp. 432-444.

WYLIE, P. B. (1976) Effects of coarse grouping and skewed marginal distributions on the Pearson product moment correlation coefficient. *Educational and Psychological Measurement,* Vol. 36, No. 1, pp.1-7.

YOUNG, F. W. (1981) Quantitative Analysis of Qualitative Data. *Psychometrika,* Vol. 46, No. 4, pp. 357-388.

7 APPENDICES

7.1 Categorisation – a Practical Example

A Practical Example for Ease of Understanding

In this study, simulation is done for 8 variables (8 survey items) and for a sample size of 200 (N = 200 or 200 respondents). Given a 5 category iteration, then the respondent's answers would either be '1', '2', '3', '4' or '5' depending on their underlying beliefs. They would choose between answering '2' or '3' to a specific survey item because there is an unobserved threshold that leads them to differentiate between these two answers. The thresholds are simulated as inputs in Step 4, and for this example of 5 categories there would be 4 thresholds, and it is the position of these thresholds that distort the respondent's answers from their true underlying belief as discussed before.

Therefore if for example:

The 10th response (sample 10) of the 5th item (survey item 5) will be '2' (C_2 = '2') if the respondent's underlying attitude on a continuous scale (represented by $Y'_{10,5}$) of '-0.9' was evaluated against their unobserved threshold set and was believed by the respondent to be larger than threshold 1 (τ_1 – say '-1.2') but smaller than threshold 2 (τ_2 – say '-0.8').

Written again in the format above for ease of understanding;

$$Y_{10,5} = C_2 \text{ if } \tau_1 < Y'_{10,5} \leq \tau_2$$

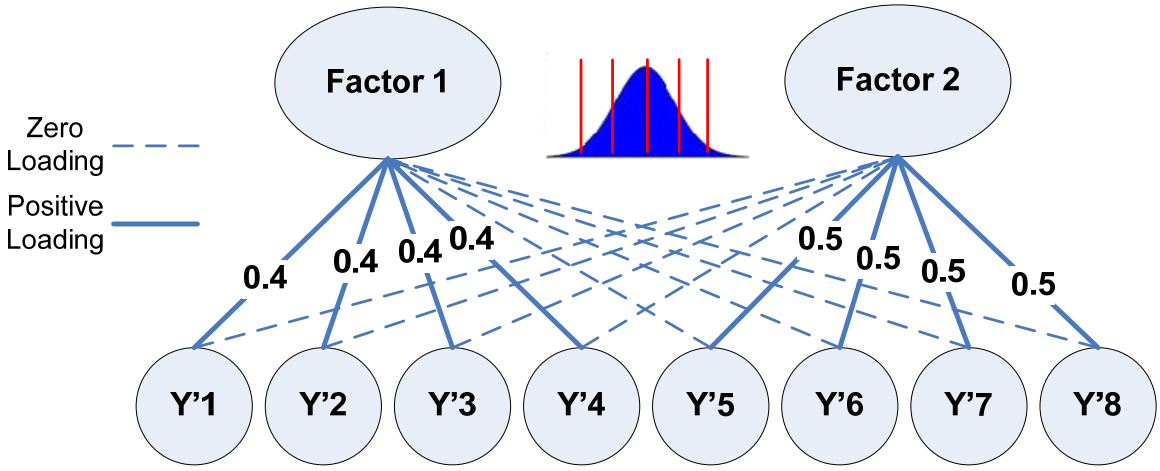
$$\therefore Y_{10,5} = \text{'2' if } -1.2 < -0.9 \leq -0.8$$

7.2 Schematics of Models

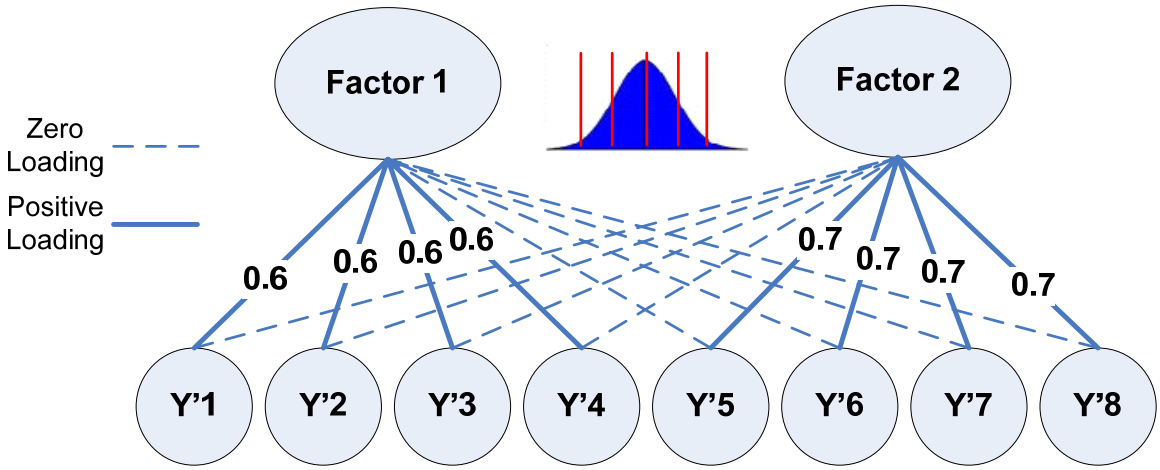
Each model's inputs, and the group to which it belonged, is highlighted in the bulleted points below.

Group 1 – Investigated the effects of 'pure' categorization on factors of various strengths.

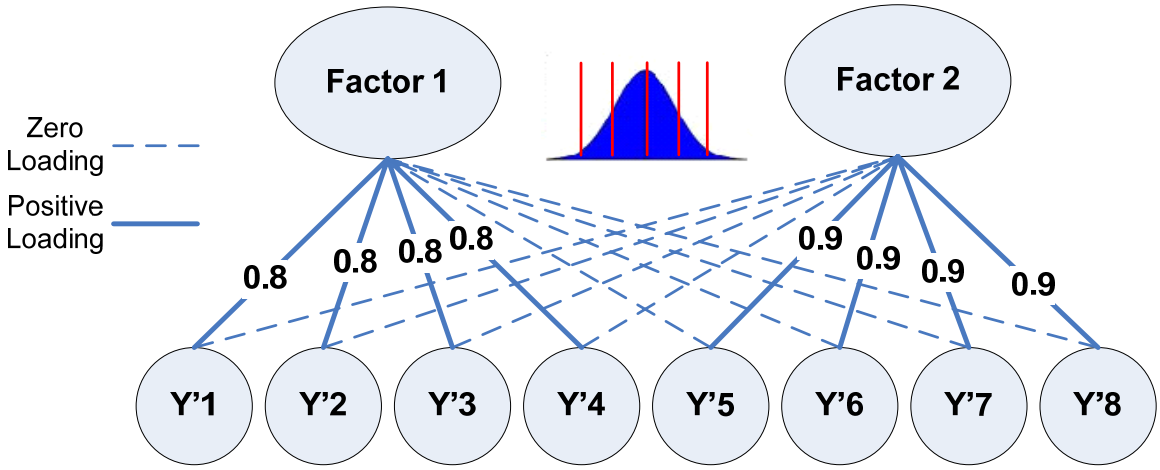
- Model 1 – Uses Factor 1 to investigate the effects of categorization with little distortion on a 0.4 Factor Loading, and uses Factor 2 to investigate the effects of categorization with little distortion on a 0.5 Factor Loading.



- Model 2 – Uses Factor 1 to investigate the effects of categorization with little distortion on a 0.6 Factor Loading, and uses Factor 2 to investigate the effects of categorization with little distortion on a 0.7 Factor Loading.

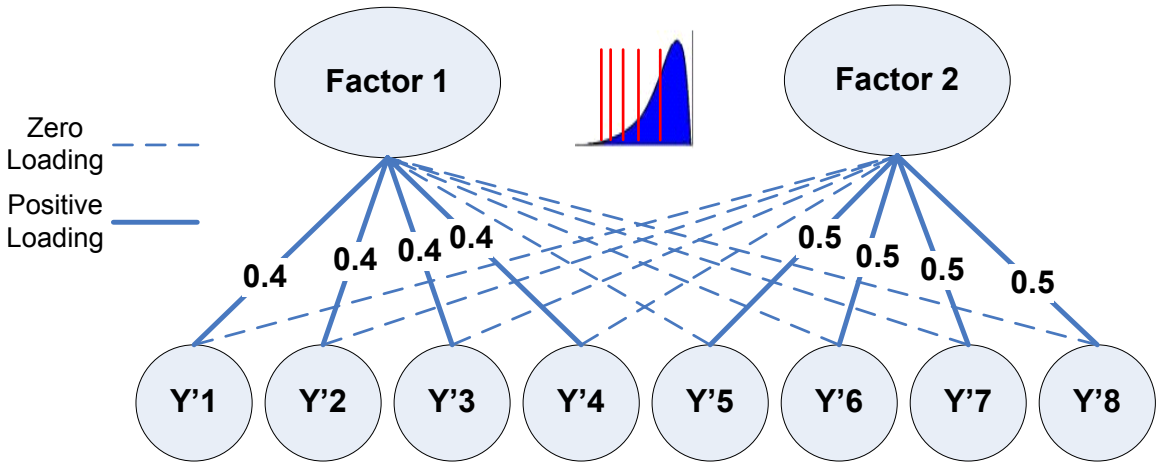


- Model 3 – Uses Factor 1 to investigate the effects of categorization with little distortion on a 0.8 Factor Loading, and uses Factor 2 to investigate the effects of categorization with little distortion on a 0.9 Factor Loading.

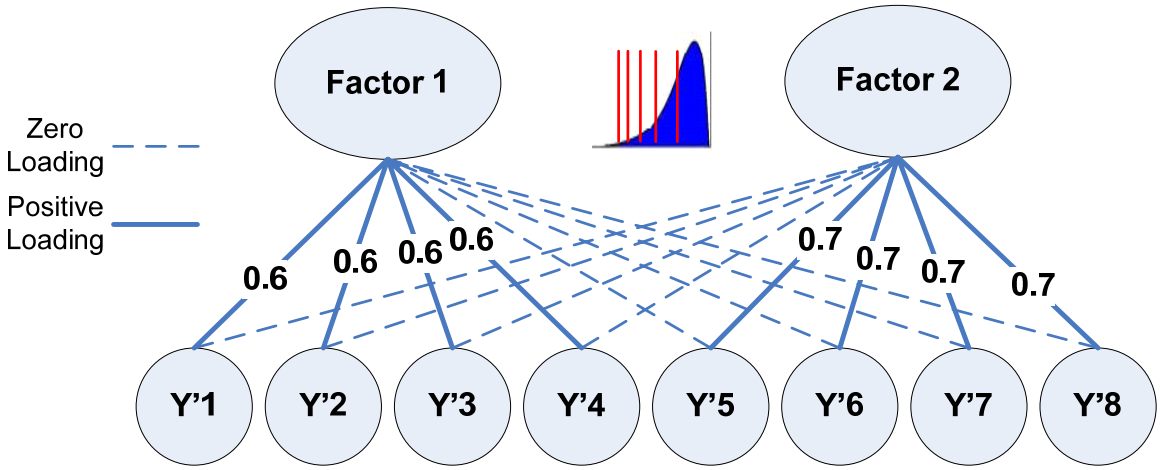


Group 2 – Investigated the effects of categorization with scale distortion on factors of various strengths.

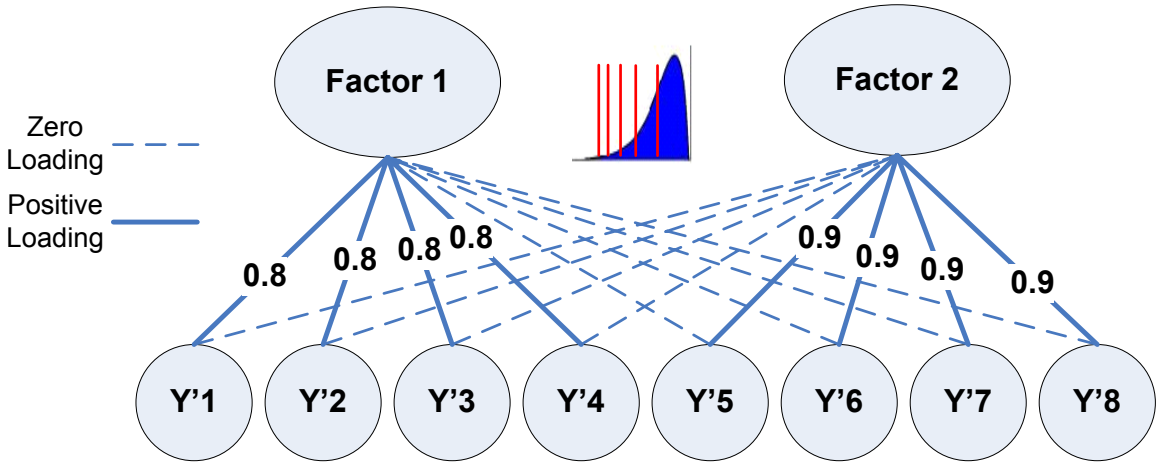
- Model 4 – Uses Factor 1 to investigate the effects of categorization with skew on a 0.4 Factor Loading, and uses Factor 2 to investigate the effects of categorization with skew on a 0.5 Factor Loading.



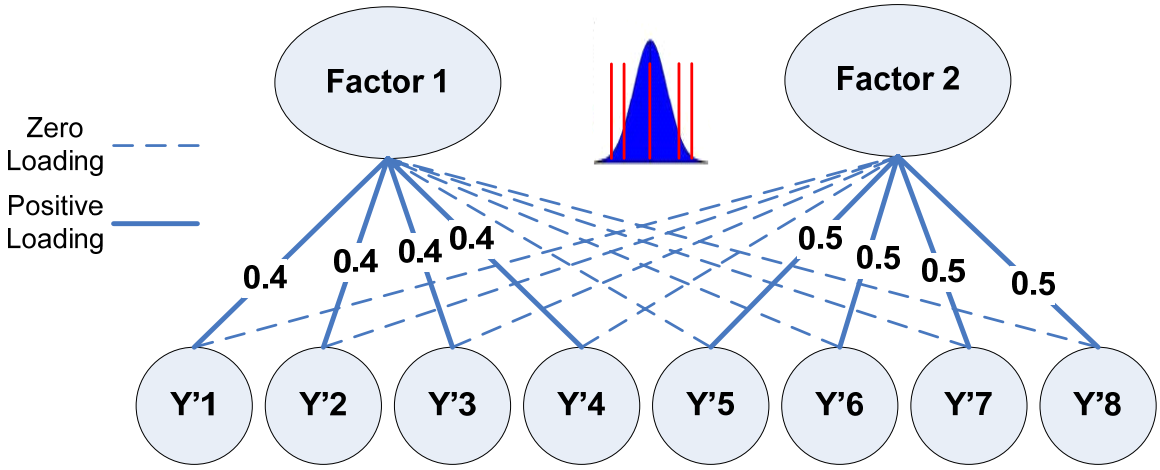
- Model 5 – Uses Factor 1 to investigate the effects of categorization with skew on a 0.6 Factor Loading, and uses Factor 2 to investigate the effects of categorization with skew on a 0.7 Factor Loading.



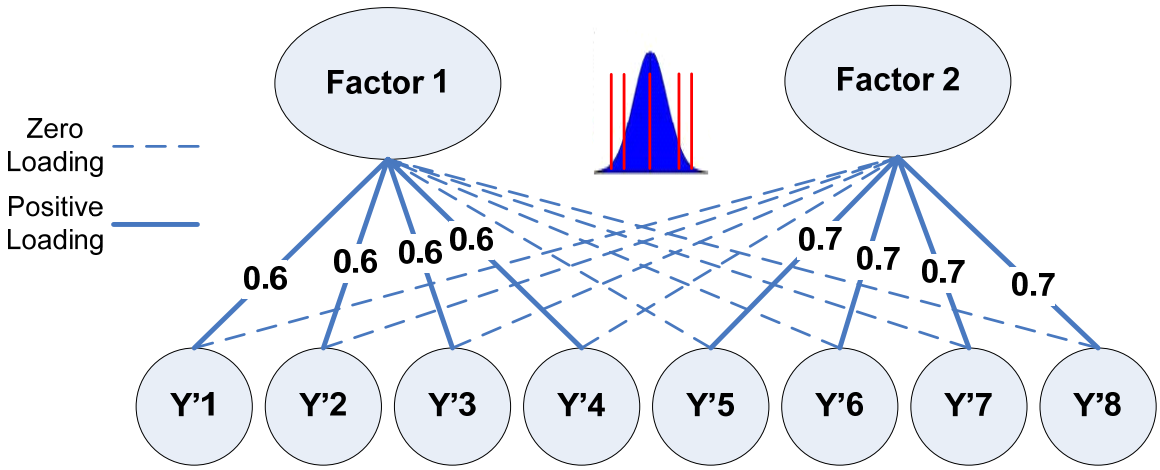
- Model 6 – Uses Factor 1 to investigate the effects of categorization with skew on a 0.8 Factor Loading, and uses Factor 2 to investigate the effects of categorization with skew on a 0.9 Factor Loading.



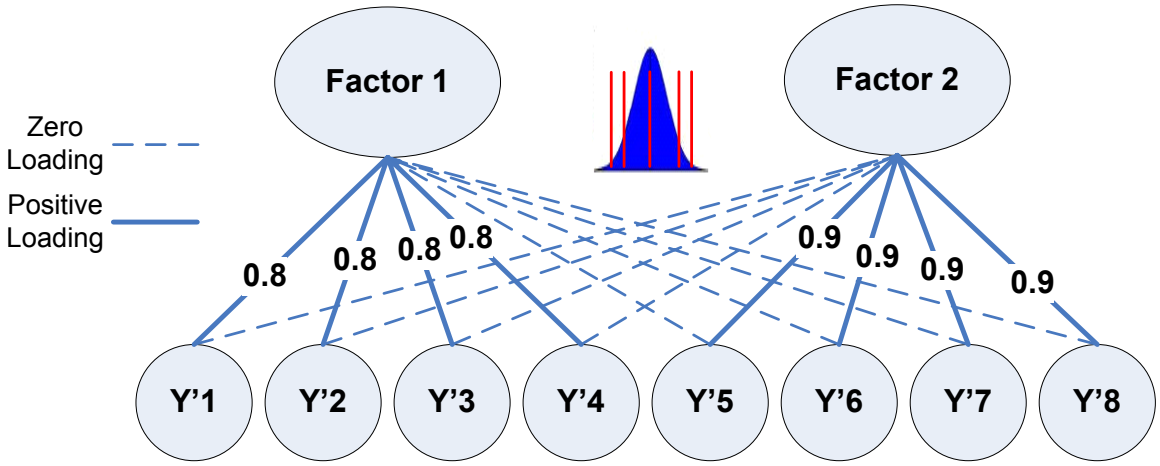
- Model 7 – Uses Factor 1 to investigate the effects of categorization with kurtosis on a 0.4 Factor Loading, and uses Factor 2 to investigate the effects of categorization with kurtosis on a 0.5 Factor Loading.



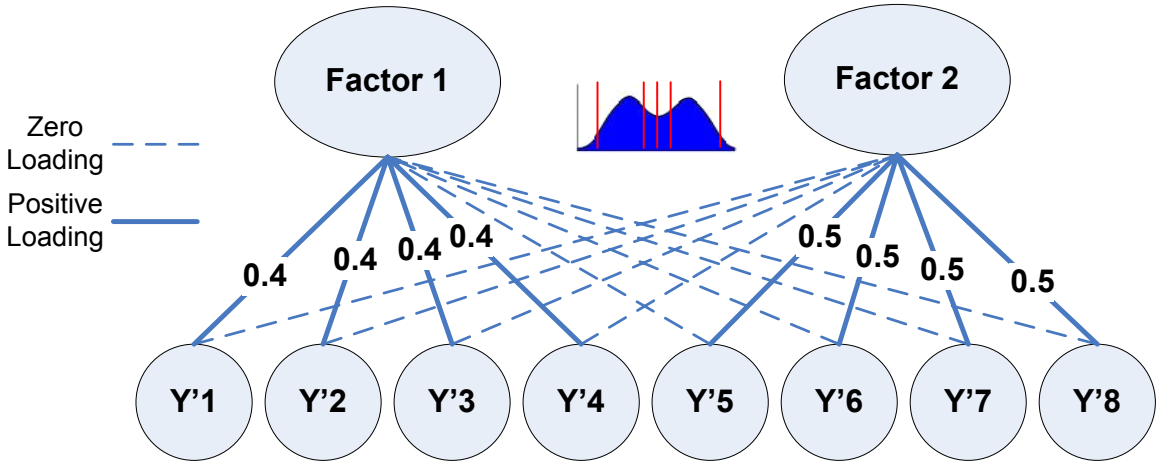
- Model 8 – Uses Factor 1 to investigate the effects of categorization with kurtosis on a 0.6 Factor Loading, and uses Factor 2 to investigate the effects of categorization with kurtosis on a 0.7 Factor Loading.



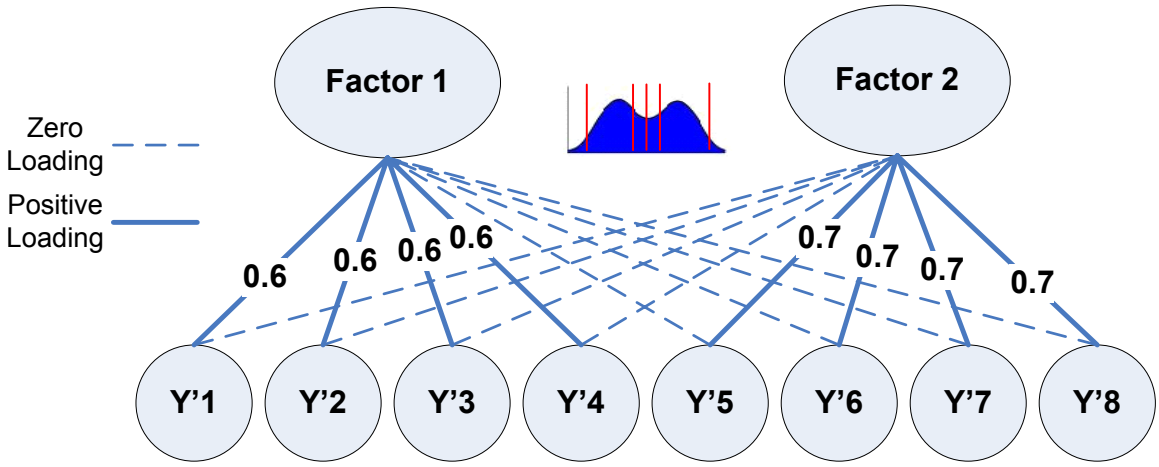
- Model 9 – Uses Factor 1 to investigate the effects of categorization with kurtosis on a 0.8 Factor Loading, and uses Factor 2 to investigate the effects of categorization with kurtosis on a 0.9 Factor Loading.



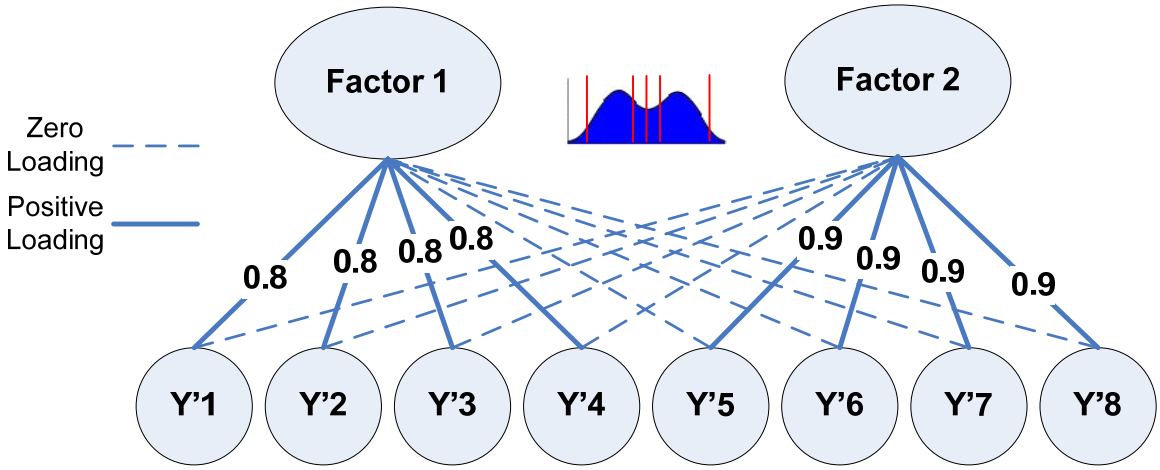
- Model 10 – Uses Factor 1 to investigate the effects of categorization with bimodality on a 0.4 Factor Loading, and uses Factor 2 to investigate the effects of categorization with bimodality on a 0.5 Factor Loading.



- Model 11 – Uses Factor 1 to investigate the effects of categorization with bimodality on a 0.6 Factor Loading, and uses Factor 2 to investigate the effects of categorization with bimodality on a 0.7 Factor Loading.



- Model 12 – Uses Factor 1 to investigate the effects of categorization with bimodality on a 0.8 Factor Loading, and uses Factor 2 to investigate the effects of categorization with bimodality on a 0.9 Factor Loading.



7.3 Factor Loading Descriptive Statistics – Little Scale Distortion

2 Category

Factor Loading	Method	μ	σ	RB	CV
0.4	CAT	0.317721	0.119502	-20.56975%	37.61214%
	SPR	0.317721	0.119502	-20.56975%	37.61214%
	NDF¹	0.207647	0.198831	-48.08830%	95.75452%
	NDF²	0.317721	0.119502	-20.56975%	37.61214%
0.5	CAT	0.397689	0.105868	-20.46219%	26.62085%
	SPR	0.397689	0.105868	-20.46219%	26.62085%
	NDF¹	0.287667	0.179056	-42.46652%	62.24407%
	NDF²	0.397689	0.105868	-20.46219%	26.62085%
0.6	CAT	0.489792	0.067076	-18.36805%	13.69480%
	SPR	0.489792	0.067076	-18.36805%	13.69480%
	NDF¹	0.394708	0.107912	-34.21531%	27.33982%
	NDF²	0.489792	0.067076	-18.36805%	13.69480%
0.7	CAT	0.573167	0.066303	-18.11903%	11.56781%
	SPR	0.573167	0.066303	-18.11903%	11.56781%
	NDF¹	0.457965	0.097149	-34.57641%	21.21323%
	NDF²	0.573167	0.066303	-18.11903%	11.56781%
0.8	CAT	0.666873	0.055893	-16.64083%	8.38142%
	SPR	0.666873	0.055893	-16.64083%	8.38142%
	NDF¹	0.53059	0.078058	-33.67621%	14.71161%
	NDF²	0.666873	0.055893	-16.64083%	8.38142%
0.9	CAT	0.776256	0.03999	-13.74933%	5.15171%
	SPR	0.776256	0.03999	-13.74933%	5.15171%
	NDF¹	0.619956	0.065018	-31.11604%	10.48750%
	NDF²	0.776256	0.03999	-13.74933%	5.15171%

Table 27 - FL Descriptive Statistics - Little Scale Distortion on 2 Categories

3 Category

Factor Loading	Method	μ	σ	RB	CV
0.4	CAT	0.342872	0.095217	-14.28204%	27.77043%
	SPR	0.342656	0.094986	-14.33594%	27.72048%
	NDF[1]	0.266874	0.165096	-33.28143%	61.86289%
	NDF[2]	0.342843	0.095426	-14.28914%	27.83360%
0.5	CAT	0.431909	0.071712	-13.61820%	16.60358%
	SPR	0.43183	0.072294	-13.63398%	16.74122%
	NDF[1]	0.348189	0.141202	-30.36216%	40.55333%
	NDF[2]	0.432072	0.071808	-13.58565%	16.61952%
0.6	CAT	0.512361	0.059906	-14.60643%	11.69222%
	SPR	0.512297	0.059898	-14.61716%	11.69201%
	NDF[1]	0.440677	0.089404	-26.55381%	20.28791%
	NDF[2]	0.512214	0.059824	-14.63096%	11.67942%
0.7	CAT	0.599645	0.050462	-14.33644%	8.41524%
	SPR	0.599595	0.050438	-14.34361%	8.41196%
	NDF[1]	0.516478	0.072183	-26.21748%	13.97598%
	NDF[2]	0.599611	0.050346	-14.34135%	8.39650%
0.8	CAT	0.688116	0.043149	-13.98553%	6.27054%
	SPR	0.688065	0.0431	-13.99186%	6.26390%
	NDF[1]	0.591904	0.06503	-26.01203%	10.98653%
	NDF[2]	0.688041	0.043144	-13.99484%	6.27049%
0.9	CAT	0.784828	0.038924	-12.79690%	4.95950%
	SPR	0.784742	0.038841	-12.80647%	4.94956%
	NDF[1]	0.669655	0.058435	-25.59385%	8.72607%
	NDF[2]	0.784887	0.039013	-12.79033%	4.97048%

Table 28 - FL Descriptive Statistics - Little Scale Distortion on 3 Categories

		4 Category			
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.364935	0.072469	-8.76614%	19.85793%
	SPR	0.35974	0.077985	-10.06505%	21.67814%
	NDF¹	0.329472	0.101994	-17.63194%	30.95668%
	NDF²	0.364969	0.072513	-8.75773%	19.86827%
0.5	**CAT**	0.460752	0.054741	-7.84965%	11.88088%
	SPR	0.453457	0.061839	-9.30855%	13.63729%
	NDF¹	0.419611	0.082724	-16.07783%	19.71438%
	NDF²	0.46086	0.054497	-7.82797%	11.82517%
0.6	**CAT**	0.551543	0.043921	-8.07621%	7.96338%
	SPR	0.544132	0.0472	-9.31134%	8.67446%
	NDF¹	0.504436	0.06457	-15.92730%	12.80047%
	NDF²	0.55143	0.044154	-8.09495%	8.00723%
0.7	**CAT**	0.641066	0.03794	-8.41921%	5.91825%
	SPR	0.633002	0.044599	-9.57109%	7.04560%
	NDF¹	0.58154	0.055042	-16.92280%	9.46485%
	NDF²	0.641008	0.037543	-8.42745%	5.85686%
0.8	**CAT**	0.731988	0.031254	-8.50152%	4.26975%
	SPR	0.724826	0.037214	-9.39675%	5.13422%
	NDF¹	0.669588	0.046039	-16.30151%	6.87578%
	NDF²	0.731773	0.031008	-8.52836%	4.23739%
0.9	**CAT**	0.823682	0.024164	-8.47981%	2.93366%
	SPR	0.818088	0.027853	-9.10130%	3.40463%
	NDF¹	0.750985	0.039524	-16.55723%	5.26294%
	NDF²	0.823279	0.024183	-8.52460%	2.93737%

Table 29 - FL Descriptive Statistics - Little Scale Distortion on 4 Categories

		5 Category			
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.379945	0.056334	-5.01384%	14.82678%
	SPR	0.374519	0.067122	-6.37036%	17.92210%
	NDF¹	0.359873	0.082611	-10.03173%	22.95556%
	NDF²	0.379951	0.056093	-5.01231%	14.76313%
0.5	**CAT**	0.470589	0.044807	-5.88221%	9.52142%
	SPR	0.46215	0.051082	-7.56992%	11.05311%
	NDF¹	0.445577	0.063718	-10.88456%	14.30015%
	NDF²	0.470583	0.044663	-5.88340%	9.49091%
0.6	**CAT**	0.56551	0.03541	-5.74841%	6.26155%
	SPR	0.558805	0.040926	-6.86575%	7.32379%
	NDF¹	0.534584	0.053086	-10.90268%	9.93036%
	NDF²	0.565332	0.035475	-5.77803%	6.27499%
0.7	**CAT**	0.660586	0.030863	-5.63052%	4.67205%
	SPR	0.652979	0.035802	-6.71726%	5.48290%
	NDF¹	0.622947	0.043497	-11.00758%	6.98242%
	NDF²	0.660467	0.030735	-5.64763%	4.65354%
0.8	**CAT**	0.755364	0.023695	-5.57944%	3.13688%
	SPR	0.74841	0.027439	-6.44875%	3.66628%
	NDF¹	0.71247	0.033474	-10.94121%	4.69828%
	NDF²	0.75447	0.024901	-5.69130%	3.30041%
0.9	**CAT**	0.846379	0.019276	-5.95791%	2.27751%
	SPR	0.840237	0.021695	-6.64033%	2.58206%
	NDF¹	0.794058	0.031056	-11.77131%	3.91107%
	NDF²	0.845661	0.019429	-6.03768%	2.29753%

Table 30 - FL Descriptive Statistics - Little Scale Distortion on 5 Categories

6 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.384206	0.048805	-3.94851%	12.70288%
	SPR	0.377476	0.057586	-5.63092%	15.25555%
	NDF¹	0.37156	0.069618	-7.11001%	18.73674%
	NDF²	0.383611	0.049487	-4.09720%	12.90028%
0.5	**CAT**	0.480783	0.036326	-3.84341%	7.55558%
	SPR	0.472392	0.044729	-5.52166%	9.46859%
	NDF¹	0.46011	0.052311	-7.97802%	11.36933%
	NDF²	0.480306	0.037015	-3.93879%	7.70645%
0.6	**CAT**	0.575497	0.029877	-4.08380%	5.19147%
	SPR	0.567569	0.035426	-5.40510%	6.24169%
	NDF¹	0.552006	0.045244	-7.99907%	8.19623%
	NDF²	0.574215	0.03554	-4.29743%	6.18924%
0.7	**CAT**	0.67199	0.025337	-4.00142%	3.77038%
	SPR	0.662583	0.032993	-5.34530%	4.97946%
	NDF¹	0.644468	0.034663	-7.93314%	5.37851%
	NDF²	0.669633	0.033796	-4.33810%	5.04701%
0.8	**CAT**	0.766989	0.019139	-4.12643%	2.49531%
	SPR	0.758717	0.024802	-5.16037%	3.26894%
	NDF¹	0.73556	0.027599	-8.05503%	3.75210%
	NDF²	0.766389	0.019484	-4.20143%	2.54238%
0.9	**CAT**	0.862147	0.015136	-4.20584%	1.75563%
	SPR	0.855049	0.018663	-4.99459%	2.18263%
	NDF¹	0.823622	0.024528	-8.48646%	2.97803%
	NDF²	0.861471	0.015346	-4.28104%	1.78136%

Table 31 - FL Descriptive Statistics - Little Scale Distortion on 6 Categories

		7 Category			
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.389131	0.041022	-2.71732%	10.54189%
	SPR	0.381708	0.05605	-4.57288%	14.68396%
	NDF¹	0.380167	0.060535	-4.95823%	15.92316%
	NDF²	0.389082	0.045554	-2.72955%	11.70795%
0.5	**CAT**	0.484561	0.03263	-3.08778%	6.73390%
	SPR	0.474582	0.042342	-5.08360%	8.92197%
	NDF¹	0.468982	0.043404	-6.20366%	9.25489%
	NDF²	0.483675	0.035583	-3.26508%	7.35689%
0.6	**CAT**	0.580195	0.026532	-3.30078%	4.57296%
	SPR	0.570943	0.032238	-4.84288%	5.64641%
	NDF¹	0.559106	0.034656	-6.81567%	6.19840%
	NDF²	0.577876	0.036889	-3.68728%	6.38350%
0.7	**CAT**	0.678679	0.021513	-3.04591%	3.16979%
	SPR	0.668367	0.029378	-4.51902%	4.39547%
	NDF¹	0.657238	0.028463	-6.10882%	4.33067%
	NDF²	0.678239	0.03188	-3.10870%	4.70039%
0.8	**CAT**	0.77564	0.017211	-3.04494%	2.21898%
	SPR	0.766407	0.02186	-4.19918%	2.85232%
	NDF¹	0.751172	0.025333	-6.10354%	3.37242%
	NDF²	0.774645	0.018097	-3.16939%	2.33620%
0.9	**CAT**	0.870809	0.012392	-3.24349%	1.42299%
	SPR	0.863193	0.016994	-4.08967%	1.96868%
	NDF¹	0.842113	0.019655	-6.43194%	2.33395%
	NDF²	0.869951	0.012527	-3.33879%	1.43998%

Table 32 - FL Descriptive Statistics - Little Scale Distortion on 7 Categories

8 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.388485	0.036111	-2.87876%	9.29542%
	SPR	0.381916	0.048288	-4.52105%	12.64372%
	NDF¹	0.381789	0.049954	-4.55287%	13.08416%
	NDF²	0.388369	0.036128	-2.90786%	9.30252%
0.5	**CAT**	0.488178	0.029015	-2.36436%	5.94346%
	SPR	0.477777	0.041236	-4.44466%	8.63080%
	NDF¹	0.474953	0.039976	-5.00931%	8.41692%
	NDF²	0.488129	0.029105	-2.37414%	5.96265%
0.6	**CAT**	0.587035	0.024012	-2.16087%	4.09038%
	SPR	0.57762	0.032043	-3.73004%	5.54743%
	NDF¹	0.574089	0.034145	-4.31850%	5.94767%
	NDF²	0.586817	0.024119	-2.19712%	4.11012%
0.7	**CAT**	0.684709	0.018551	-2.18446%	2.70936%
	SPR	0.673987	0.027386	-3.71617%	4.06323%
	NDF¹	0.667996	0.027403	-4.57206%	4.10233%
	NDF²	0.68453	0.018596	-2.20996%	2.71668%
0.8	**CAT**	0.781934	0.014584	-2.25826%	1.86510%
	SPR	0.772869	0.021415	-3.39136%	2.77082%
	NDF¹	0.76343	0.020902	-4.57122%	2.73788%
	NDF²	0.781537	0.014699	-2.30786%	1.88076%
0.9	**CAT**	0.878138	0.010425	-2.42907%	1.18718%
	SPR	0.87071	0.014807	-3.25447%	1.70055%
	NDF¹	0.856216	0.016839	-4.86494%	1.96672%
	NDF²	0.877487	0.010492	-2.50142%	1.19572%

Table 33 - FL Descriptive Statistics - Little Scale Distortion on 8 Categories

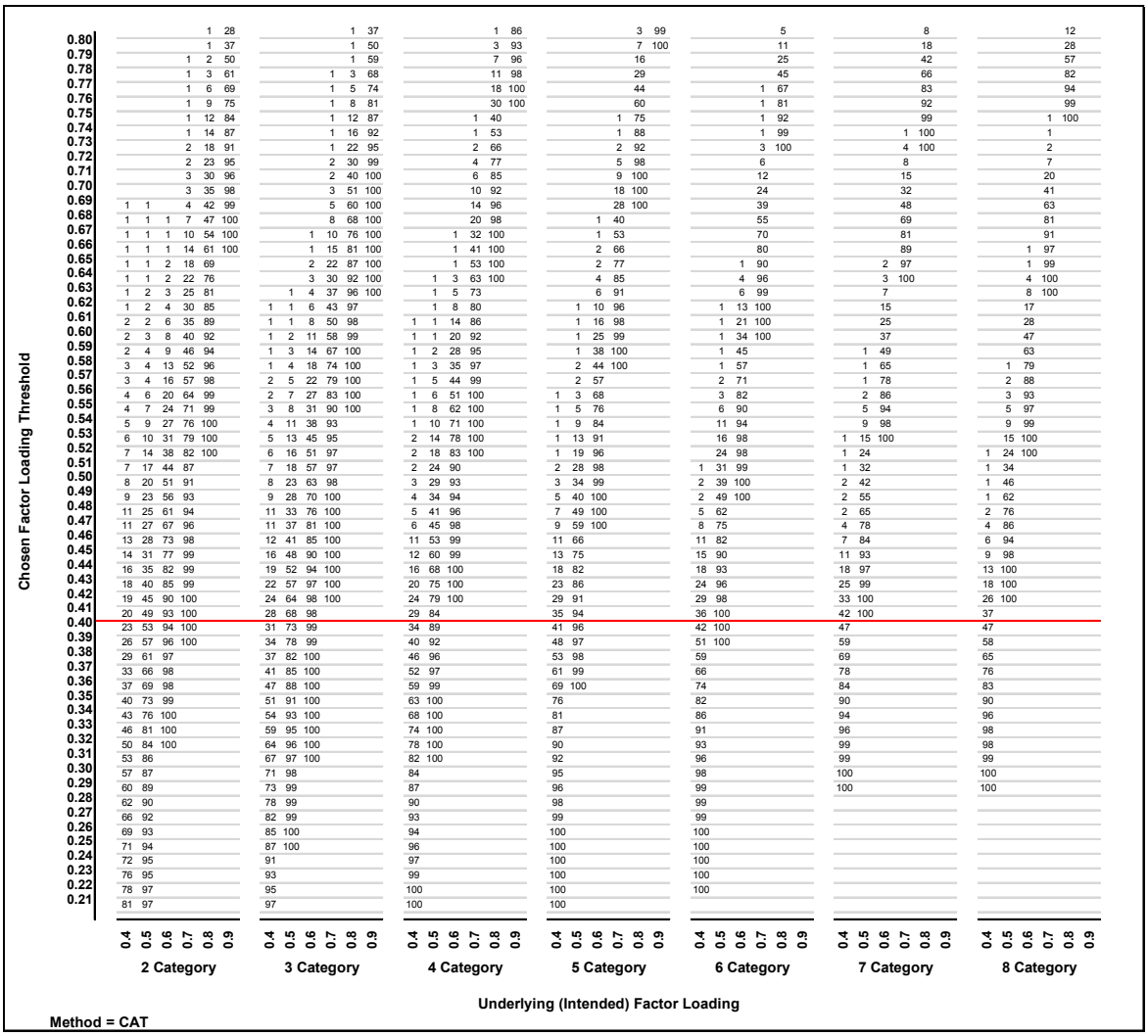
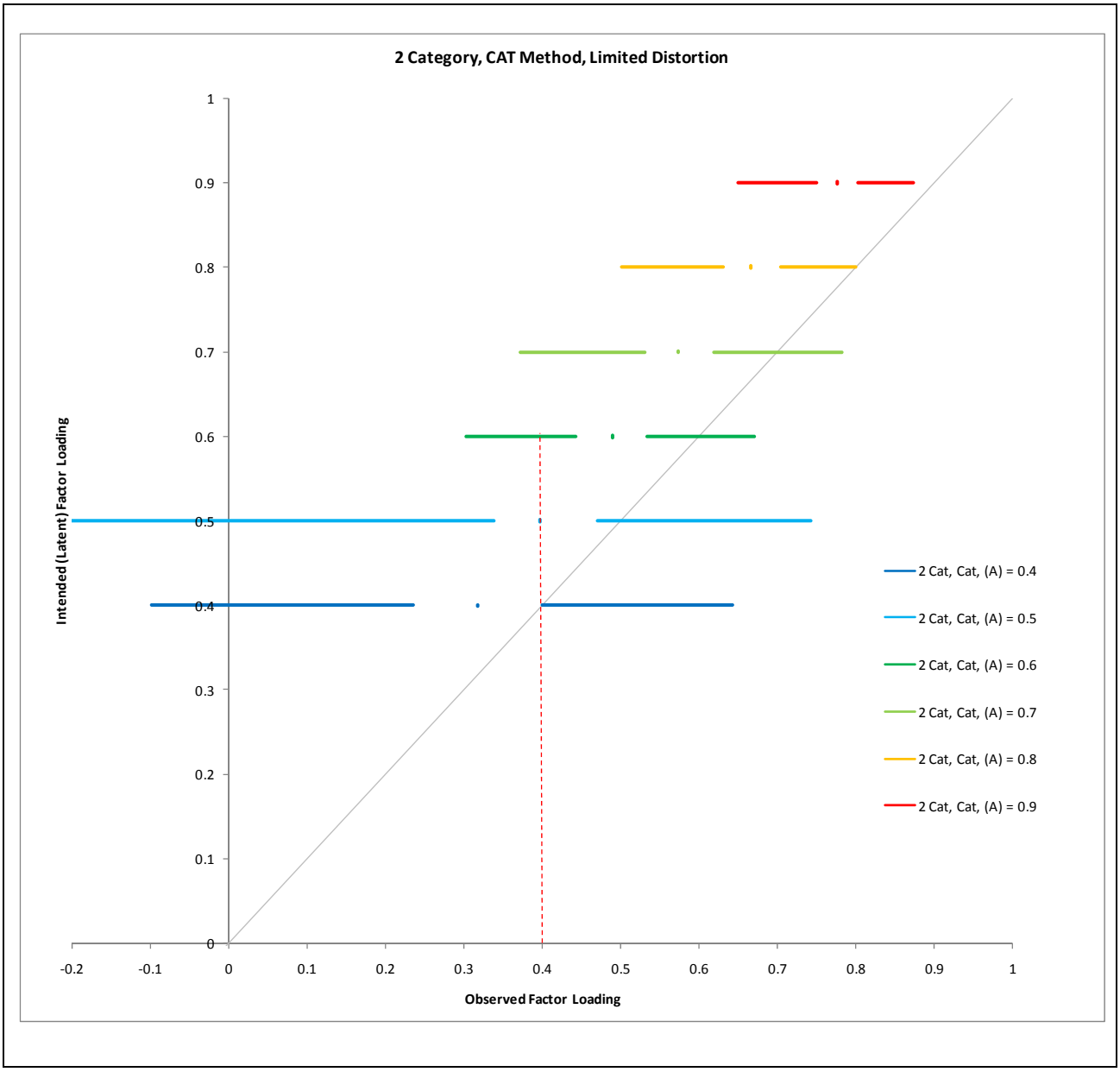
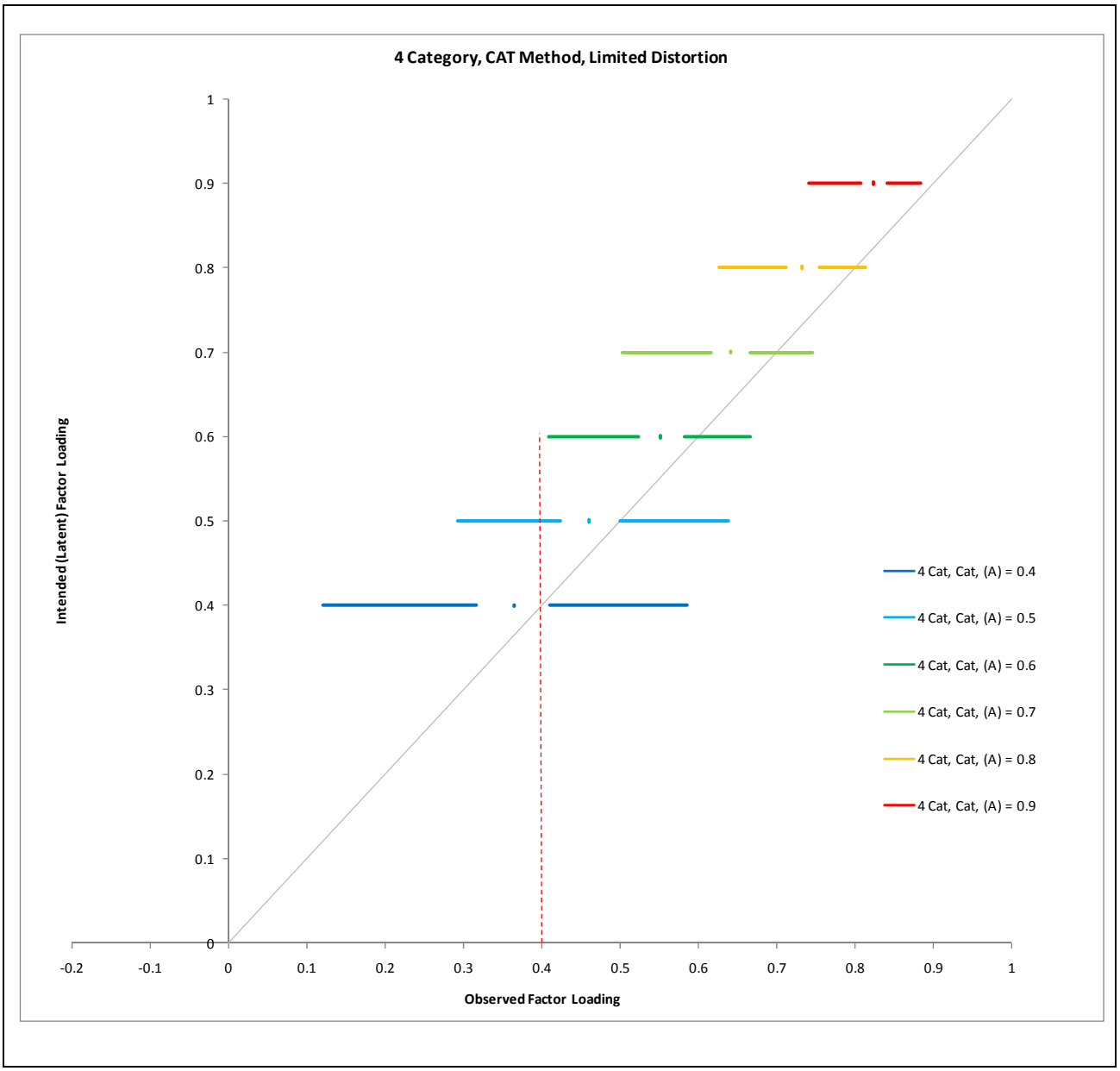
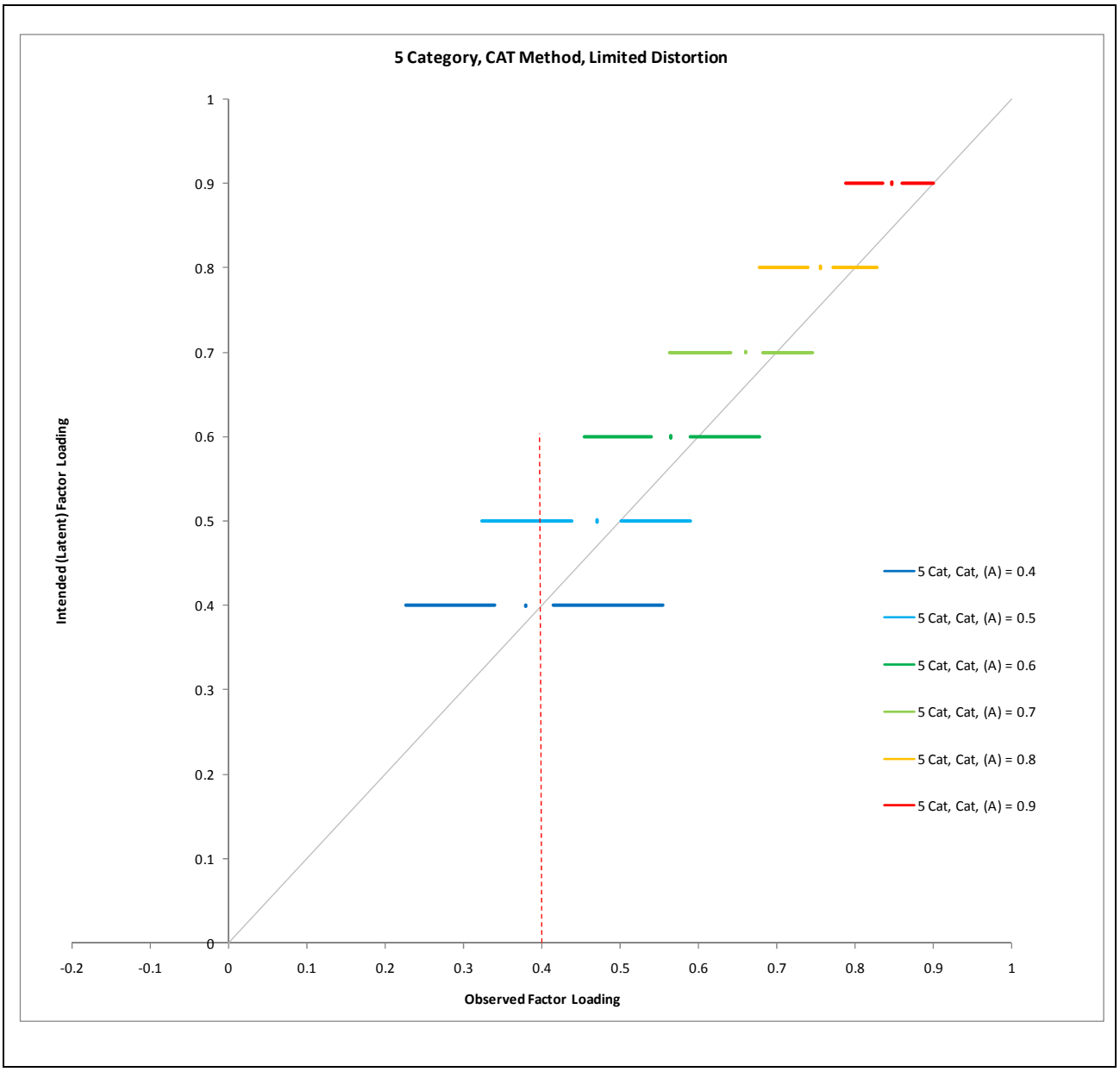
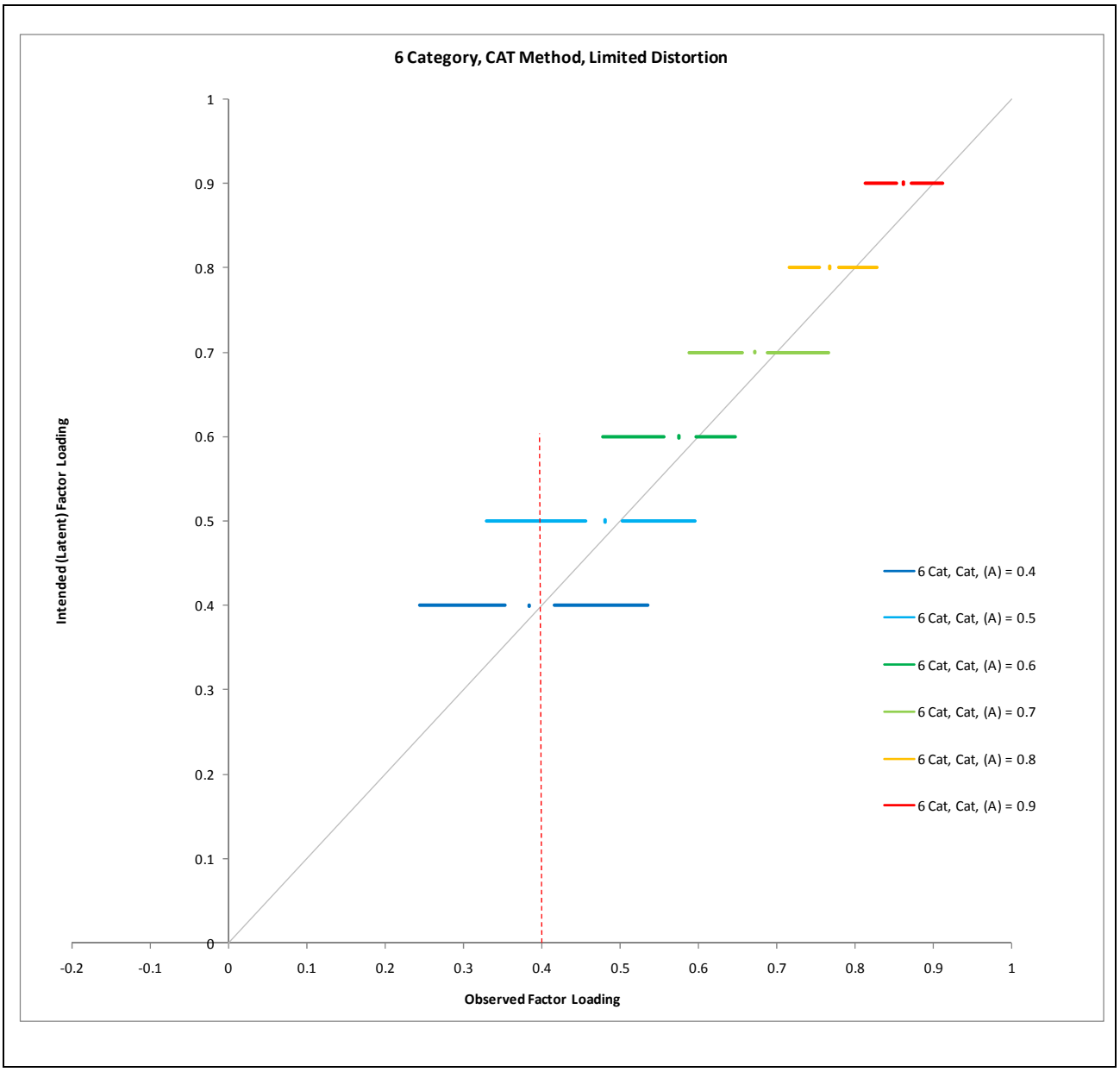


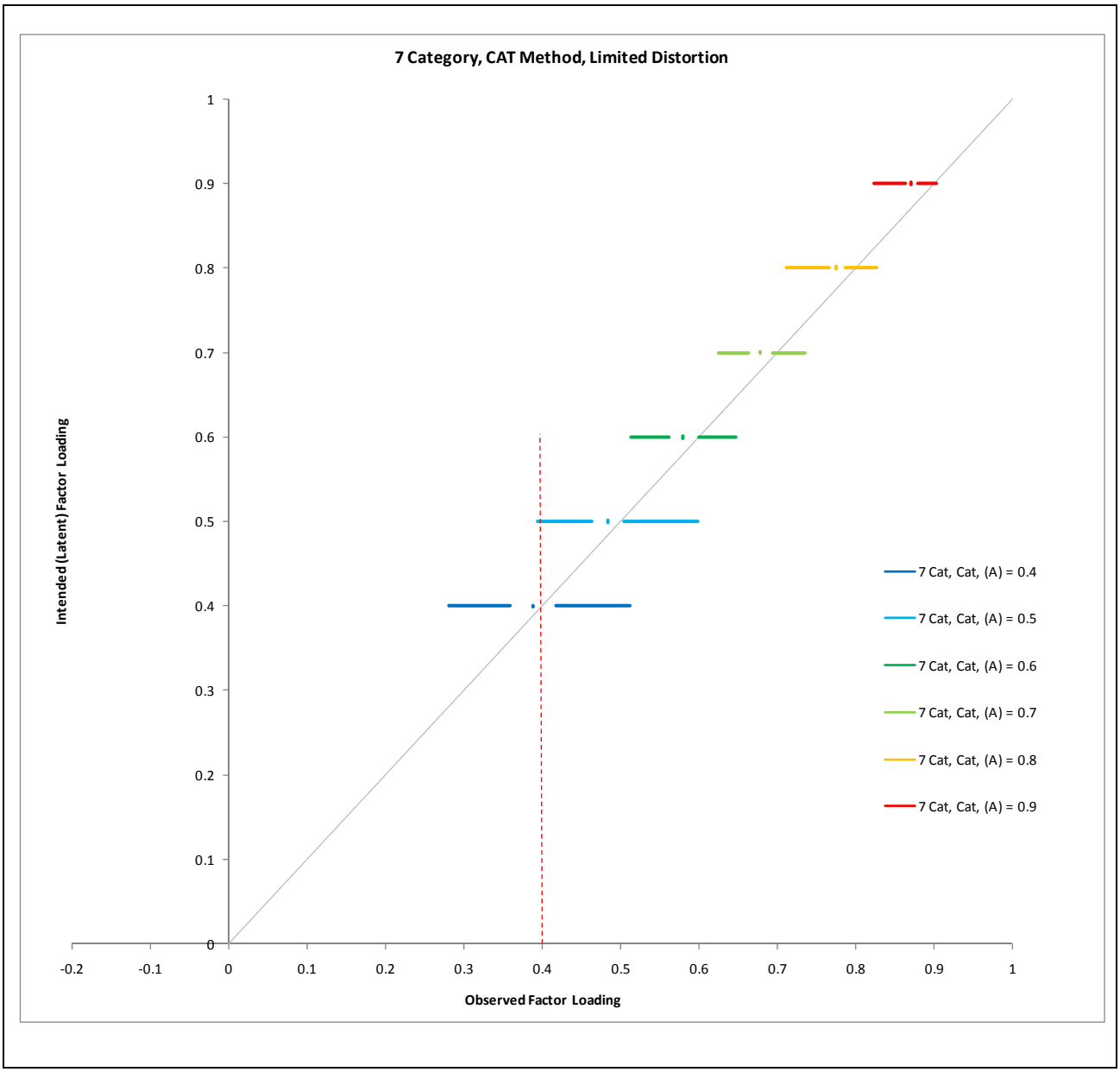
Figure 13 - Percentiles - CAT Method - Limited Distortion


2 Category, CAT Method, Limited Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2 Cat, Cat, (A) = 0.4
2 Cat, Cat, (A) = 0.5
2 Cat, Cat, (A) = 0.6
2 Cat, Cat, (A) = 0.7
2 Cat, Cat, (A) = 0.8
2 Cat, Cat, (A) = 0.9


4 Category, CAT Method, Limited Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
4 Cat, Cat, (A) = 0.4
4 Cat, Cat, (A) = 0.5
4 Cat, Cat, (A) = 0.6
4 Cat, Cat, (A) = 0.7
4 Cat, Cat, (A) = 0.8
4 Cat, Cat, (A) = 0.9


5 Category, CAT Method, Limited Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
5 Cat, Cat, (A) = 0.4
5 Cat, Cat, (A) = 0.5
5 Cat, Cat, (A) = 0.6
5 Cat, Cat, (A) = 0.7
5 Cat, Cat, (A) = 0.8
5 Cat, Cat, (A) = 0.9


6 Category, CAT Method, Limited Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
6 Cat, Cat, (A) = 0.4
6 Cat, Cat, (A) = 0.5
6 Cat, Cat, (A) = 0.6
6 Cat, Cat, (A) = 0.7
6 Cat, Cat, (A) = 0.8
6 Cat, Cat, (A) = 0.9


7 Category, CAT Method, Limited Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
7 Cat, Cat, (A) = 0.4
7 Cat, Cat, (A) = 0.5
7 Cat, Cat, (A) = 0.6
7 Cat, Cat, (A) = 0.7
7 Cat, Cat, (A) = 0.8
7 Cat, Cat, (A) = 0.9

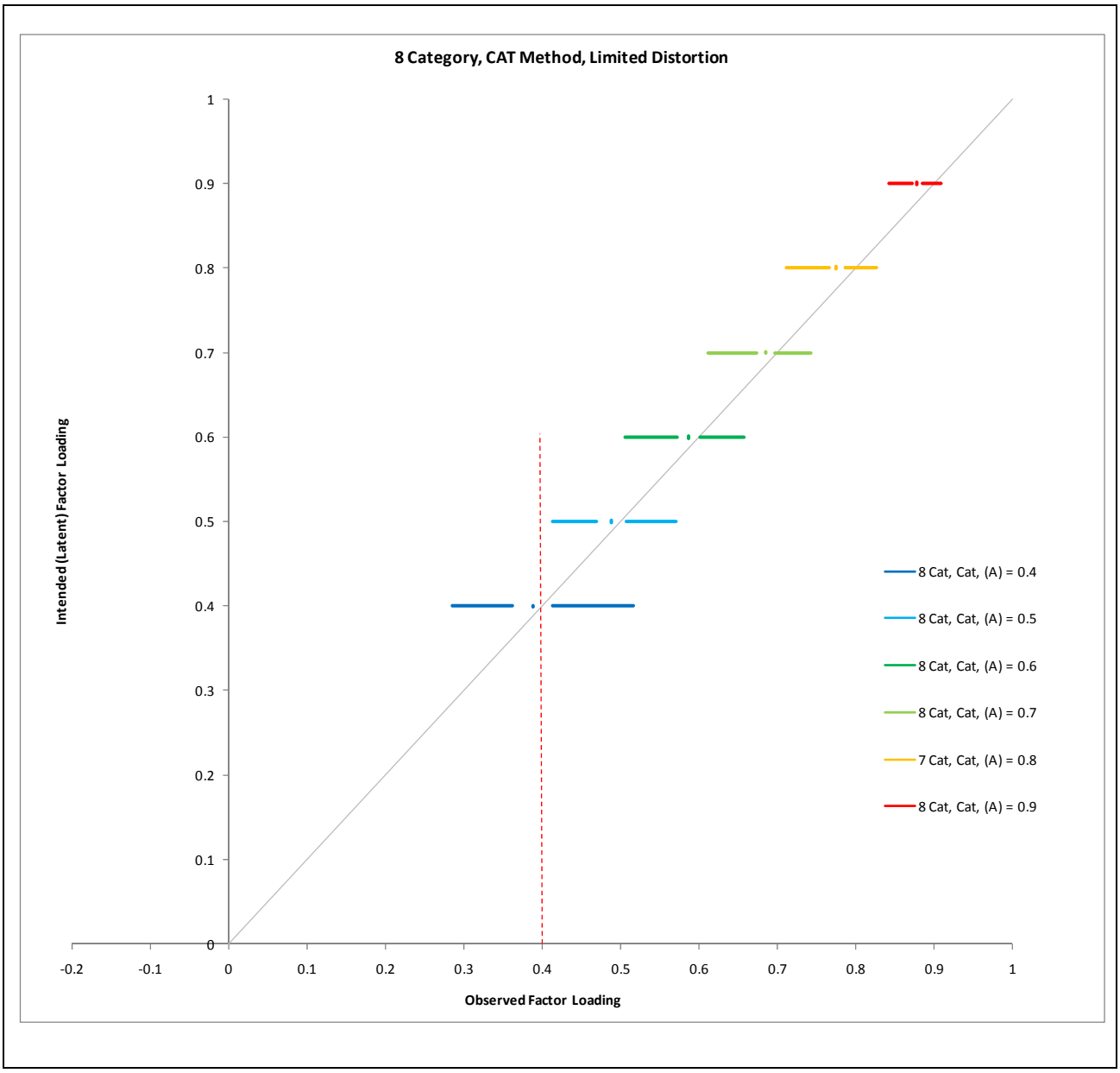


Figure 14 - Modified Box Plots Category 2, 4, 5, 6, 7 and 8 (Limited Distortion)

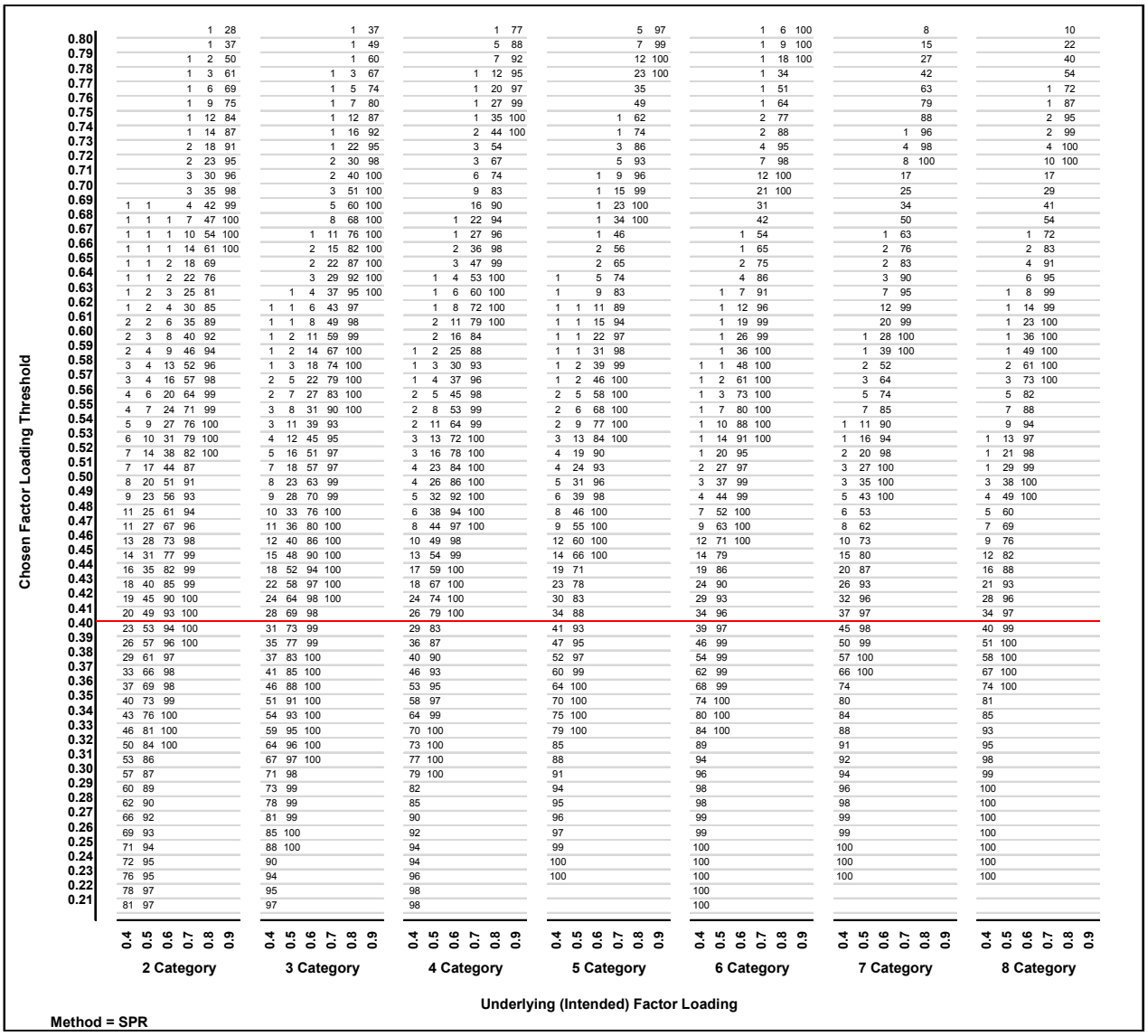


Figure 15 – Percentiles – SPR Method - Limited Distortion

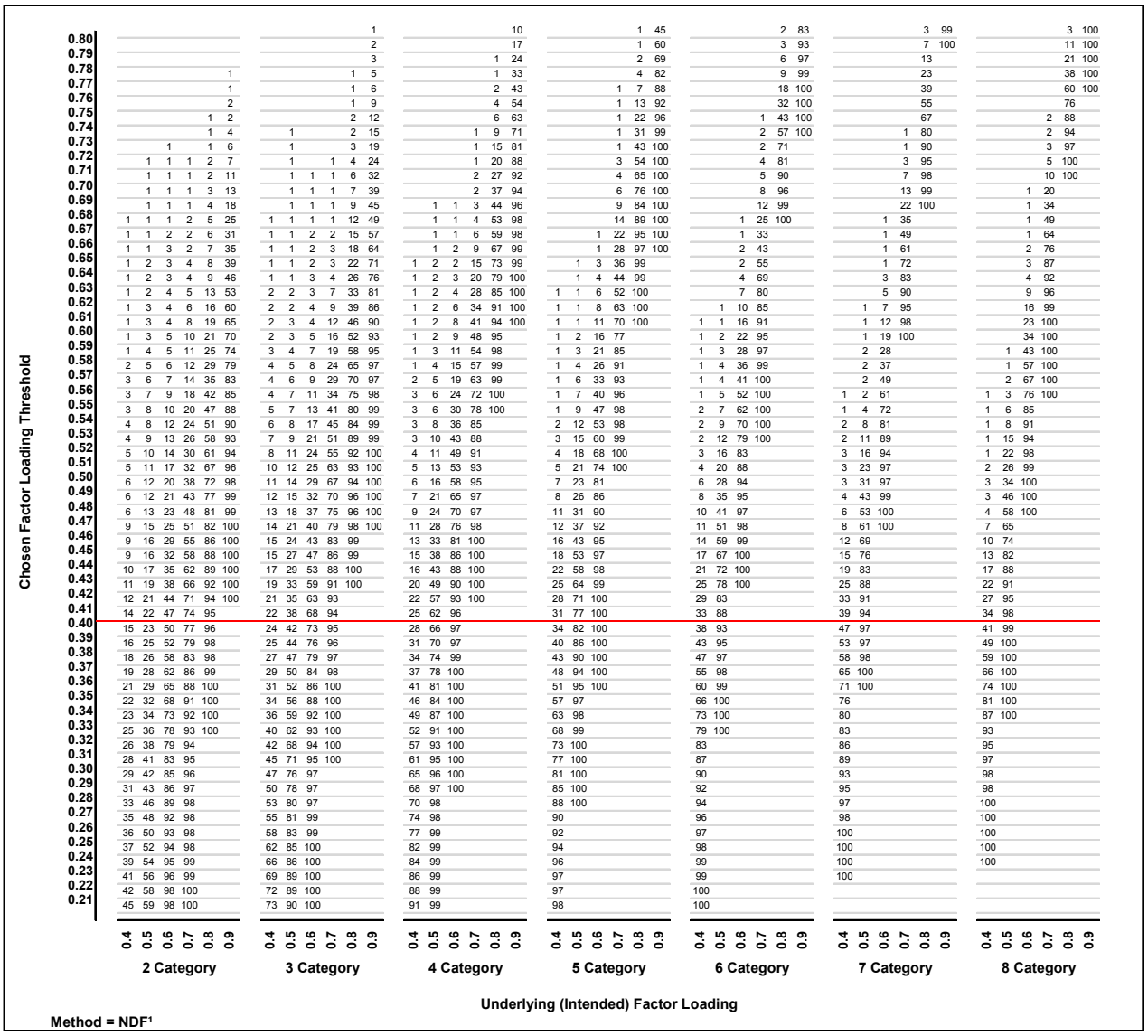


Figure 16 – Percentiles – NDF[1] Method - Limited Distortion

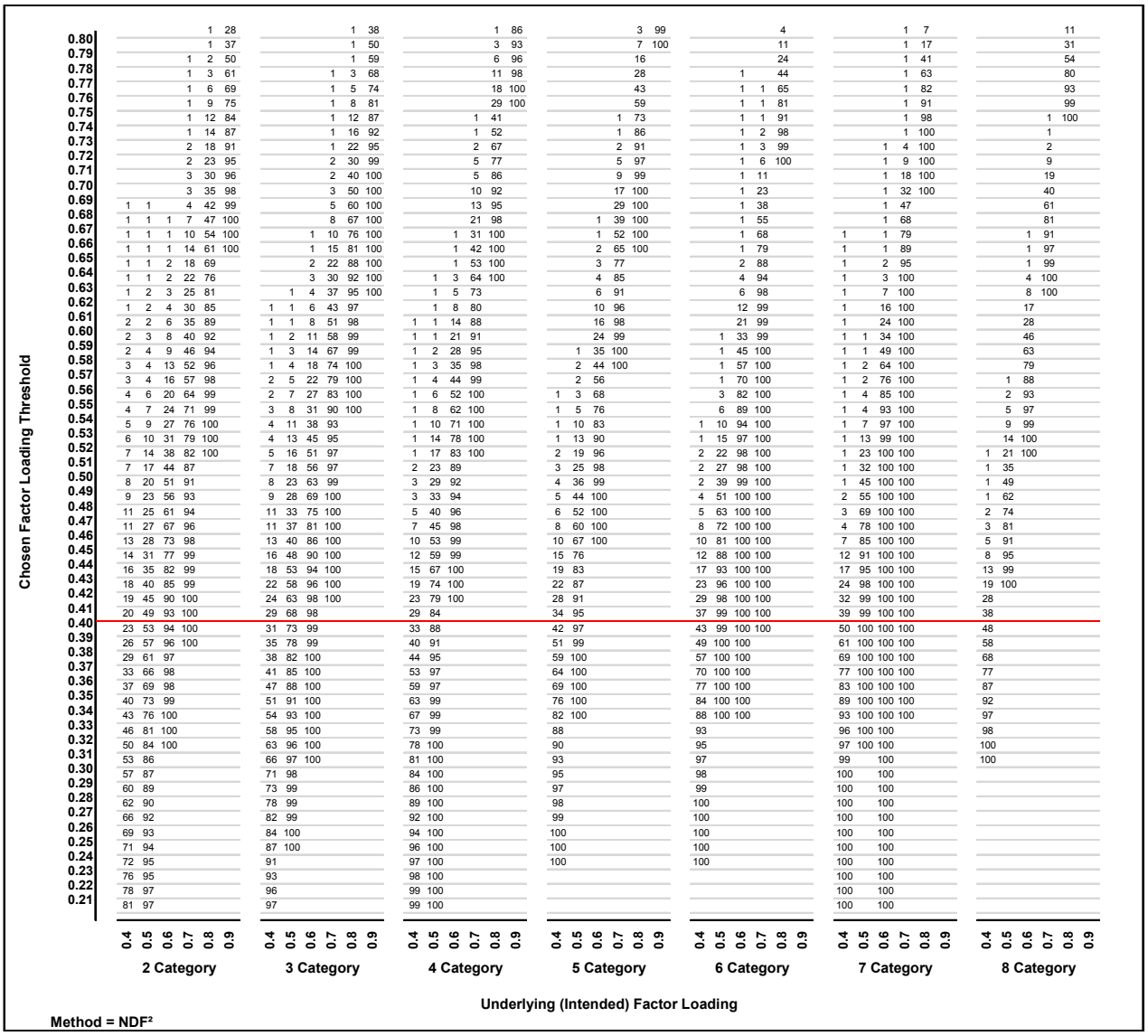


Figure 17 – Percentiles – NDF^2 Method - Limited Distortion

Relative Bias

RB		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	-25.3%	-13.8%	-9.9%	-5.0%	-3.9%	-2.7%	-2.9%
	SPR	-25.3%	-13.8%	-11.6%	-6.4%	-5.6%	-4.6%	-4.5%
	NDF[1]	-54.1%	-29.8%	-18.1%	-10.0%	-7.1%	-5.0%	-4.6%
	NDF[2]	-25.3%	-13.8%	-10.5%	-5.0%	-4.1%	-2.7%	-2.9%
0.5	**CAT**	-21.9%	-13.2%	-8.6%	-5.9%	-3.8%	-3.1%	-2.4%
	SPR	-21.9%	-13.2%	-10.2%	-7.6%	-5.5%	-5.1%	-4.4%
	NDF[1]	-51.8%	-30.2%	-16.7%	-10.9%	-8.0%	-6.2%	-5.0%
	NDF[2]	-21.9%	-13.2%	-9.5%	-5.9%	-3.9%	-3.3%	-2.4%
0.6	**CAT**	-18.4%	-14.6%	-8.1%	-5.7%	-4.1%	-3.3%	-2.2%
	SPR	-18.4%	-14.6%	-9.3%	-6.9%	-5.4%	-4.8%	-3.7%
	NDF[1]	-34.2%	-26.6%	-15.9%	-10.9%	-8.0%	-6.8%	-4.3%
	NDF[2]	-18.4%	-14.6%	-8.1%	-5.8%	-4.3%	-3.7%	-2.2%
0.7	**CAT**	-18.1%	-14.3%	-8.4%	-5.6%	-4.0%	-3.0%	-2.2%
	SPR	-18.1%	-14.3%	-9.6%	-6.7%	-5.3%	-4.5%	-3.7%
	NDF[1]	-34.6%	-26.2%	-16.9%	-11.0%	-7.9%	-6.1%	-4.6%
	NDF[2]	-18.1%	-14.3%	-8.4%	-5.6%	-4.3%	-3.1%	-2.2%
0.8	**CAT**	-16.6%	-14.0%	-8.5%	-5.6%	-4.1%	-3.0%	-2.3%
	SPR	-16.6%	-14.0%	-9.4%	-6.4%	-5.2%	-4.2%	-3.4%
	NDF[1]	-33.7%	-26.0%	-16.3%	-10.9%	-8.1%	-6.1%	-4.6%
	NDF[2]	-16.6%	-14.0%	-8.5%	-5.7%	-4.2%	-3.2%	-2.3%
0.9	**CAT**	-13.7%	-12.8%	-8.5%	-6.0%	-4.2%	-3.2%	-2.4%
	SPR	-13.7%	-12.8%	-9.1%	-6.6%	-5.0%	-4.1%	-3.3%
	NDF[1]	-31.1%	-25.6%	-16.6%	-11.8%	-8.5%	-6.4%	-4.9%
	NDF[2]	-13.7%	-12.8%	-8.5%	-6.0%	-4.3%	-3.3%	-2.5%

Figure 18 – Method RB Comparison - Limited Distortion

Coefficient of Variation

CV		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	48.1%	26.8%	19.1%	14.8%	12.7%	10.5%	9.3%
	SPR	48.1%	26.9%	21.8%	17.9%	15.3%	14.7%	12.6%
	NDF[1]	100.1%	55.3%	31.8%	23.0%	18.7%	15.9%	13.1%
	NDF[2]	48.1%	26.8%	20.2%	14.8%	12.9%	11.7%	9.3%
0.5	CAT	29.7%	16.9%	12.3%	9.5%	7.6%	6.7%	5.9%
	SPR	29.7%	16.9%	13.8%	11.1%	9.5%	8.9%	8.6%
	NDF[1]	82.5%	42.9%	21.1%	14.3%	11.4%	9.3%	8.4%
	NDF[2]	29.7%	16.9%	15.9%	9.5%	7.7%	7.4%	6.0%
0.6	CAT	13.7%	11.7%	8.0%	6.3%	5.2%	4.6%	4.1%
	SPR	13.7%	11.7%	8.7%	7.3%	6.2%	5.6%	5.5%
	NDF[1]	27.3%	20.3%	12.8%	9.9%	8.2%	6.2%	5.9%
	NDF[2]	13.7%	11.7%	8.0%	6.3%	6.2%	6.4%	4.1%
0.7	CAT	11.6%	8.4%	5.9%	4.7%	3.8%	3.2%	2.7%
	SPR	11.6%	8.4%	7.0%	5.5%	5.0%	4.4%	4.1%
	NDF[1]	21.2%	14.0%	9.5%	7.0%	5.4%	4.3%	4.1%
	NDF[2]	11.6%	8.4%	5.9%	4.7%	5.0%	4.7%	2.7%
0.8	CAT	8.4%	6.3%	4.3%	3.1%	2.5%	2.2%	1.9%
	SPR	8.4%	6.3%	5.1%	3.7%	3.3%	2.9%	2.8%
	NDF[1]	14.7%	11.0%	6.9%	4.7%	3.8%	3.4%	2.7%
	NDF[2]	8.4%	6.3%	4.2%	3.3%	2.5%	2.3%	1.9%
0.9	CAT	5.2%	5.0%	2.9%	2.3%	1.8%	1.4%	1.2%
	SPR	5.2%	4.9%	3.4%	2.6%	2.2%	2.0%	1.7%
	NDF[1]	10.5%	8.7%	5.3%	3.9%	3.0%	2.3%	2.0%
	NDF[2]	5.2%	5.0%	2.9%	2.3%	1.8%	1.4%	1.2%

Figure 19 – Method CV Comparison - Limited Distortion

Relative Bias Difference RB Diff		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	0.0000%	0.0019%	0.0000%	0.0015%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	1.7169%	1.3580%	1.6824%	1.8556%	1.6423%
	NDF¹	28.8586%	16.0073%	8.2055%	5.0194%	3.1615%	2.2409%	1.6741%
	NDF²	0.0000%	0.0038%	0.5646%	0.0000%	0.1487%	0.0122%	0.0291%
0.5	**CAT**	0.0000%	0.0143%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0310%	1.5428%	1.6877%	1.6783%	1.9958%	2.0803%
	NDF¹	29.9719%	17.0585%	8.0891%	5.0024%	4.1346%	3.1159%	2.6450%
	NDF²	0.0000%	0.0000%	0.8280%	0.0012%	0.0954%	0.1773%	0.0098%
0.6	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0107%	1.2351%	1.1173%	1.3213%	1.5421%	1.5692%
	NDF¹	15.8473%	11.9474%	7.8511%	5.1543%	3.9153%	3.5149%	2.1576%
	NDF²	0.0000%	0.0245%	0.0187%	0.0296%	0.2136%	0.3865%	0.0362%
0.7	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0072%	1.1519%	1.0867%	1.3439%	1.4731%	1.5317%
	NDF¹	16.4574%	11.8810%	8.5036%	5.3771%	3.9317%	3.0629%	2.3876%
	NDF²	0.0000%	0.0049%	0.0082%	0.0171%	0.3367%	0.0628%	0.0255%
0.8	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0063%	0.8952%	0.8693%	1.0339%	1.1542%	1.1331%
	NDF¹	17.0354%	12.0265%	7.8000%	5.3618%	3.9286%	3.0586%	2.3130%
	NDF²	0.0000%	0.0093%	0.0268%	0.1119%	0.0750%	0.1244%	0.0496%
0.9	**CAT**	0.0000%	0.0066%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0161%	0.6215%	0.6824%	0.7887%	0.8462%	0.8254%
	NDF¹	17.3667%	12.8035%	8.0774%	5.8134%	4.2806%	3.1885%	2.4359%
	NDF²	0.0000%	0.0000%	0.0448%	0.0798%	0.0752%	0.0953%	0.0723%
Overall Best Performer (Mean)	**CAT**	**0.0000%**	**0.0038%**	**0.0000%**	**0.0003%**	**0.0000%**	**0.0000%**	**0.0000%**
	SPR	**0.0000%**	**0.0119%**	**1.1939%**	**1.1336%**	**1.3081%**	**1.4778%**	**1.4637%**
	NDF¹	**20.9229%**	**13.6207%**	**8.0878%**	**5.2880%**	**3.8921%**	**3.0303%**	**2.2689%**
	NDF²	**0.0000%**	**0.0071%**	**0.2485%**	**0.0399%**	**0.1574%**	**0.1431%**	**0.0371%**

Figure 20 – Method RB % Underperformance - Limited Distortion

Coefficient of Variation Difference

CV Diff		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	0.0000%	0.0000%	0.0000%	0.0636%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0714%	2.6629%	3.1590%	2.5527%	4.1421%	3.3483%
	NDF¹	52.0240%	28.4971%	12.6562%	8.1924%	6.0339%	5.3813%	3.7887%
	NDF²	0.0000%	0.0405%	1.0174%	0.0000%	0.1974%	1.1661%	0.0071%
0.5	**CAT**	0.0000%	0.0539%	0.0000%	0.0305%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0157%	1.5221%	1.5622%	1.9130%	2.1881%	2.6873%
	NDF¹	52.7229%	26.0325%	8.7649%	4.8092%	3.8138%	2.5210%	2.4735%
	NDF²	0.0000%	0.0000%	3.5531%	0.0000%	0.1509%	0.6230%	0.0192%
0.6	**CAT**	0.0000%	0.0128%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0126%	0.7111%	1.0622%	1.0502%	1.0735%	1.4570%
	NDF¹	13.6450%	8.6085%	4.8371%	3.6688%	3.0048%	1.6254%	1.8573%
	NDF²	0.0000%	0.0000%	0.0439%	0.0134%	0.9978%	1.8105%	0.0197%
0.7	**CAT**	0.0000%	0.0187%	0.0614%	0.0185%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0155%	1.1887%	0.8294%	1.2091%	1.2257%	1.3539%
	NDF¹	9.6454%	5.5795%	3.6080%	2.3289%	1.6081%	1.1609%	1.3930%
	NDF²	0.0000%	0.0000%	0.0000%	0.0000%	1.2766%	1.5306%	0.0073%
0.8	**CAT**	0.0000%	0.0066%	0.0324%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	0.8968%	0.5294%	0.7736%	0.6333%	0.9057%
	NDF¹	6.3302%	4.7226%	2.6384%	1.5614%	1.2568%	1.1534%	0.8728%
	NDF²	0.0000%	0.0066%	0.0000%	0.1635%	0.0471%	0.1172%	0.0157%
0.9	**CAT**	0.0000%	0.0099%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	0.4710%	0.3045%	0.4270%	0.5457%	0.5134%
	NDF¹	5.3358%	3.7765%	2.3293%	1.6336%	1.2224%	0.9110%	0.7795%
	NDF²	0.0000%	0.0209%	0.0037%	0.0200%	0.0257%	0.0170%	0.0085%
Overall Best Performer (Mean)	**CAT**	**0.0000%**	**0.0170%**	**0.0156%**	**0.0188%**	**0.0000%**	**0.0000%**	**0.0000%**
	SPR	**0.0000%**	**0.0192%**	**1.2421%**	**1.2411%**	**1.3209%**	**1.6347%**	**1.7109%**
	NDF¹	**23.2839%**	**12.8695%**	**5.8056%**	**3.6991%**	**2.8233%**	**2.1255%**	**1.8608%**
	NDF²	**0.0000%**	**0.0113%**	**0.7697%**	**0.0328%**	**0.4492%**	**0.8774%**	**0.0129%**

Figure 21 – Method CV % Underperformance - Limited Distortion

7.4 Factor Loading Descriptive Statistics – Skew Distortion

2 Category

Factor Loading	Method	μ	σ	RB	CV
0.4	CAT	0.293976	0.148554	-26.50593%	50.53249%
	SPR	0.293976	0.148554	-26.50593%	50.53249%
	NDF¹	0.148198	0.203491	-62.95043%	137.30981%
	NDF²	0.293976	0.148554	-26.50593%	50.53249%
0.5	CAT	0.364412	0.141478	-27.11759%	38.82374%
	SPR	0.364412	0.141478	-27.11759%	38.82374%
	NDF¹	0.188744	0.211483	-62.25129%	112.04782%
	NDF²	0.364412	0.141478	-27.11759%	38.82374%
0.6	CAT	0.459203	0.095532	-23.46614%	20.80384%
	SPR	0.459203	0.095532	-23.46614%	20.80384%
	NDF¹	0.318366	0.154339	-46.93902%	48.47857%
	NDF²	0.459203	0.095532	-23.46614%	20.80384%
0.7	CAT	0.547817	0.084034	-21.74040%	15.33977%
	SPR	0.547817	0.084034	-21.74040%	15.33977%
	NDF¹	0.385195	0.140484	-44.97220%	36.47081%
	NDF²	0.547817	0.084034	-21.74040%	15.33977%
0.8	CAT	0.646027	0.066074	-19.24660%	10.22773%
	SPR	0.646027	0.066074	-19.24660%	10.22773%
	NDF¹	0.473417	0.0864	-40.82290%	18.25034%
	NDF²	0.646027	0.066074	-19.24660%	10.22773%
0.9	CAT	0.76027	0.057233	-15.52558%	7.52793%
	SPR	0.76027	0.057233	-15.52558%	7.52793%
	NDF¹	0.552666	0.077858	-38.59270%	14.08772%
	NDF²	0.76027	0.057233	-15.52558%	7.52793%

Table 34 - FL Descriptive Statistics - Skew Scale Distortion on 2 Categories

3 Category

Factor Loading	Method	μ	σ	RB	CV
0.4	CAT	0.301009	0.142943	-24.74774%	47.48809%
	SPR	0.298713	0.139557	-25.32179%	46.71944%
	NDF¹	0.200274	0.180357	-49.93154%	90.05497%
	NDF²	0.294554	0.148613	-26.36152%	50.45344%
0.5	CAT	0.375157	0.144549	-24.96867%	38.53018%
	SPR	0.379874	0.12925	-24.02513%	34.02438%
	NDF¹	0.24634	0.191012	-50.73198%	77.54001%
	NDF²	0.365541	0.154908	-26.89181%	42.37775%
0.6	CAT	0.486103	0.083056	-18.98279%	17.08600%
	SPR	0.476655	0.08443	-20.55753%	17.71298%
	NDF¹	0.374881	0.126172	-37.51988%	33.65665%
	NDF²	0.478657	0.087019	-20.22380%	18.17984%
0.7	CAT	0.580246	0.068245	-17.10772%	11.76140%
	SPR	0.569623	0.069128	-18.62535%	12.13576%
	NDF¹	0.437022	0.120539	-37.56822%	27.58198%
	NDF²	0.571058	0.076217	-18.42026%	13.34658%
0.8	CAT	0.678373	0.05473	-15.20332%	8.06789%
	SPR	0.663282	0.059407	-17.08981%	8.95648%
	NDF¹	0.525798	0.079507	-34.27519%	15.12117%
	NDF²	0.670116	0.06992	-16.23552%	10.43400%
0.9	CAT	0.788413	0.046694	-12.39851%	5.92255%
	SPR	0.777529	0.050859	-13.60785%	6.54116%
	NDF¹	0.609879	0.066375	-32.23567%	10.88337%
	NDF²	0.780889	0.058578	-13.23459%	7.50145%

Table 35 - FL Descriptive Statistics - Skew Scale Distortion on 3 Categories

		4 Category			
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.332324	0.122995	-16.91906%	37.01045%
	SPR	0.327054	0.121745	-18.23656%	37.22489%
	NDF¹	0.257341	0.180616	-35.66466%	70.18550%
	NDF²	0.320827	0.131539	-19.79322%	41.00007%
0.5	**CAT**	0.420118	0.092967	-15.97642%	22.12884%
	SPR	0.413384	0.092752	-17.32328%	22.43716%
	NDF¹	0.325659	0.148099	-34.86822%	45.47668%
	NDF²	0.400194	0.117155	-19.96124%	29.27451%
0.6	**CAT**	0.50568	0.074173	-15.72002%	14.66807%
	SPR	0.49651	0.074644	-17.24832%	15.03380%
	NDF¹	0.416145	0.105569	-30.64254%	25.36839%
	NDF²	0.486134	0.093003	-18.97771%	19.13111%
0.7	**CAT**	0.599879	0.060358	-14.30299%	10.06167%
	SPR	0.586496	0.061177	-16.21481%	10.43086%
	NDF¹	0.492295	0.085357	-29.67207%	17.33854%
	NDF²	0.573142	0.076489	-18.12258%	13.34561%
0.8	**CAT**	0.697419	0.048048	-12.82263%	6.88945%
	SPR	0.68178	0.051027	-14.77755%	7.48437%
	NDF¹	0.575409	0.068369	-28.07387%	11.88187%
	NDF²	0.672242	0.068589	-15.96969%	10.20295%
0.9	**CAT**	0.80738	0.035679	-10.29107%	4.41915%
	SPR	0.792467	0.041193	-11.94810%	5.19810%
	NDF¹	0.661391	0.056722	-26.51207%	8.57609%
	NDF²	0.7842	0.05244	-12.86669%	6.68711%

Table 36 - FL Descriptive Statistics - Skew Scale Distortion on 4 Categories

		5 Category			
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.351045	0.101078	-12.23879%	28.79335%
	SPR	0.35042	0.089961	-12.39511%	25.67231%
	NDF¹	0.313032	0.126252	-21.74204%	40.33215%
	NDF²	0.34852	0.09595	-12.86990%	27.53068%
0.5	**CAT**	0.444054	0.078404	-11.18927%	17.65636%
	SPR	0.441402	0.07514	-11.71961%	17.02298%
	NDF¹	0.394533	0.107467	-21.09335%	27.23913%
	NDF²	0.435808	0.079674	-12.83850%	18.28183%
0.6	**CAT**	0.53183	0.058435	-11.36168%	10.98748%
	SPR	0.528372	0.052309	-11.93794%	9.90008%
	NDF¹	0.477722	0.073352	-20.37975%	15.35446%
	NDF²	0.520387	0.06071	-13.26891%	11.66625%
0.7	**CAT**	0.630315	0.050538	-9.95501%	8.01783%
	SPR	0.620336	0.047454	-11.38053%	7.64966%
	NDF¹	0.563166	0.064791	-19.54778%	11.50485%
	NDF²	0.610686	0.056803	-12.75921%	9.30151%
0.8	**CAT**	0.73211	0.04225	-8.48626%	5.77104%
	SPR	0.718198	0.039384	-10.22530%	5.48377%
	NDF¹	0.647529	0.050123	-19.05886%	7.74061%
	NDF²	0.711248	0.047785	-11.09402%	6.71854%
0.9	**CAT**	0.840837	0.026781	-6.57365%	3.18502%
	SPR	0.823793	0.031395	-8.46744%	3.81101%
	NDF¹	0.737156	0.045202	-18.09374%	6.13201%
	NDF²	0.8169	0.038332	-9.23337%	4.69237%

Table 37 - FL Descriptive Statistics - Skew Scale Distortion on 5 Categories

6 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.363526	0.084886	-9.11849%	23.35082%
	SPR	0.360786	0.078073	-9.80338%	21.63971%
	NDF¹	0.342391	0.0939	-14.40222%	27.42490%
	NDF²	0.35902	0.081473	-10.24496%	22.69324%
0.5	**CAT**	0.457242	0.064333	-8.55160%	14.06987%
	SPR	0.455248	0.059029	-8.95032%	12.96640%
	NDF¹	0.427076	0.078829	-14.58475%	18.45788%
	NDF²	0.44942	0.064034	-10.11590%	14.24823%
0.6	**CAT**	0.549386	0.050279	-8.43570%	9.15189%
	SPR	0.543884	0.046268	-9.35269%	8.50705%
	NDF¹	0.512971	0.057909	-14.50476%	11.28890%
	NDF²	0.541067	0.053088	-9.82209%	9.81178%
0.7	**CAT**	0.648432	0.043865	-7.36692%	6.76485%
	SPR	0.638986	0.041033	-8.71627%	6.42161%
	NDF¹	0.598332	0.053746	-14.52394%	8.98255%
	NDF²	0.633487	0.048643	-9.50190%	7.67855%
0.8	**CAT**	0.749416	0.03582	-6.32297%	4.77975%
	SPR	0.734571	0.034857	-8.17860%	4.74518%
	NDF¹	0.688669	0.042047	-13.91635%	6.10556%
	NDF²	0.733072	0.039111	-8.36603%	5.33520%
0.9	**CAT**	0.852136	0.021568	-5.31828%	2.53105%
	SPR	0.836169	0.024671	-7.09236%	2.95046%
	NDF¹	0.77615	0.034503	-13.76108%	4.44539%
	NDF²	0.833139	0.031508	-7.42895%	3.78182%

Table 38 - FL Descriptive Statistics - Skew Scale Distortion on 6 Categories

7 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.35911	0.092	-10.22261%	25.61886%
	SPR	0.35512	0.08578	-11.21998%	24.15525%
	NDF¹	0.333783	0.105333	-16.55429%	31.55746%
	NDF²	0.356991	0.08856	-10.75222%	24.80741%
0.5	**CAT**	0.455599	0.068562	-8.88028%	15.04876%
	SPR	0.451295	0.07621	-9.74102%	16.88692%
	NDF¹	0.425842	0.084783	-14.83151%	19.90951%
	NDF²	0.448867	0.078269	-10.22657%	17.43695%
0.6	**CAT**	0.545691	0.056539	-9.05152%	10.36095%
	SPR	0.542508	0.048337	-9.58206%	8.90995%
	NDF¹	0.506471	0.063273	-15.58825%	12.49297%
	NDF²	0.541771	0.053322	-9.70481%	9.84222%
0.7	**CAT**	0.646509	0.044863	-7.64153%	6.93932%
	SPR	0.637377	0.040281	-8.94621%	6.31976%
	NDF¹	0.595239	0.048868	-14.96586%	8.20984%
	NDF²	0.634693	0.044533	-9.32963%	7.01651%
0.8	**CAT**	0.751267	0.036214	-6.09164%	4.82033%
	SPR	0.735884	0.033734	-8.01450%	4.58421%
	NDF¹	0.684782	0.042484	-14.40222%	6.20400%
	NDF²	0.734295	0.039291	-8.21313%	5.35081%
0.9	**CAT**	0.861347	0.021869	-4.29482%	2.53892%
	SPR	0.842686	0.025638	-6.36826%	3.04242%
	NDF¹	0.778474	0.033466	-13.50293%	4.29892%
	NDF²	0.83957	0.032115	-6.71442%	3.82521%

Table 39 - FL Descriptive Statistics - Skew Scale Distortion on 7 Categories

8 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.364378	0.086018	-8.90552%	23.60692%
	SPR	0.362546	0.07334	-9.36357%	20.22917%
	NDF¹	0.345591	0.092695	-13.60226%	26.82227%
	NDF²	0.359471	0.077777	-10.13229%	21.63644%
0.5	**CAT**	0.459428	0.064971	-8.11441%	14.14182%
	SPR	0.45856	0.060923	-8.28801%	13.28569%
	NDF¹	0.432078	0.076861	-13.58434%	17.78863%
	NDF²	0.455179	0.066564	-8.96416%	14.62360%
0.6	**CAT**	0.548821	0.053213	-8.52987%	9.69586%
	SPR	0.546686	0.044196	-8.88573%	8.08435%
	NDF¹	0.514462	0.057341	-14.25630%	11.14581%
	NDF²	0.546319	0.050139	-8.94686%	9.17752%
0.7	**CAT**	0.647703	0.044533	-7.47104%	6.87554%
	SPR	0.640117	0.041447	-8.55477%	6.47498%
	NDF¹	0.600896	0.05341	-14.15774%	8.88833%
	NDF²	0.63424	0.048171	-9.39424%	7.59510%
0.8	**CAT**	0.751675	0.034252	-6.04064%	4.55677%
	SPR	0.737387	0.032237	-7.82667%	4.37185%
	NDF¹	0.694443	0.041668	-13.19467%	6.00017%
	NDF²	0.735894	0.038621	-8.01320%	5.24817%
0.9	**CAT**	0.860032	0.022004	-4.44087%	2.55849%
	SPR	0.842913	0.026022	-6.34304%	3.08711%
	NDF¹	0.789529	0.034122	-12.27451%	4.32188%
	NDF²	0.839834	0.030033	-6.68514%	3.57609%

Table 40 - FL Descriptive Statistics - Skew Scale Distortion on 7 Categories

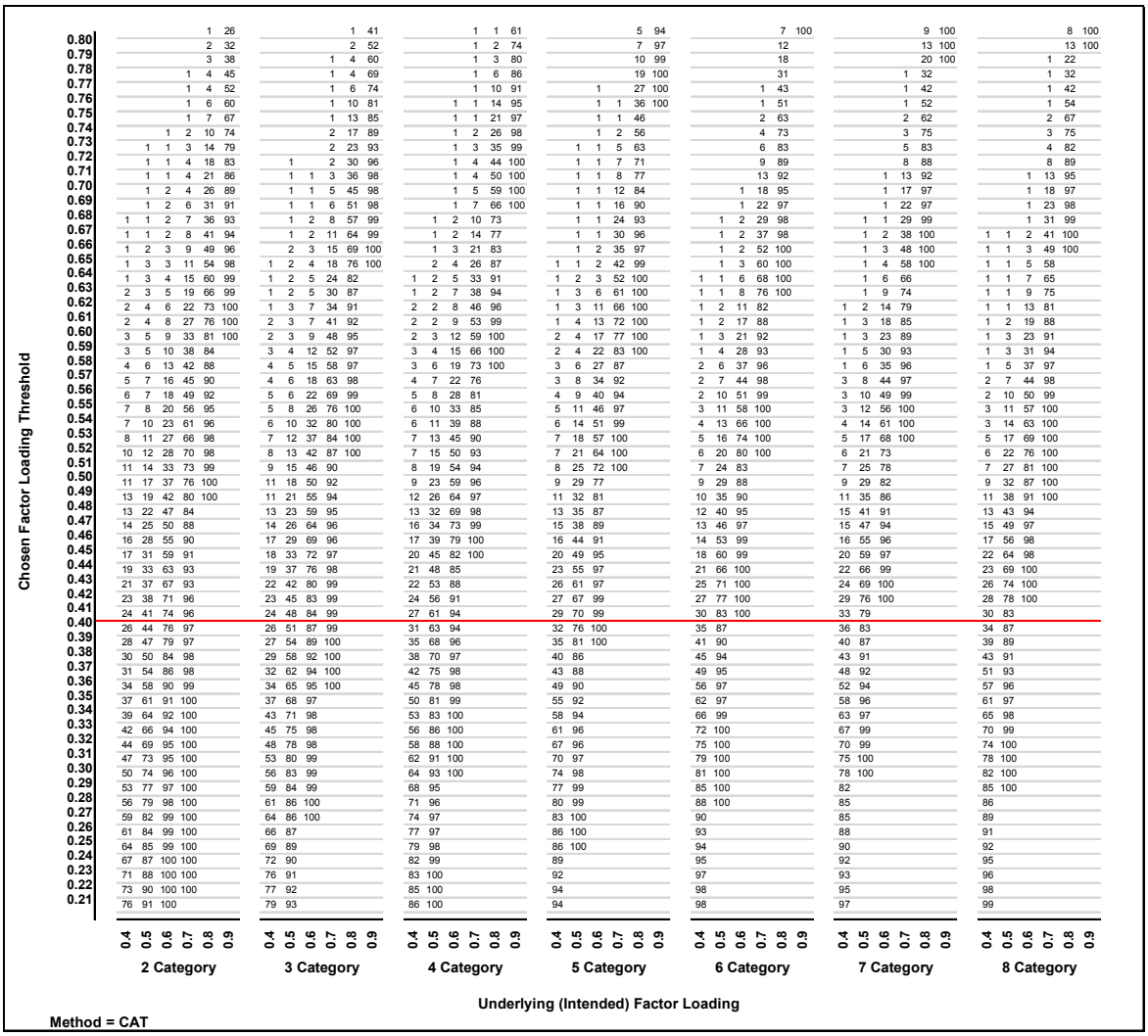
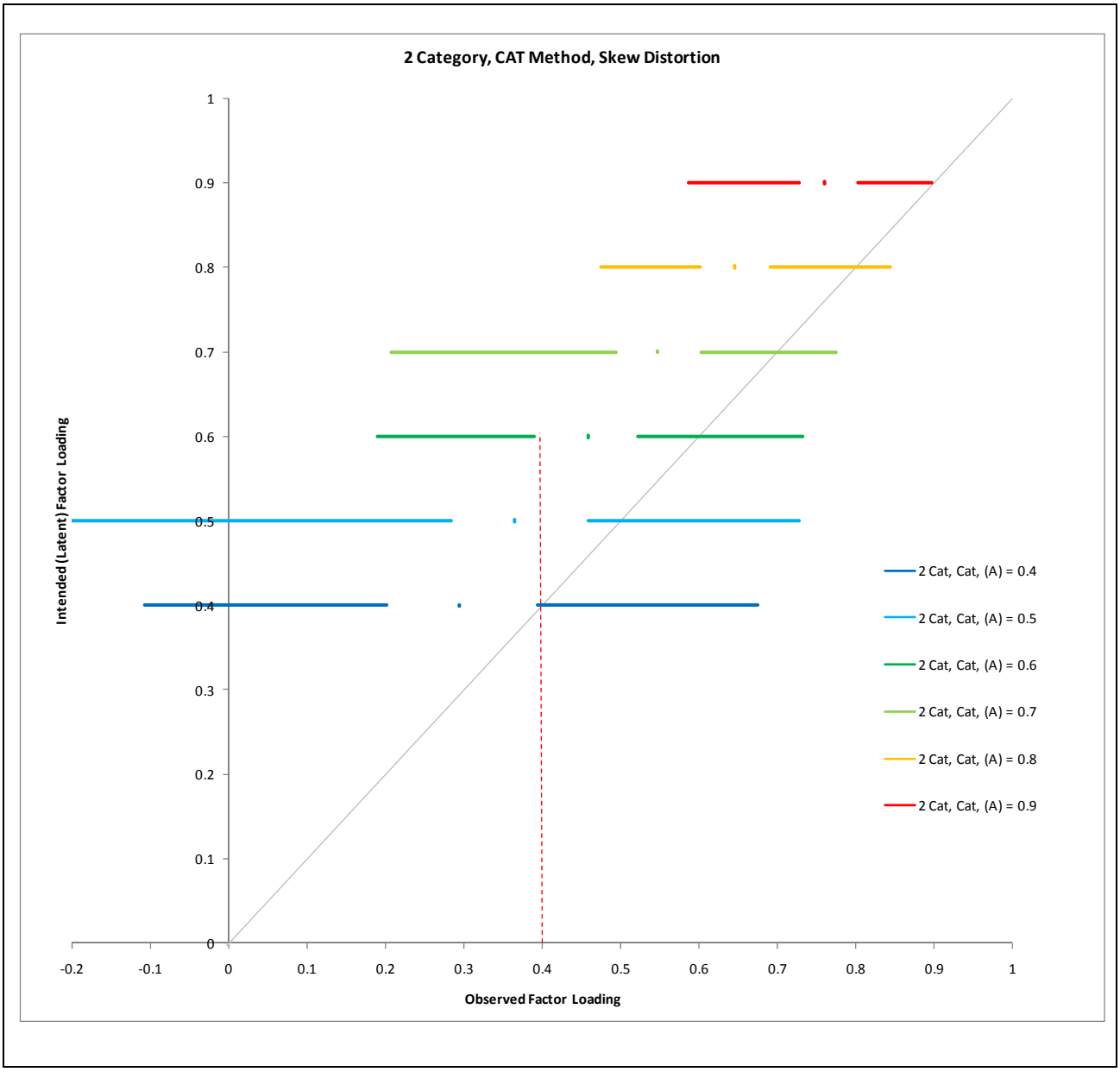
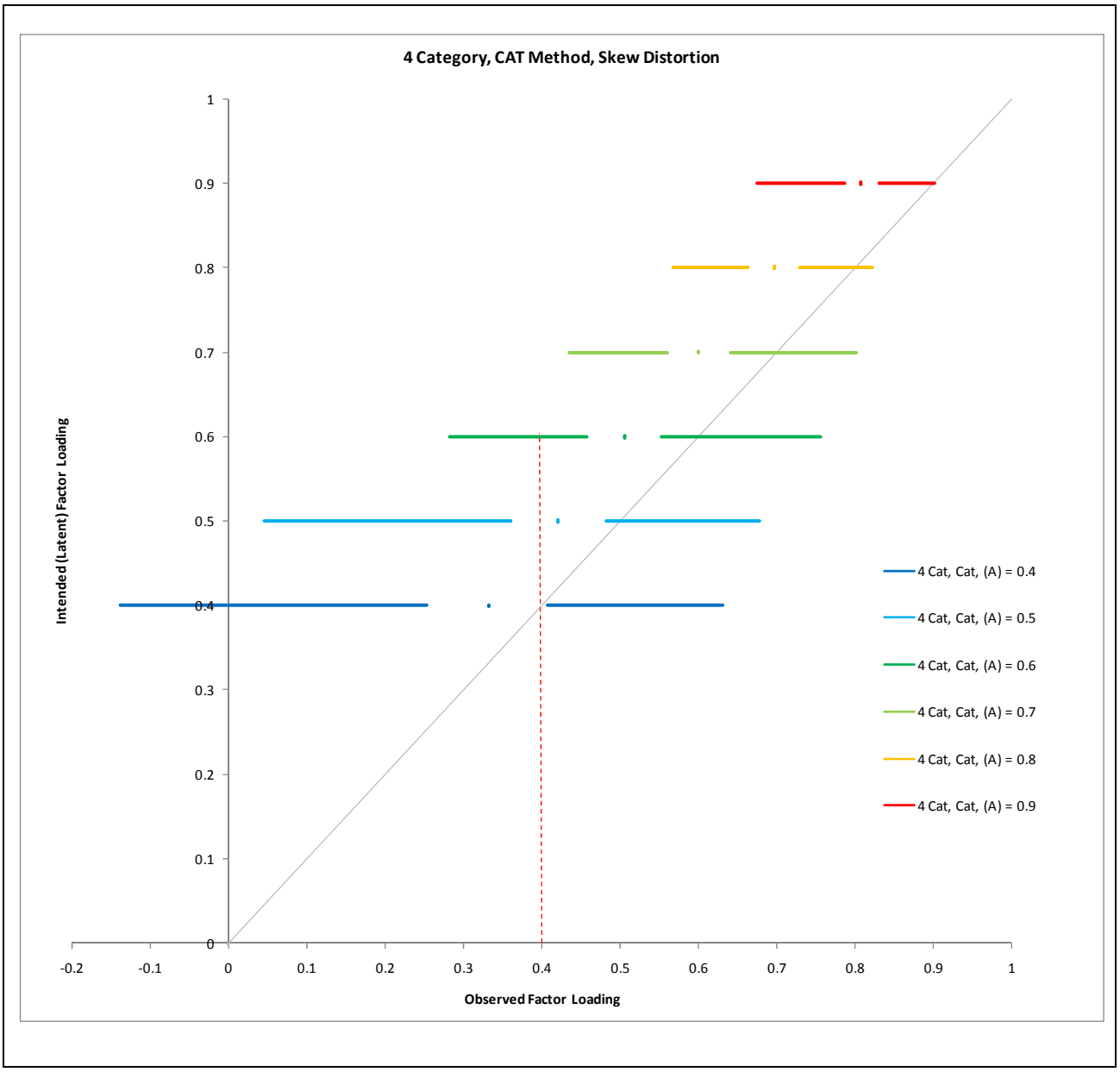
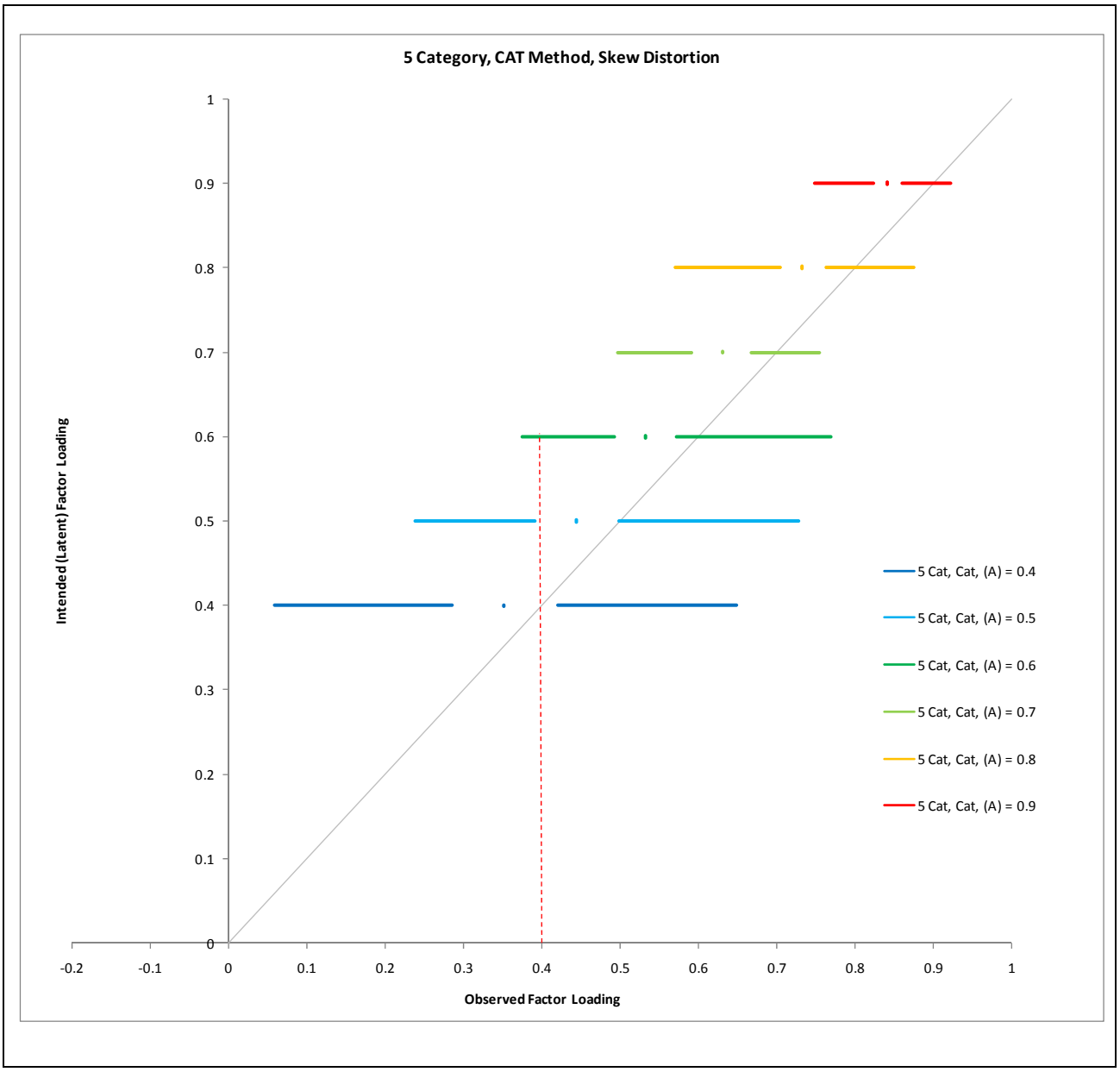
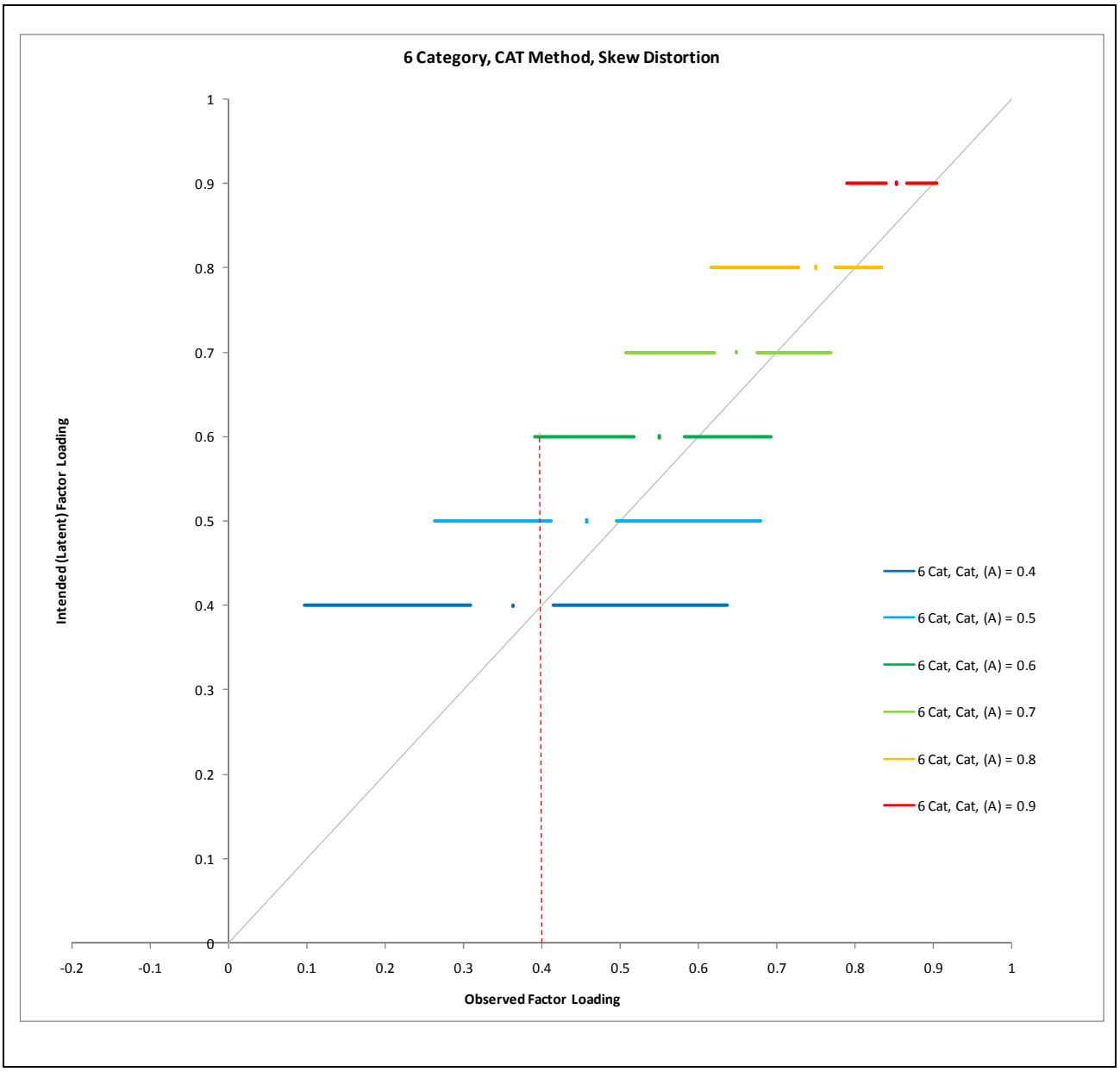


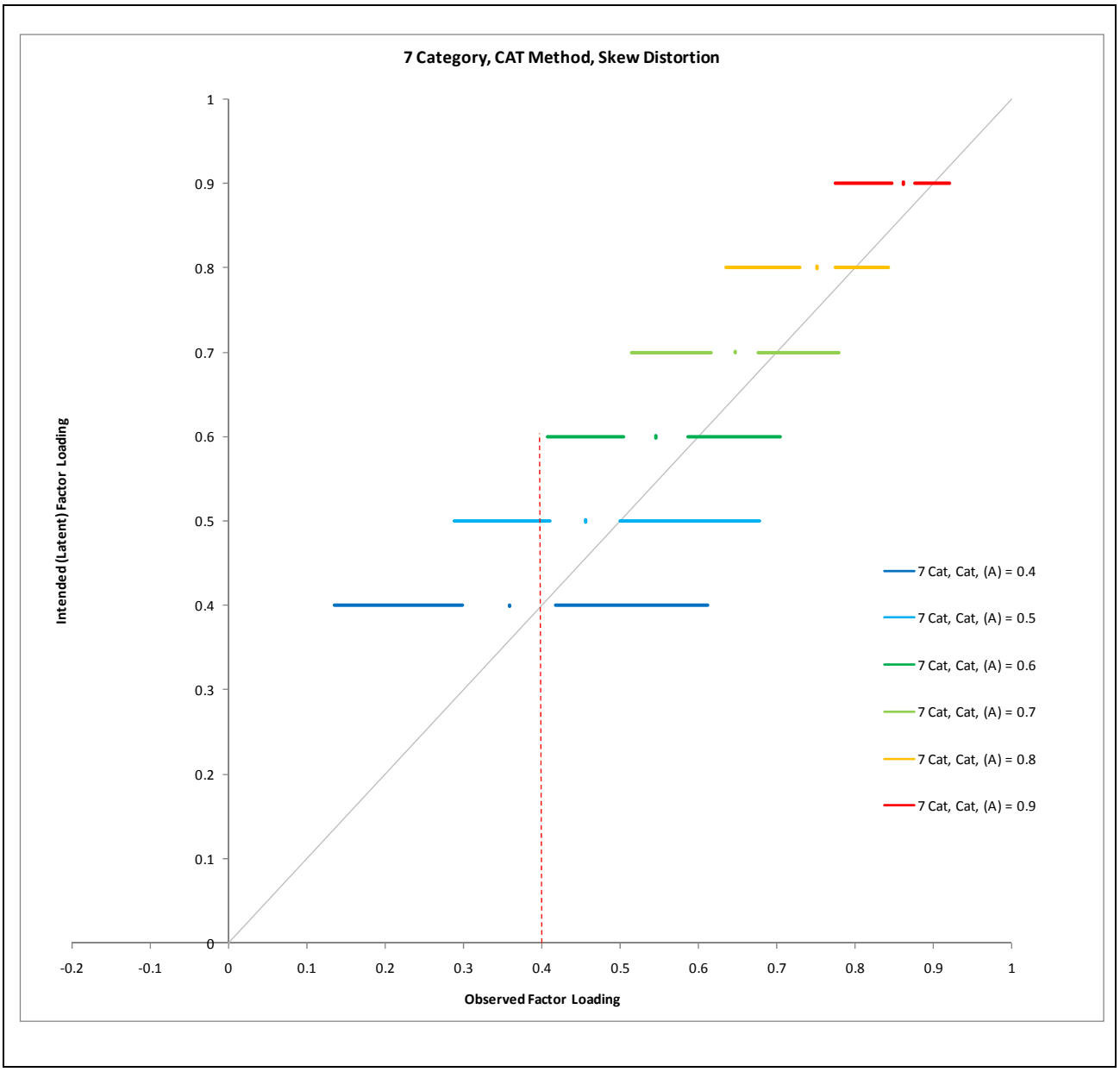
Figure 22 - Percentiles – CAT Method - Skewed Distortion


2 Category, CAT Method, Skew Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2 Cat, Cat, (A) = 0.4
2 Cat, Cat, (A) = 0.5
2 Cat, Cat, (A) = 0.6
2 Cat, Cat, (A) = 0.7
2 Cat, Cat, (A) = 0.8
2 Cat, Cat, (A) = 0.9


4 Category, CAT Method, Skew Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
4 Cat, Cat, (A) = 0.4
4 Cat, Cat, (A) = 0.5
4 Cat, Cat, (A) = 0.6
4 Cat, Cat, (A) = 0.7
4 Cat, Cat, (A) = 0.8
4 Cat, Cat, (A) = 0.9


5 Category, CAT Method, Skew Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
5 Cat, Cat, (A) = 0.4
5 Cat, Cat, (A) = 0.5
5 Cat, Cat, (A) = 0.6
5 Cat, Cat, (A) = 0.7
5 Cat, Cat, (A) = 0.8
5 Cat, Cat, (A) = 0.9


6 Category, CAT Method, Skew Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
6 Cat, Cat, (A) = 0.4
6 Cat, Cat, (A) = 0.5
6 Cat, Cat, (A) = 0.6
6 Cat, Cat, (A) = 0.7
6 Cat, Cat, (A) = 0.8
6 Cat, Cat, (A) = 0.9


7 Category, CAT Method, Skew Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
7 Cat, Cat, (A) = 0.4
7 Cat, Cat, (A) = 0.5
7 Cat, Cat, (A) = 0.6
7 Cat, Cat, (A) = 0.7
7 Cat, Cat, (A) = 0.8
7 Cat, Cat, (A) = 0.9

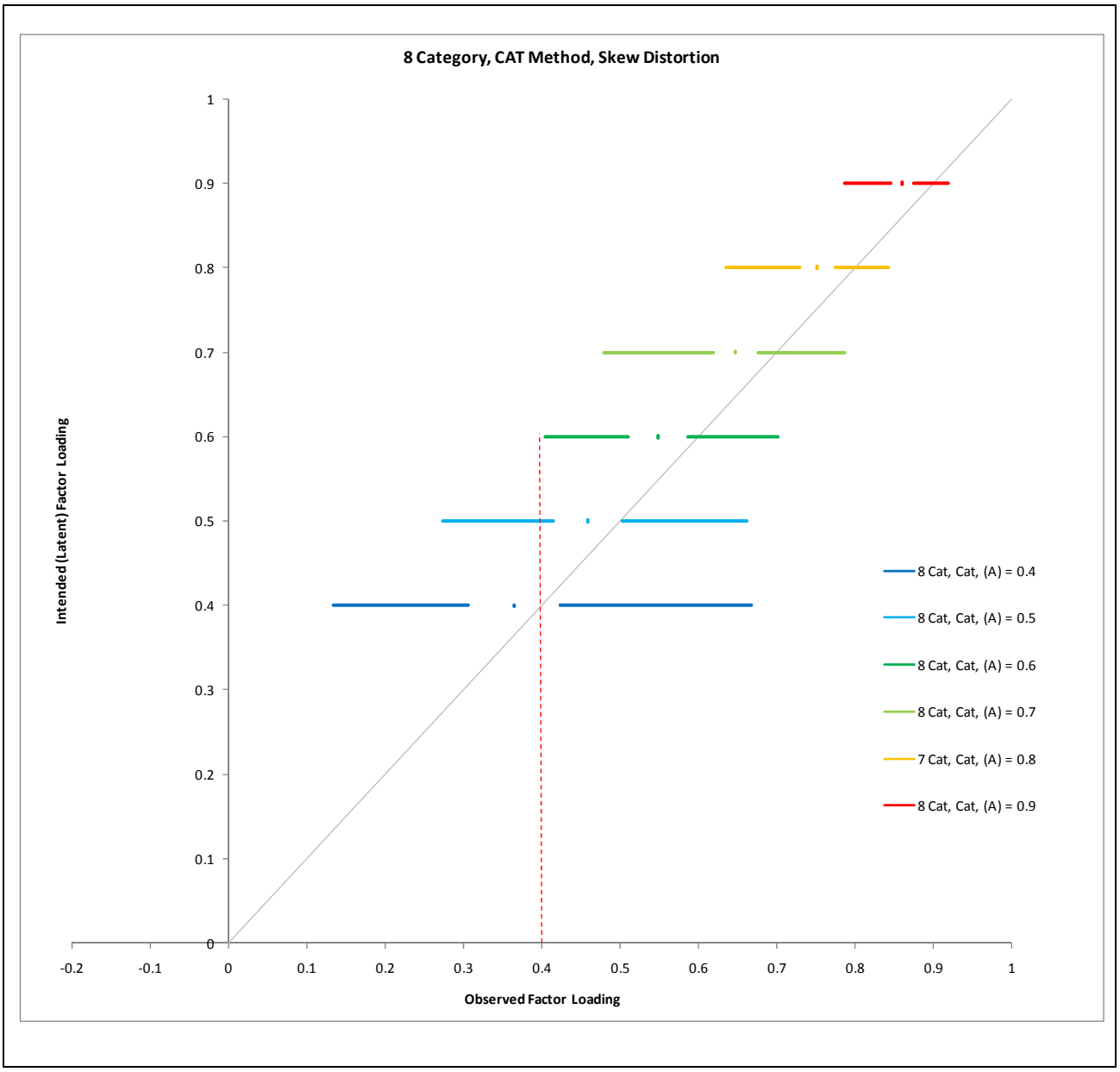


Figure 23 - Modified Box Plots Category 2, 4, 5, 6, 7 and 8 (Skewed Distortion)

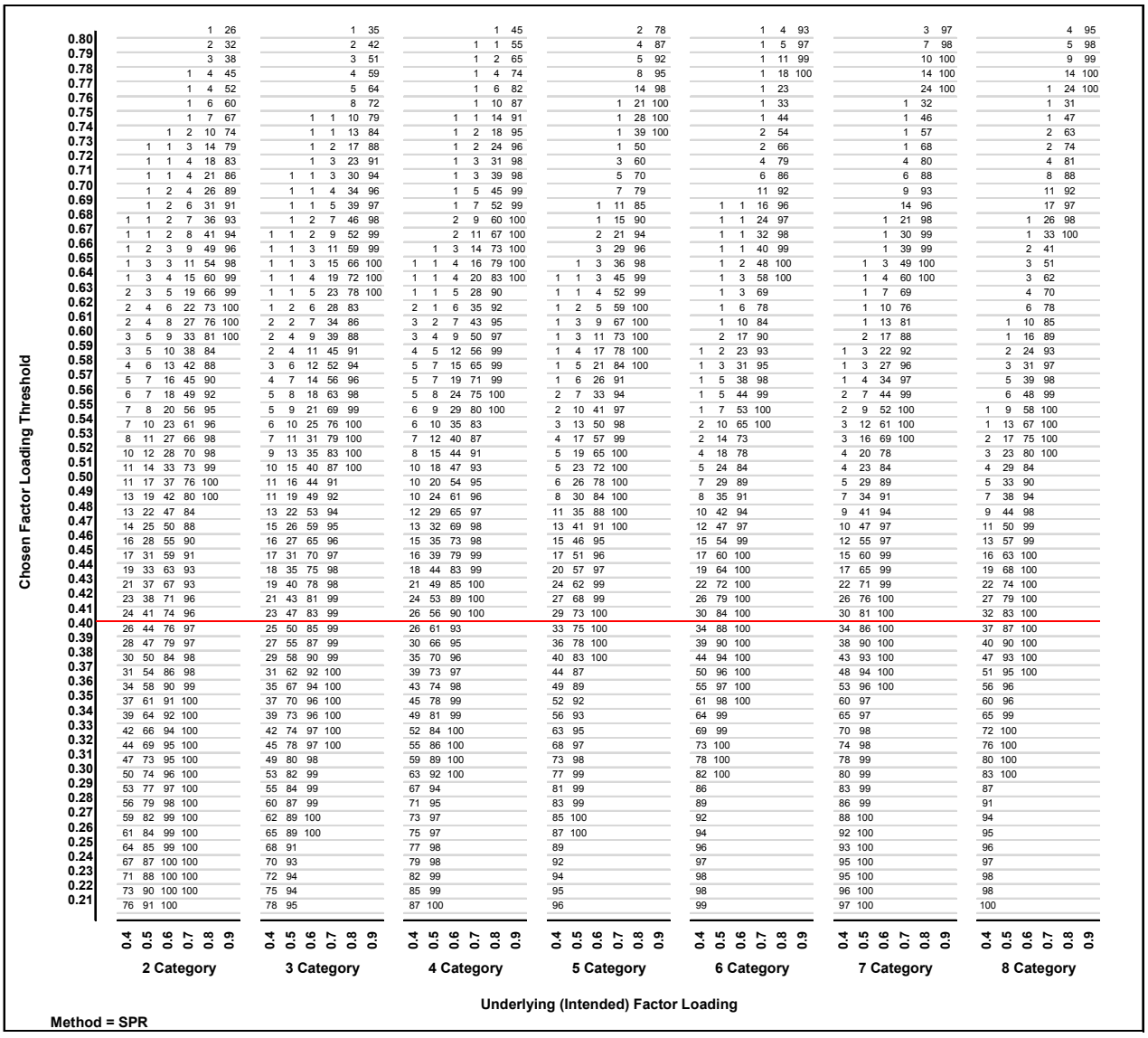


Figure 24 – Percentiles – SPR Method – Skewed Distortion

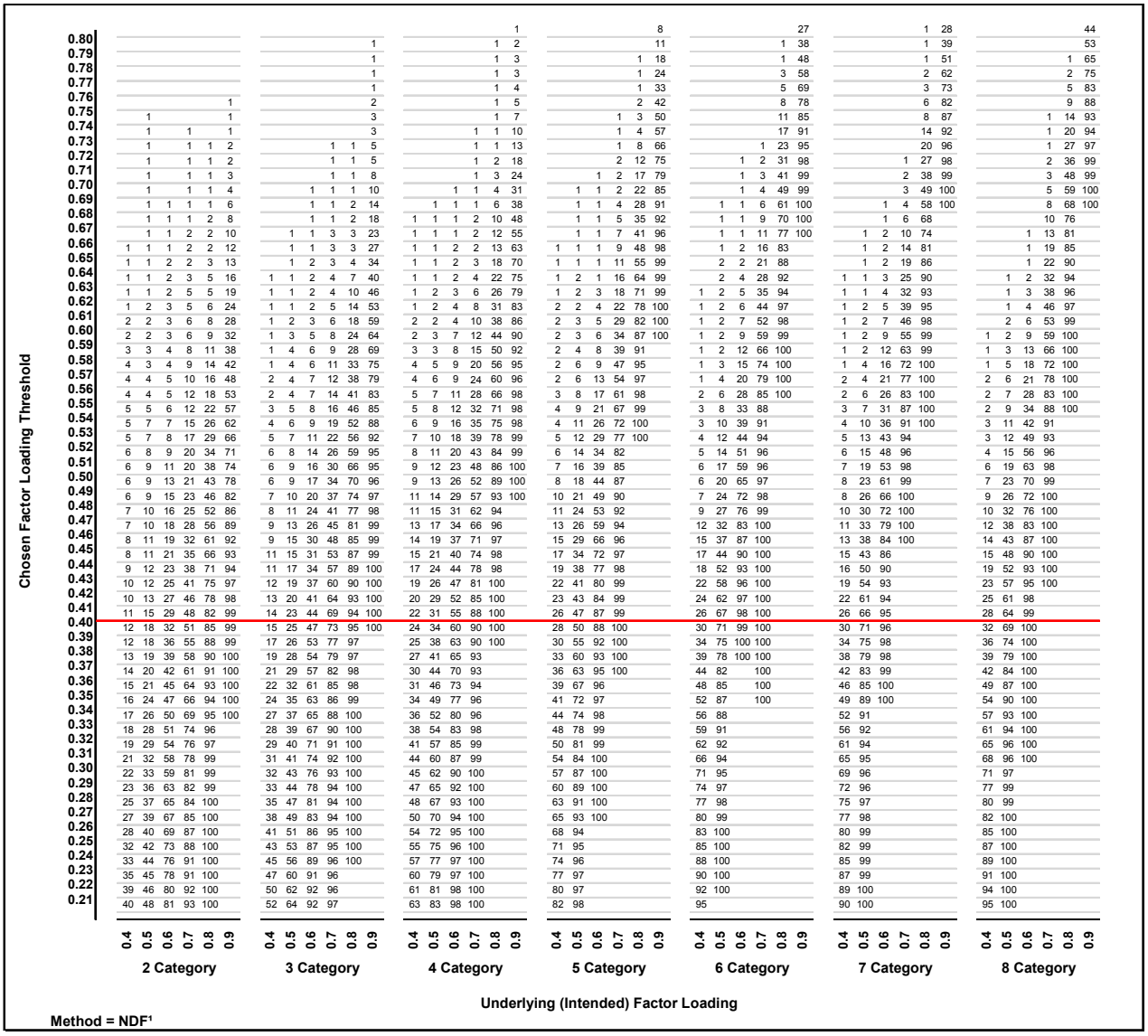


Figure 25 – Percentiles – NDF[1] Method - Skewed Distortion

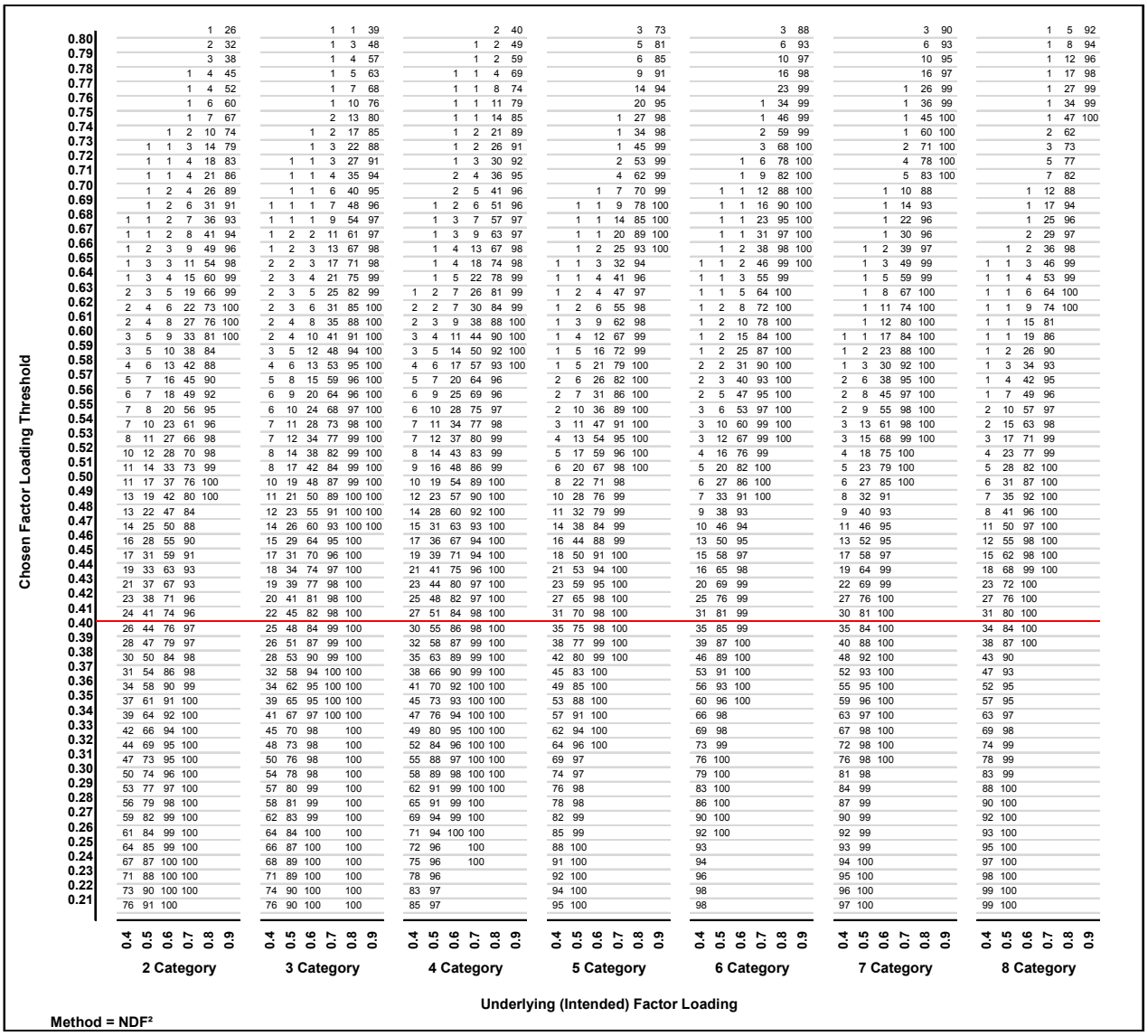


Figure 26 – Percentiles – NDF² Method - Skewed Distortion

Relative Bias

RB		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	-26.5%	-24.7%	-16.9%	-12.2%	-9.1%	-10.2%	-8.9%
	SPR	-26.5%	-25.3%	-18.2%	-12.4%	-9.8%	-11.2%	-9.4%
	NDF[1]	-63.0%	-49.9%	-35.7%	-21.7%	-14.4%	-16.6%	-13.6%
	NDF[2]	-26.5%	-26.4%	-19.8%	-12.9%	-10.2%	-10.8%	-10.1%
0.5	CAT	-27.1%	-25.0%	-16.0%	-11.2%	-8.6%	-8.9%	-8.1%
	SPR	-27.1%	-24.0%	-17.3%	-11.7%	-9.0%	-9.7%	-8.3%
	NDF[1]	-62.3%	-50.7%	-34.9%	-21.1%	-14.6%	-14.8%	-13.6%
	NDF[2]	-27.1%	-26.9%	-20.0%	-12.8%	-10.1%	-10.2%	-9.0%
0.6	CAT	-23.5%	-19.0%	-15.7%	-11.4%	-8.4%	-9.1%	-8.5%
	SPR	-23.5%	-20.6%	-17.2%	-11.9%	-9.4%	-9.6%	-8.9%
	NDF[1]	-46.9%	-37.5%	-30.6%	-20.4%	-14.5%	-15.6%	-14.3%
	NDF[2]	-23.5%	-20.2%	-19.0%	-13.3%	-9.8%	-9.7%	-8.9%
0.7	CAT	-21.7%	-17.1%	-14.3%	-10.0%	-7.4%	-7.6%	-7.5%
	SPR	-21.7%	-18.6%	-16.2%	-11.4%	-8.7%	-8.9%	-8.6%
	NDF[1]	-45.0%	-37.6%	-29.7%	-19.5%	-14.5%	-15.0%	-14.2%
	NDF[2]	-21.7%	-18.4%	-18.1%	-12.8%	-9.5%	-9.3%	-9.4%
0.8	CAT	-19.2%	-15.2%	-12.8%	-8.5%	-6.3%	-6.1%	-6.0%
	SPR	-19.2%	-17.1%	-14.8%	-10.2%	-8.2%	-8.0%	-7.8%
	NDF[1]	-40.8%	-34.3%	-28.1%	-19.1%	-13.9%	-14.4%	-13.2%
	NDF[2]	-19.2%	-16.2%	-16.0%	-11.1%	-8.4%	-8.2%	-8.0%
0.9	CAT	-15.5%	-12.4%	-10.3%	-6.6%	-5.3%	-4.3%	-4.4%
	SPR	-15.5%	-13.6%	-11.9%	-8.5%	-7.1%	-6.4%	-6.3%
	NDF[1]	-38.6%	-32.2%	-26.5%	-18.1%	-13.8%	-13.5%	-12.3%
	NDF[2]	-15.5%	-13.2%	-12.9%	-9.2%	-7.4%	-6.7%	-6.7%

Figure 27 – Method RB Comparison - Skewed Distortion

Coefficient of Variance

CV		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	50.5%	47.5%	37.0%	28.8%	23.4%	25.6%	23.6%
	SPR	50.5%	46.7%	37.2%	25.7%	21.6%	24.2%	20.2%
	NDF¹	137.3%	90.1%	70.2%	40.3%	27.4%	31.6%	26.8%
	NDF²	50.5%	50.5%	41.0%	27.5%	22.7%	24.8%	21.6%
0.5	CAT	38.8%	38.5%	22.1%	17.7%	14.1%	15.0%	14.1%
	SPR	38.8%	34.0%	22.4%	17.0%	13.0%	16.9%	13.3%
	NDF¹	112.0%	77.5%	45.5%	27.2%	18.5%	19.9%	17.8%
	NDF²	38.8%	42.4%	29.3%	18.3%	14.2%	17.4%	14.6%
0.6	CAT	20.8%	17.1%	14.7%	11.0%	9.2%	10.4%	9.7%
	SPR	20.8%	17.7%	15.0%	9.9%	8.5%	8.9%	8.1%
	NDF¹	48.5%	33.7%	25.4%	15.4%	11.3%	12.5%	11.1%
	NDF²	20.8%	18.2%	19.1%	11.7%	9.8%	9.8%	9.2%
0.7	CAT	15.3%	11.8%	10.1%	8.0%	6.8%	6.9%	6.9%
	SPR	15.3%	12.1%	10.4%	7.6%	6.4%	6.3%	6.5%
	NDF¹	36.5%	27.6%	17.3%	11.5%	9.0%	8.2%	8.9%
	NDF²	15.3%	13.3%	13.3%	9.3%	7.7%	7.0%	7.6%
0.8	CAT	10.2%	8.1%	6.9%	5.8%	4.8%	4.8%	4.6%
	SPR	10.2%	9.0%	7.5%	5.5%	4.7%	4.6%	4.4%
	NDF¹	18.3%	15.1%	11.9%	7.7%	6.1%	6.2%	6.0%
	NDF²	10.2%	10.4%	10.2%	6.7%	5.3%	5.4%	5.2%
0.9	CAT	7.5%	5.9%	4.4%	3.2%	2.5%	2.5%	2.6%
	SPR	7.5%	6.5%	5.2%	3.8%	3.0%	3.0%	3.1%
	NDF¹	14.1%	10.9%	8.6%	6.1%	4.4%	4.3%	4.3%
	NDF²	7.5%	7.5%	6.7%	4.7%	3.8%	3.8%	3.6%

Figure 28 – Method CV Comparison - Skewed Distortion

Relative Bias Difference

RB Diff		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.5740%	1.3175%	0.1563%	0.6849%	0.9974%	0.4580%
	NDF¹	36.4445%	25.1838%	18.7456%	9.5033%	5.2837%	6.3317%	4.6967%
	NDF²	0.0000%	1.6138%	2.8742%	0.6311%	1.1265%	0.5296%	1.2268%
0.5	**CAT**	0.0000%	0.9435%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	1.3469%	0.5303%	0.3987%	0.8607%	0.1736%
	NDF¹	35.1337%	26.7069%	18.8918%	9.9041%	6.0332%	5.9512%	5.4699%
	NDF²	0.0000%	2.8667%	3.9848%	1.6492%	1.5643%	1.3463%	0.8497%
0.6	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	1.5747%	1.5283%	0.5763%	0.9170%	0.5305%	0.3559%
	NDF¹	23.4729%	18.5371%	14.9225%	9.0181%	6.0691%	6.5367%	5.7264%
	NDF²	0.0000%	1.2410%	3.2577%	1.9072%	1.3864%	0.6533%	0.4170%
0.7	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	1.5176%	1.9118%	1.4255%	1.3494%	1.3047%	1.0837%
	NDF¹	23.2318%	20.4605%	15.3691%	9.5928%	7.1570%	7.3243%	6.6867%
	NDF²	0.0000%	1.3125%	3.8196%	2.8042%	2.1350%	1.6881%	1.9232%
0.8	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	1.8865%	1.9549%	1.7390%	1.8556%	1.9229%	1.7860%
	NDF¹	21.5763%	19.0719%	15.2512%	10.5726%	7.5934%	8.3106%	7.1540%
	NDF²	0.0000%	1.0322%	3.1471%	2.6078%	2.0431%	2.1215%	1.9726%
0.9	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	1.2093%	1.6570%	1.8938%	1.7741%	2.0734%	1.9022%
	NDF¹	23.0671%	19.8372%	16.2210%	11.5201%	8.4428%	9.2081%	7.8336%
	NDF²	0.0000%	0.8361%	2.5756%	2.6597%	2.1107%	2.4196%	2.2443%
Overall Best Performer (Mean)	**CAT**	**0.0000%**	**0.1573%**	**0.0000%**	**0.0000%**	**0.0000%**	**0.0000%**	**0.0000%**
	SPR	**0.0000%**	**1.1270%**	**1.6194%**	**1.0535%**	**1.1633%**	**1.2816%**	**0.9599%**
	NDF¹	**27.1544%**	**21.6329%**	**16.5669%**	**10.0185%**	**6.7632%**	**7.2771%**	**6.2612%**
	NDF²	**0.0000%**	**1.4837%**	**3.2765%**	**2.0432%**	**1.7276%**	**1.4597%**	**1.4389%**

Figure 29 – Method RB % Underperformance - Skewed Distortion

Coefficient of Variation Difference

CV Diff		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	0.0000%	0.7687%	0.0000%	3.1210%	1.7111%	1.4636%	3.3777%
	SPR	0.0000%	0.0000%	0.2144%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	86.7773%	43.3355%	33.1751%	14.6598%	5.7852%	7.4022%	6.5931%
	NDF²	0.0000%	3.7340%	3.9896%	1.8584%	1.0535%	0.6522%	1.4073%
0.5	**CAT**	0.0000%	4.5058%	0.0000%	0.6334%	1.1035%	0.0000%	0.8561%
	SPR	0.0000%	0.0000%	0.3083%	0.0000%	0.0000%	1.8382%	0.0000%
	NDF¹	73.2241%	43.5156%	23.3478%	10.2162%	5.4915%	4.8608%	4.5029%
	NDF²	0.0000%	8.3534%	7.1457%	1.2589%	1.2818%	2.3882%	1.3379%
0.6	**CAT**	0.0000%	0.0000%	0.0000%	1.0874%	0.6448%	1.4510%	1.6115%
	SPR	0.0000%	0.6270%	0.3657%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	27.6747%	16.5706%	10.7003%	5.4544%	2.7818%	3.5830%	3.0615%
	NDF²	0.0000%	1.0938%	4.4630%	1.7662%	1.3047%	0.9323%	1.0932%
0.7	**CAT**	0.0000%	0.0000%	0.0000%	0.3682%	0.3432%	0.6196%	0.4006%
	SPR	0.0000%	0.3744%	0.3692%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	21.1310%	15.8206%	7.2769%	3.8552%	2.5609%	1.8901%	2.4134%
	NDF²	0.0000%	1.5852%	3.2839%	1.6518%	1.2569%	0.6968%	1.1201%
0.8	**CAT**	0.0000%	0.0000%	0.0000%	0.2873%	0.0346%	0.2361%	0.1849%
	SPR	0.0000%	0.8886%	0.5949%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	8.0226%	7.0533%	4.9924%	2.2568%	1.3604%	1.6198%	1.6283%
	NDF²	0.0000%	2.3661%	3.3135%	1.2348%	0.5900%	0.7666%	0.8763%
0.9	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.6186%	0.7789%	0.6260%	0.4194%	0.5035%	0.5286%
	NDF¹	6.5598%	4.9608%	4.1569%	2.9470%	1.9143%	1.7600%	1.7634%
	NDF²	0.0000%	1.5789%	2.2680%	1.5074%	1.2508%	1.2863%	1.0176%
Overall Best Performer (Mean)	**CAT**	**0.0000%**	**0.8791%**	**0.0000%**	**0.9162%**	**0.6395%**	**0.6284%**	**1.0718%**
	SPR	**0.0000%**	**0.4181%**	**0.4386%**	**0.1043%**	**0.0699%**	**0.3903%**	**0.0881%**
	NDF¹	**37.2316%**	**21.8761%**	**13.9416%**	**6.5649%**	**3.3157%**	**3.5193%**	**3.3271%**
	NDF²	**0.0000%**	**3.1186%**	**4.0773%**	**1.5462%**	**1.1230%**	**1.1204%**	**1.1421%**

Figure 30 – Method CV % Underperformance - Skewed Distortion

7.5 Factor Loading Descriptive Statistics – Kurtotic Distortion

2 Category

Factor Loading	Method	μ	σ	RB	CV
0.4	CAT	0.259235	0.175293	-35.19124%	67.61930%
	SPR	0.259235	0.175293	-35.19124%	67.61930%
	NDF¹	0.11813	0.23015	-70.46745%	194.82744%
	NDF²	0.259235	0.175293	-35.19124%	67.61930%
0.5	CAT	0.327312	0.176902	-34.53766%	54.04698%
	SPR	0.327312	0.176902	-34.53766%	54.04698%
	NDF¹	0.133366	0.240744	-73.32684%	180.51400%
	NDF²	0.327312	0.176902	-34.53766%	54.04698%
0.6	CAT	0.439534	0.104023	-26.74429%	23.66667%
	SPR	0.439534	0.104023	-26.74429%	23.66667%
	NDF¹	0.266636	0.17848	-55.56061%	66.93744%
	NDF²	0.439534	0.104023	-26.74429%	23.66667%
0.7	CAT	0.533584	0.095599	-23.77377%	17.91649%
	SPR	0.533584	0.095599	-23.77377%	17.91649%
	NDF¹	0.324176	0.177284	-53.68910%	54.68755%
	NDF²	0.533584	0.095599	-23.77377%	17.91649%
0.8	CAT	0.633339	0.079064	-20.83264%	12.48374%
	SPR	0.633339	0.079064	-20.83264%	12.48374%
	NDF¹	0.440291	0.111455	-44.96368%	25.31390%
	NDF²	0.633339	0.079064	-20.83264%	12.48374%
0.9	CAT	0.748204	0.058967	-16.86627%	7.88110%
	SPR	0.748204	0.058967	-16.86627%	7.88110%
	NDF¹	0.5112	0.088864	-43.19996%	17.38337%
	NDF²	0.748204	0.058967	-16.86627%	7.88110%

Table 41 - FL Descriptive Statistics - Kurtotic Scale Distortion on 2 Categories

3 Category

Factor Loading	Method	μ	σ	RB	CV
0.4	CAT	0.327744	0.114443	-18.06398%	34.91830%
	SPR	0.327727	0.114155	-18.06832%	34.83232%
	NDF¹	0.23655	0.180837	-40.86253%	76.44752%
	NDF²	0.297179	0.160728	-25.70517%	54.08454%
0.5	CAT	0.399746	0.102477	-20.05085%	25.63556%
	SPR	0.399698	0.102438	-20.06041%	25.62897%
	NDF¹	0.284353	0.178876	-43.12943%	62.90634%
	NDF²	0.371995	0.146308	-25.60090%	39.33066%
0.6	CAT	0.4916	0.064221	-18.06663%	13.06360%
	SPR	0.491621	0.064162	-18.06317%	13.05116%
	NDF¹	0.401197	0.115476	-33.13380%	28.78285%
	NDF²	0.474919	0.087916	-20.84681%	18.51173%
0.7	CAT	0.572716	0.062709	-18.18346%	10.94939%
	SPR	0.572607	0.062764	-18.19902%	10.96115%
	NDF¹	0.466498	0.099934	-33.35743%	21.42209%
	NDF²	0.554251	0.080627	-20.82134%	14.54704%
0.8	CAT	0.671376	0.049626	-16.07803%	7.39168%
	SPR	0.671464	0.049516	-16.06703%	7.37435%
	NDF¹	0.550483	0.076728	-31.18964%	13.93837%
	NDF²	0.655601	0.071203	-18.04989%	10.86071%
0.9	CAT	0.768643	0.043301	-14.59523%	5.63338%
	SPR	0.76858	0.043302	-14.60217%	5.63399%
	NDF¹	0.630018	0.06744	-29.99796%	10.70443%
	NDF²	0.760671	0.054697	-15.48099%	7.19057%

Table 42 - FL Descriptive Statistics - Kurtotic Scale Distortion on 3 Categories

4 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.328955	0.109782	-17.76129%	33.37300%
	SPR	0.328853	0.109717	-17.78676%	33.36364%
	NDF¹	0.221721	0.186199	-44.56981%	83.97899%
	NDF²	0.303878	0.137471	-24.03040%	45.23890%
0.5	**CAT**	0.398522	0.103322	-20.29553%	25.92615%
	SPR	0.398366	0.103076	-20.32678%	25.87457%
	NDF¹	0.28789	0.165949	-42.42202%	57.64334%
	NDF²	0.372925	0.136108	-25.41507%	36.49752%
0.6	**CAT**	0.489317	0.068411	-18.44722%	13.98088%
	SPR	0.48936	0.06842	-18.44005%	13.98162%
	NDF¹	0.403166	0.102786	-32.80562%	25.49468%
	NDF²	0.471672	0.088593	-21.38807%	18.78281%
0.7	**CAT**	0.570555	0.062445	-18.49210%	10.94453%
	SPR	0.570346	0.062503	-18.52201%	10.95870%
	NDF¹	0.466168	0.093086	-33.40453%	19.96825%
	NDF²	0.556576	0.081765	-20.48914%	14.69069%
0.8	**CAT**	0.667797	0.050566	-16.52533%	7.57201%
	SPR	0.667744	0.050332	-16.53197%	7.53756%
	NDF¹	0.543828	0.073731	-32.02152%	13.55773%
	NDF²	0.654178	0.074648	-18.22770%	11.41090%
0.9	**CAT**	0.765789	0.044984	-14.91229%	5.87418%
	SPR	0.765724	0.045092	-14.91960%	5.88876%
	NDF¹	0.62638	0.064265	-30.40218%	10.25977%
	NDF²	0.757075	0.057451	-15.88058%	7.58852%

Table 43 - FL Descriptive Statistics - Kurtotic Scale Distortion on 4 Categories

5 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.365165	0.073827	-8.70874%	20.21736%
	SPR	0.361219	0.078751	-9.69525%	21.80145%
	NDF¹	0.330722	0.109673	-17.31959%	33.16169%
	NDF²	0.353366	0.091438	-11.65850%	25.87621%
0.5	**CAT**	0.453105	0.058815	-9.37907%	12.98034%
	SPR	0.448557	0.060107	-10.28853%	13.40001%
	NDF¹	0.408464	0.089995	-18.30727%	22.03246%
	NDF²	0.440258	0.074994	-11.94830%	17.03411%
0.6	**CAT**	0.546972	0.046355	-8.83803%	8.47476%
	SPR	0.540484	0.05059	-9.91940%	9.36012%
	NDF¹	0.499607	0.065608	-16.73211%	13.13185%
	NDF²	0.529247	0.063363	-11.79212%	11.97235%
0.7	**CAT**	0.636273	0.04068	-9.10389%	6.39348%
	SPR	0.630092	0.04236	-9.98679%	6.72283%
	NDF¹	0.581251	0.06253	-16.96410%	10.75776%
	NDF²	0.617239	0.052377	-11.82296%	8.48576%
0.8	**CAT**	0.733788	0.032864	-8.27651%	4.47873%
	SPR	0.724383	0.03486	-9.45209%	4.81240%
	NDF¹	0.668329	0.04853	-16.45886%	7.26141%
	NDF²	0.718819	0.044475	-10.14762%	6.18723%
0.9	**CAT**	0.83304	0.025698	-7.44001%	3.08488%
	SPR	0.822285	0.02909	-8.63496%	3.53769%
	NDF¹	0.757764	0.039295	-15.80395%	5.18560%
	NDF²	0.822194	0.034682	-8.64508%	4.21827%

Table 44 - FL Descriptive Statistics - Kurtotic Scale Distortion on 5 Categories

6 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.375229	0.060797	-6.19279%	16.20269%
	SPR	0.370139	0.066366	-7.46519%	17.92997%
	NDF¹	0.35486	0.082608	-11.28490%	23.27896%
	NDF²	0.366122	0.075714	-8.46948%	20.67986%
0.5	**CAT**	0.467416	0.052271	-6.51675%	11.18296%
	SPR	0.46135	0.055939	-7.72992%	12.12499%
	NDF¹	0.441992	0.071516	-11.60161%	16.18040%
	NDF²	0.457498	0.067719	-8.50049%	14.80209%
0.6	**CAT**	0.560452	0.043264	-6.59133%	7.71955%
	SPR	0.554883	0.04652	-7.51942%	8.38372%
	NDF¹	0.528004	0.05624	-11.99939%	10.65146%
	NDF²	0.547007	0.052965	-8.83212%	9.68277%
0.7	**CAT**	0.654736	0.034704	-6.46628%	5.30044%
	SPR	0.64783	0.037552	-7.45283%	5.79660%
	NDF¹	0.615702	0.047719	-12.04251%	7.75037%
	NDF²	0.642308	0.043183	-8.24177%	6.72310%
0.8	**CAT**	0.751154	0.026563	-6.10579%	3.53629%
	SPR	0.740836	0.02933	-7.39555%	3.95905%
	NDF¹	0.706697	0.041016	-11.66282%	5.80386%
	NDF²	0.736447	0.037117	-7.94414%	5.04004%
0.9	**CAT**	0.847207	0.020581	-5.86586%	2.42932%
	SPR	0.834953	0.023851	-7.22749%	2.85661%
	NDF¹	0.79364	0.031893	-11.81778%	4.01853%
	NDF²	0.831845	0.028717	-7.57282%	3.45225%

Table 45 - FL Descriptive Statistics - Kurtotic Scale Distortion on 6 Categories

7 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.381181	0.052903	-4.70483%	13.87871%
	SPR	0.374848	0.062536	-6.28807%	16.68308%
	NDF¹	0.367493	0.072623	-8.12676%	19.76162%
	NDF²	0.372537	0.064044	-6.86586%	17.19143%
0.5	**CAT**	0.4753	0.042079	-4.94004%	8.85317%
	SPR	0.468676	0.048653	-6.26483%	10.38085%
	NDF¹	0.457601	0.056537	-8.47979%	12.35504%
	NDF²	0.465827	0.050121	-6.83456%	10.75947%
0.6	**CAT**	0.571356	0.033085	-4.77407%	5.79059%
	SPR	0.564798	0.038245	-5.86705%	6.77149%
	NDF¹	0.547605	0.045244	-8.73251%	8.26213%
	NDF²	0.561701	0.040873	-6.38316%	7.27662%
0.7	**CAT**	0.666153	0.028071	-4.83533%	4.21387%
	SPR	0.659407	0.029851	-5.79905%	4.52700%
	NDF¹	0.638273	0.038754	-8.81817%	6.07177%
	NDF²	0.65442	0.035975	-6.51139%	5.49720%
0.8	**CAT**	0.76141	0.021796	-4.82379%	2.86257%
	SPR	0.751944	0.026124	-6.00694%	3.47415%
	NDF¹	0.728082	0.031404	-8.98970%	4.31330%
	NDF²	0.747046	0.029878	-6.61926%	3.99952%
0.9	**CAT**	0.858752	0.017379	-4.58313%	2.02381%
	SPR	0.851699	0.019613	-5.36676%	2.30277%
	NDF¹	0.820242	0.025959	-8.86198%	3.16480%
	NDF²	0.844404	0.025156	-6.17736%	2.97913%

Table 46 - FL Descriptive Statistics - Kurtotic Scale Distortion on 7 Categories

8 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.381094	0.055608	-4.72649%	14.59161%
	SPR	0.378412	0.058798	-5.39697%	15.53818%
	NDF¹	0.369715	0.066495	-7.57129%	17.98546%
	NDF²	0.375728	0.068225	-6.06788%	18.15817%
0.5	**CAT**	0.473795	0.043366	-5.24102%	9.15282%
	SPR	0.470248	0.045869	-5.95040%	9.75416%
	NDF¹	0.459562	0.056808	-8.08770%	12.36133%
	NDF²	0.466918	0.051135	-6.61645%	10.95162%
0.6	**CAT**	0.572073	0.03747	-4.65449%	6.54990%
	SPR	0.568467	0.038209	-5.25551%	6.72133%
	NDF¹	0.553715	0.04436	-7.71422%	8.01127%
	NDF²	0.56552	0.039635	-5.74671%	7.00867%
0.7	**CAT**	0.663641	0.03186	-5.19411%	4.80083%
	SPR	0.659835	0.032121	-5.73785%	4.86800%
	NDF¹	0.638804	0.039951	-8.74228%	6.25397%
	NDF²	0.656876	0.037581	-6.16052%	5.72111%
0.8	**CAT**	0.763101	0.024515	-4.61239%	3.21258%
	SPR	0.753106	0.02556	-5.86180%	3.39397%
	NDF¹	0.7315	0.03056	-8.56249%	4.17777%
	NDF²	0.752268	0.029045	-5.96645%	3.86094%
0.9	**CAT**	0.863802	0.01784	-4.02200%	2.06534%
	SPR	0.85114	0.020406	-5.42890%	2.39750%
	NDF¹	0.82429	0.026759	-8.41217%	3.24633%
	NDF²	0.849469	0.024238	-5.61457%	2.85327%

Table 47 - FL Descriptive Statistics - Kurtotic Scale Distortion on 8 Categories

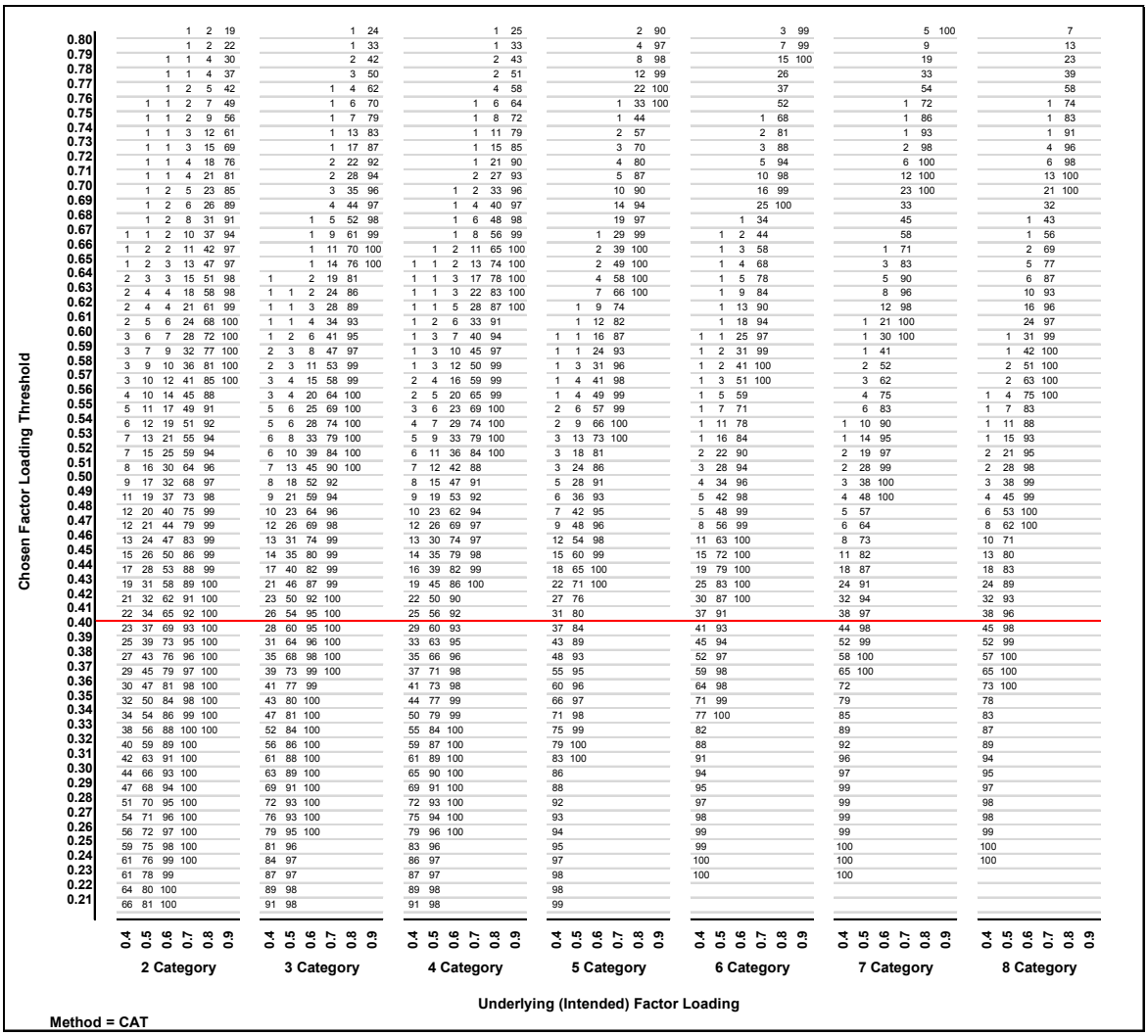
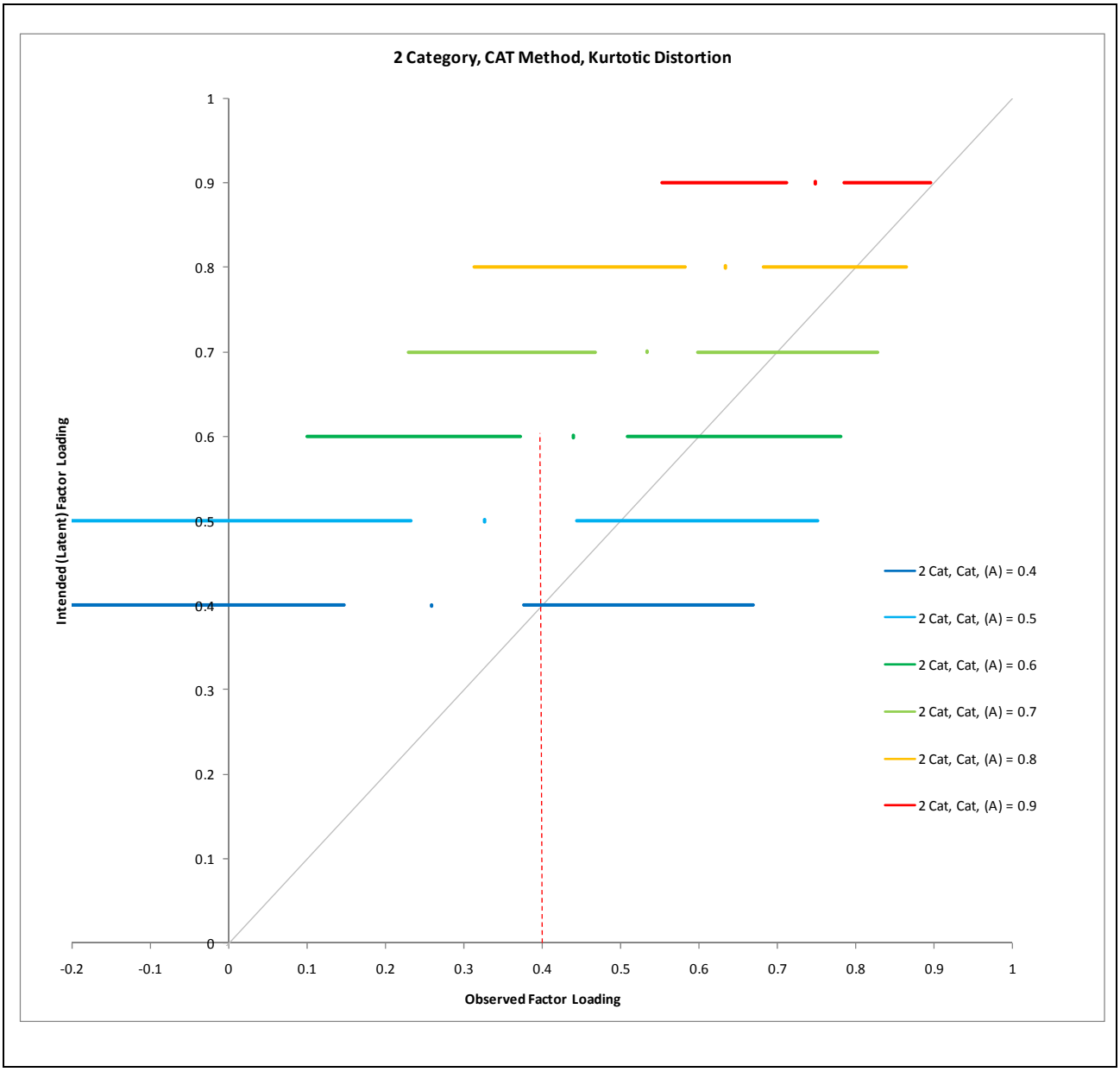
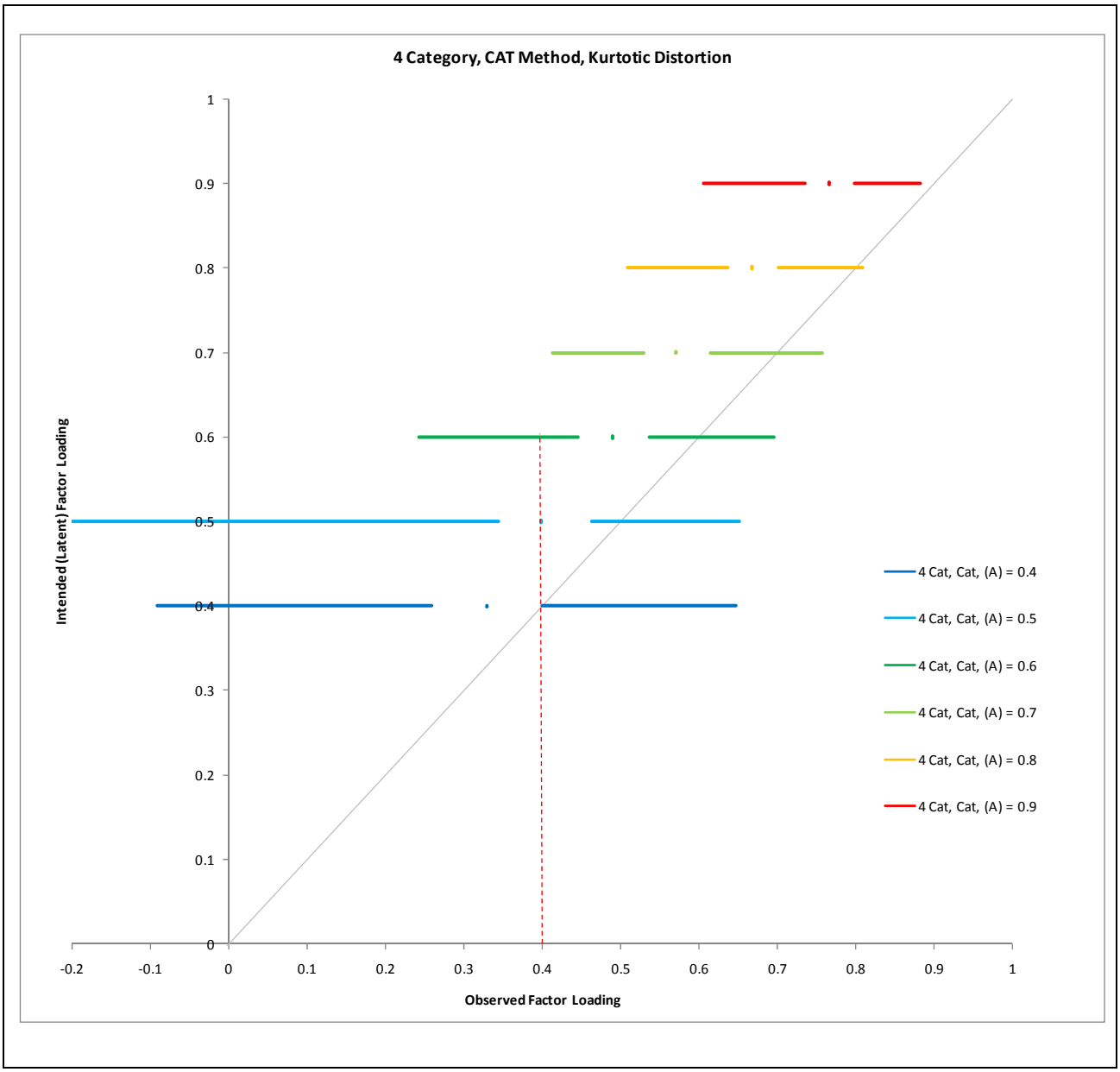
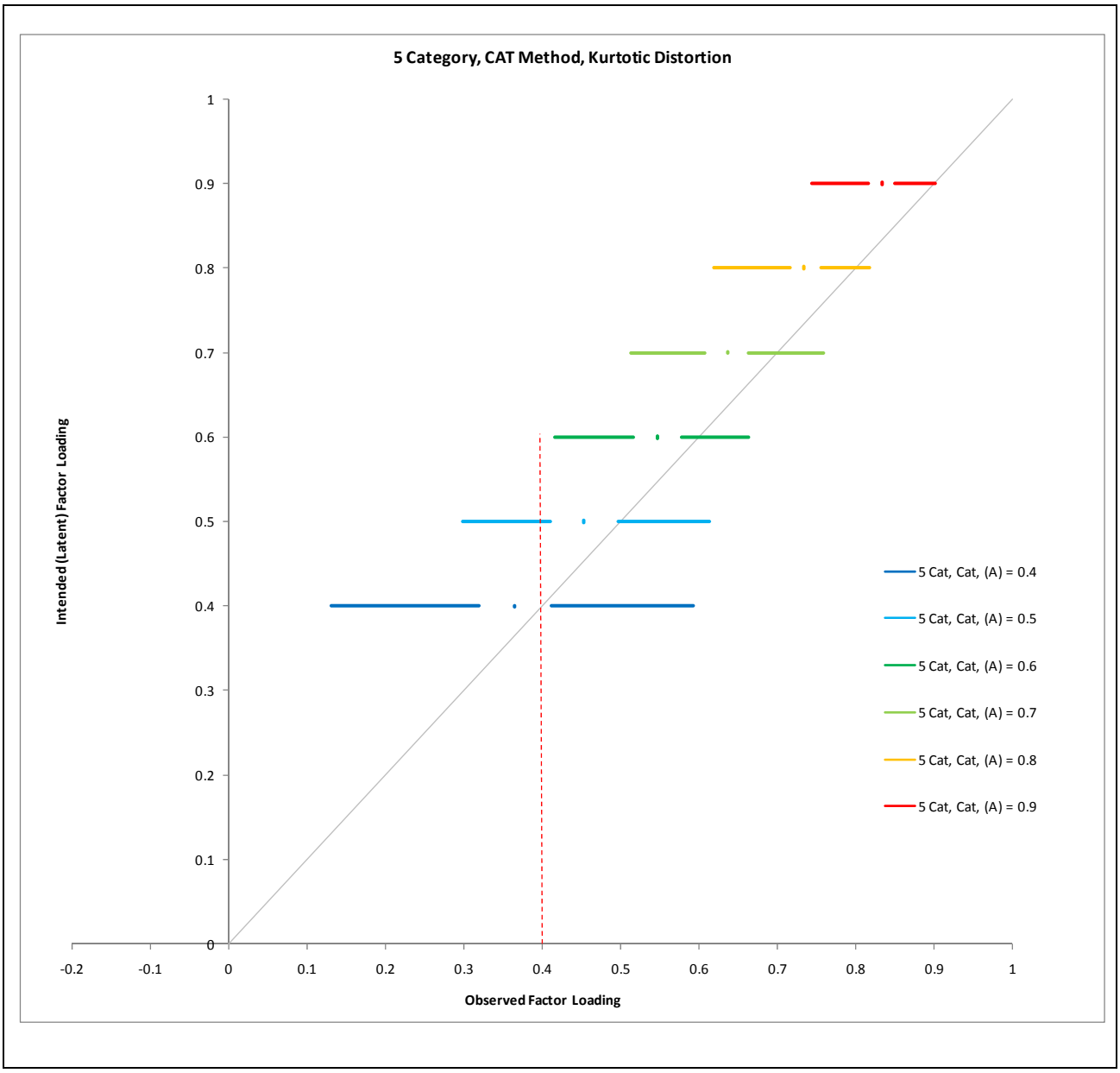
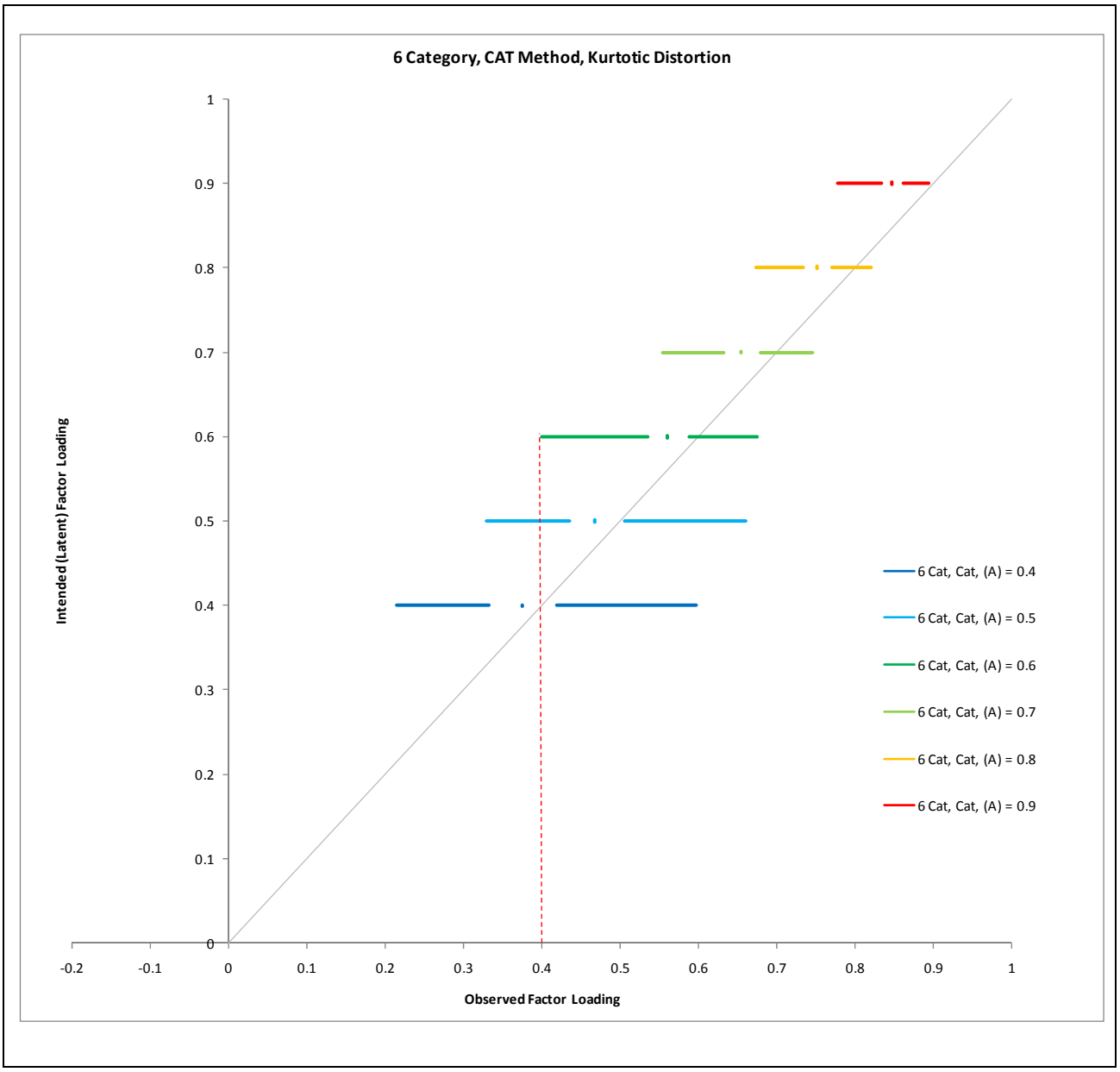


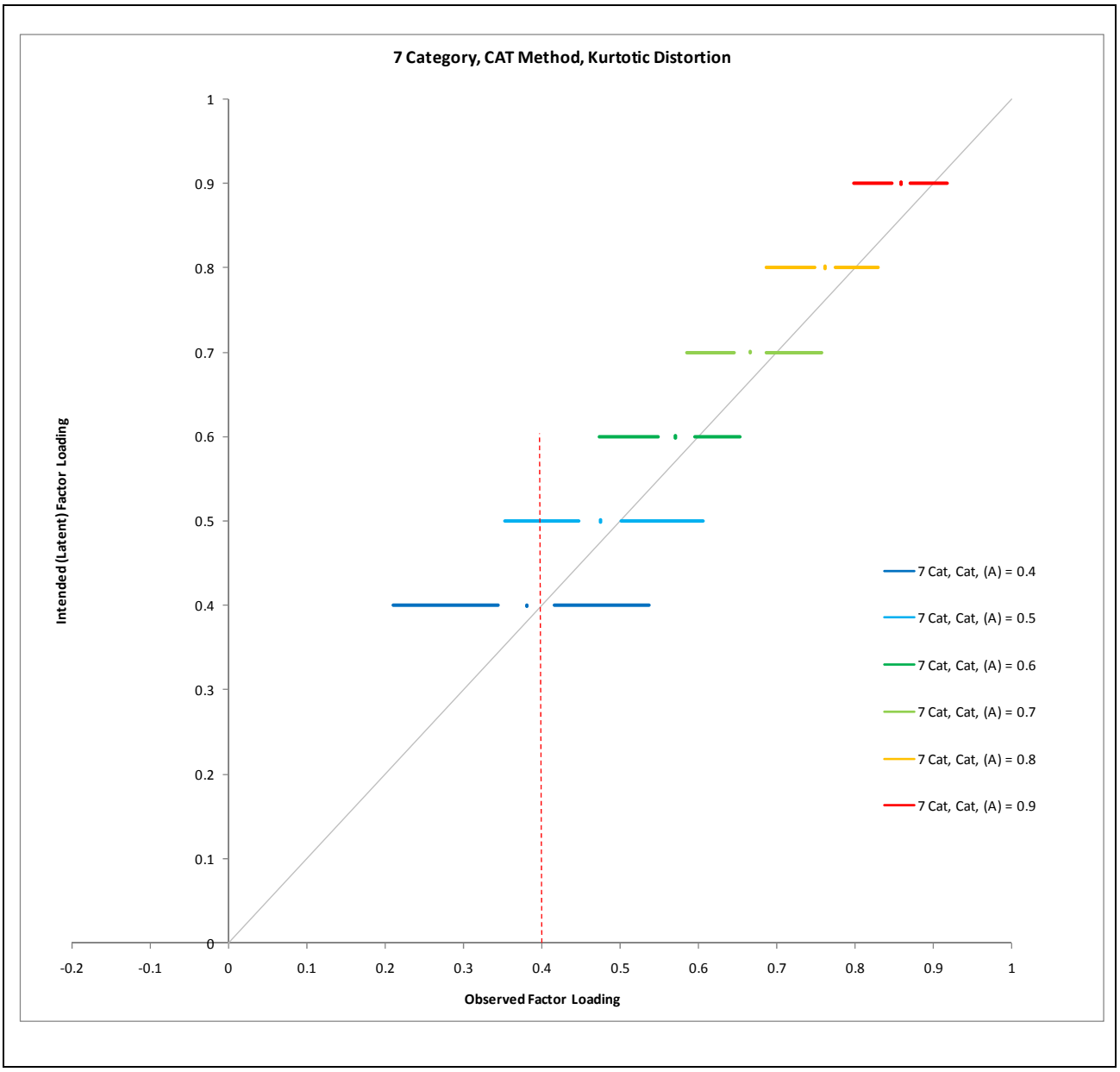
Figure 31 - Percentiles – CAT Method - Kurtotic Distortion


2 Category, CAT Method, Kurtotic Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2 Cat, Cat, (A) = 0.4
2 Cat, Cat, (A) = 0.5
2 Cat, Cat, (A) = 0.6
2 Cat, Cat, (A) = 0.7
2 Cat, Cat, (A) = 0.8
2 Cat, Cat, (A) = 0.9


4 Category, CAT Method, Kurtotic Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
4 Cat, Cat, (A) = 0.4
4 Cat, Cat, (A) = 0.5
4 Cat, Cat, (A) = 0.6
4 Cat, Cat, (A) = 0.7
4 Cat, Cat, (A) = 0.8
4 Cat, Cat, (A) = 0.9


5 Category, CAT Method, Kurtotic Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
5 Cat, Cat, (A) = 0.4
5 Cat, Cat, (A) = 0.5
5 Cat, Cat, (A) = 0.6
5 Cat, Cat, (A) = 0.7
5 Cat, Cat, (A) = 0.8
5 Cat, Cat, (A) = 0.9


6 Category, CAT Method, Kurtotic Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
6 Cat, Cat, (A) = 0.4
6 Cat, Cat, (A) = 0.5
6 Cat, Cat, (A) = 0.6
6 Cat, Cat, (A) = 0.7
6 Cat, Cat, (A) = 0.8
6 Cat, Cat, (A) = 0.9


7 Category, CAT Method, Kurtotic Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
7 Cat, Cat, (A) = 0.4
7 Cat, Cat, (A) = 0.5
7 Cat, Cat, (A) = 0.6
7 Cat, Cat, (A) = 0.7
7 Cat, Cat, (A) = 0.8
7 Cat, Cat, (A) = 0.9

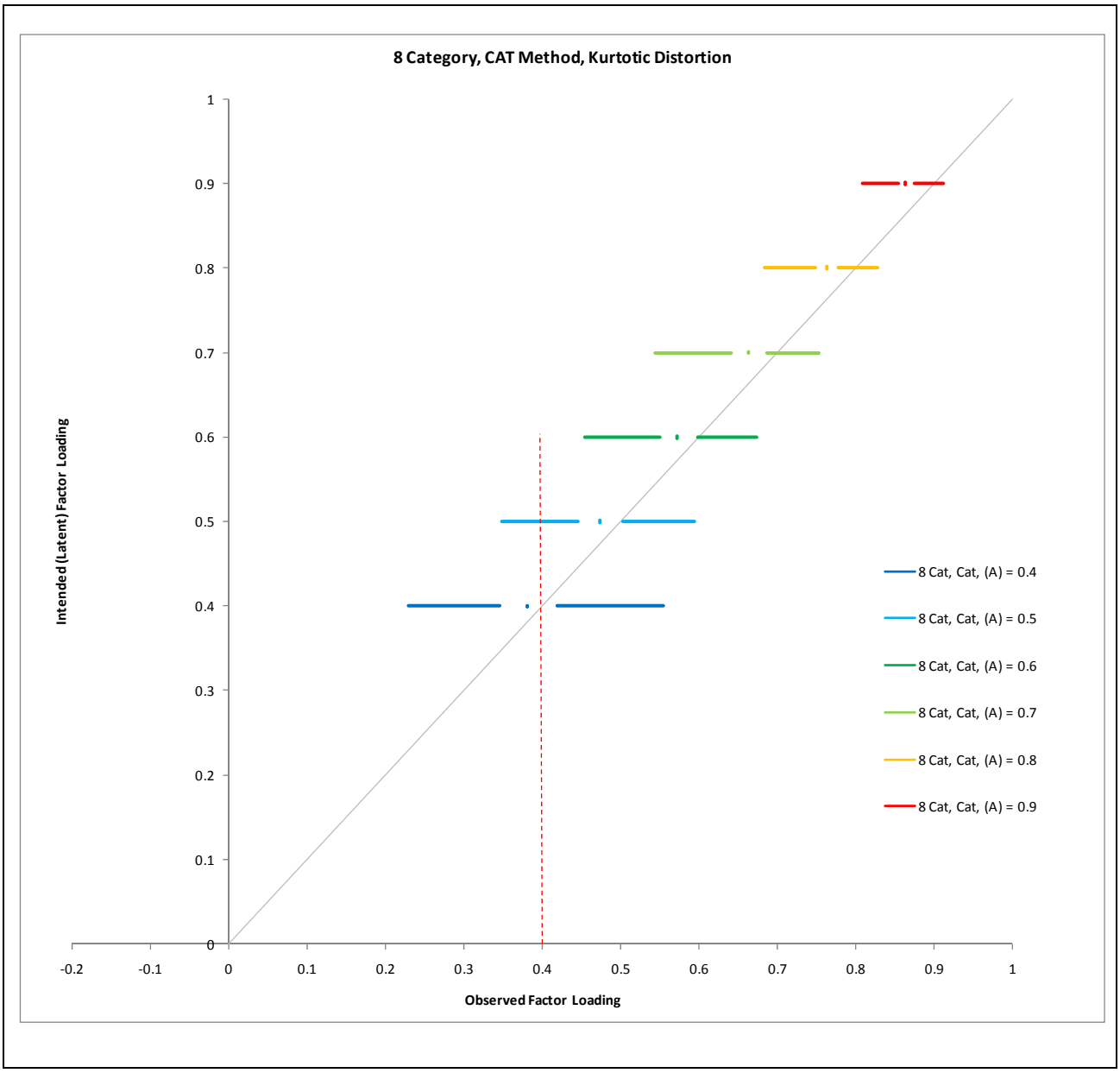


Figure 32 - Modified Box Plots Category 2, 4, 5, 6, 7 and 8 (Kurtotic Distortion)

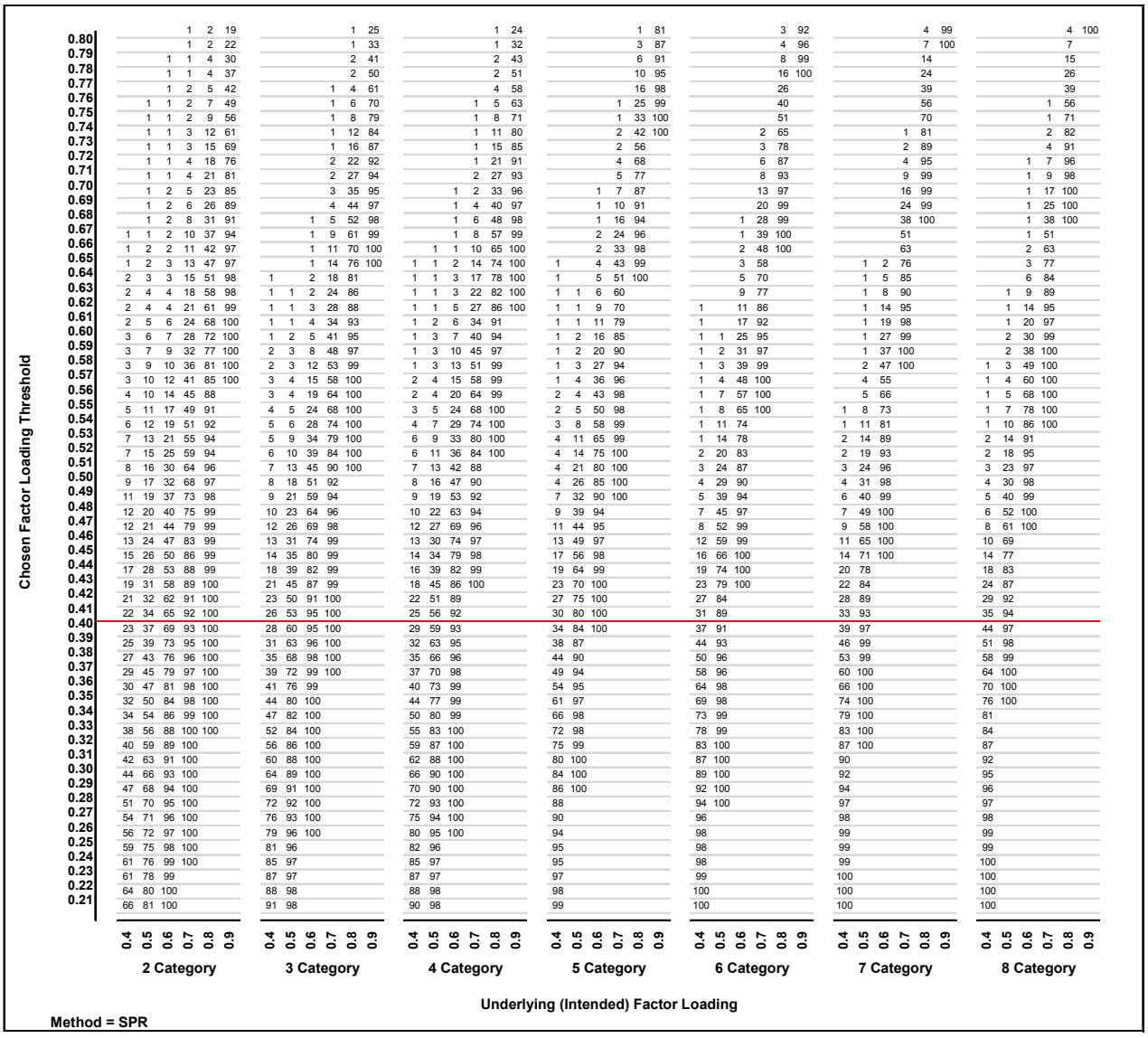


Figure 33 – Percentiles – SPR Method - Kurtotic Distortion

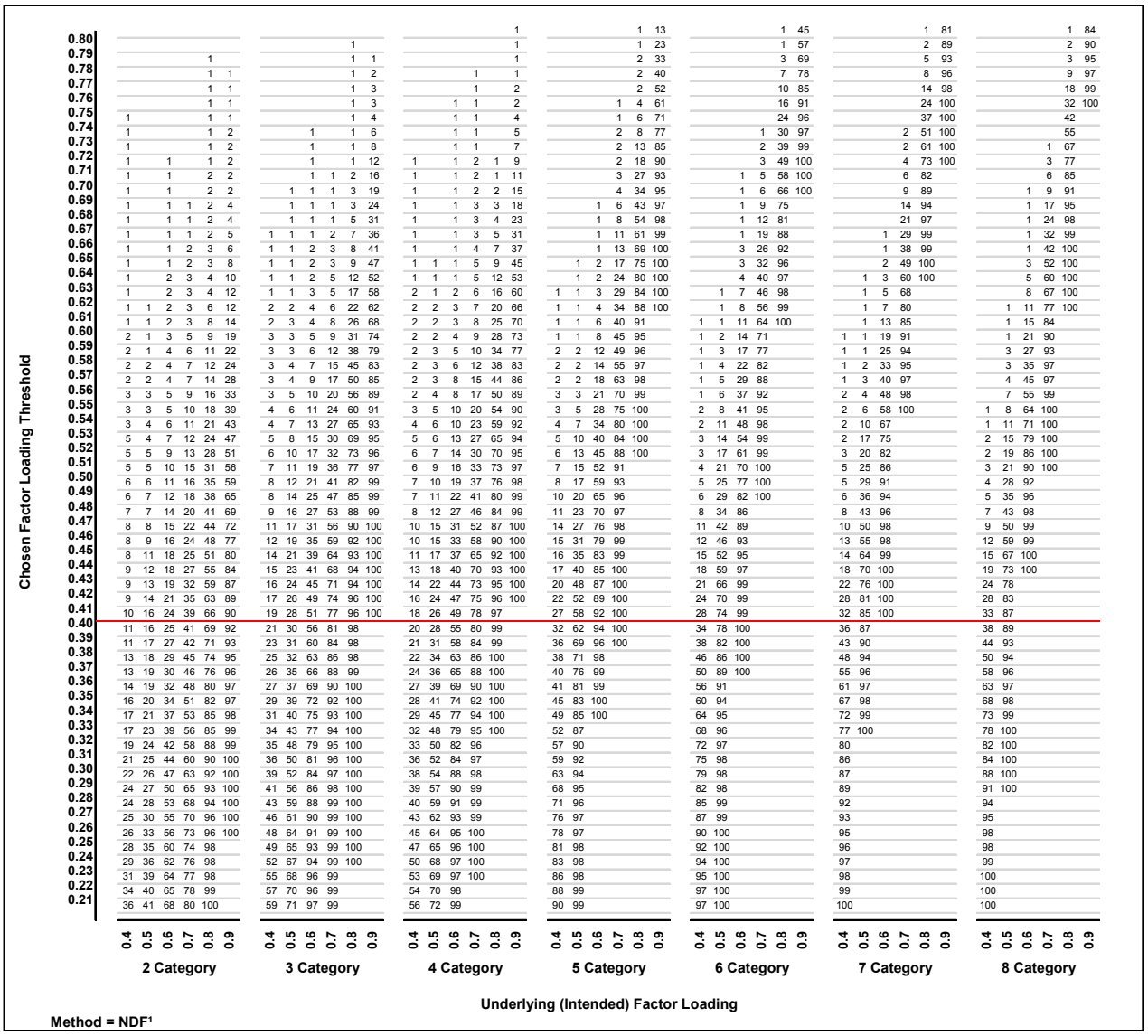


Figure 34 – Percentiles – NDF[1] Method - Kurtotic Distortion

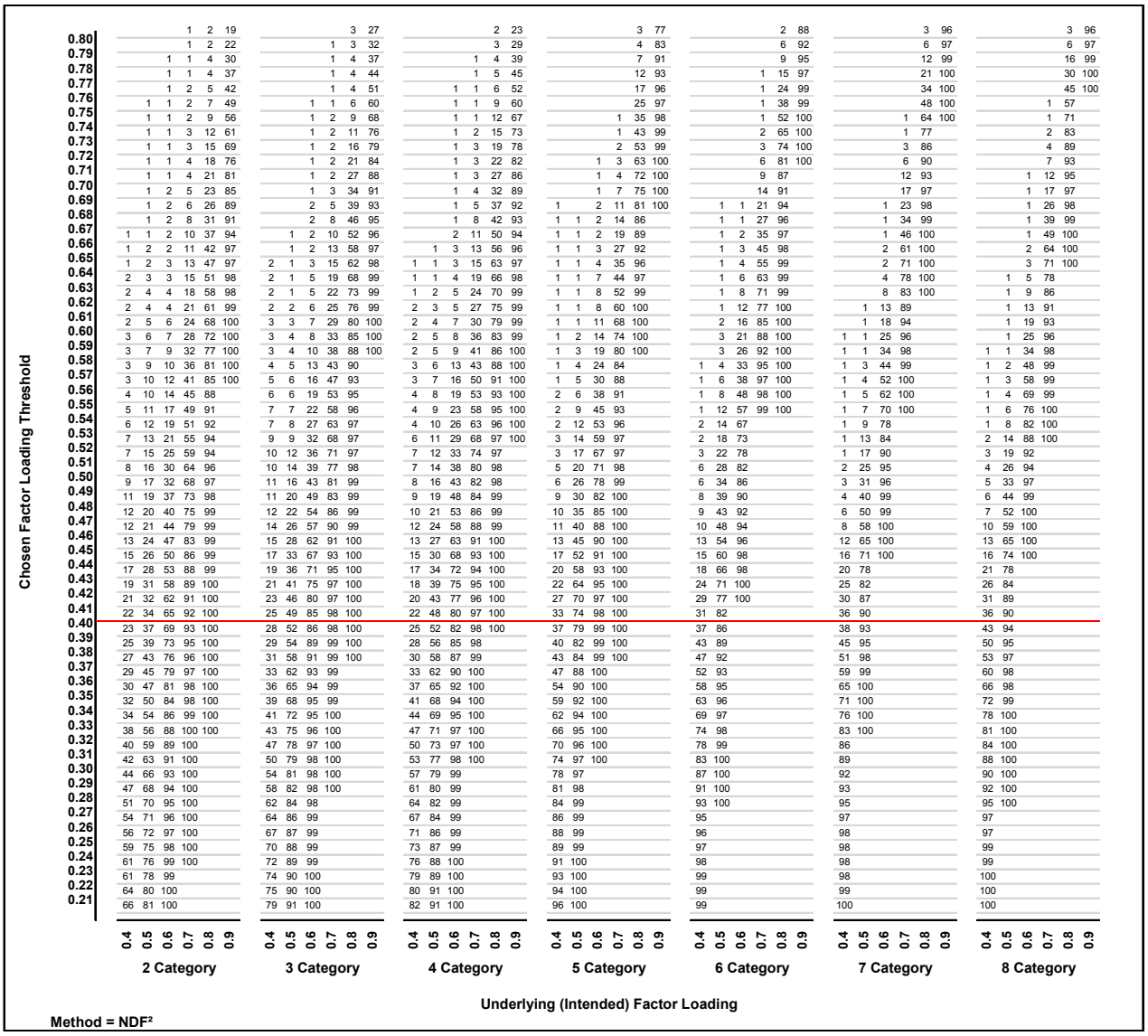


Figure 35 – Percentiles – NDF^2 Method - Kurtotic Distortion

Relative Bias

RB		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	-35.2%	-18.1%	-17.8%	-8.7%	-6.2%	-4.7%	-4.7%
	SPR	-35.2%	-18.1%	-17.8%	-9.7%	-7.5%	-6.3%	-5.4%
	NDF[1]	-70.5%	-40.9%	-44.6%	-17.3%	-11.3%	-8.1%	-7.6%
	NDF[2]	-35.2%	-25.7%	-24.0%	-11.7%	-8.5%	-6.9%	-6.1%
0.5	CAT	-34.5%	-20.1%	-20.3%	-9.4%	-6.5%	-4.9%	-5.2%
	SPR	-34.5%	-20.1%	-20.3%	-10.3%	-7.7%	-6.3%	-6.0%
	NDF[1]	-73.3%	-43.1%	-42.4%	-18.3%	-11.6%	-8.5%	-8.1%
	NDF[2]	-34.5%	-25.6%	-25.4%	-11.9%	-8.5%	-6.8%	-6.6%
0.6	CAT	-26.7%	-18.1%	-18.4%	-8.8%	-6.6%	-4.8%	-4.7%
	SPR	-26.7%	-18.1%	-18.4%	-9.9%	-7.5%	-5.9%	-5.3%
	NDF[1]	-55.6%	-33.1%	-32.8%	-16.7%	-12.0%	-8.7%	-7.7%
	NDF[2]	-26.7%	-20.8%	-21.4%	-11.8%	-8.8%	-6.4%	-5.7%
0.7	CAT	-23.8%	-18.2%	-18.5%	-9.1%	-6.5%	-4.8%	-5.2%
	SPR	-23.8%	-18.2%	-18.5%	-10.0%	-7.5%	-5.8%	-5.7%
	NDF[1]	-53.7%	-33.4%	-33.4%	-17.0%	-12.0%	-8.8%	-8.7%
	NDF[2]	-23.8%	-20.8%	-20.5%	-11.8%	-8.2%	-6.5%	-6.2%
0.8	CAT	-20.8%	-16.1%	-16.5%	-8.3%	-6.1%	-4.8%	-4.6%
	SPR	-20.8%	-16.1%	-16.5%	-9.5%	-7.4%	-6.0%	-5.9%
	NDF[1]	-45.0%	-31.2%	-32.0%	-16.5%	-11.7%	-9.0%	-8.6%
	NDF[2]	-20.8%	-18.0%	-18.2%	-10.1%	-7.9%	-6.6%	-6.0%
0.9	CAT	-16.9%	-14.6%	-14.9%	-7.4%	-5.9%	-4.6%	-4.0%
	SPR	-16.9%	-14.6%	-14.9%	-8.6%	-7.2%	-5.4%	-5.4%
	NDF[1]	-43.2%	-30.0%	-30.4%	-15.8%	-11.8%	-8.9%	-8.4%
	NDF[2]	-16.9%	-15.5%	-15.9%	-8.6%	-7.6%	-6.2%	-5.6%

Figure 36 – Method RB Comparison - Kurtotic Distortion

Coefficient of Variance

CV		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	67.6%	34.9%	33.4%	20.2%	16.2%	13.9%	14.6%
	SPR	67.6%	34.8%	33.4%	21.8%	17.9%	16.7%	15.5%
	NDF¹	194.8%	76.4%	84.0%	33.2%	23.3%	19.8%	18.0%
	NDF²	67.6%	54.1%	45.2%	25.9%	20.7%	17.2%	18.2%
0.5	CAT	54.0%	25.6%	25.9%	13.0%	11.2%	8.9%	9.2%
	SPR	54.0%	25.6%	25.9%	13.4%	12.1%	10.4%	9.8%
	NDF¹	180.5%	62.9%	57.6%	22.0%	16.2%	12.4%	12.4%
	NDF²	54.0%	39.3%	36.5%	17.0%	14.8%	10.8%	11.0%
0.6	CAT	23.7%	13.1%	14.0%	8.5%	7.7%	5.8%	6.5%
	SPR	23.7%	13.1%	14.0%	9.4%	8.4%	6.8%	6.7%
	NDF¹	66.9%	28.8%	25.5%	13.1%	10.7%	8.3%	8.0%
	NDF²	23.7%	18.5%	18.8%	12.0%	9.7%	7.3%	7.0%
0.7	CAT	17.9%	10.9%	10.9%	6.4%	5.3%	4.2%	4.8%
	SPR	17.9%	11.0%	11.0%	6.7%	5.8%	4.5%	4.9%
	NDF¹	54.7%	21.4%	20.0%	10.8%	7.8%	6.1%	6.3%
	NDF²	17.9%	14.5%	14.7%	8.5%	6.7%	5.5%	5.7%
0.8	CAT	12.5%	7.4%	7.6%	4.5%	3.5%	2.9%	3.2%
	SPR	12.5%	7.4%	7.5%	4.8%	4.0%	3.5%	3.4%
	NDF¹	25.3%	13.9%	13.6%	7.3%	5.8%	4.3%	4.2%
	NDF²	12.5%	10.9%	11.4%	6.2%	5.0%	4.0%	3.9%
0.9	CAT	7.9%	5.6%	5.9%	3.1%	2.4%	2.0%	2.1%
	SPR	7.9%	5.6%	5.9%	3.5%	2.9%	2.3%	2.4%
	NDF¹	17.4%	10.7%	10.3%	5.2%	4.0%	3.2%	3.2%
	NDF²	7.9%	7.2%	7.6%	4.2%	3.5%	3.0%	2.9%

Figure 37 – Method CV Comparison - Kurtotic Distortion

Relative Bias Difference RB Diff		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0043%	0.0255%	0.9865%	1.2724%	1.5832%	0.6705%
	NDF¹	35.2762%	22.7985%	26.8085%	8.6109%	5.0921%	3.4219%	2.8448%
	NDF²	0.0000%	7.6412%	6.2691%	2.9498%	2.2767%	2.1610%	1.3414%
0.5	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0096%	0.0312%	0.9095%	1.2132%	1.3248%	0.7094%
	NDF¹	38.7892%	23.0786%	22.1265%	8.9282%	5.0849%	3.5398%	2.8467%
	NDF²	0.0000%	5.5501%	5.1195%	2.5692%	1.9837%	1.8945%	1.3754%
0.6	**CAT**	0.0000%	0.0035%	0.0072%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	0.0000%	1.0814%	0.9281%	1.0930%	0.6010%
	NDF¹	28.8163%	15.0706%	14.3656%	7.8941%	5.4081%	3.9584%	3.0597%
	NDF²	0.0000%	2.7836%	2.9480%	2.9541%	2.2408%	1.6091%	1.0922%
0.7	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0156%	0.0299%	0.8829%	0.9865%	0.9637%	0.5437%
	NDF¹	29.9153%	15.1740%	14.9124%	7.8602%	5.5762%	3.9828%	3.5482%
	NDF²	0.0000%	2.6379%	1.9970%	2.7191%	1.7755%	1.6761%	0.9664%
0.8	**CAT**	0.0000%	0.0110%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	0.0066%	1.1756%	1.2898%	1.1832%	1.2494%
	NDF¹	24.1310%	15.1226%	15.4962%	8.1823%	5.5570%	4.1659%	3.9501%
	NDF²	0.0000%	1.9829%	1.7024%	1.8711%	1.8384%	1.7955%	1.3541%
0.9	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0069%	0.0073%	1.1950%	1.3616%	0.7836%	1.4069%
	NDF¹	26.3337%	15.4027%	15.4899%	8.3639%	5.9519%	4.2788%	4.3902%
	NDF²	0.0000%	0.8858%	0.9683%	1.2051%	1.7070%	1.5942%	1.5926%
Overall Best Performer (Mean)	**CAT**	**0.0000%**	**0.0024%**	**0.0012%**	**0.0000%**	**0.0000%**	**0.0000%**	**0.0000%**
	SPR	**0.0000%**	**0.0061%**	**0.0168%**	**1.0385%**	**1.1753%**	**1.1553%**	**0.8635%**
	NDF¹	**30.5436%**	**17.7745%**	**18.1998%**	**8.3066%**	**5.4450%**	**3.8913%**	**3.4399%**
	NDF²	**0.0000%**	**3.5802%**	**3.1674%**	**2.3781%**	**1.9703%**	**1.7884%**	**1.2870%**

Figure 38 – Method RB % Underperformance - Kurtotic Distortion

Coefficient of Variation Difference

CV Diff		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	0.0000%	0.0860%	0.0094%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	0.0000%	1.5841%	1.7273%	2.8044%	0.9466%
	NDF¹	127.2081%	41.6152%	50.6154%	12.9443%	7.0763%	5.8829%	3.3938%
	NDF²	0.0000%	19.2522%	11.8753%	5.6588%	4.4772%	3.3127%	3.5666%
0.5	**CAT**	0.0000%	0.0066%	0.0516%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	0.0000%	0.4197%	0.9420%	1.5277%	0.6013%
	NDF¹	126.4670%	37.2774%	31.7688%	9.0521%	4.9974%	3.5019%	3.2085%
	NDF²	0.0000%	13.7017%	10.6230%	4.0538%	3.6191%	1.9063%	1.7988%
0.6	**CAT**	0.0000%	0.0124%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	0.0007%	0.8854%	0.6642%	0.9809%	0.1714%
	NDF¹	43.2708%	15.7317%	11.5138%	4.6571%	2.9319%	2.4715%	1.4614%
	NDF²	0.0000%	5.4606%	4.8019%	3.4976%	1.9632%	1.4860%	0.4588%
0.7	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0118%	0.0142%	0.3294%	0.4962%	0.3131%	0.0672%
	NDF¹	36.7711%	10.4727%	9.0237%	4.3643%	2.4499%	1.8579%	1.4531%
	NDF²	0.0000%	3.5976%	3.7462%	2.0923%	1.4227%	1.2833%	0.9203%
0.8	**CAT**	0.0000%	0.0173%	0.0345%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0000%	0.0000%	0.3337%	0.4228%	0.6116%	0.1814%
	NDF¹	12.8302%	6.5640%	6.0202%	2.7827%	2.2676%	1.4507%	0.9652%
	NDF²	0.0000%	3.4864%	3.8733%	1.7085%	1.5037%	1.1370%	0.6484%
0.9	**CAT**	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	SPR	0.0000%	0.0006%	0.0146%	0.4528%	0.4273%	0.2790%	0.3322%
	NDF¹	9.5023%	5.0711%	4.3856%	2.1007%	1.5892%	1.1410%	1.1810%
	NDF²	0.0000%	1.5572%	1.7143%	1.1334%	1.0229%	0.9553%	0.7879%
Overall Best Performer (Mean)	**CAT**	**0.0000%**	**0.0204%**	**0.0159%**	**0.0000%**	**0.0000%**	**0.0000%**	**0.0000%**
	SPR	**0.0000%**	**0.0021%**	**0.0049%**	**0.6675%**	**0.7799%**	**1.0861%**	**0.3833%**
	NDF¹	**59.3416%**	**19.4553%**	**18.8879%**	**5.9835%**	**3.5521%**	**2.7177%**	**1.9438%**
	NDF²	**0.0000%**	**7.8426%**	**6.1057%**	**3.0241%**	**2.3348%**	**1.6801%**	**1.3634%**

Figure 39 – Method CV % Underperformance - Kurtotic Distortion

7.6 Factor Loading Descriptive Statistics – Bimodal Distortion

2 Category

Factor Loading	Method	μ	σ	RB	CV
0.4	CAT	0.305088	0.13322	-23.72797%	43.66611%
	SPR	0.305088	0.13322	-23.72797%	43.66611%
	NDF¹	0.222461	0.185294	-44.38477%	83.29271%
	NDF²	0.305088	0.13322	-23.72797%	43.66611%
0.5	CAT	0.391965	0.113818	-21.60691%	29.03773%
	SPR	0.391965	0.113818	-21.60691%	29.03773%
	NDF¹	0.286082	0.187994	-42.78353%	65.71327%
	NDF²	0.391965	0.113818	-21.60691%	29.03773%
0.6	CAT	0.485757	0.07306	-19.04046%	15.04043%
	SPR	0.485757	0.07306	-19.04046%	15.04043%
	NDF¹	0.389033	0.114077	-35.16117%	29.32310%
	NDF²	0.485757	0.07306	-19.04046%	15.04043%
0.7	CAT	0.5716	0.062369	-18.34287%	10.91136%
	SPR	0.5716	0.062369	-18.34287%	10.91136%
	NDF¹	0.46069	0.090637	-34.18713%	19.67425%
	NDF²	0.5716	0.062369	-18.34287%	10.91136%
0.8	CAT	0.6638	0.052307	-17.02501%	7.87991%
	SPR	0.6638	0.052307	-17.02501%	7.87991%
	NDF¹	0.530709	0.074602	-33.66144%	14.05702%
	NDF²	0.6638	0.052307	-17.02501%	7.87991%
0.9	CAT	0.776139	0.042109	-13.76228%	5.42542%
	SPR	0.776139	0.042109	-13.76228%	5.42542%
	NDF¹	0.62219	0.0604	-30.86778%	9.70772%
	NDF²	0.776139	0.042109	-13.76228%	5.42542%

Table 48 - FL Descriptive Statistics - Bimodal Scale Distortion on 2 Categories

3 Category

Factor Loading	Method	μ	σ	RB	CV
0.4	CAT	0.317781	0.120199	-20.55477%	37.82453%
	SPR	0.317436	0.120397	-20.64094%	37.92804%
	NDF[1]	0.214959	0.187856	-46.26019%	87.39150%
	NDF[2]	0.316032	0.122161	-20.99202%	38.65468%
0.5	CAT	0.411983	0.08996	-17.60338%	21.83577%
	SPR	0.412156	0.090057	-17.56881%	21.85026%
	NDF[1]	0.294434	0.179725	-41.11325%	61.04077%
	NDF[2]	0.410674	0.090056	-17.86519%	21.92889%
0.6	CAT	0.494476	0.072189	-17.58736%	14.59918%
	SPR	0.494447	0.072257	-17.59219%	14.61365%
	NDF[1]	0.404552	0.096144	-32.57473%	23.76560%
	NDF[2]	0.492611	0.073383	-17.89821%	14.89675%
0.7	CAT	0.583727	0.059638	-16.61038%	10.21678%
	SPR	0.583728	0.059611	-16.61022%	10.21213%
	NDF[1]	0.474676	0.089649	-32.18907%	18.88628%
	NDF[2]	0.581587	0.060231	-16.91615%	10.35626%
0.8	CAT	0.676015	0.049094	-15.49808%	7.26224%
	SPR	0.675968	0.049183	-15.50394%	7.27596%
	NDF[1]	0.548358	0.069284	-31.45524%	12.63485%
	NDF[2]	0.676022	0.049083	-15.49726%	7.26061%
0.9	CAT	0.790467	0.039927	-12.17030%	5.05105%
	SPR	0.790453	0.039853	-12.17191%	5.04181%
	NDF[1]	0.643531	0.053749	-28.49658%	8.35228%
	NDF[2]	0.790477	0.039911	-12.16918%	5.04901%

Table 49 - FL Descriptive Statistics - Bimodal Scale Distortion on 3 Categories

4 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.360987	0.072985	-9.75336%	20.21824%
	SPR	0.362329	0.071787	-9.41768%	19.81256%
	NDF¹	0.32881	0.102974	-17.79752%	31.31730%
	NDF²	0.355662	0.0819	-11.08450%	23.02736%
0.5	**CAT**	0.457881	0.058964	-8.42387%	12.87767%
	SPR	0.458522	0.058457	-8.29556%	12.74897%
	NDF¹	0.42049	0.081793	-15.90198%	19.45193%
	NDF²	0.448165	0.065697	-10.36709%	14.65919%
0.6	**CAT**	0.550027	0.045477	-8.32890%	8.26811%
	SPR	0.551352	0.044753	-8.10805%	8.11692%
	NDF¹	0.503136	0.065388	-16.14403%	12.99606%
	NDF²	0.541669	0.051769	-9.72184%	9.55739%
0.7	**CAT**	0.644827	0.038391	-7.88180%	5.95365%
	SPR	0.645982	0.03786	-7.71683%	5.86081%
	NDF¹	0.588865	0.052686	-15.87640%	8.94698%
	NDF²	0.633244	0.043642	-9.53654%	6.89184%
0.8	**CAT**	0.741532	0.030762	-7.30845%	4.14844%
	SPR	0.742451	0.030104	-7.19363%	4.05462%
	NDF¹	0.675718	0.041171	-15.53531%	6.09289%
	NDF²	0.742551	0.029846	-7.18113%	4.01937%
0.9	**CAT**	0.854733	0.020626	-5.02969%	2.41316%
	SPR	0.85508	0.020284	-4.99106%	2.37217%
	NDF¹	0.776582	0.030429	-13.71311%	3.91828%
	NDF²	0.853984	0.019851	-5.11292%	2.32450%

Table 50 - FL Descriptive Statistics - Bimodal Scale Distortion on 4 Categories

		5 Category			
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.355535	0.080968	-11.11629%	22.77363%
	SPR	0.362285	0.072606	-9.42874%	20.04122%
	NDF¹	0.32601	0.10664	-18.49759%	32.71062%
	NDF²	0.355289	0.082136	-11.17781%	23.11824%
0.5	**CAT**	0.45408	0.060726	-9.18391%	13.37330%
	SPR	0.459591	0.057612	-8.08179%	12.53559%
	NDF¹	0.414967	0.093892	-17.00660%	22.62630%
	NDF²	0.452989	0.06564	-9.40223%	14.49051%
0.6	**CAT**	0.543675	0.050914	-9.38744%	9.36486%
	SPR	0.551846	0.045981	-8.02561%	8.33213%
	NDF¹	0.504541	0.064642	-15.90981%	12.81201%
	NDF²	0.543456	0.052703	-9.42398%	9.69782%
0.7	**CAT**	0.638998	0.041231	-8.71461%	6.45248%
	SPR	0.647198	0.037462	-7.54317%	5.78839%
	NDF¹	0.590248	0.049728	-15.67888%	8.42486%
	NDF²	0.635592	0.042241	-9.20109%	6.64597%
0.8	**CAT**	0.735592	0.033245	-8.05099%	4.51943%
	SPR	0.743631	0.030128	-7.04608%	4.05145%
	NDF¹	0.677659	0.042748	-15.29262%	6.30813%
	NDF²	0.743775	0.029882	-7.02809%	4.01764%
0.9	**CAT**	0.850705	0.023301	-5.47721%	2.73907%
	SPR	0.856909	0.020109	-4.78791%	2.34667%
	NDF¹	0.779243	0.028593	-13.41744%	3.66932%
	NDF²	0.855784	0.019522	-4.91286%	2.28122%

Table 51 - FL Descriptive Statistics - Bimodal Scale Distortion on 5 Categories

6 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.365303	0.069374	-8.67414%	18.99092%
	SPR	0.372502	0.061489	-6.87456%	16.50709%
	NDF¹	0.352319	0.088838	-11.92028%	25.21529%
	NDF²	0.368566	0.071028	-7.85844%	19.27139%
0.5	**CAT**	0.46503	0.05373	-6.99404%	11.55419%
	SPR	0.471691	0.048606	-5.66175%	10.30467%
	NDF¹	0.44304	0.066135	-11.39196%	14.92759%
	NDF²	0.464288	0.056361	-7.14246%	12.13914%
0.6	**CAT**	0.556253	0.044278	-7.29115%	7.96003%
	SPR	0.564552	0.038889	-5.90806%	6.88843%
	NDF¹	0.530013	0.049939	-11.66449%	9.42221%
	NDF²	0.554793	0.046524	-7.53456%	8.38579%
0.7	**CAT**	0.653602	0.03553	-6.62833%	5.43608%
	SPR	0.662492	0.030684	-5.35834%	4.63160%
	NDF¹	0.623185	0.040988	-10.97358%	6.57712%
	NDF²	0.653994	0.035447	-6.57234%	5.42015%
0.8	**CAT**	0.751643	0.028258	-6.04460%	3.75954%
	SPR	0.76009	0.025034	-4.98881%	3.29358%
	NDF¹	0.713415	0.034767	-10.82315%	4.87336%
	NDF²	0.760803	0.024638	-4.89957%	3.23836%
0.9	**CAT**	0.864076	0.018769	-3.99153%	2.17209%
	SPR	0.869749	0.015614	-3.36122%	1.79527%
	NDF¹	0.81458	0.023914	-9.49106%	2.93571%
	NDF²	0.867962	0.015054	-3.55975%	1.73444%

Table 52 - FL Descriptive Statistics - Bimodal Scale Distortion on 6 Categories

		7 Category			
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.360006	0.077048	-9.99857%	21.40189%
	SPR	0.372362	0.062149	-6.90948%	16.69050%
	NDF¹	0.349519	0.091573	-12.62019%	26.19976%
	NDF²	0.367779	0.067185	-8.05534%	18.26771%
0.5	**CAT**	0.460153	0.056848	-7.96936%	12.35409%
	SPR	0.471923	0.048385	-5.61540%	10.25267%
	NDF¹	0.439248	0.078943	-12.15035%	17.97227%
	NDF²	0.464333	0.05424	-7.13339%	11.68130%
0.6	**CAT**	0.549685	0.048673	-8.38576%	8.85463%
	SPR	0.564528	0.039379	-5.91204%	6.97550%
	NDF¹	0.53026	0.053096	-11.62337%	10.01326%
	NDF²	0.558045	0.044808	-6.99254%	8.02953%
0.7	**CAT**	0.646681	0.038858	-7.61705%	6.00883%
	SPR	0.662669	0.03079	-5.33296%	4.64633%
	NDF¹	0.620149	0.043471	-11.40734%	7.00982%
	NDF²	0.655851	0.035714	-6.30701%	5.44539%
0.8	**CAT**	0.744186	0.030687	-6.97674%	4.12354%
	SPR	0.760163	0.025032	-4.97965%	3.29294%
	NDF¹	0.71181	0.036109	-11.02379%	5.07287%
	NDF²	0.760899	0.024624	-4.88765%	3.23624%
0.9	**CAT**	0.857679	0.021654	-4.70234%	2.52475%
	SPR	0.869899	0.015753	-3.34455%	1.81093%
	NDF¹	0.815411	0.023727	-9.39872%	2.90979%
	NDF²	0.868171	0.015072	-3.53650%	1.73606%

Table 53 - FL Descriptive Statistics - Bimodal Scale Distortion on 7 Categories

8 Category					
Factor Loading	**Method**	**μ**	**σ**	**RB**	**CV**
0.4	**CAT**	0.361537	0.073087	-9.61565%	20.21559%
	SPR	0.373788	0.059046	-6.55300%	15.79668%
	NDF¹	0.353738	0.079276	-11.56542%	22.41095%
	NDF²	0.370556	0.066897	-7.36101%	18.05318%
0.5	**CAT**	0.460687	0.058675	-7.86256%	12.73632%
	SPR	0.472491	0.04997	-5.50177%	10.57583%
	NDF¹	0.44724	0.065842	-10.55192%	14.72184%
	NDF²	0.466462	0.053812	-6.70769%	11.53624%
0.6	**CAT**	0.551462	0.048051	-8.08959%	8.71336%
	SPR	0.56636	0.03899	-5.60660%	6.88425%
	NDF¹	0.536307	0.053046	-10.61542%	9.89103%
	NDF²	0.559935	0.044118	-6.67743%	7.87905%
0.7	**CAT**	0.647939	0.038294	-7.43727%	5.91005%
	SPR	0.664369	0.02974	-5.09010%	4.47640%
	NDF¹	0.627118	0.040352	-10.41170%	6.43450%
	NDF²	0.658158	0.032321	-5.97749%	4.91083%
0.8	**CAT**	0.745267	0.031281	-6.84165%	4.19728%
	SPR	0.761481	0.025576	-4.81493%	3.35875%
	NDF¹	0.715413	0.034698	-10.57332%	4.85005%
	NDF²	0.762856	0.026144	-4.64302%	3.42708%
0.9	**CAT**	0.859334	0.021122	-4.51848%	2.45797%
	SPR	0.871745	0.015501	-3.13942%	1.77820%
	NDF¹	0.817853	0.022742	-9.12742%	2.78073%
	NDF²	0.870631	0.015223	-3.26327%	1.74855%

Table 54 - FL Descriptive Statistics - Bimodal Scale Distortion on 8 Categories

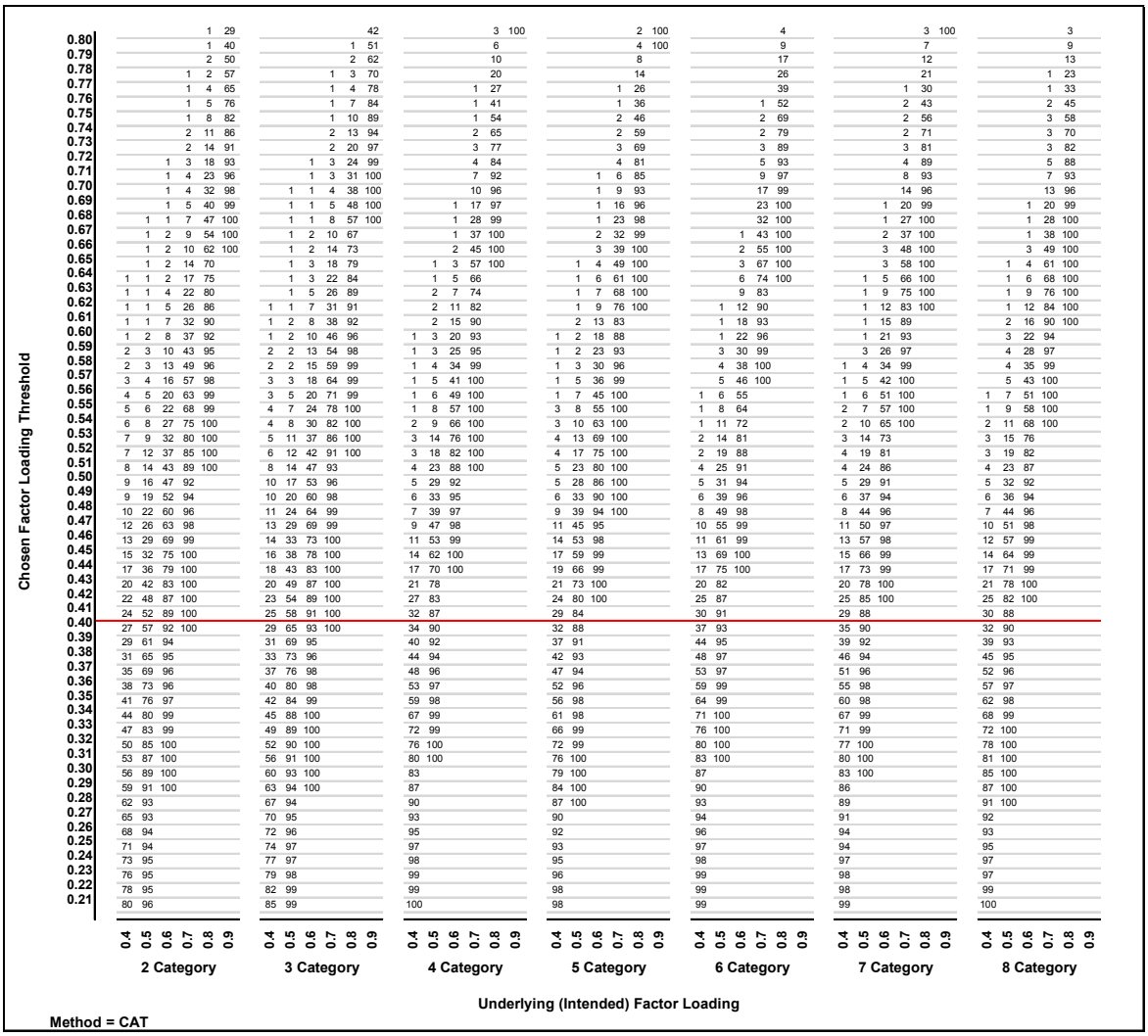
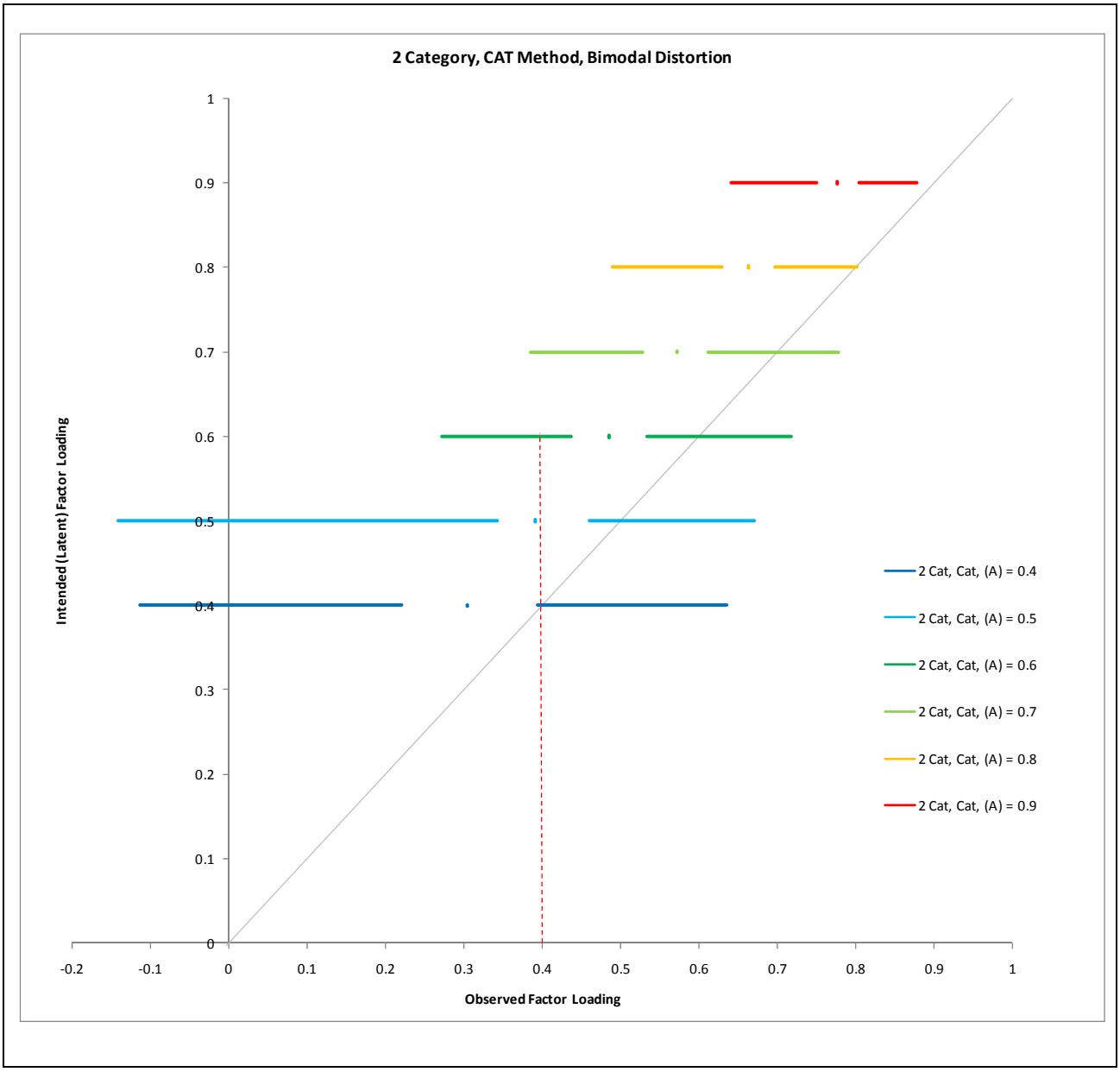
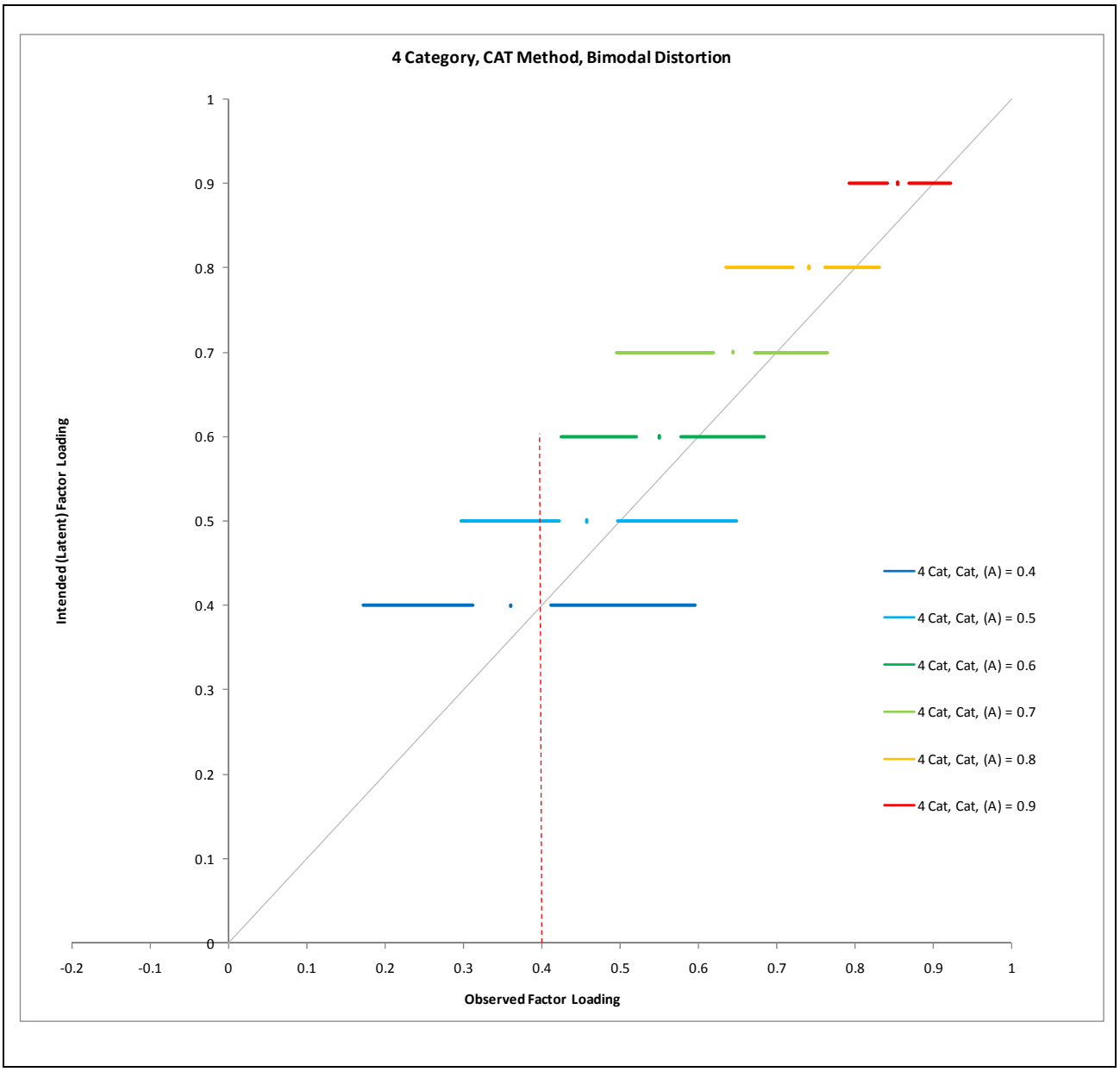
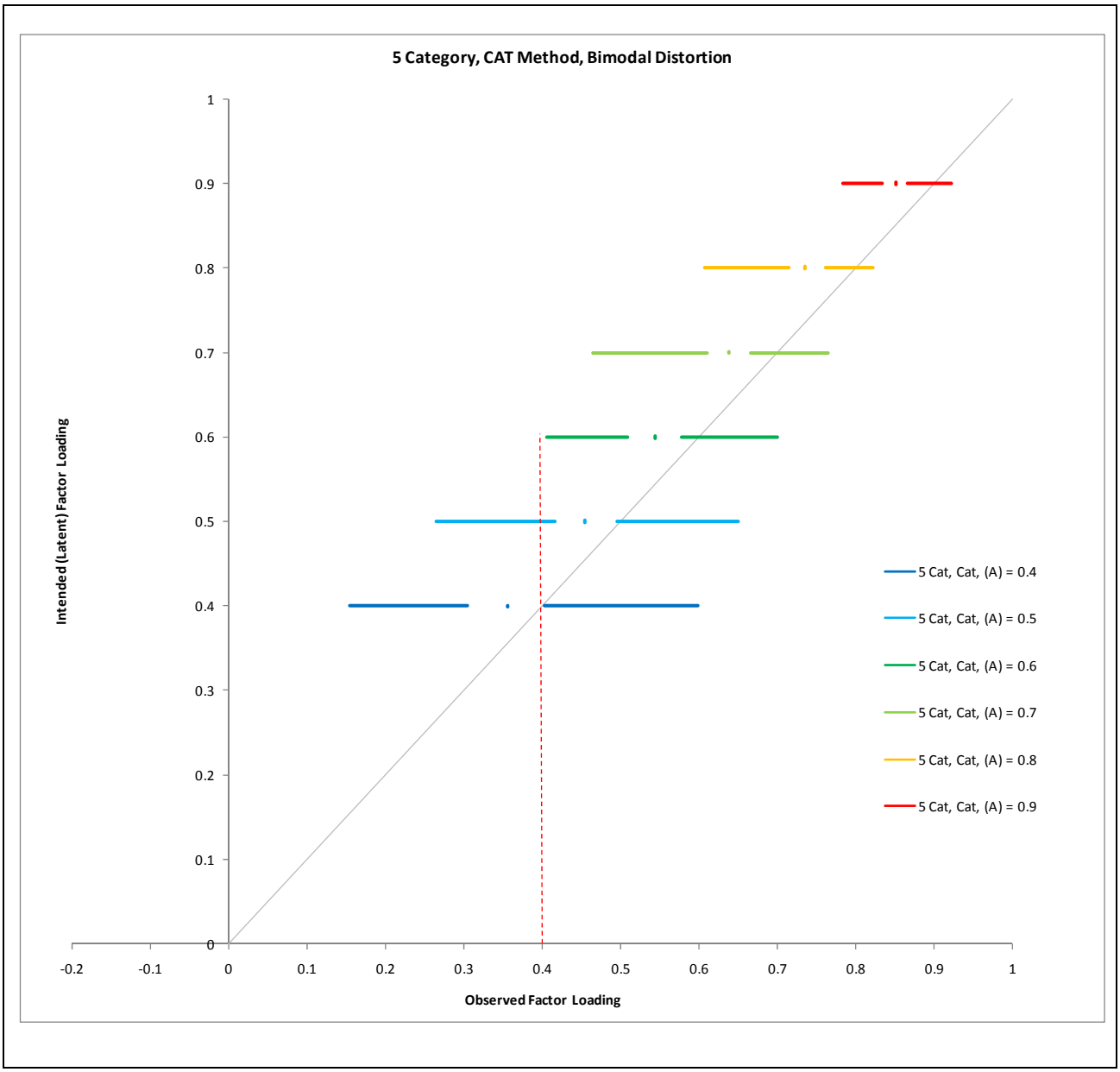
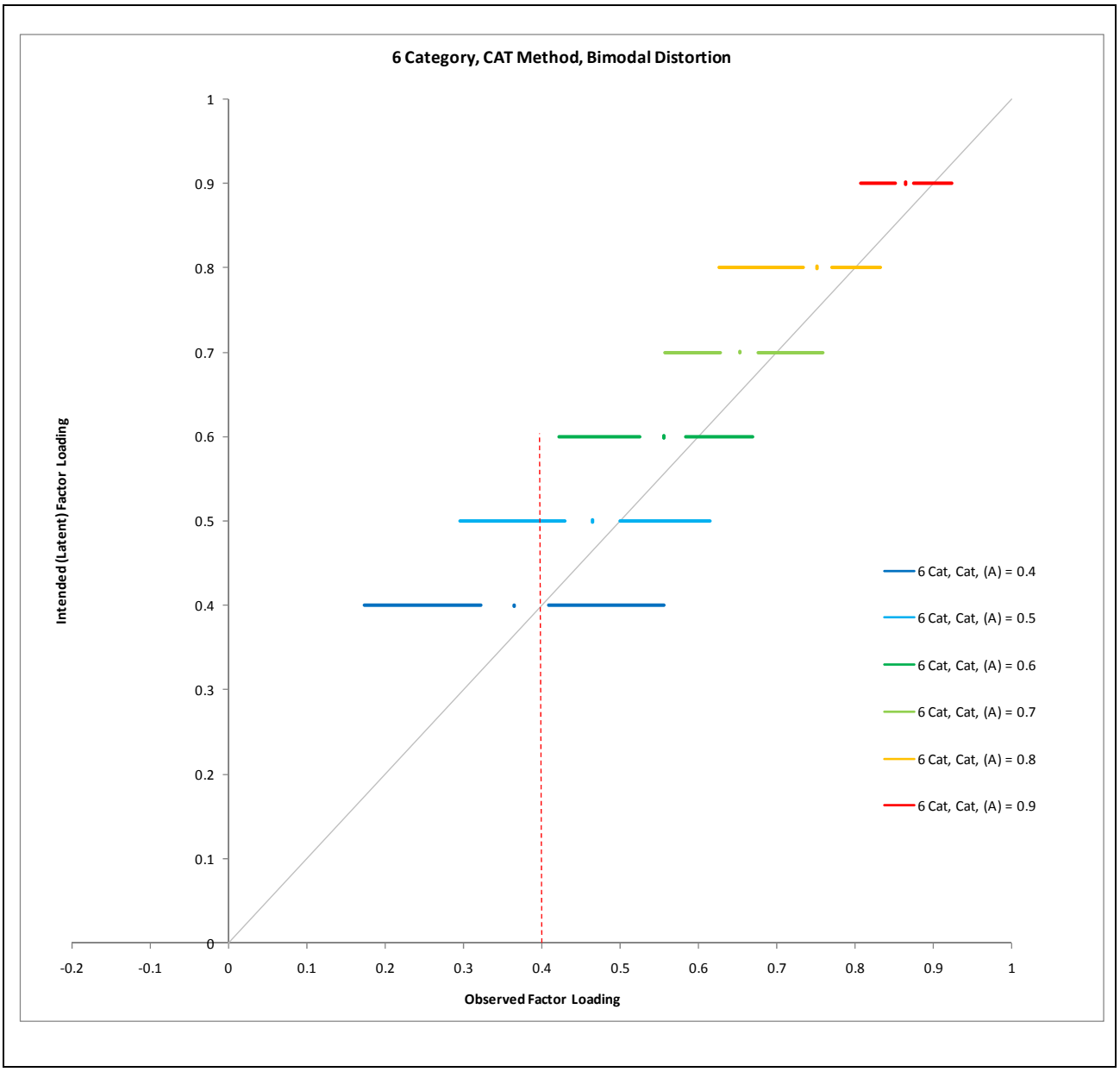


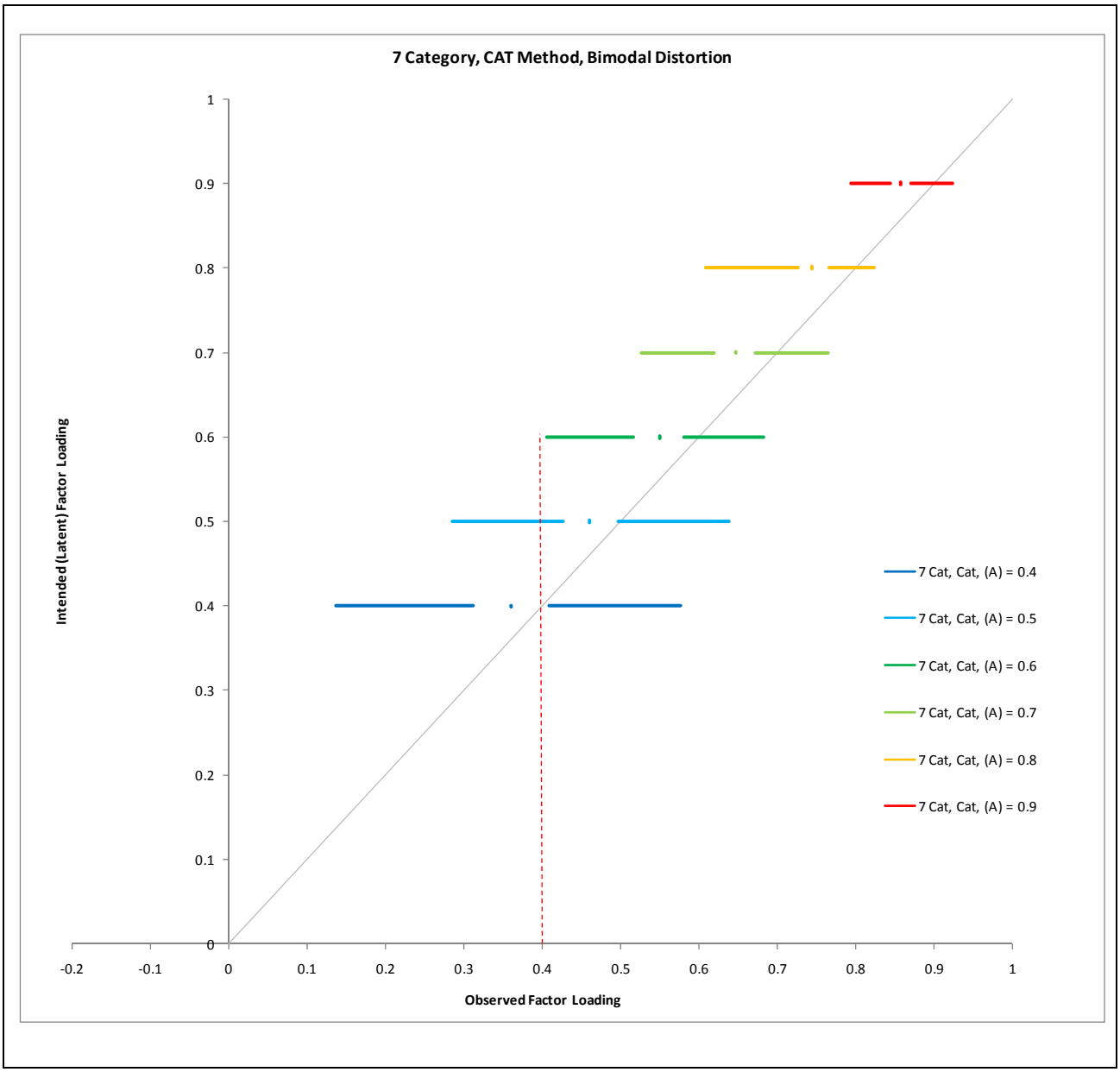
Figure 40 - Percentiles – CAT Method - Bimodal Distortion


2 Category, CAT Method, Bimodal Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2 Cat, Cat, (A) = 0.4
2 Cat, Cat, (A) = 0.5
2 Cat, Cat, (A) = 0.6
2 Cat, Cat, (A) = 0.7
2 Cat, Cat, (A) = 0.8
2 Cat, Cat, (A) = 0.9


4 Category, CAT Method, Bimodal Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
4 Cat, Cat, (A) = 0.4
4 Cat, Cat, (A) = 0.5
4 Cat, Cat, (A) = 0.6
4 Cat, Cat, (A) = 0.7
4 Cat, Cat, (A) = 0.8
4 Cat, Cat, (A) = 0.9


5 Category, CAT Method, Bimodal Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
5 Cat, Cat, (A) = 0.4
5 Cat, Cat, (A) = 0.5
5 Cat, Cat, (A) = 0.6
5 Cat, Cat, (A) = 0.7
5 Cat, Cat, (A) = 0.8
5 Cat, Cat, (A) = 0.9


6 Category, CAT Method, Bimodal Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
6 Cat, Cat, (A) = 0.4
6 Cat, Cat, (A) = 0.5
6 Cat, Cat, (A) = 0.6
6 Cat, Cat, (A) = 0.7
6 Cat, Cat, (A) = 0.8
6 Cat, Cat, (A) = 0.9


7 Category, CAT Method, Bimodal Distortion
Intended (Latent) Factor Loading
Observed Factor Loading
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
7 Cat, Cat, (A) = 0.4
7 Cat, Cat, (A) = 0.5
7 Cat, Cat, (A) = 0.6
7 Cat, Cat, (A) = 0.7
7 Cat, Cat, (A) = 0.8
7 Cat, Cat, (A) = 0.9

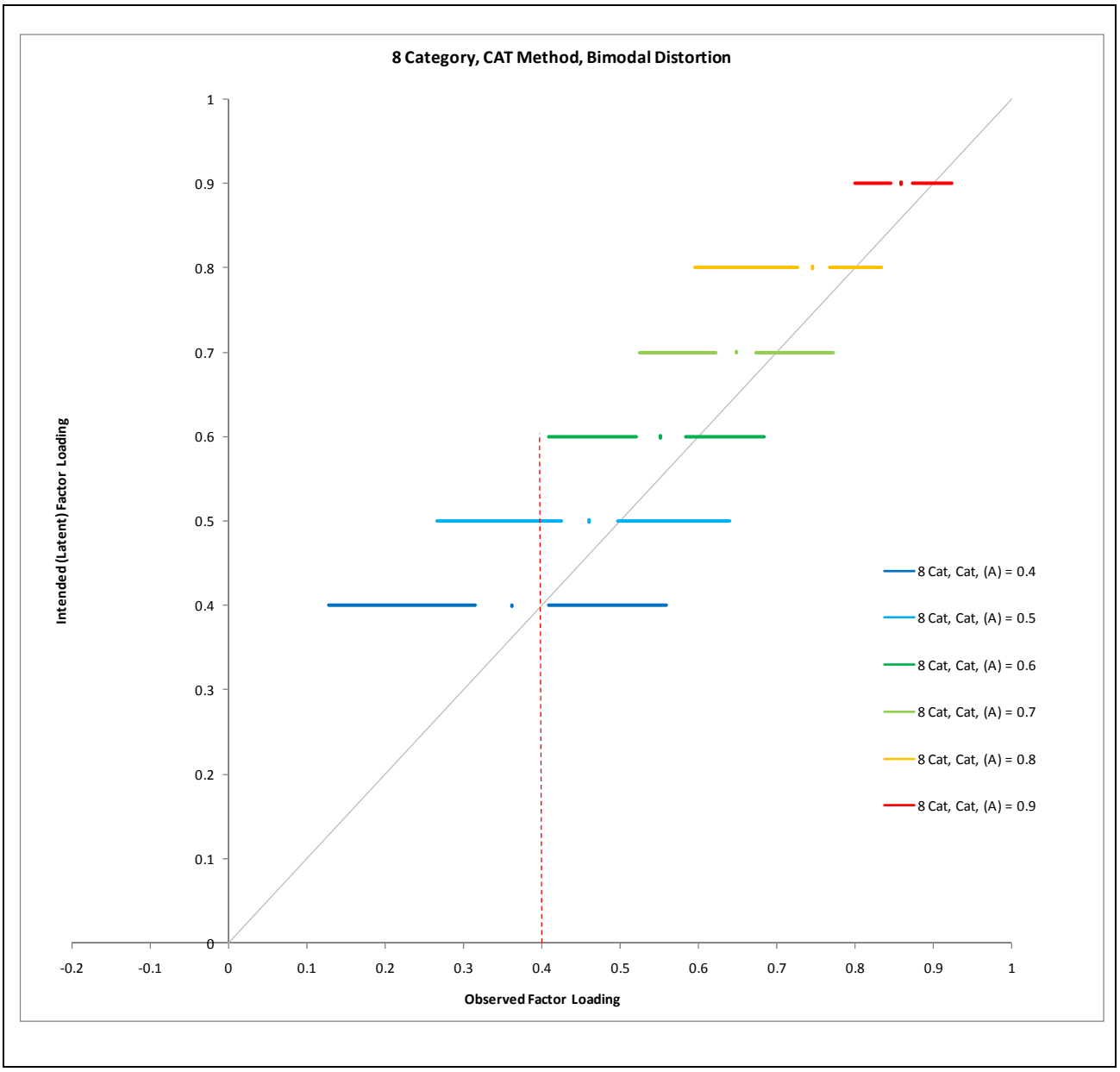


Figure 41 - Modified Box Plots Category 2, 4, 5, 6, 7 and 8 (Bimodal Distortion)

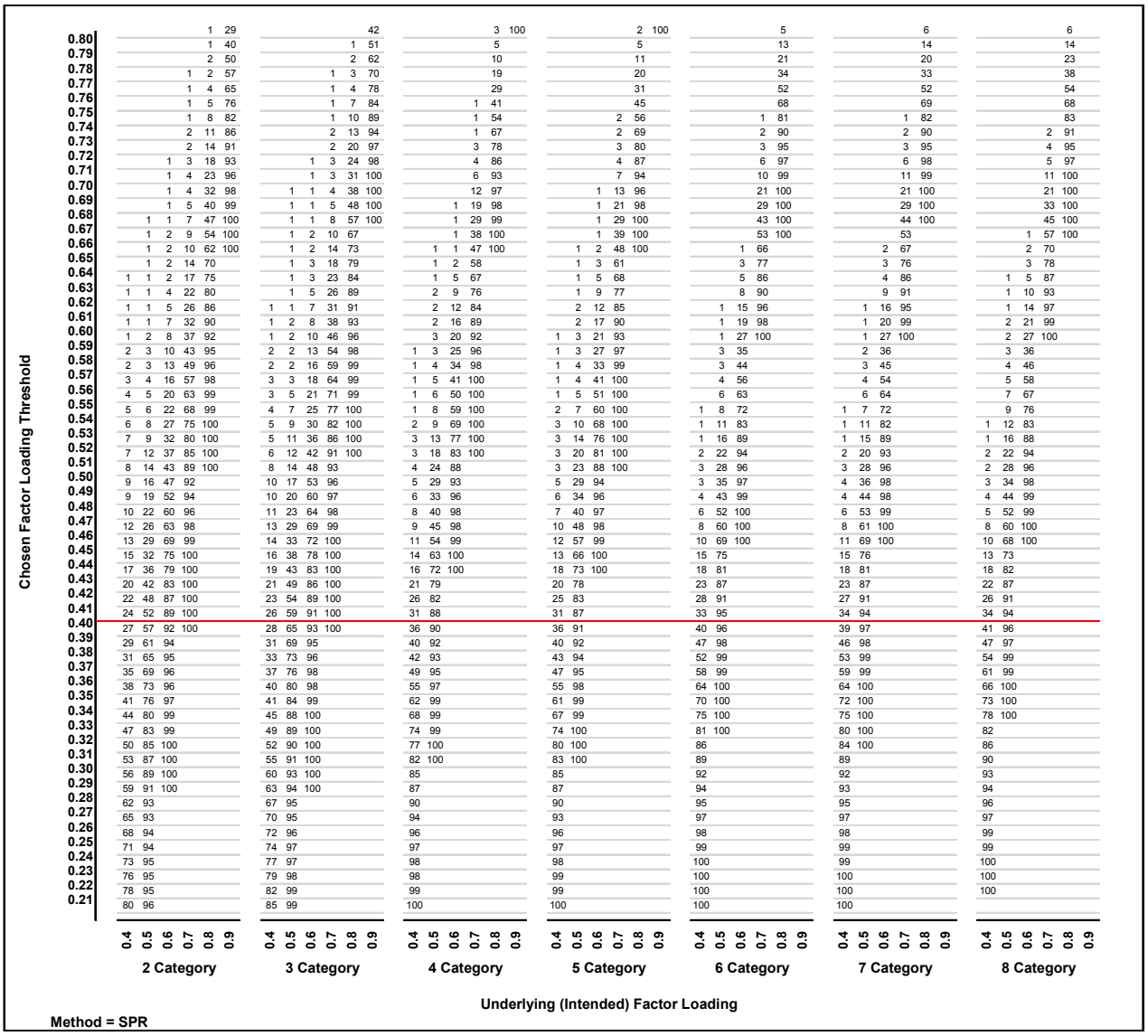


Figure 42 - Percentiles – SPR Method - Bimodal Distortion

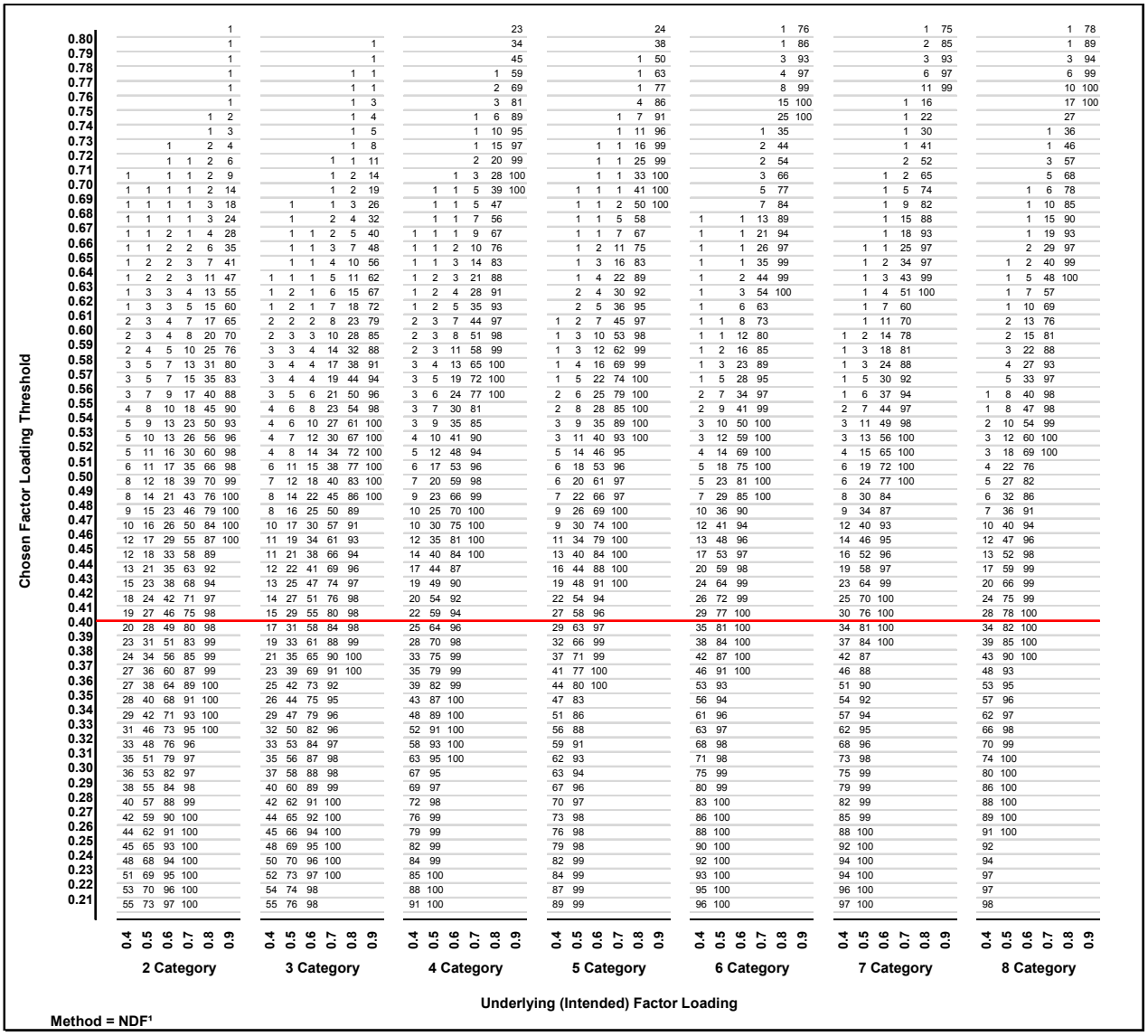


Figure 43 - Percentiles – NDF[1] Method - Bimodal Distortion

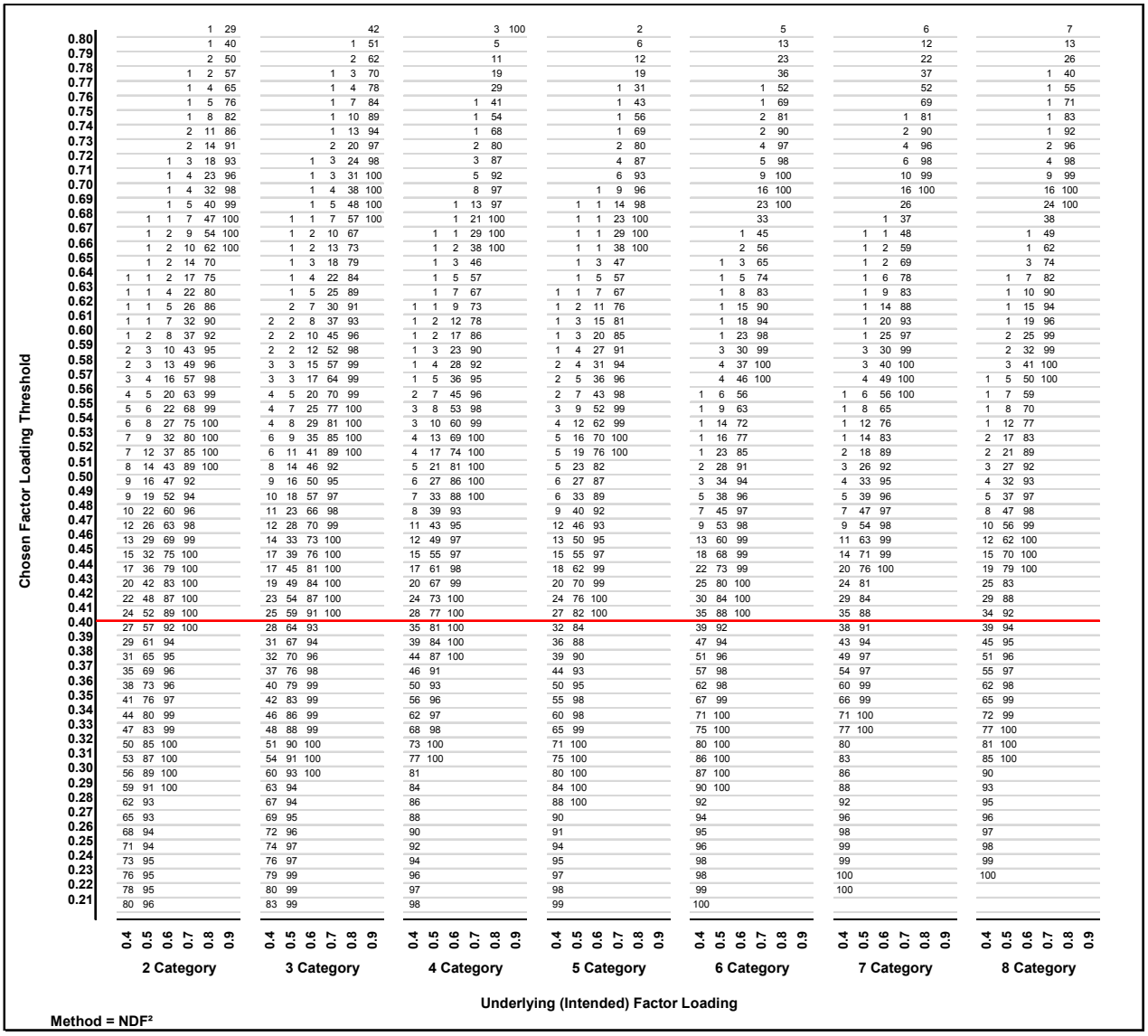


Figure 44 - Percentiles – NDF^2 Method - Bimodal Distortion

Relative Bias

RB		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	-23.7%	-20.6%	-9.8%	-11.1%	-8.7%	-10.0%	-9.6%
	SPR	-23.7%	-20.6%	-9.4%	-9.4%	-6.9%	-6.9%	-6.6%
	NDF[1]	-44.4%	-46.3%	-17.8%	-18.5%	-11.9%	-12.6%	-11.6%
	NDF[2]	-23.7%	-21.0%	-11.1%	-11.2%	-7.9%	-8.1%	-7.4%
0.5	CAT	-21.6%	-17.6%	-8.4%	-9.2%	-7.0%	-8.0%	-7.9%
	SPR	-21.6%	-17.6%	-8.3%	-8.1%	-5.7%	-5.6%	-5.5%
	NDF[1]	-42.8%	-41.1%	-15.9%	-17.0%	-11.4%	-12.2%	-10.6%
	NDF[2]	-21.6%	-17.9%	-10.4%	-9.4%	-7.1%	-7.1%	-6.7%
0.6	CAT	-19.0%	-17.6%	-8.3%	-9.4%	-7.3%	-8.4%	-8.1%
	SPR	-19.0%	-17.6%	-8.1%	-8.0%	-5.9%	-5.9%	-5.6%
	NDF[1]	-35.2%	-32.6%	-16.1%	-15.9%	-11.7%	-11.6%	-10.6%
	NDF[2]	-19.0%	-17.9%	-9.7%	-9.4%	-7.5%	-7.0%	-6.7%
0.7	CAT	-18.3%	-16.6%	-7.9%	-8.7%	-6.6%	-7.6%	-7.4%
	SPR	-18.3%	-16.6%	-7.7%	-7.5%	-5.4%	-5.3%	-5.1%
	NDF[1]	-34.2%	-32.2%	-15.9%	-15.7%	-11.0%	-11.4%	-10.4%
	NDF[2]	-18.3%	-16.9%	-9.5%	-9.2%	-6.6%	-6.3%	-6.0%
0.8	CAT	-17.0%	-15.5%	-7.3%	-8.1%	-6.0%	-7.0%	-6.8%
	SPR	-17.0%	-15.5%	-7.2%	-7.0%	-5.0%	-5.0%	-4.8%
	NDF[1]	-33.7%	-31.5%	-15.5%	-15.3%	-10.8%	-11.0%	-10.6%
	NDF[2]	-17.0%	-15.5%	-7.2%	-7.0%	-4.9%	-4.9%	-4.6%
0.9	CAT	-13.8%	-12.2%	-5.0%	-5.5%	-4.0%	-4.7%	-4.5%
	SPR	-13.8%	-12.2%	-5.0%	-4.8%	-3.4%	-3.3%	-3.1%
	NDF[1]	-30.9%	-28.5%	-13.7%	-13.4%	-9.5%	-9.4%	-9.1%
	NDF[2]	-13.8%	-12.2%	-5.1%	-4.9%	-3.6%	-3.5%	-3.3%

Figure 45 – Method RB Comparison - Bimodal Distortion

Coefficient of Variance

CV		Category						
Factor Loading	Method	2 CAT	3 CAT	4 CAT	5 CAT	6 CAT	7 CAT	8 CAT
0.4	CAT	43.7%	37.8%	20.2%	22.8%	19.0%	21.4%	20.2%
	SPR	43.7%	37.9%	19.8%	20.0%	16.5%	16.7%	15.8%
	NDF¹	83.3%	87.4%	31.3%	32.7%	25.2%	26.2%	22.4%
	NDF²	43.7%	38.7%	23.0%	23.1%	19.3%	18.3%	18.1%
0.5	CAT	29.0%	21.8%	12.9%	13.4%	11.6%	12.4%	12.7%
	SPR	29.0%	21.9%	12.7%	12.5%	10.3%	10.3%	10.6%
	NDF¹	65.7%	61.0%	19.5%	22.6%	14.9%	18.0%	14.7%
	NDF²	29.0%	21.9%	14.7%	14.5%	12.1%	11.7%	11.5%
0.6	CAT	15.0%	14.6%	8.3%	9.4%	8.0%	8.9%	8.7%
	SPR	15.0%	14.6%	8.1%	8.3%	6.9%	7.0%	6.9%
	NDF¹	29.3%	23.8%	13.0%	12.8%	9.4%	10.0%	9.9%
	NDF²	15.0%	14.9%	9.6%	9.7%	8.4%	8.0%	7.9%
0.7	CAT	10.9%	10.2%	6.0%	6.5%	5.4%	6.0%	5.9%
	SPR	10.9%	10.2%	5.9%	5.8%	4.6%	4.6%	4.5%
	NDF¹	19.7%	18.9%	8.9%	8.4%	6.6%	7.0%	6.4%
	NDF²	10.9%	10.4%	6.9%	6.6%	5.4%	5.4%	4.9%
0.8	CAT	7.9%	7.3%	4.1%	4.5%	3.8%	4.1%	4.2%
	SPR	7.9%	7.3%	4.1%	4.1%	3.3%	3.3%	3.4%
	NDF¹	14.1%	12.6%	6.1%	6.3%	4.9%	5.1%	4.9%
	NDF²	7.9%	7.3%	4.0%	4.0%	3.2%	3.2%	3.4%
0.9	CAT	5.4%	5.1%	2.4%	2.7%	2.2%	2.5%	2.5%
	SPR	5.4%	5.0%	2.4%	2.3%	1.8%	1.8%	1.8%
	NDF¹	9.7%	8.4%	3.9%	3.7%	2.9%	2.9%	2.8%
	NDF²	5.4%	5.0%	2.3%	2.3%	1.7%	1.7%	1.7%

Figure 46 – Method CV Comparison - Bimodal Distortion

Relative Bias Difference RB Diff		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	0.0000%	0.0000%	0.3357%	1.6875%	1.7996%	3.0891%	3.0627%
	SPR	0.0000%	0.0862%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	20.6568%	25.7054%	8.3798%	9.0688%	5.0457%	5.7107%	5.0124%
	NDF²	0.0000%	0.4373%	1.6668%	1.7491%	0.9839%	1.1459%	0.8080%
0.5	**CAT**	0.0000%	0.0346%	0.1283%	1.1021%	1.3323%	2.3540%	2.3608%
	SPR	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	21.1766%	23.5444%	7.6064%	8.9248%	5.7302%	6.5350%	5.0502%
	NDF²	0.0000%	0.2964%	2.0715%	1.3204%	1.4807%	1.5180%	1.2059%
0.6	**CAT**	0.0000%	0.0000%	0.2209%	1.3618%	1.3831%	2.4737%	2.4830%
	SPR	0.0000%	0.0048%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	16.1207%	14.9874%	8.0360%	7.8842%	5.7564%	5.7113%	5.0088%
	NDF²	0.0000%	0.3109%	1.6138%	1.3984%	1.6265%	1.0805%	1.0708%
0.7	**CAT**	0.0000%	0.0002%	0.1650%	1.1714%	1.2700%	2.2841%	2.3472%
	SPR	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	15.8443%	15.5789%	8.1596%	8.1357%	5.6152%	6.0744%	5.3216%
	NDF²	0.0000%	0.3059%	1.8197%	1.6579%	1.2140%	0.9741%	0.8874%
0.8	**CAT**	0.0000%	0.0008%	0.1273%	1.0229%	1.1450%	2.0891%	2.1986%
	SPR	0.0000%	0.0067%	0.0125%	0.0180%	0.0892%	0.0920%	0.1719%
	NDF¹	16.6364%	15.9580%	8.3542%	8.2645%	5.9236%	6.1361%	5.9303%
	NDF²	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
0.9	**CAT**	0.0000%	0.0011%	0.0386%	0.6893%	0.6303%	1.3578%	1.3791%
	SPR	0.0000%	0.0027%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	17.1055%	16.3274%	8.7220%	8.6295%	6.1298%	6.0542%	5.9880%
	NDF²	0.0000%	0.0000%	0.1219%	0.1250%	0.1985%	0.1919%	0.1239%
Overall Best Performer (Mean)	**CAT**	**0.0000%**	**0.0061%**	**0.1693%**	**1.1725%**	**1.2600%**	**2.2746%**	**2.3052%**
	SPR	**0.0000%**	**0.0167%**	**0.0021%**	**0.0030%**	**0.0149%**	**0.0153%**	**0.0287%**
	NDF¹	**17.9234%**	**18.6836%**	**8.2097%**	**8.4846%**	**5.7002%**	**6.0369%**	**5.3852%**
	NDF²	**0.0000%**	**0.2251%**	**1.2156%**	**1.0418%**	**0.9173%**	**0.8184%**	**0.6827%**

Figure 47 – Method RB % Underperformance - Bimodal Distortion

Coefficient of Variation Difference

CV Diff		Category						
Factor Loading	**Method**	**2 CAT**	**3 CAT**	**4 CAT**	**5 CAT**	**6 CAT**	**7 CAT**	**8 CAT**
0.4	**CAT**	0.0000%	0.0000%	0.4057%	2.7324%	2.4838%	4.7114%	4.4189%
	SPR	0.0000%	0.1035%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	39.6266%	49.5670%	11.5047%	12.6694%	8.7082%	9.5093%	6.6143%
	NDF²	0.0000%	0.8301%	3.2148%	3.0770%	2.7643%	1.5772%	2.2565%
0.5	**CAT**	0.0000%	0.0000%	0.1287%	0.8377%	1.2495%	2.1014%	2.1605%
	SPR	0.0000%	0.0145%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	36.6755%	39.2050%	6.7030%	10.0907%	4.6229%	7.7196%	4.1460%
	NDF²	0.0000%	0.0931%	1.9102%	1.9549%	1.8345%	1.4286%	0.9604%
0.6	**CAT**	0.0000%	0.0000%	0.1512%	1.0327%	1.0716%	1.8791%	1.8291%
	SPR	0.0000%	0.0145%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	14.2827%	9.1664%	4.8791%	4.4799%	2.5338%	3.0378%	3.0068%
	NDF²	0.0000%	0.2976%	1.4405%	1.3657%	1.4974%	1.0540%	0.9948%
0.7	**CAT**	0.0000%	0.0046%	0.0928%	0.6641%	0.8045%	1.3625%	1.4336%
	SPR	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
	NDF¹	8.7629%	8.6741%	3.0862%	2.6365%	1.9455%	2.3635%	1.9581%
	NDF²	0.0000%	0.1441%	1.0310%	0.8576%	0.7886%	0.7991%	0.4344%
0.8	**CAT**	0.0000%	0.0016%	0.1291%	0.5018%	0.5212%	0.8873%	0.8385%
	SPR	0.0000%	0.0153%	0.0352%	0.0338%	0.0552%	0.0567%	0.0000%
	NDF¹	6.1771%	5.3742%	2.0735%	2.2905%	1.6350%	1.8366%	1.4913%
	NDF²	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%	0.0683%
0.9	**CAT**	0.0000%	0.0092%	0.0887%	0.4578%	0.4377%	0.7887%	0.7094%
	SPR	0.0000%	0.0000%	0.0477%	0.0654%	0.0608%	0.0749%	0.0296%
	NDF¹	4.2823%	3.3105%	1.5938%	1.3881%	1.2013%	1.1737%	1.0322%
	NDF²	0.0000%	0.0072%	0.0000%	0.0000%	0.0000%	0.0000%	0.0000%
Overall Best Performer (Mean)	**CAT**	**0.0000%**	**0.0026%**	**0.1660%**	**1.0378%**	**1.0947%**	**1.9551%**	**1.8984%**
	SPR	**0.0000%**	**0.0246%**	**0.0138%**	**0.0165%**	**0.0193%**	**0.0219%**	**0.0049%**
	NDF¹	**18.3012%**	**19.2162%**	**4.9734%**	**5.5925%**	**3.4411%**	**4.2734%**	**3.0414%**
	NDF²	**0.0000%**	**0.2287%**	**1.2661%**	**1.2092%**	**1.1474%**	**0.8098%**	**0.7857%**

Figure 48 – Method CV % Underperformance - Bimodal Distortion