MODELLING UNFOLDING RESPONSE DATA
WITHIN THE STRUCTURAL EQUATION
MODELLING FRAMEWORK

LIM JIE XIN

SCHOOL OF HUMANITIES AND SOCIAL SCIENCES

2015
MODELLING UNFOLDING RESPONSE DATA
WITHIN THE STRUCTURAL EQUATION MODELLING FRAMEWORK

LIM JIE XIN

School of Humanities and Social Sciences

A thesis submitted to the Nanyang Technological University in partial fulfilment of the requirement for the degree of Master of Arts

2015
I would never have completed this thesis without the guidance and support of my supervisor, my friends and colleagues and family.

I would like to express my deepest gratitude to my supervisor, Dr. Ringo Ho, for his excellent guidance and patience. I would like to thank him for the support he has provided throughout the years and patiently reading and corrected my codes and writings. I am thankful to have him as my supervisor.

I would like to thank my friends and colleagues at the Graduate Students Office. Special thanks to Edward Halim and Fan Qianqian for proofreading earlier versions of the paper and for their feedbacks, Pony Chew for her advice on statistics, and Ma Sirui and Liu Hengshuang for the constant encouragement.

Lastly, I would also like to thank my parents and my brothers for their support and unconditional love.
## Contents

Acknowledgements ................................................................. ii

1 Introduction ........................................................................ 1
   1.1 Purpose and Scope of Study ........................................... 2
      1.1.1 Significance of Study ........................................... 2
   1.2 Overview of Thesis ....................................................... 3

2 Literature Review .............................................................. 5
   2.1 Review of Thurstone’s and Likert’s Scaling Methods ........... 6
      2.1.1 Thurstone’s Scaling Method .................................. 6
      2.1.2 Likert’s Scaling Method ...................................... 8
   2.2 Argument for Unfolding Response Process ......................... 11
   2.3 Empirical Evidence for Unfolding Response Process .......... 12
   2.4 A Review of Models for Unfolding Responses ................... 14
      2.4.1 Non-IRT Models .............................................. 14
      2.4.2 Parametric Item Response Theory (IRT) Models ....... 16
      2.4.3 Nonparametric IRT Models ................................ 21
      2.4.4 Common (Linear) Factor Model as Approximation .... 22
      2.4.5 Quadratic Factor Model .................................... 24
   2.5 Summary and Conclusion ............................................. 25

3 Proposed Model and Research Question ................................ 27
3.1 Proposed Model: OCRUM ........................................ 28
  3.1.1 The Effect of $t_j$ and $\tau_j \cdot c$ on the Response Functions .... 30
3.2 Estimation of Latent Interaction Effect .......................... 31
  3.2.1 Semi-parametric Approach ................................. 31
  3.2.2 Indicator-product Approach ............................ 35
  3.2.3 Bayesian Approach .................................. 35
  3.2.4 Distribution-analytic Approach ......................... 36
3.3 Aim and Scope of Research .................................. 37

4 Study 1: Monte Carlo Simulation ................................. 39
  4.1 Overview of Data Generation Procedures ...................... 40
    4.1.1 Choice of Thresholds Values ($\tau_c$) .................. 40
    4.1.2 Generation of Observed Responses ($Y_j$) ............... 40
  4.2 Implementation in Mplus ................................... 41
    4.2.1 Model Identification ................................ 42
    4.2.2 Model Fitting in Mplus ............................ 42
    4.2.3 Estimating Factor Scores in Mplus ...................... 43
  4.3 Overview of Analysis .................................... 43
    4.3.1 Parameter Recovery ................................ 43
    4.3.2 Assessing Model-Data Fit ............................ 44
  4.4 Study 1a ................................................. 46
    4.4.1 Method ............................................ 49
    4.4.2 Results ........................................... 50
    4.4.3 Summary .......................................... 65
  4.5 Study 1b ................................................. 66
    4.5.1 Method ............................................ 67
    4.5.2 Results ........................................... 68
    4.5.3 Summary .......................................... 71
5 Study 2: Re-analysis of Empirical Data 74
  5.1 General Method ................................................. 74
    5.1.1 Empirical Datasets ........................................ 74
    5.1.2 Starting Values in OCRUM .................................. 75
    5.1.3 GGUM2004 .................................................. 75
    5.1.4 Analysis .................................................... 75
  5.2 Empirical Dataset 1: Attitude Toward Capital Punishment Scale ..... 76
    5.2.1 Results ..................................................... 77
  5.3 Empirical Dataset 2: Order Scale ................................ 80
    5.3.1 Results ..................................................... 84
  5.4 Discussion ..................................................... 86

6 Conclusion 92
  6.1 Limitations of Model ........................................... 93
  6.2 Future Directions ............................................... 94

References 96

Appendix A MATLAB Threshold Generating Script 106

Appendix B Mplus OCRUM Syntax for Unidimensional Data 109

Appendix C MATLAB Study 1a Data Simulation Script 112

Appendix D Mplus Linear Factor Model Syntax 120

Appendix E MATLAB Study 1b Data Simulation Script 122

Appendix F Mplus OCRUM Syntax for Independent Cluster Pattern Data 134
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Item Location ($\delta_j$) Used in Simulation</td>
<td>48</td>
</tr>
<tr>
<td>4.2</td>
<td>Study 1a Successful Replications</td>
<td>51</td>
</tr>
<tr>
<td>4.3</td>
<td>Mean and Standard Deviation of Pearson Correlation between $\hat{\theta}$ and Simulated $\theta$</td>
<td>61</td>
</tr>
<tr>
<td>4.4</td>
<td>Mean and Standard Deviation of Spearman Rank Correlation between $\hat{\theta}$ and Simulated $\theta$</td>
<td>61</td>
</tr>
<tr>
<td>4.5</td>
<td>Effect of Independent Variables on Pearson Correlation between $\hat{\theta}$ and Simulated $\theta$</td>
<td>62</td>
</tr>
<tr>
<td>4.6</td>
<td>Effect of Independent Variables on Spearman Rank Correlation between $\hat{\theta}$ and Simulated $\theta$</td>
<td>62</td>
</tr>
<tr>
<td>4.7</td>
<td>Model Comparison using AIC</td>
<td>63</td>
</tr>
<tr>
<td>4.8</td>
<td>Model Comparison using BIC</td>
<td>64</td>
</tr>
<tr>
<td>4.9</td>
<td>Model Comparison using sBIC</td>
<td>64</td>
</tr>
<tr>
<td>4.10</td>
<td>Model Comparison using Likelihood-ratio Test</td>
<td>65</td>
</tr>
<tr>
<td>4.11</td>
<td>Study 1b Successful Replications</td>
<td>69</td>
</tr>
<tr>
<td>4.12</td>
<td>Mean and Standard Deviation of $\rho$ Parameter Bias</td>
<td>69</td>
</tr>
<tr>
<td>4.13</td>
<td>Effect of Independent Variables on Absolute $\rho$ Parameter Bias</td>
<td>70</td>
</tr>
<tr>
<td>4.14</td>
<td>Mean and Standard Deviation of $\rho$ Standard Error Bias of the Independent Clusters Model</td>
<td>71</td>
</tr>
<tr>
<td>4.15</td>
<td>Effect of Independent Variables on Absolute $\rho$ Standard Error Bias of the Independent Clusters Model</td>
<td>72</td>
</tr>
</tbody>
</table>
5.1 Thurstone’s (1932) Attitude Toward Capital Punishment 78
5.2 Pairwise correlation of the item locations for the Attitude Toward Capital Punishment scale 79
5.3 Chernyshenko et al.’s (2007) Order Scale Polytomous Response Endorsement Rate 84
5.4 Chernyshenko et al.’s (2007) Order Scale 85
5.5 Pairwise correlation of the item locations for the Order scale 86
List of Figures

2.1 An example of the Unfolding Response Function ........................................ 9
2.2 An example of the Dominance Response Function ....................................... 11

3.1 Relationship between latent response, thresholds, and observed response 30
3.2 Response function of a hypothetical dichotomous item, holding $\tau_{j,c}$ constant 32
3.3 Response function of a hypothetical dichotomous item, holding $t_j$ constant 32
3.4 Response function of a hypothetical polytomous item, holding $\tau_{j,c}$ constant 33
3.5 Response function of a hypothetical polytomous item, holding $t_j$ constant 34

4.1 $\delta_j$ parameter bias for test length ($J$) = 5 ........................................ 53
4.2 $\delta_j$ parameter bias for test length ($J$) = 10 ........................................ 54
4.3 $\delta_j$ parameter bias for test length ($J$) = 20 ........................................ 55
4.4 $\delta_j$ standard error bias for test length ($J$) = 5 ........................................ 56
4.5 $\delta_j$ standard error bias for test length ($J$) = 10 ...................................... 57
4.6 $\delta_j$ standard error bias for test length ($J$) = 20 ...................................... 58

5.1 Theoretical response function of item CP06 ........................................ 81
5.2 Theoretical response function of item CP18 ........................................ 82
5.3 Relationship between OCRUM and GGUM $\hat{\theta}$ for the Attitude Toward Capital Punishment Scale .......................................................... 83
5.4 Theoretical response function of item ORD23 ........................................ 87
5.5 Theoretical response function of item ORD37 ........................................ 88
5.6 Relationship between OCRUM and GGUM $\hat{\theta}$ for the Ideal Point IRT Order Scale


89
Abstract

The dominance and the unfolding response mechanisms have been proposed to describe the way individuals respond to items. Methods to model the unfolding response data have been proposed from various frameworks. However, these methods are less acquainted by researchers in the social sciences and some require specialised software for implementation. This thesis addressed this issue by proposing to model the unfolding response data within the general SEM framework through Mplus which is widely used by applied researchers. Two simulation studies were designed to examine the proposed model. Study 1a studied the effects of sample size, test length, item locations, and response options on the parameter recovery and model-data fit using simulated data. Study 1b investigated the ability of the model and the estimation method implemented in Mplus in recovering the correlation between factors. Two empirical datasets were also analysed to illustrate the application of the proposed method (OCRUM) and also to compare the performance of OCRUM with the Generalised Graded Unfolding Model (GGUM), one of the widely used methods for the analysis of unfolding response data. This thesis concluded with limitation of the proposed model and areas for future research were identified.
Chapter 1

Introduction

Two response mechanisms have been proposed to explain how individuals respond to survey items: the dominance and the unfolding response mechanism. However, existing methods of analysing the response data resulted from the unfolding response mechanism (referred to as unfolding response data) are less acquainted by social scientists in general and some of the methods have relatively less modelling flexibility in terms of model expansion. This study proposed a model and its implementation in the Structural Equation Modelling (SEM) framework that is a general modelling framework commonly used by social scientists to address these issues.

The unfolding response process assumes that endorsement (i.e. agreement) on an item has an inverted relationship with the distance between the individual and the item (i.e. smaller the distance, higher the endorsement and vice versa; Thurstone, 1928). Mathematically, this process assumes a single-peaked response function (Andrich, 1996; Roberts, Laughlin, & Wedell, 1999; Stark, Chernyshenko, Drasgow, & Williams, 2006; Thurstone & Chave, 1929).

On the theoretical level, the unfolding response process would most likely to be observed in tests of typical behaviour, such as personality and attitude (Tay, Drasgow, Rounds, & Williams, 2009). In such tests, individuals are usually instructed to rate statements on the extend of them in describing the individuals. Such instruction would likely to elicit an introspection in the respondents where they compare themselves with
the statement. Hence, an item is only endorsed when it is a close description of the person and vice versa. Analysis of empirical response data on attitudes and personality also showed evidence of the single-peaked response function (e.g. Chernyshenko, Stark, Chan, Drasgow, & Williams, 2001; Roberts et al., 1999; Stark et al., 2006; Tay et al., 2009).

Methods have been proposed to model the unfolding response data from various frameworks such as the parametric and non-parametric Item Response Theory (IRT) framework, and the multidimensional scaling framework. However, these methods are less acquainted by researchers in the social science and require specialised software for implementation.

1.1 Purpose and Scope of Study

This thesis proposes to model the unfolding response data in the general SEM framework which is commonly used by applied researchers. We first proposed the mathematical model (Ordered Categorical Response Unfolding Model; OCRUM) and discussed its implementation in Mplus version 7.2, one of the widely used SEM software. The feasibility and performance of the proposed method was investigated through two simulation studies and two sets of empirical response data.

The two simulation studies examined the effects of sample size, test length, item locations, and response options on the parameter recovery and model-data fit. The analysis of empirical response data aimed to illustrate the application of the proposed method (OCRUM) and also to compare the performance of OCRUM with the Generalised Graded Unfolding Model (GGUM), which is one of the widely used methods for the analysis of unfolding response data.

1.1.1 Significance of Study

By leveraging on the flexibility of the SEM framework, this thesis contributes to the SEM methodology by allowing more modelling possibilities involving the proposed
model (OCRUM). By flexibility, we refer to the ease of model modification and extension. Relative to other frameworks, models are easier to specify and modify in the SEM framework.

Furthermore, the question whether a set of data conform to either dominance or unfolding response process remains an empirical question to be addressed. If the proposed model is feasible, we could approach this problem from a model comparison perspective. By fitting the data on both a dominance response model (e.g. linear common factor model) and an unfolding response model (e.g. OCRUM), it is possible to then examine the model fit to determine which response process is more likely to be underlying the set of data. If a dominance response model fits the data better than the unfolding response model, we can infer from the fit statistics that the dominance response process is more suitable than the unfolding response process, and vice versa, to characterise a given data.

1.2 Overview of Thesis

The modelling of unfolding response data within the Structural Equation Modelling (SEM) framework has not been explored extensively in the literature. In Chapter 2, the theoretical postulation and empirical evidence for the unfolding response mechanism are presented. We also review some existing methods to model unfolding response data and some of the potential drawbacks of each of the method.

In Chapter 3, we introduce our proposed model, the Ordered Categorical Response Unfolding Model (OCRUM), and describe the estimation procedure.

In Chapter 4, we report the method and results of two simulation studies which provided some insights into our research question on the feasibility of the model and the estimation method in modelling simulated data by investigating the parameter recovery. The performance of various fit-statistics were also investigated and reported.

In Chapter 5, we compare the performance of the proposed model with the Generalised Graded Unfolding Model (GGUM) in terms of the similarities and differences in the estimated item and person location using two sets of empirical response data.
In Chapter 6, we summarise the findings in Chapter 4 and Chapter 5 and we identify the major limitations of this thesis and OCRUM in analysing empirical data. Finally, we conclude with an outline of suggestions for future studies.
Chapter 2

Literature Review

The use of rating scale (e.g. to measure the degree of agreement) is ubiquitous for non-cognitive measures such as attitude and personality, where respondents are presented with explicit statements (i.e. items) and are instructed to rate the extent of accuracy of the items in describing the respondents. A notable example of this method is the NEO-PI-R (Costa & McCrae, 1992), a well-known measure of the five domains of personality, where individuals are instructed to rate each statement with a rating scale consisting of five response options: "strongly disagree", "disagree", "neutral", "agree", and "strongly agree".

The development of this method can be traced back to Thurstone’s (1928) and Likert’s (1932) pioneering work on attitude measurement (Krosnick, Judd, & Wittenbrink, 2005). One of the primary differences in the development was the assumption made regarding the response process which was stated explicitly by Thurstone and implied by Likert. Assumption on the response processes is an important issue as it determines how items are selected in the scale construction process and also the computation of scores from the responses (Stark et al., 2006). Furthermore, it is critical to understand the response process and to utilise appropriate statistical models which conform to the assumptions of the response process to analyse the data (Reise, 2010).
2.1 Review of Thurstone’s and Likert’s Scaling Methods

Scaling methods, in general, are used to either scale individuals, items, or both (McIver & Carmines, 1981). Thurstone’s (1928) and Likert’s (1932) methods are two different types of scaling method. Thurstone’s method scales both the items and individuals on a psychological continuum. It consists of a two-step process that first scales the items along a psychological continuum and then scales respondents on the same continuum. Likert’s (1932) method was proposed as an alternative to Thurstone’s method to alleviate the need of item scaling. Essentially, Likert’s method is a subject-centered method which only scales the individuals (Torgerson, 1958; as cited in McIver & Carmines, 1981).

2.1.1 Thurstone’s Scaling Method

Thurstone’s method has its roots in psychophysical scaling. Psychophysical scaling was developed to investigate the relationship between physical stimuli and the psychological judgement of the stimuli (Himmelfarb, 1993). In a series of papers, Thurstone and his colleagues (e.g. Thurstone, 1927, 1928, 1929; Thurstone & Chave, 1929) extended the method of psychophysical scaling to the measurement of attitude.

The process of Thurstone’s method commences with writing an item pool consisting of negative, positive, and also neutral (intermediate) statements that reflect a particular psychological construct (e.g. attitude towards capital punishment). A group of judges will be instructed to scale the items on the corresponding psychological continuum through one of the several methods: method of equal-appearing intervals (Thurstone & Chave, 1929), method of successive intervals (Saffir, 1937), or method of paired-comparisons (Thurstone, 1927). The process of item scaling will conclude with a scale value for each item that represents the location of the items on the psychological continuum.

To illustrate the process of item scaling, using the method of equal-appearing interval as an example, judges are instructed to place each items into one of a number of ordered intervals (usually 11 intervals), with one extreme representing negative valance and the other extreme representing positive valance, while the middle represents neutrality.
Judges are to sort the items into the piles according to the content of the item, but not their own opinion. The scale value (item location) of each item is calculated using the median of the interval scores assigned by the judges (Thurstone & Chave, 1929).

Once all the statements from the item pool have been scaled, selected items can be administered to individuals using a binary (e.g. agree-disagree) response format. The individual can be scaled on the psychological continuum by taking the mean or the median of the scale values of the items that the individual endorsed. Similar to item location, the location of the individuals on the continuum is known as the person location (person’s ideal point).

Two methods to eliminate inadequately written items were suggested in Thurstone and Chave (1929) using data from the item scaling procedure (particularly the interval methods) and person scaling procedure, respectively. An item which has a large variability in the interval scores assigned by the judges is considered as an ambiguous item (e.g. an item judged as positive by some, and neutral or negative by others). Thurstone and Chave (1929) quantified item ambiguity using the Q-value (i.e. inter-quartile range) statistics, with larger inter-quartile range indicating higher ambiguity.

An item that does not differentiate or discriminate individuals along the psychological continuum is referred to as an irrelevant item (Thurstone & Chave, 1929). For example, if an item depicting a neutral attitude towards capital punishment is endorsed by individuals with positive and negative attitude towards capital punishment, this item should be eliminated for failing to discriminate individuals with different attitudes. The concept of irrelevant items in Thurstone’s method is based on the assumption regarding the underlying response process. Thurstone (1928) assumed that endorsement of an item has an inverted relationship with the distance between the individual and the item: A individual will endorse all or most of the items that are located close to the individual person location, and do not endorse items which are located further away. Hence, it is expected that individuals with neutral stance towards capital punishment are more likely to endorse a neutral item than individuals with strong positive or negative positions about capital punishment.
This response process, which was coined as the *unfolding* or *ideal-point response process* by Coombs (1964), implies a single-peaked response function (Andrich, 1996; Roberts et al., 1999; Stark et al., 2006; Thurstone & Chave, 1929). The basic tenets of the *unfolding response process* are that individuals only endorse items which are located closer to their person location on the same latent psychological continuum, and that items are not endorsed because the individuals are located further above or below the items (Coombs, 1964; Thurstone, 1928).

To illustrate, when presented with an item describing moderate conscientiousness (e.g. I write notes to myself only if I have too many things to do at once; an example taken from Chernyshenko, Stark, Drasgow, & Roberts, 2007), both individuals with extremely high and low degree of conscientiousness will not endorse this item because their locations are further away from the item’s location. By the characteristics of Conscientiousness as a personality trait (e.g. Chernyshenko et al., 2007; Weiner & Greene, 2008), a person with high conscientiousness tend to be organised and neat (Chernyshenko et al., 2007), which is in contrast to a low conscientious person. In relation to the example item given above, a high conscientious person will always write notes regardless of the amount of tasks at hand, while a low conscientious person will almost never write notes, therefore according to the *unfolding response process*, we would expect that both such individuals will not endorse a moderate item. In contrast, an individual with a moderate degree of consciousness, whose location is closer to the item, will endorse the item. When the endorsement of this item is plotted against the person location (on the X-axis), the theoretical non-monotonic single-peaked response function implied by the *unfolding response process* will resemble that of shown in figure 2.1.

### 2.1.2 Likert’s Scaling Method

Likert (1932) developed an alternative procedure to Thurstone’s method in scaling individuals to relinquish the need for item scaling. Even though Likert (1932) did not provide a mathematical model for his scaling method, psychometricians have justified
his method within the framework of classical test theory after his initial proposal (McIver & Carmines, 1981; Roberts et al., 1999). In general, Likert’s method can be described as having its roots in psychometrics, that is, the study of mental and psychological testing (Himmelfarb, 1993).

Similar to Thurstone’s method, an item pool consisting of items reflecting a psychological construct is developed. However, unlike Thurstone’s requirement of having items on all the varying locations along the psychological continuum, items written under Likert’s method only consist of positive and negative items, while neutral items are avoided (Roberts et al., 1999). Neutral items are likely to be doubled-barrelled (e.g. I don’t believe in capital punishment but I’m not sure it isn’t necessary) which Likert (1932) argued would not be effective to differentiate individuals with different person locations as individuals from both ends of the psychological continuum (i.e. individuals with starkly different attitude) may give the same rating on the item.

Each item are then presented to the respondents with instructions to express their
opinion of the item with a rating scale (e.g. strongly disagree, disagree, neutral/undecided, agree, strongly agree). Each rating point is assigned an integer, usually from 1 to 5 for a five-point scale, with the strongest agreement receiving a value of 5. The ratings on each items are then summed up for each individual to represent their location on the latent psychological continuum.

Likert (1932) also proposed to examine the corrected item-total correlation (i.e. correlation between a target item with the total scores computed excluding the target item) as a means to eliminate ambiguous and non-discriminating item. Likert (1932) proposed that item with zero or low corrected item-total correlation indicates an undifferentiating item (i.e. an item which does not measure the same psychological construct as the other items in the scale).

Likert’s argument regarding the effectiveness of neutral items, the procedure of item analysis (e.g item-total correlation), and person scaling are based on the assumption of the dominance response process, which implies a cumulative (monotonically increasing) response function (Coombs, 1964), as opposed to the single-peaked response function as suggested by Thurstone (1927). The basic idea of the dominance response process is that individuals are more likely to endorse items which are located below their person location and not endorse items which are located above their person location (Coombs, 1964).

The dominance response process can be illustrated clearly in the context of cognitive testing. When presented with a mathematics item (e.g. \( \int_{1}^{2} \frac{1}{x^2+1} \, dx \)), individuals with higher mathematical ability (i.e. higher person location) will have a higher likelihood of getting this item correct than individuals with lower mathematical ability (i.e. lower person location). In the context of cognitive testing, the item location can be interpreted as the difficulty of the item. In the context of personality measurement, the dominance response process suggests that an extreme conscientious person (i.e. a person who is extremely organised) will endorse an item measuring moderate degree of conscientious (e.g. I arrange my books by their sizes) more than a person with moderate or low conscientiousness. When the endorsement is plotted against the person location (on the
X-axis), the theoretical monotonic increasing curve implied by the *dominance response process* will resemble that of shown in figure 2.2.

![Graph](image)

Figure 2.2: An example of the *Dominance* response function where the expected response of agreement increases as respondents latent trait increases. Note that the specific form of the expected response depends on the fitted mathematical model.

### 2.2 Argument for Unfolding Response Process

As detailed in the previous section, two different underlying response processes have been proposed in the literature: the *unfolding response process* and the *dominance response process*. In fact, Thurstone and Chave (1929) acknowledged that both response processes are plausible depending on the scale or measurement. From their perspective, the occurrence of *unfolding response process* and the *dominance response process* is a product of the scale construction. They believed that a scale could be constructed to induce either one of the response processes, but construct could be better measured if "the scale is intentionally constructed so that a person is more likely to [endorse] the [items] at one part of the scale than at any other part" (Thurstone & Chave, 1929, p.80), implying their preference for the unfolding response process.

In contrast to Thurstone and Chave (1929), contemporary researchers have proposed...
that the test of maximal behaviour and typical behaviour (using a self-report format) would prompt different response processes (Tay et al., 2009) and the unfolding response process is more likely to occur when individuals respond to items in test of typical behaviour such as as attitude and personality (e.g. Andrich, 1996; Drasgow, Chernyshenko, & Stark, 2010; Roberts, 1995; Stark et al., 2006).

Test of maximal behaviour, such as mental or physical ability testing, aims to measure a person’s maximum capacity. Respondents in a test of ability would be motivated to get an item correct, and items which are below the person’s location (i.e. relatively easy items) would have higher probability of being answered correctly, whereas items located above the person’s location (i.e. relatively difficult items) would have lower probability of being answered correctly. Hence a dominance response process is most likely to be observed in this type of tests where an objective or socially correct answer is present (Tay et al., 2009).

On the other hand, in tests of typical behaviour, such as personality or attitude, individuals are typically asked to rate the extend of the statement in describing themselves. This is likely to elicit an introspection in the respondents in which people compare the item with themselves to determine whether the description of the items matches the perception of themselves. Therefore, it is highly likely for people to only endorse items which are closer to their location on the latent trait continuum (Drasgow et al., 2010; Stark et al., 2006; Tay et al., 2009). This response process is consistent with the tenet of the unfolding response process.

2.3 Empirical Evidence for Unfolding Response Process

A review of literature on the examination of the relationship between trait and item endorsement in empirical response data from several personality and vocational interest measurements showed that some items in the scales were better described by the unfolding response process. For an instance, Roberts et al. (1999) visually inspected the response function of 10 items measuring attitudes toward abortion and found that items measuring
moderate attitude showed a single-peaked function. However, this pattern was less
evident for more negative and positive items. In addition to attitude, investigation of
responses on personality scales also supported the notion of the *unfolding response
process*. Chernyshenko et al. (2001) found that as a whole, some of the scales in the
Fifth Edition of the Sixteen Personality Factor Questionnaire (16PF; Conn & Rieke,
1994) and Big Five Factor markers (Goldberg, 1997) did not fit dominance response
models (i.e. models that assume a dominance response process) well. Additionally,
the visual inspection of response function of some items also showed non-monotonic
pattern. Nevertheless, they did not attempt to compare the fit with the unfolding response
models (i.e. models which assume an unfolding response process).

Stark et al. (2006) addressed this limitation by fitting the item responses of the Fifth
Edition of the Sixteen Personality Factor Questionnaire (16PF; Conn & Rieke, 1994) to
both dominance and unfolding response models. Their findings echoed Chernyshenko
et al. (2001) that some scales fitted the dominance response model while some fitted the
unfolding response model better. Though some items exclusively display a monotonic
response function, a number of items in some of the scales showed the response function
as assumed by the unfolding response process (i.e. single-peaked response function).
Using similar approach as Stark et al. (2006), Tay et al. (2009) also found that unfolding
item response models fitted the response function across three vocational interest inventories
better than dominance response models according to the chi-square goodness-of-fit statistics
and graphical fit plots.

These findings have given some empirical evidences to support the theoretical speculations
as illustrated in the previous section that item responses on non-cognitive measures are
likely to display the response function as predicted by the unfolding response process.
The next section introduces the various models which researchers have utilised to model
unfolding data.
2.4 A Review of Models for Unfolding Responses

Unfolding response models, which refer to the mathematical models describing the unfolding response function in the response data, have only been established in the last few decades. Although Thurstone (1928) described the theoretical response function corresponding to the unfolding response process, it was not proposed as a response model formally.

The unidimensional unfolding model was later introduced by Coombs (1964). However, his model was devised to analyse paired-comparison data or preferential choice data, which generally involves the ranking of items from least to most preferred. This thesis concerns only with modelling the data derived from the method of direct self-report where respondents are presented with items one at a time and are instructed to rate the extent of each of the item in describing themselves (either with dichotomous or polytomous response options). In the following section, methods and models for analysing unfolding responses were reviewed.

2.4.1 Non-IRT Models

Contemporary development of the unfolding response models have their roots in the item response theory (IRT) framework, which in essence, is a class of mathematical models which defines the probabilistic relationship between the endorsements (responses), the individuals, and the items characteristics (de Ayala, 2009). However, methods and models to analyse unfolding response data not based on the IRT framework have been developed as well.

Multidimensional Unfolding Model

The multidimensional unfolding model was developed to address the limitation of unidimensionality of Coombs’ (1964) unfolding model (McIver & Carmines, 1981). Being an extension of the unidimensional unfolding model, the multidimensional unfolding model was developed to analyse paired-comparison and preferential choice data within
the multidimensional scaling (MDS) framework. However, as posited in Javaras (2004), the response data from a direct self-report method can be treated as ranking of the items with ties to fit into the MDS framework.

MDS is a collection of data analysis techniques that aims to represent a set of stimuli (e.g. items and/or individuals) as coordinates in a multidimensional space so that similar stimuli are placed closer to each other while dissimilar stimuli are located further apart. By the virtue of MDS, both individuals and the items are represented as points in the multidimensional space. The distance between the individual and each item has an inverse relationship with their preference to the items, such that a smaller distance between person and an item represents a higher preference to the item (Takane, 2007), thus implying the response function assumed by the unfolding response process (i.e. a single-peaked function).

As noted by Busing (2010), the multidimensional unfolding technique has not been used much due to the technical problem of producing degenerative solution which refers to the statistical artefact of a perfect "data-model fit", as quantified using the least squares loss function, albeit a non-interpretable unfolding solution.

**Correspondence Analysis**

Recently, the Correspondence Analysis (CA) with row principal normalisation was proposed as an alternative to the Unfolding IRT models (Polak, Heiser, & de Rooij, 2009). Row principal normalisation was preferred over other normalisation (e.g. column principal, symmetrical normalisation) because interpretation of the CA solution would conform to the essence of the unfolding response process: a smaller individual-item distance in the CA solution corresponds to a higher rating on the item by that individual (Polak et al., 2009).

CA was initially developed to analyse contingency table (crosstabulation) data, however, it is also suitable for other types of data provided that the elements in the table represents strength of association between the rows and the columns (e.g. person by item response data; Polak, 2011). CA is essentially a special application of MDS on categorical data.
using the chi-squared distance (Everitt, 2005), therefore, similar to MDS, a CA solution entails representing the individuals and items on a multidimensional space. However, unlike MDS, a variant of CA, known as the constrained (canonical) correspondence analysis (CCA), has the ability to include predicting variables in the analysis by constraining the individuals’ location (ideal point) to be a linear combination of the predictors (Polak, 2011).

Simulation studies showed that CA was able recover both the ordering of the person and item location as indicated by high (more than .90) Pearson and Spearman rank correlation between the simulated (true population) and estimated values in all the simulated conditions, especially when the items were distributed equidistantly along the unidimensional latent continuum as compared to when items were clustered at the two ends of the latent continuum (Polak, 2011). However, the performance of CA is unknown for data with more than one underlying latent traits (i.e. multidimensional data).

Tangent to the development of MDS as unfolding technique, IRT models have been developed to model the unfolding data. The following sub-section introduces the IRT approach of modelling unfolding response data which focuses on modelling the probability of endorsing the response options as a function of the item characteristics and person location (i.e. latent trait).

### 2.4.2 Parametric Item Response Theory (IRT) Models

As introduced in the previous section, IRT models attempt to describe the relationship between individual, items, and observed data (endorsement on items) mathematically (de Ayala, 2009). A parametric IRT model is one that assumes a specific form (shape) of the item (or categorical) response function (Sijtsma, 1998) and when the response functions are constructed based on interpretable parameters (Maydeu-Olivares, 2005).

#### Simple Squared Logistic and Hyperbolic Cosine Model

Integrating the probabilistic concept of the item response model framework and the unfolding mechanism, Andrich (1988) developed the Simple Squared Logistic Model
(SSLM) to analyse dichotomous data collected through the direct response method. The SSLM is expressed as

\[ P(Y_{ij} = y | \theta_i) = \frac{\exp[-y(\theta_i - \delta_j)^2]}{\gamma_{ij}} \]  

(2.1)

where

\[ \gamma_{ij} = 1 + \exp[-y(\theta_i - \delta_j)^2], \]

\( Y_{ij} \) is the observed response on item \( j \) from person \( i \), and

\( y = \{0, 1\} \), with \( y = 0 \) representing disagreement, and \( y = 1 \) representing agreement,

\( \theta_i \) is the person location of person \( i \), and

\( \delta_j \) is the item location of item \( j \).

The single-peaked function of agreeing an item is expressed as the squared difference (or distance) of the individual’s and the item’s location. A limitation of this model is that the maximum probability can only reach .5, which may not be appropriate empirically.

Recalled that according to the principle of the unfolding response process, there are two possible reasons as to why an individual do not endorse (disagreeing) an item: 1) the individual is located below the item (disagree from below; DB), and 2) the individual is located above the item (disagree from above; DA). Instead of modelling the descriptive form of the response function using the traditional squared difference expression, Andrich and Luo (1993) developed the Hyperbolic Cosine Model (HCM) to explicitly model the probability of endorsement, DB and DA for dichotomous response data. The sum of the probability of latent responses of DB and DA is represented as the probability of the observed non-endorsement in the data. Based on the Rating Scale IRT model (Andrich, 1978, 1979), the HCM takes the form of

\[ P(Y_{ij} = 0 | \theta_i) = \frac{2 \cosh(\theta_i - \delta_j)}{\gamma_{ij}} \]

\[ P(Y_{ij} = 1 | \theta_i) = \frac{\exp(-\tau_{1ij})}{\gamma_{ij}} \]  

(2.2)

where
\( \cosh() \) is the hyperbolic cosine function,

\[
\gamma_{ij} = \exp(-\tau_{1j}) + 2 \cosh(\theta_i - \delta_j),
\]

\( Y_{ij} \) is the observed response on item \( j \) from person \( i \).

\( Y_{ij} \) can take on two values: \( \{0, 1\} \), 0 representing disagreement, and 1 representing agreement.

\( \theta_i \) is the person location of person \( i \),

\( \delta_j \) is the item location of item \( j \), and

\( \tau_{1j} \) represents the threshold for item \( j \).

The HCM was later generalised as the general hyperbolic cosine model (GHCM) to model polytomous data by Andrich (1996), which is defined as

\[
P(Y_{ij} = y; y < m | \theta_i) = \frac{\exp(-\sum_{k=0}^{y-1} \tau_{jk}) \cosh((m - y)(\theta_i - \delta_j))}{\gamma_{ij}},
\]

\[
P(Y_{ij} = m | \theta_i) = \frac{\exp(-\sum_{k=0}^{m} \tau_{jk})}{\gamma_{ij}}
\]

where

\[
\gamma_{ij} = \sum_{k=0}^{m-1} \left[ \exp(-\sum_{k=0}^{y} \tau_{jk}) \cosh((m - k)(\theta_i - \delta_j)) + \exp(-\sum_{k=0}^{m} \tau_{jk}) \right],
\]

\( Y_{ij} \) is the observed response for item \( j \) and subject \( i \), and

\( y = \{0, 1, ..., m\} \), with \( y = 0 \) representing the strongest level of disagreement, and \( y = m \) representing the strongest level of agreement,

\( \theta_i \) is the person location of person \( i \),

\( \delta_j \) is the item location of item \( j \), and

\( \tau_{jk} \) is the relative location of response category \( k \) of item \( j \).

When \( m = 1 \) (in the case of dichotomous rating scale), GHCM (equation 2.3) is reduced to HCM (equation 2.2). For detailed description of HCM and GHCM, readers can refer to Andrich and Luo (1993) and Andrich (1996).

**Generalised Graded Unfolding Model**

Similar to GHCM (Andrich, 1996), both Graded Unfolding Model (GUM; Roberts & Laughlin, 1996) and Generalised Graded Unfolding Model (GGUM; Roberts, Donoghue,
model the response function of the observed responses as a function of the distance of the person and item location and with the underlying latent responses. However, a difference between GHCM and both GUM and GGUM is on the modelling of the (categorical) response function of the strongest level of agreement: GHCM assumes a single latent response for this response option, however, GUM and GGUM assumes two latent response for this response option (i.e. strongly agree from below and from above) as well as for all other response options.

GGUM is based on Muraki’s (1992) generalised partial credit model (GPCM) for its generality, but Roberts et al. (2000) acknowledged that other models could be used instead. GGUM is defined as

\[
P(Y_{ij} = y | \theta_i) = \frac{\exp\{\alpha_j[y(\theta_i - \delta_j) - \sum_{k=0}^{y} \tau_{jk}]\} + \exp\{\alpha_j[(S_j - y)(\theta_i - \delta_j) - \sum_{k=0}^{\omega} \tau_{jk}]\} \gamma_{ij}}{\gamma_{ij}} \tag{2.4}
\]

where

\[
\gamma_{ij} = \sum_{\omega=0}^{m_j} (\exp\{\alpha_j[\omega(\theta_i - \delta_j) - \sum_{k=0}^{\omega} \tau_{jk}]\} + \exp\{\alpha_j[(S_j - \omega)(\theta_i - \delta_j) - \sum_{k=0}^{\omega} \tau_{jk}]\})
\]

\(Y_{ij}\) is the observed response for item \(j\) and subject \(i\), and \(y = \{0, 1, ..., m\}\), with \(y = 0\) representing the strongest level of disagreement, and \(m\) representing the strongest level of agreement,

\(S_j = 2m_j + 1\),

\(\theta_i\) is the person location of person \(i\),

\(\delta_j\) is the item location of item \(j\),

\(\alpha_j\) is the discrimination parameter of item \(j\), and

\(\tau_{jk}\) is the relative location of response category \(k\) of item \(j\).

GUM is a special case of GGUM, such that when all the \(\alpha_j\) are fixed as 1 and \(\tau_{jk}\) constrained to be equal across items, the model is reduced to GUM.
Normal PDF Model

The common limitation of the HCM, GHCM, GUM, and GGUM reviewed previously is on the strict assumption of unidimensionality of the response data. To address this limitation, a multidimensional unfolding IRT model was proposed to model multidimensionality in dichotomous data (Maydeu-Olivares, Hernández, & McDonald, 2006). Multidimensionality refers to the observation that responses to an item is dependent on more than one latent trait. Hence, instead of a single ideal point (location) for item or person, if a two-dimensional space was found to fit the data the best, there would be an ideal line (Maydeu-Olivares et al., 2006).

The Normal PDF model uses the normal probability density function (PDF) as a link function to model the item response function and to model the density of the person locations (latent traits; Maydeu-Olivares et al., 2006). The model is based on the normal ogive model (e.g. McDonald, 1997, as cited in Maydeu-Olivares et al., 2006), however, instead of using a standard normal cumulative distribution function as a link function, the link function is replaced by a normal PDF. To be in line with the discussion of previous unidimensional models, the unidimensional form of the Normal PDF model can expressed as

\[ P(Y_{ij} = 1|\theta_i) = \exp\left(-\frac{(\alpha_j + \beta_j \theta_i)^2}{2}\right) \]  

(2.5)

where

- \(\alpha_j\) is the intercept (constrained to be negative) of item \(j\),
- \(\beta_j\) is the loading (slope) of item \(j\), and
- \(\theta_i\) is the person location of person \(i\).

Equation 2.5 is readily generalised to a multidimensional (i.e. items measuring more than one dimension or factor) model. Interested readers can refer to Maydeu-Olivares et al. (2006) for a more detailed description of this model.
2.4.3 Nonparametric IRT Models

In contrast to the parametric IRT models, the nonparametric models do not assume a specific form of the item response function (Sijtsma, 1998). Due to the non-parametric nature, only the ordering (ranking) of the items and person can be established, but it is more flexible from a practical standpoint when researchers do not wish to assume a particular form for the response function (Sijtsma, 1998).

Multiple Unidimensional Unfolding Model

The Multiple Unidimensional Unfolding (MUDFOLD) model belongs to a class of nonparametric models. MUDFOLD assumes a single-peaked response function, but unlike parametric IRT unfolding models, it does not assume a specific shape for the response function. By virtue of a nonparametric model, only the order of the items and person can be estimated; this is different from parametric models such as GGUM where the item and person location are metric (i.e. distance between items and person can be estimated; Van Schuur & Post, 1998).

The purpose of a MUDFOLD analysis is to form an unfolding scale consisting of a subset of $j^*$ items (where $j^* \leq j$) from a set of $j$ items. Given that the unfolding response process is true, the observed responses of three adjacent items: \{1, 0, 1\} (assuming the items are measured on a dichotomous format) would be considered as a violation to the assumption of the response process, and this response pattern would be considered as an observed error pattern $O$. The iterative process of forming the unfolding scale is by maximizing the H-coefficient as calculated as $H = 1 - \frac{O}{E}$, where $E$ is the expected error pattern. Iteratively, an item is included into the scale if the individual item H-coefficient is statistically different from 0, and is above a pre-specified H-coefficient threshold (usually .30), and if the item is able to increase the overall H-coefficient of the scale (Polak, 2011; Van Schuur, 1992; Van Schuur & Post, 1998).
Maximum Likelihood Formula Scoring Model with Ideal Point Constraint

The Maximum Likelihood Formula Scoring (MFS) model construct its response function by using a linear combination of orthogonal functions (e.g. polynomials, trigonometric Levine, 1984). The general formulation of the MFS model is

$$P(y_{ij} = m|\theta_i) = \sum_k \alpha_{jkm} h_k(\theta_i)$$

(2.6)

where

$\alpha_{jkm}$ is the weight for item $j$, function $k$, and response option $m$, and

$h_k$ are orthogonal functions, such as orthogonal polynomials and trigonometric functions.

The flexibility of the formation of response function allows the model to form a variety of shape depending on the number and type of orthogonal function used. Inequality constraints on the first and second order derivatives of the orthogonal functions can be imposed to specify the shape of the response functions. To model the unfolding response process, response option of "Agree" can be constrained to be bell shaped in dichotomous data (e.g. Stark et al., 2006) or a bell shape on the middle option and monotonicity constraints on the other options for polytomous data (e.g. Chernyshenko et al., 2001).

2.4.4 Common (Linear) Factor Model as Approximation

It has been found that the unfolding response function may not be evident for extreme positive and negative items because of insufficient observations to the right of (i.e. above) the extreme positive items and to the left of (i.e. below) the negative items (Stark et al., 2006). As a result, the estimated empirical response functions of these extreme items, which are likely to be monotonically increasing or decreasing, might be able to be approximated well using dominance response model (Stark et al., 2006), such as the linear common factor model.

For example, most contemporary psychological scales were developed and scored based on some form of Likert’s (1932) procedure (i.e. items with negative valence
were first reversed and ratings were summed to calculate scores for each respondent). Consider the three items measuring Attitude towards Censorship (Rosander & Thurstone, 1931):

1. I doubt if censorship is wise.
2. Censorship is a very difficult problem and I am not sure how far I think it should go.
3. Everything that is printed for publication should first be examined by government censors.

It is likely that item 2 presented in the list above would be omitted from the final measurement scale because it represents an ambivalent attitude towards censorship which is impossible to score using Likert’s method. Hence, the use of Likert’s methods would result in a scale with extreme items (e.g. item 1 and 3) which reflect a clear positive or negative position (Andrich, 1996). Therefore, this enable the use of linear factor model to approximate the response function, even though unfolding might be the underlying response mechanism.

A unidimensional common factor model can be expressed as

\[ Y_{ij} = \mu_j + \lambda_j \theta_i + \varepsilon_{ij} \]  \hspace{1cm} (2.7)

where \( Y_{ij} \) is the responses on item \( j \) (\( j = 1, 2, 3, ..., m \)) for person \( i \) (\( i = 1, 2, 3, ..., p \)), \( \mu_j \) is the intercept for item \( j \), \( \lambda_j \) is the factor loading for item \( j \), and \( \theta_i \) is the common factor, and \( \varepsilon_{ij} \) is the measurement error of item \( j \) for person \( i \).

Hubert (2006, as cited in Stark et al., 2006) argued that the dominance response model is a special case of the unfolding response model when the item locations are allowed to be located at the extreme ends of the latent (construct) continuum. However, there may be situations where this approximation is not optimal. As shown by research by Chernyshenko et al. (2001), Stark et al. (2006), and Tay et al. (2009), even though
contemporary personality and interests scales had been developed using Likert’s (1932) procedures, some items were still found to display the single-peaked response functions.

Moreover, it has been documented that factor analysis of unidimensional unfolding response data which include items from across the latent continuum would erroneously produce a two-dimensional solution. This observation has been termed the extra-factor phenomenon (Davison, 1977; Maraun & Rossi, 2001; Tay & Drasgow, 2012; Van Schuur & Kiers, 1994).

### 2.4.5 Quadratic Factor Model

In explaining the extra-factor phenomenon in their paper, Maraun and Rossi (2001) has showed that the unidimensional quadratic factor model is mathematically equivalent to a unidimensional unfolding model, suggesting the application of a quadratic factor model to model unfolding response data.

A quadratic factor model can be expressed as

\[
Y_{ij} = a_j + \alpha_j \theta_i + \beta_j \theta_i^2 + \varepsilon_{ij} \tag{2.8}
\]

where

- \(Y_{ij}\) is the observed response for item \(j\) for person \(i\),
- \(a_j\) is the intercept of \(Y_{ij}\),
- \(\alpha_j\) is the factor loading for the linear component,
- \(\beta_j\) is the factor loading for the quadratic component, and
- \(\varepsilon_{ij}\) is the residual.

Equation 2.8 is also equivalent to a unidimensional unfolding model as shown by Maraun and Rossi (2001), which can be expressed as

\[
Y_{ij} = a_j + t_j (\theta_i - \delta_j)^2 + \varepsilon_{ij} \tag{2.9}
\]

\(^{1}\)Maraun and Rossi (2001) has showed that a uni-dimensional unfolding model is mathematically equivalent to a uni-dimensional quadratic factor model, which is not distinguishable from a two-dimensional linear factor model at the covariance structure level (McDonald, 1967). Therefore, application of linear factor analysis on unidimensional unfolding data produces a two-dimensional factor solution.
where

\[ Y_{ij} \] is the observed response for item \( j \) for person \( i \).

\( a_j \) is the intercept of \( Y_{ij} \),

\( t_j \) is the curvature parameter describing the slope’s rate of change,

\( \theta_i \) is the person location of person \( i \),

\( \delta_j \) is the item location of item \( j \), and

\( \epsilon_{ij} \) is the residual.

As shown in equation 2.9, if \( t_j \) is a negative value, it essentially expresses the unfolding response process which is characterised by a single-peaked function: the smaller the difference between the person location \( \theta_i \) and item location \( \delta_j \), the larger the observed response \( (Y_{ij}; \text{higher endorsement on item } j \text{ for person } i) \).

Even though Maraun and Rossi (2001) suggested that an unfolding model can be re-parameterised as a quadratic factor model, estimation of parameters were not discussed in their paper. As shown in later sections, the model proposed in this thesis was primarily built upon Maraun and Rossi’s (2001) work. Interested readers are directed to refer to Maraun and Rossi (2001) for the mathematical proof of the equivalence of equation 2.8 and equation 2.9.

### 2.5 Summary and Conclusion

The list of models briefly reviewed in the previous sections is not exhaustive. They have their strengths and limitations in modelling unfolding response data. MDS, albeit having a simplicity in computation process, faced the problem of degenerative solution (e.g. Busing, 2010). On the other hand, other models were mainly developed from the parametric and nonparametric IRT framework which may be less acquainted by social scientists and these models usually require specialised software to implement, except for a few dominance item response models that have been formulated as common factor model, such as the 2-parameter item response model (Kamata & Bauer, 2008) and the (generalised) partial credit model (Huggins-Manley & Algina, 2015) which can be
implemented in general Structural Equation Modelling software. Even though Maraun and Rossi (2001) has suggested the possibility of reformulated an unfolding model as a quadratic factor model, implementation of the method in general modelling software has not been explored or discussed extensively in the literature.

Furthermore, social sciences researchers are usually interested in investigating the relationship between latent variables, and not just the measurement of the latent variables (i.e. attitude, personality; Lu, Thomas, & Zumbo, 2005; Usami, 2011). A common method is to estimate the latent trait after fitting a measurement model (e.g. IRT models) and use it directly in a regression, either as a predictor or outcome variable. However, this method has been found to produce attenuation in path coefficient estimates (Lu et al., 2005; Usami, 2011). Even though Usami (2011) has incorporated the GGUM in the structural equation modelling framework so that both GGUM parameters and structural coefficients are estimated simultaneously to avoid parameter bias, the model could only accommodate person location as the outcome variables. Similarly, this is also the limitation of the constrained (canonical) correspondence analysis (CCA) where person location is linearly regressed on the predictors.

In brief, existing analysis and models for unfolding response data are either less acquainted by social scientists in general or limited in the flexibility in terms of the ability in modelling relationships with other manifest or latent variables. In response to these limitations, building on existing literature, this study proposed the Ordered Categorical Response Unfolding Model (OCRUM), which

1. models the non-monotonic (single-peaked) relationship between the endorsement (observed response) and the latent trait (as assumed by the unfolding response process),

2. estimates item locations and person locations, and

3. can be implemented within the Structural Equation Modelling (SEM) framework which is a general modelling framework that is widely used by social scientists.
Chapter 3

Proposed Model and Research Question

Structural equation modelling (SEM) is a generalisation and extension of commonly used statistical models such as regression and factor analysis. The flexibility of SEM allows researchers to model the relationships among observed and unobserved variables with two broad categories of models: structural model and measurement model (Hoyle, 2012).

The structural model is a class of model which attempts to explain the directional relationships among variables, be it observed or unobserved (Hoyle, 2012). On the other hand, the measurement model deals with the relationship among observed variables (i.e. items) and latent (unobserved) variables (i.e. factors). Confirmatory factor analysis is a common example of the implementation of a measurement model (Brown, 2006).

This research proposed an extension and implementation of the unfolding model in the form of a quadratic factor model as proposed by Maraun and Rossi (2001). This research extended the proposed unfolding model to accommodate categorical data. The model proposed in this study is termed the Ordered Categorical Response Unfolding Model (OCRUM). OCRUM models the single-peaked relationship between the item endorsement and latent trait (as assumed by the unfolding response process) as a function of the squared difference between item location and person location.
OCRUM can be classified as a non-linear measurement model which is linear in coefficients but nonlinear in latent traits (McDonald, 1982). The need of including a non-linear function in a factor analysis (common factor) model have suggested by Barlett (1953; as cited in Wall & Amemiya, 2007) to improve data-model fit if the linear model does not fit the empirical data well. Works have since been done to develop models and estimate non-linearity between item responses and the latent variable (e.g. Etezadi-Amoli & McDonald, 1983; Gibson, 1951; McDonald, 1962). Nevertheless, the rationale of modelling the non-linearity was not grounded on the assumption of the unfolding response process.

### 3.1 Proposed Model: OCRUM

In line with the idea of the unfolding response mechanism, adapting the work from the literature (e.g. Davison, 1977; Maraun & Rossi, 2001; Van Schuur & Kiers, 1994), the proposed model, Ordered Categorical Response Unfolding Model (OCRUM), is defined as

\[ Y_{ij}^\ast = a_j + t_j(\theta_i - \delta_j)^2 + \varepsilon_{ij} \]  

(3.1)

Equation 3.1 can be re-expressed by expanding the equation as

\[ Y_{ij}^\ast = \mu_j + \beta_j\theta_i + t_j\theta_i^2 + \varepsilon_{ij} \]  

(3.2)

where

- \( \mu_j = a_j + t_j\delta_j^2 \),
- \( \beta_j = -2\delta_j t_j \),
- \( t_j \) is the curvature parameter (factor loading),
- \( Y_{ij}^\ast \) is the (latent) response of subject \( i \) on item \( j \),
- \( a_j \) is the conditional mean of \( Y_{ij}^\ast \) when \( \theta_i = \delta_j \),
- \( \theta_i \) is the person location (latent trait) for subject \( i \), and
\( \varepsilon_{ij} \) is the measurement error of subject \( i \) on item \( j \).

Equation 3.2 effectively describes the endorsement on item \( j \) as a function of the squared difference of the latent trait (\( \theta_i \)) and the item location (\( \delta_j \)). It is also assumed that \( \theta_i \) and \( \delta_j \) are located on the same continuum but each \( \delta_j \) does not vary across individuals. When \( t_j \) is constrained to be a negative value, this results in a single-peaked function.

Items measured on a rating scale that results in dichotomous or polytomous data are considered as ordinal data (Stevens, 1960, as cited in Bovaird & Koziol, 2012). However equation 3.2 is a measurement model which assumes \( Y_{ij}^* \) as a continuous variable. In social science research, it is not uncommon to use rating scale (e.g. 5-point Likert scale) as a questionnaire response format. Response data resulted from rating scales is considered ordinal in terms of its scale of measurement. Therefore, it is more appropriate to model the data with a \textit{ordered categorical} measurement model.

The approach for modelling ordered categorical data in structural equation model (SEM) is similar to modelling the continuous measurement model, except for that item thresholds (\( \tau \)) are estimated in place of item intercepts (intercepts are constrained to be zero) and the measurement errors are not estimated due to identification reasons (Bovaird & Koziol, 2012).

Muthén (1983, 1984) posited that underlying the \textit{observed} categorical responses \( Y_j \) there exist a latent (\textit{unobserved}) continuous distribution \( Y_j^* \) which is related as such

\[
Y_j = \begin{cases} 
0 & \text{if } Y_j^* \leq \tau_{j:1} \\
c & \text{if } \tau_{j:c} < Y_j^* \leq \tau_{j:c+1} \\
m & \text{if } Y_j^* > \tau_{j:m} 
\end{cases} \tag{3.3}
\]

where

- \( Y_j^* \) is the underlying latent response for item \( j \),
- \( Y_j \) is the observed categorical response for item \( j \),
- \( c \) is the observed value for the category, where \( c = \{0, 1, ..., m\} \), and 0 is the strongest level of disagreement, and \( m \) represents the value of the strongest level of agreement, and

29
\(\tau_{j,c}\) is the category \(c\) threshold for item \(j\), and the total number of thresholds to be estimated for each item equals to \(m\).

The relationship between \(Y_j, Y_j^+\), and \(\tau_{j,c}\) is presented graphically in Figure 3.1. In essence, an observed categorical response \(Y_j\) is determined by both the latent response \(Y_j^+\) the threshold \(\tau_{j,c}\). Refer to Figure 3.1, when a latent response \(Y_j^+\) is between \(\tau_2\) and \(\tau_3\), the observed response \(Y_j\) will be "D" (Disagree). When \(Y_j^+\) is larger than (i.e. to the right of) \(\tau_5\), the observed response \(Y_j\) will be "SA" (Strongly Agree).

![Figure 3.1: The curve is an illustration of the latent response \(Y_j^+\) with five thresholds with six observed response options (the standard normal distribution for \(Y_j^+\) was used purely for the convenience).](image)

3.1.1 The Effect of \(t_j\) and \(\tau_{j,c}\) on the Response Functions

The response functions in OCRUM is jointly affected by the curvature parameter \(t_j\) and threshold parameters \(\tau_{j,c}\). Figure 3.2 illustrates the effect of \(t_j\) on the response functions of a dichotomous item when the \(\tau_{j,c}\) value is held constant at -1. Similarly, Figure 3.4 illustrates the effect of \(t_j\) on the response functions of a six-category item when the \(\tau_{j,c}\) value is held constant (\(\tau_{j,1} = -3.8; \tau_{j,2} = -2.8; \tau_{j,3} = -1.8; \tau_{j,4} = -0.8, \tau_{j,5} = 0.2\), with an inter-threshold distance of 1). Both sets of figures showed that as \(t_j\) increases,
the rate of change in the probability of endorsement for a particular category increases (i.e. the curves become steeper).

The effect of $\tau_j \cdot c$ on the response functions is shown in Figure 3.3 for a dichotomous item and Figure 3.5 for a polytomous (six-category) item. For these sets of figures, the $t_j$ was held constant at -1. As observed in Figure 3.3, the smaller the thresholds parameters, the maximum value of the probability reaches its upper bound. The $\tau_j \cdot c$ values in Figure 3.5 were chosen so that the inter-threshold distance was equal. This distance was 0.25 in Figure 3.5b and 0.50 in Figure 3.5a. As inter-threshold increases, the probability of endorsement for the middle categories were more differentiated (compare the response functions in Figure 3.5a, Figure 3.5b, and Figure 3.4a). Similar to the dichotomous item, the probability of the strongest level of agreement (i.e. category 5) reaches its upper limit as $\tau_{j,5}$ gets smaller.

### 3.2 Estimation of Latent Interaction Effect

OCRUM (see equation 3.2) involves a quadratic term ($\theta_i^2$) which can be thought as a latent interaction of the person location ($\theta_i$) with itself. This section gives an overview of the possible methods to estimate the latent interaction effect: the semi-parametric approach, indicator-product approach, Bayesian approach, and the distribution-analytic approach.

#### 3.2.1 Semi-parametric Approach

The semi-parametric approach was based on the indirect application of the structural equation mixture model (SEMM) to model the nonlinear relationships between latent variables. This approach does not require the distributional assumption of normality for the exogenous and error terms and the form of the nonlinear relationship is not necessary to be specified a priori (Bauer, 2005). SEMM is a type of finite mixture model which estimates a mixture of structure equation models for latent groups, and the number and nature of these latent groups (classes) are not interpreted substantively.
Figure 3.2: Response function of a hypothetical dichotomous item as a function of $\theta_i - \delta_i$ and $t_j$, while $\tau_j \cdot c = -1$.

Figure 3.3: Response function of a hypothetical dichotomous item as a function of $\theta_i - \delta_i$ and $\tau_j \cdot c$, while $t_j = -1$. 

(a) $t_j = -1$

(b) $t_j = -1.5$

(c) $t_j = -2$
Figure 3.4: Response function of a hypothetical polytomous (6 categories) item as a function of $\theta_i - \delta_i$ and $t_j$, while $\tau_j \cdot c = -3.8, -2.8, -1.8, -0.8, 0.2$
Figure 3.5: Response function of a hypothetical polytomous (6 categories) item as a function of $\theta_i - \delta_i$ and $\tau_j \cdot c$, while $t_j = -1$.

(a) $\tau_j \cdot c = -2.80, -2.30, -1.80, -1.30, -0.80$

(b) $\tau_j \cdot c = -2.30, -2.05, -1.80, -1.55, -1.30$
but only to construct the semi-parametric estimation of the regression weights in the aggregated population (Bauer, 2005).

In general, the parameters estimation using the semi-parametric approach will produce higher bias and is less efficient than the parametric approaches, provided the assumptions of the parametric approaches were not violated. Furthermore, the semi-parametric approach is well suited for exploratory research where the form of the nonlinear relationship is unknown (Bauer, 2005). Therefore, when the functional form of the relationship is known (or assumed in the case of this paper), parametric approaches might be preferred over semi-parametric approach.

### 3.2.2 Indicator-product Approach

The product-indicator approach (e.g. Kenny & Judd, 1984) has been the primary method to model interaction between latent variables. To model interaction between two latent variables, a third latent interaction variable has to be specified in the model. This is done by forming new indicators for the interaction variables by taking the cross-product of indicators of the two latent variables. This approach can be differentiated as the constrained and unconstrained product-indicator approach based on the parameters in the model (Marsh, Wen, Nagengast, & Hau, 2012).

A drawback of the constrained approach is the imposition of complicated parameter constraints which require the assumption of multivariate normality on the indicators. Even though the unconstrained approach relax the requirement of parameter constraints, the results may be less precise under small sample size and when the indicators are normally distributed (Marsh, Wen, Hau, & Nagengast, 2013). However, both approaches shared the same tedious process of producing cross-products of indicators and in deciding subjectively on the pairings of indicators (Marsh et al., 2012).

### 3.2.3 Bayesian Approach

Unlike Frequentist statistics which assume parameters as fixed across studies, Bayesian views parameters as random variables which vary across studies, and the distribution of
the parameters (posterior distribution) are estimated conditionally on prior information regarding their distribution (prior distribution) and the data (Marsh et al., 2012).

The Bayesian approach provides a solution to the limitation of the product-indicator approach stated above by relinquishing the need to create product indicators to model latent interaction. Moreover, the simulation-based estimation method (e.g., Markov Chain Monte Carlo [MCMC] method) takes into account the non-normality of the outcome variables (and the indicators for latent outcome variables) due to the nonlinear effect in the estimation procedure (Marsh et al., 2012). Nonetheless, the implementation is more statistically and practically demanding as it requires research to specify prior knowledge about the parameter values in the model. On top of that, even though some basic Bayesian analysis can be conducted in newer versions of Mplus, specialised software such as WinBUGS is required for the estimation of relatively complex models with nonlinear (e.g., interaction) effects (Marsh et al., 2013).

3.2.4 Distribution-analytic Approach

Similar to the Bayesian approach, the distribution-analytical approach explicitly model the non-normal distribution of the indicators. Creation of product-indicators is not required. It has also been found to produce more precise estimates than the product-indicator approach. Two methods developed under this approach are the quasi-maximum likelihood (QML; Klein & Muthén, 2007) estimation and the latent moderated structural (LMS; Klein & Moosbrugger, 2000) approach. The LMS method is currently implemented in Mplus (Muthén & Muthén, 2013).

The LMS method is based on two statistical ideas: mixture distributions and conditional distributions (Klein & Moosbrugger, 2000). The (non-normal) distribution of observed variable are constructed by a series of normal distributions with varying mean and variance (hence, mixture distributions). A conditional distribution of a variable is the distribution of that variable while holding other variable constant at a particular value (Kelava et al., 2011). The LMS utilises the Cholesky decomposition to decompose the variance-covariance matrix of the first-order latent predictor variables into orthogonal
parts, which enables the partition of the observed outcome variables into their respective linear and nonlinear components (Kelava et al., 2011; Klein & Moosbrugger, 2000).

The multivariate distribution of the observed variables are then represented using a mixture distribution conditioned on the non-linear components (by using their means, variances, and covariances to determine the normal distributions in the mixture). A numerical integration is required to approximate the mixture distribution and the Expectation-maximization algorithm (Dempster, Laird, & Rubin, 1977, as cited in Kelava et al., 2011) is used to estimate the regression coefficients (Kelava et al., 2011; Klein & Moosbrugger, 2000; Marsh et al., 2012).

A few drawbacks of the LMS method is that the null model is undefined, hence the general fit statistics commonly used to examine data-model fit are not applicable. Nevertheless, chi-square difference tests can be used to compare nested model, while information criteria can be used to compare non-nested models (Marsh et al., 2012). Furthermore, computational time increases with more nonlinear effects specified in the model (Marsh et al., 2013).

3.3 Aim and Scope of Research

This thesis proposed to model the unfolding response process by extending the common factor model by including a item location parameter and a latent interaction (quadratic; \( \theta^2_i \)) effect in the measurement model. This model can be used to model categorical (both dichotomous and polytomous) and continuous response data. The interaction effect of the latent factor (\( \theta^2_i \)) was proposed to be estimated using the LMS approach for its precision over the product-indicator approach and simplicity over the Bayesian approach.

The proposed modelling approach was examined by two simulations studies. Specifically, the LMS method as implemented in Mplus in estimating the model parameters with the proposed model were investigated.

The first study consisted of two Monte Carlo simulation studies to examine the
recovery of model parameters and model-data fit under various conditions. To illustrate the application of the proposed approach, empirical response data from the Attitude Towards Capital Punishment Scale (Thurstone, 1932) and the Order Scale (Chernyshenko et al., 2007) was re-analysed using the OCRUM in Study 2.

As mentioned by Andrich (1988), results obtained from a new proposed model should be accompanied by a comparison of results from traditional methods in the development of any new models to investigate the similarities and differences in analysis outcomes. Therefore, a comparison between OCRUM and GGUM was done in Study 2 using the two empirical datasets. The next two chapters describe the method and report the results of the two simulation studies (Study 1) and the re-analysis of the two empirical datasets (Study 2).
Chapter 4

Study 1: Monte Carlo Simulation

The general research question that Study 1 wished to answer was the effect of different conditions, which reflected possible real-world situations, on the recovery of model parameters and data-model fit. Two Monte Carlo simulation studies (Study 1a and Study 1b) were incorporated in Study 1 to explore the different sets of conditions. The two studies were exploratory because the effects of the conditions on the model estimates were largely unknown.

A slight digression is in order to briefly discuss the concepts of Monte Carlo simulation. Monte Carlo simulation studies allows researchers to investigate the effect of conditions, such as different degree of violation of statistical assumptions, and their potential interactions on modelling results such as parameter estimates, standard error, and fit statistics. Simulation studies are valuable when statistical theory is not available to provide predictions on the influence of the conditions (Bandalos & Gagné, 2012).

Monte Carlo studies are essentially experimental studies which conditions or variables, such as sample size, are manipulated. In brief, such studies involve specifying true population values for a given model and generate numerous datasets for each combination of conditions (i.e. replication). The data are then fitted to the model and the results are analysed to understand the effects of the conditions (Bandalos & Gagné, 2012).
4.1 Overview of Data Generation Procedures

An overview of the procedure to generate the observed responses for the simulation study is given in this section. The method was adapted from Muthén and Kaplan (1985).

4.1.1 Choice of Thresholds Values ($\tau_c$)

To generate the observed responses required for the simulation studies, we first had to determine the thresholds values ($\tau_c$) to be used to generate the observed categorical responses ($Y_j$) from the latent continuous responses ($Y^*_j$). The relationship between $\tau_c$, $Y_j$, and $Y^*_j$ was outlined in the previous chapter and illustrated in Figure 3.1.

The $\tau_c$ values were generated as followed: a single $Y^*$ variable was simulated from Equation 3.2, with $a = 0$, $t = -1$, and $\delta = 0$, while $\theta$ and $\varepsilon$ were simulated from the standard normal distribution ($\mu = 0; \sigma = 1$) with a large sample size ($N = 1,000,000$).

- the value at the 50th percentile of the $Y^*$ distribution was used as the threshold to generate the dichotomous data. The value is given in equation 4.1
- the value at the 5th, 23rd, 50th, 77th, and the 95th percentile, respectively, of the $Y^*$ distribution was used as the threshold to generate the polytomous data, thus simulating data collected from a 6-point rating scale. The values are given in equation 4.2

The derivation of the $\tau_c$ values and the data generation MATLAB script can be found in Appendix A on page 106 and Appendix C on page 112.

4.1.2 Generation of Observed Responses ($Y_j$)

The observed response data were simulated using these steps:

1. the population "true" values of $\delta_j$ were chosen based on the design of each study: the cross-over between test length ($J$) and item location ranges ($D$). The values are given in Table 4.1.
2. $Y^*_j$ were simulated from the proposed model as specified in Equation 3.2, where $a_j$ was fixed as 0, and $t_j$ was fixed as -1, and $d_j$ corresponded to each of the value in Table 4.1. $\theta_i$ and $\varepsilon_{ij}$ were simulated from standard normal distribution ($\mu = 0; \sigma = 1$).

3. the continuous $Y^*_j$ values were transformed into dichotomous \{0; 1\} and polytomous \{0; 1; 2; 3; 4; 5\} $Y_j$ values through equation 4.1 and equation 4.2 respectively

- For dichotomous responses, $\tau_c = \{-0.7392\}$, specifically

$$Y_j = \begin{cases} 0 & \text{if } Y^*_j \leq -0.7392 \\ 1 & \text{if } Y^*_j > -0.7392 \end{cases}$$  \quad (4.1)

- For polytomous responses, $\tau_c = \{-4.1667; -1.8781; -0.7392; 0.1889; 1.2253\}$, specifically

$$Y_j = \begin{cases} 0 & \text{if } Y^*_j \leq -4.1667 \\ 1 & \text{if } -4.1667 < Y^*_j \leq -1.8781 \\ 2 & \text{if } -1.8781 < Y^*_j \leq -0.7392 \\ 3 & \text{if } -0.7392 < Y^*_j \leq 0.1889 \\ 4 & \text{if } 0.1889 < Y^*_j \leq 1.2253 \\ 5 & \text{if } Y^*_j > 1.2253 \end{cases}$$  \quad (4.2)

4.2 Implementation in Mplus

As described in the previous section, the quadratic effect (i.e. interaction effect of the person location [$\theta_i$] with itself) was estimated by the LMS method which is implemented in Mplus version 7.2.\footnote{Muthén (2008) described that the LMS was implemented with a slightly different algorithm in Mplus as compared to the procedure described in Klein and Moosbrugger (2000)} This section briefly introduced the model identification procedure, the relevant Mplus syntax in specifying the model, and the fit statistics that Mplus provides for the analysis. An example of the Mplus syntax for OCRUM can be found in
4.2.1 Model Identification

The proposed model, in effect, is an extension of a linear common factor model, therefore identification of the model is identical to that of a confirmatory factor analysis (CFA) model in the Structural Equation Modelling (SEM) framework: by fixing factor loading of one of the observed variables (i.e. indicators) to 1, or by fixing the variance of the common factor to a fixed value, typically 1 (Kenny & Milan, 2012). We chose to fix the variance of the latent factor (person location; $\theta_i$) as 1 for identification.

4.2.2 Model Fitting in Mplus

To model the quadratic effect through the Distribution-Analytic approach in Mplus (i.e. LMS), users can specify the following options under the `ANALYSIS` command in the syntax: `TYPE IS RANDOM; ALGORITHM IS INTEGRATION;`. By specifying `TYPE IS RANDOM`, random effect is estimated, and `ALGORITHM IS INTEGRATION` instructs Mplus to use the maximum likelihood estimation method with robust standard errors (i.e. `ESTIMATOR IS MLR`) via a numerical integration algorithm (Muthén & Muthén, 2013, p.68).

The MLR estimator is one of the full-information maximum likelihood (FIML) estimators implemented in Mplus (Muthén, 2005). An FIML estimates parameters that "maximise the likelihood of the sample response patterns" instead of using the observed variance-covariance matrix (Bovaird & Koziol, 2012). Through the estimation based on empirical response patterns rather than summary statistics (i.e. variance and covariance), FIML takes into account the discrete properties of categorical data, hence FIML is also an alternative approach to the limited-information estimators (e.g. weighted least squares) to modelling categorical responses in the measurement model (Bovaird & Koziol, 2012). `CATEGORICAL ARE ALL` can be specify under the `VARIABLE` syntax command to note that all variables are ordered categorical (i.e.dichotomous or polytomous). In this study, we chose to use probit link as the link function. This can be specified under the
To state the quadratic term, users can use the | symbol to define the random effect variables in the model along with the XWITH to construct the quadratic effect under the MODEL; command. The BY statement defines the linear part of the quadratic function while the ON statement defines the quadratic part of the function. Using the MODEL CONSTRAINT: command, users can imposed the linear constraints as required in the proposed model. The new command specifies the new parameters to be used in the constrains (Muthén & Muthén, 2013, p.68).

4.2.3 Estimating Factor Scores in Mplus

Factor scores (person locations; ˆθ) are estimated using the Expected A Posteriori (EAP) method in Mplus when categorical data and MLE method were involved in the parameter estimation (Mplus, n.d.). Factor scores can be obtained in Mplus through the use of SAVE IS FScores; under the SAVEDATA command.

4.3 Overview of Analysis

4.3.1 Parameter Recovery

To evaluate the feasibility of the proposed model and the LMS method in analysing the simulated realistic data, recovery of the parameter estimates and the corresponding standard errors are evaluated using parameter bias and standard error bias. A higher bias value represents higher deviation of the estimates from the true "population" value of the parameter and empirical standard error.

Parameter Bias

Parameter estimate bias is one of the common dependent variables in simulation studies. Bias is defined as the difference between the estimated parameter and the population parameter. Parameter estimate bias are usually calculated as raw bias,
where \( \hat{\gamma}_{pq} \) is the \( q \)th sample (i.e. replication) estimate of the \( p \)th true (population) parameter value \( \gamma_p \) and \( r \) is the number of replications in a simulation (Bandalos & Leite, 2013).

**Standard Error Bias**

Standard error bias can also be calculated similar to the calculation of parameter bias. The empirical standard errors \( SE(\hat{\gamma}_p)_q \) are calculated as the standard deviation of the parameter estimates over the number of replications in a simulation (Bandalos & Leite, 2013). The empirical standard error serves as the benchmark for assessing the standard errors estimated by a given estimation method (i.e. LMS in this study). Standard error bias is calculated as

\[
Bias(\hat{SE}(\hat{\gamma}_p)) = \frac{\sum_{q=1}^{r}(\hat{SE}(\hat{\gamma}_p)_q - SE(\hat{\gamma}_p))}{r}
\]  

(4.4)

where \( \hat{SE}(\hat{\gamma}_p)_q \) is the estimated standard error of parameter \( \hat{\gamma}_p \) in the \( q \)th replication, and \( SE(\hat{\gamma}_p) \) is the empirical standard error of the parameter \( \hat{\gamma}_p \). \( SE(\hat{\gamma}_p) \) is determined by calculating the standard deviation of the parameters estimates \( (\hat{\gamma}_p) \) over the \( q \) replications (Bandalos & Gagné, 2012; Bandalos & Leite, 2013).

### 4.3.2 Assessing Model-Data Fit

Due to the application of the full-information maximum likelihood (FIML) estimator which involves numerical integration with categorical data, \( \chi^2 \) and other fit statistics for the fitted model are not defined (Muthén, 2010), therefore, only a handful of model fit statistics are currently available in Mplus: Loglikelihood value, Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and Sample-size Adjusted BIC.
Comparing Nested Model

The information on loglikelihood value can be used to conduct the likelihood ratio tests to test the relative model-data fit of two nested model. A statistical significant test suggest that the less restrictive model is preferred over the more restrictive model.

The standard likelihood ratio test (Bollen, 1989) is computed as such

$$T_D = -2(LL_0 - LL_1), T_D \sim \chi^2(p_1 - p_0)$$

(4.5)

where

$T_D$ is the difference statistics which follows a $\chi^2$ distribution with the degree of freedom $df = p_1 - p_0$, where

$p_1$ is the number of estimated parameters in the less restrictive model,

$p_0$ is the number of estimated parameters in the more restrictive model,

$LL_1$ is the loglikelihood of the less restrictive model, and

$LL_0$ is the loglikelihood of the more restrictive model.

Comparing Nested and Non-Nested Model

Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), and the Sample-size Adjusted BIC (sBIC) can be used to select the better fitting model of two models, regardless whether the two models are nested. In general, for a given comparison, the model with a smaller information criteria is preferred (West, Taylor, & Wu, 2012). Even though sBIC is reported in Mplus, it is not a common index to be used (Muthén, 2006). Nevertheless, the current study included all three information criteria.
Softwares

Mplus version 7.2 (Muthén & Muthén, 2013) was used for model fitting. MATLAB R2013a (MATLAB, 2013) was used to simulate the observed responses and to extract the relevant statistics such as estimated parameters and their standard errors from the output for further analysis. SPSS version 22 (IBM Corp, 2013) was used to analyse the statistics.

4.4 Study 1a

The goal of this study was to examine the ability of the proposed model (Ordered Categorical Response Unfolding Model; OCRUM) and LMS approach in recovering the model parameters under a range of conditions (independent variables): sample sizes (N), test length (J), item location ranges (D), and response options (R).

Sample Size (N) and Test Length (J)

The three sample sizes reflected a small (N = 150), moderate (N = 300), and relatively larger sample size (N = 600). A short test was defined as test length of 5 items (e.g. the Satisfaction with Life Scale; Diener, Emmons, Larsen, & Griffin, 1985). An 10-item test represents a typical test (e.g. individual scales in the 16PF Questionnaire Conn & Rieke, 1994), and 20-item test is a long test (e.g. the Beck Depression Inventory; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961).

Item Location Ranges (D)

Andrich (1988) simulated person locations from a normal distribution with a mean ($\mu$) of 0, and standard deviation ($\sigma$) of 2.5 with items locations ranging between -3.325 and 3.325. Hence all items are within 1.33 standard deviation (SD) of the person distribution. Likewise, Roberts (1995) simulated person location from a normal distribution of $\mu = 0$, $\sigma = 2.0$ with items locations ranging between -4.25 to 4.25. Hence all items are within 2.13 SD of the person distribution. Taken the research design from the literature,
we decided to incorporate the two item location ranges: ±1.5 and ±2.5. On top of that, a scale with item location (δ) ranges within ±1.5 (i.e. 1.5 SD) and within ±2.5 (i.e. 2.5 SD) reflects the difference in coverage of items measuring different attitude or valance on a latent continuum.

The distance between each adjacent item were equidistant regardless of the number of items and the item location ranges (D). To illustrate, for a given simulated dataset with test length J = 5, when item location range D = ±1.5, the item were simulated to be located at -1.50; -0.75; 0.00; 0.75; 1.50; when item location range D = ±2.5, the item were simulated to be located at -2.50; -1.25; 0.00; 1.25; 2.50. Notice that the distance between any adjacent items within the same item location range were the same (i.e. equidistant; with rounding errors). The items with negative item location values represent an opposite attitude or valence compared to the items with positive item location values. Table 4.1 (see page 48) consists of the item location values used in the crossed conditions of test length and item location range.

In this study, we categorised the items as intermediate items and extreme items based on their δj (item location) values in relation to the distribution of the θi on the same latent continuum. We define an item with |δj| ≤ 1.5 as an intermediate item and an extreme item when |δj| > 1.5.

The rationale of the definition is outlined as followed. We opined that, given that θ has a standard normal distribution (mean [µ] = 0, [σ] = 1), an item is considered as extreme when the proportion of people around that particular item is small. The proportion of people, or the Overlap value, is quantify as \( \frac{n_{δj±1}}{N} \), where \( n_{δj±1} \) is the number of samples located between \( δj−1 \) and \( δj+1 \), and N is the total sample size. In this paper, an Overlap value smaller than .30 (i.e. 30% of the total sample) was considered as small.

Given that θ follows a standard normal distribution (i.e. z distribution), the theoretical proportion can be derived for each of the item. This can be obtained by first obtaining the cumulative probability (pz) for both \( z_1 = δj − 1 \) and \( z_2 = δj + 1 \). The Overlap value is obtained by taking the difference of the two proportion (pz2 − pz1).
<table>
<thead>
<tr>
<th>D</th>
<th>J</th>
<th>Extreme</th>
<th>Intermediate</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>±2.5</td>
<td>10</td>
<td>-2.5</td>
<td>-1.50; -0.75; 0.00; 0.75; 1.50</td>
<td>'a' represents no item located at the &quot;extreme&quot; and is defined with the Overlap value described in subsection 4.4. J = Number of item per factor; D = Item location ranges. For example, comparing the information on the 1st row and the 4th row, there is no extreme item on the 1st row, and there are two extreme items (one on each side of the continuum) on the 4th row.</td>
</tr>
<tr>
<td>±2.5</td>
<td>20</td>
<td>2.50; -2.24; -1.97; -1.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>±2.5</td>
<td>10</td>
<td>-2.5</td>
<td>-1.50; -1.17; -0.83; -0.45; 0.45; 0.83; 1.17; 1.50</td>
<td></td>
</tr>
<tr>
<td>±2.5</td>
<td>20</td>
<td>2.50; -1.94; -1.34; -0.75; -0.17; 0.17; 0.75; 1.34; 1.94; 2.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Using an item with item location $\delta = 0$ as an example, $z_1 = 0 - 1 = -1$, which has a cumulative probability (proportion) of $p_{z_1} = .16$, and $z_2 = 0 + 1 = 1$, which has a cumulative probability (proportion) of $p_{z_2} = .84$. Therefore the Overlap value $= p_{z_2} - p_{z_1} = .84 - .16 = .68$. Hence an item with $\delta = 0$ has an Overlap value of .68 (i.e. 68% of the total sample). Therefore, using the same calculation, we can derived that $|\delta_j| = 1.5$ has an Overlap value of .30. Hence any item with $|\delta_j| \leq 1.5$ is defined as an intermediate item and an extreme item is when $|\delta_j| > 1.5$ in the paper.

Response Options (R)

Considering the common use of dichotomous response format such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989, 2001) and polytomous response format in tests such as NEO-PI-R (Costa & McCrae, 1992) in psychological testing setting, we focused on both dichotomous and polytomous response in this study.

We chose to simulate a 6-point, instead of a 5-point, rating scale for the polytomous responses based on Dalal, Carter, and Lake’s (2014) findings. They found that even-numbered rating scale was preferable when the unfolding response was assumed to be the underlying response mechanism because a middle response option (i.e. the neutral option) did not have a plausible theoretical explanation under the unfolding response mechanism. An alternative reason provided by Dalal et al. (2014) for not using an odd-numbered rating scale was that the middle response option was likely to be used inappropriately (i.e. the option was not chosen to reflect neutrality).

4.4.1 Method

Research Design

The study used a between-subjects design, outlined by $3 \times 3 \times 2 \times 2$ (sample size: N = 150; 300; 600) x 3 (test length: J = 5; 10; 20) x 2 (item location range: D = ±1.5; ±2.5) x 2 (response option: R = dichotomous; polytomous). Two hundred replications were done
for each of the 36 conditions. The dependent variables of interest were: the $\delta_j$ parameter bias, $\delta_j$ standard error (SE) bias, the correlation between the simulated (true) $\theta$ and $\hat{\theta}$, and the "power" of the various fit-statistics in identifying the correct model.

$\delta_j$ was the only parameter of interest in this study because it is the only OCRUM parameter (see equation 3.2) that has a straight-forward interpretation and implication for the item (i.e. location of items on the latent continuum). $\beta_j$, which is commonly known as the item factor loadings, was not investigated as its value is determined by the $\delta_j$ and the $t_j$ parameter (see equation 3.2). Therefore we focused our investigation on $\delta_j$ within the limited space of this thesis.

**Data Generation**

The process of simulating the categorical response data for Study 1a was outlined in subsection 4.1. The item location $\delta_j$ used to generate the responses was described in subsection 4.4 can be found in Table 4.1. The MATLAB data generation script can be found in Appendix C.

**Procedure**

Recalled that the two aims of this study were to investigate the effect of the independent variables, not only on the parameter estimates, but also on data-model fit. All simulated datasets were fitted onto both a single-factor linear common factor model (equation 2.7) and OCRUM (equation 3.2) in Mplus. To obtain comparable parameter estimates, MLR was used as the estimator and probit link was used as the link function when fitting both models (see Appendix B for Mplus syntax for the OCRUM model and Appendix D for the Mplus syntax for a linear common factor model).

**4.4.2 Results**

**Successful Replications**

The number of successful replication is reported in Table 4.2. The second value represents the non-convergence, which is defined as the number of datasets which failed
to fit the model in Mplus. The third value represents the number of datasets with convergence but with error, which is defined as the number of datasets which fitted the model but with at least one parameter fixed (i.e. not estimated) in Mplus.

Referring to Table 4.2, in the case of dichotomous response option (R: Dichotomous), Mplus had difficulty in fitting OCRUM when sample size is small (N = 150) and when item location range is large (±2.5). However this issue in small sample size and large item location range was alleviated as the test length was larger (J = 20). In contrast to the dichotomous response option condition, most of the datasets were fitted without error in the polytomous response option condition (R: Polytomous), regardless of sample size (N), test length (J), and item location ranges (D).

Table 4.2: Study 1a Successful Replications

<table>
<thead>
<tr>
<th>J</th>
<th>D ±1.5</th>
<th>±2.5</th>
<th>±1.5</th>
<th>±2.5</th>
<th>±1.5</th>
<th>±2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>14/0/5</td>
<td>45/4/151</td>
<td>194/0/6</td>
<td>89/7/104/</td>
<td>197/0/3</td>
<td>120/3/77</td>
</tr>
<tr>
<td>300</td>
<td>195/0/5</td>
<td>61/6/133</td>
<td>199/0/1</td>
<td>159/0/41</td>
<td>200/0/0</td>
<td>184/2/14</td>
</tr>
<tr>
<td>600</td>
<td>199/0/1</td>
<td>104/3/93</td>
<td>200/0/0</td>
<td>188/0/12</td>
<td>200/0/0</td>
<td>198/0/2</td>
</tr>
</tbody>
</table>

R: Polytomous

<table>
<thead>
<tr>
<th>J</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>198/2/0</td>
<td>187/3/10</td>
<td>200/0/0</td>
</tr>
<tr>
<td>300</td>
<td>199/1/0</td>
<td>200/0/0</td>
<td>199/1/0</td>
</tr>
<tr>
<td>600</td>
<td>196/4/0</td>
<td>199/0/1</td>
<td>200/0/0</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges; R = Response Options. The first value represents the number of successful convergence. The second value represents the number of non-convergence and the third value represents the number of convergence but with error.

Parameter Bias

Due to the design of the study, the δj values used for each crossed conditions of item length (J) and item location range (D) were different. Taking this into consideration and due to the varying number of δj in each condition, we believe that a visualisation of the parameter bias of each δj would be more informative than conducting hypothesis testing with a summary statistics (e.g. average of bias across items).
An examination of Figure 4.1, Figure 4.2 and Figure 4.3 showed several consistent patterns. Comparing dichotomous and polytomous datasets, the parameter bias were higher for the dichotomous datasets (i.e. the solid lines) in most of the conditions. As expected, the more extreme an item (i.e. the items not between the two vertical dotted lines), the larger the parameter bias. However, it was worth taking note that these biases might be mitigated by increasing the sample size: the parameter bias decreased (towards 0) as the sample size increased from 150 to 300.

**Standard Error Bias**

Similar to the rationale stated in the previous section, hypothesis testing was not used to analyse the standard error bias. Instead plots of the standard error bias were presented (see Figure 4.4, Figure 4.5, and Figure 4.6).

A very important pattern should be observed in Figure 4.4. The standard error bias of the conditions of test length ($J = 5$, and item location range ($D$) = $\pm 2.5$ were relatively higher than all other conditions especially for the Dichotomous response option (R) condition, which in practice, researchers should be wary of the inaccurate rejecting or fail to reject the null hypothesis that the item location is different from 0.

It was also found that regardless of the conditions, polytomous responses (i.e. datasets) would have ignorable standard error bias (bias were close to 0), except for the $J = 5$ condition. However in this condition, as the sample size ($N$) is larger than 300, the bias became smaller. In contrast to the polytomous responses, extreme items with dichotomous response had relatively larger standard error bias (i.e. standard error was over- or under-estimated).

**Recovery of Person Location ($\theta_i$)**

The mean and standard deviation of the Pearson correlation between the $\hat{\theta}$ and simulated $\theta$ are presented in Table 4.3. Table 4.4 contains the mean and standard deviation of the Spearman rank correlation between the $\hat{\theta}$ and simulated $\theta$. For both tables, the mean correlation coefficient were higher than .70, with Spearman rank correlation
Figure 4.1: $\delta_j$ parameter bias for each item was plotted for the test length ($J) = 5$ conditions. $N =$ sample size, and $D =$ item location ranges (the ± sign was omitted for brevity). The solid line represents the R = Dichotomous condition, while the dashed line represents the R = Polytomous condition. Items to the left vertical dotted line on the left and items to the right of the vertical dotted line on the right represent extreme items. Items between the two vertical dotted lines are the intermediate items.
Figure 4.2: $\delta_j$ parameter bias for each item was plotted for the test length ($J$) = 10 conditions. $N =$ sample size, and $D =$ item location ranges (the $\pm$ sign was omitted for brevity). The solid line represents the $R =$ Dichotomous condition, while the dashed line represents the $R =$ Polytomous condition. Items to the left vertical dotted line on the left and items to the right of the vertical dotted line on the right represent extreme items. Items between the two vertical dotted lines are the intermediate items.
Figure 4.3: $\delta_j$ parameter bias for each item was plotted for the test length ($J$) = 20 conditions. $N =$ sample size, and $D =$ item location ranges (the $\pm$ sign was omitted for brevity). The solid line represents the $R =$ Dichotomous condition, while the dashed line represents the $R =$ Polytomous condition. Items to the left vertical dotted line on the left and items to the right of the vertical dotted line on the right represent extreme items. Items between the two vertical dotted lines are the intermediate items.
Figure 4.4: $\delta_j$ standard error bias for each item was plotted for the test length ($J$) = 5 conditions. $N = \text{sample size}$, and $D = \text{item location ranges (the } \pm \text{ sign was omitted for brevity})$. The solid line represents the $R = \text{Dichotomous condition}$, while the dashed line represents the $R = \text{Polytomous condition}$. Items to the left vertical dotted line on the left and items to the right of the vertical dotted line on the right represent extreme items. Items between the two vertical dotted lines are the intermediate items.
Figure 4.5: $\delta_j$ standard error bias for each item was plotted for the test length (J) = 10 conditions. N = sample size, and D = item location ranges (the ± sign was omitted for brevity). The solid line represents the R = Dichotomous condition, while the dashed line represents the R = Polytomous condition. Items to the left vertical dotted line on the left and items to the right of the vertical dotted line on the right represent extreme items. Items between the two vertical dotted lines are the intermediate items.
Figure 4.6: $\delta_j$ standard error bias for each item was plotted for the test length ($J$) = 20 conditions. $N =$ sample size, and $D =$ item location ranges (the ± sign was omitted for brevity). The solid line represents the $R =$ Dichotomous condition, while the dashed line represents the $R =$ Polytomous condition. Items to the left vertical dotted line on the left and items to the right of the vertical dotted line on the right represent extreme items. Items between the two vertical dotted lines are the intermediate items.
higher than the Pearson correlation in general. The standard deviation of the correlations in each condition is smaller than 0.01.

Two between-subjects ANOVA were conducted to examine the effect of sample size (N), test length (J), item location ranges (D), and response option (R) on the Pearson correlation and Spearman rank correlation between the estimated and simulated $\theta$. The ANOVA results were reported in Table 4.5 and Table 4.6.

200 replications were done for each independent cell. The number of replications can be understood as the number of samples per cell. Therefore, in regards to the "sample" size, this study contained a relatively large number sample size. It is known that with a large sample size, trivial effects (e.g. small difference between groups) could also reach statistical significance (i.e. $p < \alpha$ level) due to the relatively smaller standard error.

To overcome the potential problem of interpreting trivial effects as substantial, effect sizes were reported for this study instead of the $p$-values of statistical tests conducted. Effect size can be defined as a quantified magnitude of an effect (Kelley & Preacher, 2012). We contend that effect sizes of the interaction and/or main effects of the independent variables would provide more information than $p$-values. We chose to report the partial eta-squared ($\eta^2_p$) as an effect size index, which is calculated as

$$\eta^2_p = \frac{SS_{Effect}}{SS_{Effect} + SS_{Error}}$$

(4.6)

where

$SS_{Effect}$ is the sum of squares of the (main or interaction) effect in question, and

$SS_{Error}$ is the sums of squares of the error term

$\eta^2_p$ is interpreted as the variance in each of the effects and the associated error that is accounted for by that effect. Following Cohen (1988), $\eta^2_p = 0.02$ (2% variance explained) is interpreted as small effect, $\eta^2_p = 0.13$ (13% variance explained) is interpreted as medium effect, while $\eta^2_p = 0.26$ (26% variance explained) is interpreted as a large effect.

In both analyses, there was a large interaction effect between test length and response option (J*R). The $\eta^2_p$ was 0.452 for the analysis on Pearson correlation, and $\eta^2_p$ was
0.647 for the analysis on Spearman rank correlation. This could be interpreted as the effect of response option on the correlation differ at at least one level of test length. An inspection of the descriptive statistics table (Table 4.3 and Table 4.4) showed that in dichotomous response option condition, as test length gets longer, then correlation between the estimated and simulated \( \theta \) increases. However, in the polytomous response option condition, the both Pearson and Spearman rank correlation were consistently above .95 regardless of the test length.

Tests of simple effects of test length (J) at the two level of response option (R) were conducted for both Pearson and Spearman correlation separately. ANOVA showed that the effect of test length at polytomous response option condition was small for both Pearson \( [F(2, 6410) = 207.446, \eta^2_p = 0.061] \) and Spearman correlation \( [F(2, 6410) = 419.642, \eta^2_p = 0.116] \), confirming the observations from the descriptive statistics. This is in contrast with the large effect of test length at dichotomous response option level, where \( F(2, 6410) = 5814.374, \eta^2_p = 0.645 \) and \( F(2, 6410) = 12663.488, \eta^2_p = 0.798 \) for Pearson and Spearman correlation as the dependent variable respectively.

More specifically on the Dichotomous response option condition, the mean Pearson correlation difference between 5-item \( (M_{\rho_P} = .777) \) and 10-item \( (M_{\rho_P} = .882) \) was -0.105. The mean Pearson correlation difference between 5-item and 20-item \( (M_{\rho_P} = .938) \) was -0.161. The mean Pearson correlation difference between 10-item and 20-item was -0.056.

The mean Spearman correlation difference between 5-item \( (M_{\rho_{Sp}} = .793) \) and 10-item \( (M_{\rho_{Sp}} = .908) \) was -0.114. The mean Spearman correlation difference between 5-item and 20-item \( (M_{\rho_{Sp}} = .958) \) was -0.164. The mean Spearman correlation difference between 10-item and 20-item was -0.050.

**Model Comparison**

The simulated data was generated according to equation 3.2, therefore, fitting the simulated data on OCRUM should produce a lower AIC, BIC, and sBIC values as compared to fitting to a linear factor model (i.e. a wrong model). Similarly, a test of
Table 4.3: Mean and Standard Deviation of Pearson Correlation between $\hat{\theta}$ and Simulated $\theta$

<table>
<thead>
<tr>
<th>J</th>
<th>D</th>
<th>±1.5</th>
<th>±2.5</th>
<th>±1.5</th>
<th>±2.5</th>
<th>±1.5</th>
<th>±2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>±1.5</td>
<td>.748</td>
<td>.783</td>
<td>.842</td>
<td>.904</td>
<td>.916</td>
<td>.956</td>
</tr>
<tr>
<td></td>
<td>±2.5</td>
<td>.066</td>
<td>.046</td>
<td>.068</td>
<td>.036</td>
<td>.060</td>
<td>.034</td>
</tr>
<tr>
<td>300</td>
<td>±1.5</td>
<td>.761</td>
<td>.793</td>
<td>.856</td>
<td>.911</td>
<td>.918</td>
<td>.957</td>
</tr>
<tr>
<td></td>
<td>±2.5</td>
<td>.045</td>
<td>.038</td>
<td>.049</td>
<td>.022</td>
<td>.038</td>
<td>.026</td>
</tr>
<tr>
<td>600</td>
<td>±1.5</td>
<td>.767</td>
<td>.811</td>
<td>.863</td>
<td>.917</td>
<td>.923</td>
<td>.960</td>
</tr>
<tr>
<td></td>
<td>±2.5</td>
<td>.036</td>
<td>.025</td>
<td>.031</td>
<td>.014</td>
<td>.024</td>
<td>.016</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges; R = Response Options. Values represent mean. Standard deviations are reported in brackets.

Table 4.4: Mean and Standard Deviation of Spearman Rank Correlation between $\hat{\theta}$ and Simulated $\theta$

<table>
<thead>
<tr>
<th>J</th>
<th>D</th>
<th>±1.5</th>
<th>±2.5</th>
<th>±1.5</th>
<th>±2.5</th>
<th>±1.5</th>
<th>±2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>±1.5</td>
<td>.790</td>
<td>.775</td>
<td>.889</td>
<td>.912</td>
<td>.947</td>
<td>.965</td>
</tr>
<tr>
<td></td>
<td>±2.5</td>
<td>.057</td>
<td>.052</td>
<td>.042</td>
<td>.029</td>
<td>.028</td>
<td>.012</td>
</tr>
<tr>
<td>300</td>
<td>±1.5</td>
<td>.802</td>
<td>.785</td>
<td>.898</td>
<td>.919</td>
<td>.948</td>
<td>.966</td>
</tr>
<tr>
<td></td>
<td>±2.5</td>
<td>.042</td>
<td>.046</td>
<td>.030</td>
<td>.016</td>
<td>.020</td>
<td>.009</td>
</tr>
<tr>
<td>600</td>
<td>±1.5</td>
<td>.809</td>
<td>.799</td>
<td>.905</td>
<td>.924</td>
<td>.953</td>
<td>.967</td>
</tr>
<tr>
<td></td>
<td>±2.5</td>
<td>.034</td>
<td>.027</td>
<td>.020</td>
<td>.011</td>
<td>.013</td>
<td>.006</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges; R = Response Options. Values represent mean. Standard deviations are reported in brackets.
Table 4.5: Effect of Independent Variables on Pearson Correlation between $\hat{\theta}$ and Simulated $\theta$

<table>
<thead>
<tr>
<th>Degree of freedom</th>
<th>Sum of Squares</th>
<th>$\eta^2_{ip}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>2</td>
<td>0.064</td>
</tr>
<tr>
<td>$J$</td>
<td>2</td>
<td>7.696</td>
</tr>
<tr>
<td>$D$</td>
<td>1</td>
<td>0.788</td>
</tr>
<tr>
<td>$R$</td>
<td>1</td>
<td>18.723</td>
</tr>
<tr>
<td>$N*J$</td>
<td>4</td>
<td>0.017</td>
</tr>
<tr>
<td>$N*D$</td>
<td>2</td>
<td>0.001</td>
</tr>
<tr>
<td>$J*D$</td>
<td>2</td>
<td>0.032</td>
</tr>
<tr>
<td>$N*R$</td>
<td>2</td>
<td>0.046</td>
</tr>
<tr>
<td>$J*R$</td>
<td>2</td>
<td>4.284</td>
</tr>
<tr>
<td>$D*R$</td>
<td>1</td>
<td>0.598</td>
</tr>
<tr>
<td>$N<em>J</em>D$</td>
<td>4</td>
<td>0.003</td>
</tr>
<tr>
<td>$N<em>J</em>R$</td>
<td>4</td>
<td>0.010</td>
</tr>
<tr>
<td>$N<em>D</em>R$</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>$J<em>D</em>R$</td>
<td>2</td>
<td>0.028</td>
</tr>
<tr>
<td>$N<em>J</em>D*R$</td>
<td>4</td>
<td>0.005</td>
</tr>
<tr>
<td>Residual</td>
<td>6410</td>
<td>5.194</td>
</tr>
</tbody>
</table>

$N$ = Sample size; $J$ = Number of item per factor; $D$ = Item location ranges; $R$ = Response Options. Interaction effects are denoted by the asterisk symbol (*), for example, the interaction between $N$ and $J$ is denoted as $N*J$.

Table 4.6: Effect of Independent Variables on Spearman Rank Correlation between $\hat{\theta}$ and Simulated $\theta$

<table>
<thead>
<tr>
<th>Degree of freedom</th>
<th>Sum of Squares</th>
<th>$\eta^2_{ip}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>2</td>
<td>0.053</td>
</tr>
<tr>
<td>$J$</td>
<td>2</td>
<td>8.002</td>
</tr>
<tr>
<td>$D$</td>
<td>1</td>
<td>0.039</td>
</tr>
<tr>
<td>$R$</td>
<td>1</td>
<td>12.708</td>
</tr>
<tr>
<td>$N*J$</td>
<td>4</td>
<td>0.014</td>
</tr>
<tr>
<td>$N*D$</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>$J*D$</td>
<td>2</td>
<td>0.088</td>
</tr>
<tr>
<td>$N*R$</td>
<td>2</td>
<td>0.030</td>
</tr>
<tr>
<td>$J*R$</td>
<td>2</td>
<td>4.562</td>
</tr>
<tr>
<td>$D*R$</td>
<td>1</td>
<td>0.010</td>
</tr>
<tr>
<td>$N<em>J</em>D$</td>
<td>4</td>
<td>0.002</td>
</tr>
<tr>
<td>$N<em>J</em>R$</td>
<td>4</td>
<td>0.010</td>
</tr>
<tr>
<td>$N<em>D</em>R$</td>
<td>2</td>
<td>0.001</td>
</tr>
<tr>
<td>$J<em>D</em>R$</td>
<td>2</td>
<td>0.067</td>
</tr>
<tr>
<td>$N<em>J</em>D*R$</td>
<td>4</td>
<td>0.002</td>
</tr>
<tr>
<td>Residual</td>
<td>6410</td>
<td>2.486</td>
</tr>
</tbody>
</table>

$N$ = Sample size; $J$ = Number of item per factor; $D$ = Item location ranges; $R$ = Response Options. Interaction effects are denoted by the asterisk symbol (*), for example, the interaction between $N$ and $J$ is denoted as $N*J$. 
nested model should also produce a significant likelihood ratio tests, pointing towards
the use of a more complex (i.e. less restrictive) model (i.e. OCRUM) instead of a simpler
(i.e. more restrictive) model (i.e. linear factor model).

The ratio of number of times each of the four statistics (i.e. AIC, BIC, sBIC, and
likelihood ratio test) indicated OCRUM as the better fitting model and total number
of dataset in each condition were presented in Table 4.7, Table 4.8, Table 4.9, and
Table 4.10 respectively. A value of 1 represent the statistics identified OCRUM as the
better fitting model in all simulated dataset under the specified conditions. In contrast,
a value of 0 represent the statistics failed to identify OCRUM as the better fitting model
in the simulated dataset under the specified conditions.

It is clear that AIC, sBIC, and the likelihood ratio test indicated OCRUM as the
correct model in almost all cross-over conditions of sample size (N), test length (J), item
location ranges (D), and response options (R). The BIC statistics might be problematic
in cases when the response option is dichotomous with small sample size (N = 150) and
short test length (J = 5) as BIC only managed to identify OCRUM as the better fitting
model with a 50% rate.

Table 4.7: Model Comparison using AIC

<table>
<thead>
<tr>
<th></th>
<th>J 5</th>
<th>J 10</th>
<th>J 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>D ±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
</tr>
<tr>
<td>150</td>
<td>0.986</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>300</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>600</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

R: Dichotomous

<table>
<thead>
<tr>
<th></th>
<th>J 5</th>
<th>J 10</th>
<th>J 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>D ±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±2.5</td>
</tr>
<tr>
<td>150</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>300</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>600</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

R: Polytomous

<table>
<thead>
<tr>
<th></th>
<th>J 5</th>
<th>J 10</th>
<th>J 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>D ±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
</tr>
<tr>
<td>150</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>300</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>600</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges;
R = Response Options. Values represent proportion of successful identification of
the correct model as the better fitting model.
Table 4.8: Model Comparison using BIC

<table>
<thead>
<tr>
<th>J</th>
<th>D</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>150</th>
<th>N</th>
<th>300</th>
<th>600</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.448</td>
<td>0.489</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>300</td>
<td>0.897</td>
<td>0.918</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>600</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

RO: Polytomous

<table>
<thead>
<tr>
<th>J</th>
<th>150</th>
<th>N</th>
<th>300</th>
<th>600</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges; R = Response Options. Values represent proportion of successful identification of the correct model as the better fitting model.

Table 4.9: Model Comparison using sBIC

<table>
<thead>
<tr>
<th>J</th>
<th>D</th>
<th>5</th>
<th>10</th>
<th>20</th>
<th>150</th>
<th>N</th>
<th>300</th>
<th>600</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td>0.993</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>300</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>600</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

RO: Polytomous

<table>
<thead>
<tr>
<th>J</th>
<th>150</th>
<th>N</th>
<th>300</th>
<th>600</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges; R = Response Options. Values represent proportion of successful identification of the correct model as the better fitting model.
Table 4.10: Model Comparison using Likelihood-ratio Test

<table>
<thead>
<tr>
<th>R: Dichotomous</th>
<th>5</th>
<th>10</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>±1.5</td>
<td>±1.5</td>
<td>±1.5</td>
</tr>
<tr>
<td>D</td>
<td>±2.5</td>
<td>±2.5</td>
<td>±2.5</td>
</tr>
<tr>
<td>150 N</td>
<td>0.979</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>300 N</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>600 N</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R: Polytomous</th>
</tr>
</thead>
<tbody>
<tr>
<td>150 N</td>
</tr>
<tr>
<td>300 N</td>
</tr>
<tr>
<td>600 N</td>
</tr>
</tbody>
</table>

*Note.* N = Sample size; J = Number of item per factor; D = Item location ranges; R = Response Options. Values represent proportion of successful identification of the correct model as the better fitting model.

### 4.4.3 Summary

One of the goals of this study was to examine the ability of the proposed model (Ordered Categorical Response Unfolding Model; OCRUM) and LMS approach in recovering the model parameters under a range of conditions (independent variables): sample sizes (N), test length (J), item location ranges (D), and response options (R).

Taken together, OCRUM may not perform well when analysing dichotomous response data with small sample size (N = 150) pairing with short test (J = 5) while including extreme items. In general, OCRUM performed well with polytomous response data, and the parameter and standard error bias decreased slightly as a function of sample size and test length; as sample size and test length increase, the bias decreases.

Another evidence pointing towards better performance of OCRUM with polytomous response data is the correlation between the estimated and simulated θ. It has been found that the correlations were consistently above .95 in all conditions when the responses were polytomous. This is in contrast with the dichotomous response data while showed a slightly lower correlation when the test was shorted (J = 5).

Another goal of the study was to investigate the application on AIC, BIC, sBIC, and the likelihood-ratio test in the model comparison procedure. Based on the analysis, AIC, sBIC, and the likelihood-ratio test were found to perform well under different levels of
sample size, test length, response options, and item location ranges, while BIC may not perform well in dichotomous dataset with small sample size (N = 150).

4.5 Study 1b

It is not uncommon for scales to measure more than one dimension. For instance, there are 16 scales in the 16PF instrument (Conn & Rieke, 1994). Another well-known example of a scale measuring more than one factor is the NEO-PI-R (Costa & McCrae, 1992) which measures five major domains of personality. It is also likely that these domains or scales are correlated with each other and the estimation of correlation coefficient should be within an acceptable range of error.

This study wished to extend Study 1a by examining the capability of the model and estimator to recover the parameters of the proposed model (OCRUM) when there is more than one factor involved in a measurement. The parameter of interest in this study was the correlation between the factors (ρ). This study only considered a two-factor independent clusters factor model, where each item is an indicator of one, and only one factor (i.e. without cross-loadings; McDonald, 1999). To illustrate, considering a four-item scale measuring two factors, where two items measure a single factor, OCRUM (refer to equation 3.2) could be modified as the following

\[
Y^*_1 = a_1 + t_{11} \ast (\theta_{i1} - \delta_{11})^2 + 0 \ast (\theta_{i2} - 0)^2 + \varepsilon_{i1}
\]

\[
Y^*_2 = a_2 + t_{21} \ast (\theta_{i1} - \delta_{21})^2 + 0 \ast (\theta_{i2} - 0)^2 + \varepsilon_{i2}
\]

\[
Y^*_3 = a_3 + 0 \ast (\theta_{i1} - 0)^2 + t_{32} \ast (\theta_{i2} - \delta_{32})^2 + \varepsilon_{i3}
\]

\[
Y^*_4 = a_4 + 0 \ast (\theta_{i1} - 0)^2 + t_{42} \ast (\theta_{i2} - \delta_{42})^2 + \varepsilon_{i4}
\]

(4.7)

where \(Y^*_1\) and \(Y^*_2\) are the indicators of Factor 1, and \(Y^*_3\) and \(Y^*_4\) are the indicators of Factor 2. It is also assumed that the items are unidimensional, hence the values for \(\delta_{j2}\) in \(Y^*_1\) and \(Y^*_2\) and \(\delta_{j1}\) in \(Y^*_3\) and \(Y^*_4\) equal to zero. Equation 4.7 can be expressed in general as

\[
Y^*_{ij} = a_j + t_{j1} \ast (\theta_{i1} - \delta_{j1})^2 + t_{j2} \ast (\theta_{i2} - \delta_{j2})^2 + \varepsilon_{ij}
\]

(4.8)
where 
\( Y_{ij}^* \) is the latent response of person \( i \) for item \( j \),
\( t_{j1} \) is the curvature parameter for the \( j \)th item which measures the 1st factor,
\( t_{j2} \) is the curvature parameter for the \( j \)th item which measures the 2nd factor,
\( \delta_{j1} \) is the item location parameter for the \( j \)th item on the 1st dimension (factor), and
\( \delta_{j2} \) is the item location parameter for the \( j \)th item on the 2nd dimension (factor).

4.5.1 Method

Research Design

The study used a between-subjects design, outlined by 2 (sample size: \( N = 300; 600 \)) x 2 (items per factor: \( J = 10; 20 \)) x 2 (item location range: \( D = \pm 1.5; \pm 2.5 \)) x 3 (factor correlations: \( \rho = 0.3; 0.5; 0.7 \)). Polytomous responses, simulating responses from a 6-point rating scale, was simulated for all 18 conditions. Two hundred replications were done for each independent cell. The dependent variables reported in this study were: the \( \rho \) parameter bias and \( \rho \) standard error bias.

The \( \delta_j \) parameter bias and standard error bias were not reported as the pattern were found to be similar with the result of Study 1a. Due to the constrain of the page limit of this thesis, the results on the parameter and SE bias of \( \delta_j \) in Study 1b were not reported.

Previous simulations have found a relatively higher parameter bias in smaller sample size (\( N = 150 \)) and also in dichotomous responses (\( R = 2 \)), hence these two conditions were not considered in Study 1b. The three levels of \( \rho \) was selected to represent medium (\( \rho = .30 \)), large (\( \rho = .50 \)), and relatively larger (\( \rho = .70 \)) relationship according to Cohen’s (1992) table of effect size magnitude.

Data Generation

The process of simulating the categorical response data for Study 1b was as outlined in subsection 4.1 with slight modification. Instead of generating \( \Theta \) from a standard normal distribution, the two correlated \( \Theta \)s were simulated from a multivariate normal
distribution, with $\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ and with the specific $\rho$ values depending on the conditions. The data generation MATLAB script can be found in Appendix E on page 122.

**Procedure**

All simulated datasets were fitted with OCRUM (see equation 4.8) in Mplus. The Mplus syntax can be found in Appendix F.

**4.5.2 Results**

**Successful Replications**

The number of successful replication (i.e. convergence) is reported in Table 4.11 as the first value in each cell. The second value represents the number of non-convergence, which is defined as the number of datasets which failed to fit the model in Mplus. The third value represents the number of replications with convergence but with error, which is defined as the number of datasets which fitted the model but with at least one fixed parameter in Mplus.

As expected from the conditions of moderate sample size ($N \leq 300$) and polytomous response options, almost all of the replications converged successfully. The convergence rate was also similar across the different correlations between the two latent traits.

**Parameter Bias**

The mean of the $\rho$ parameter bias across the different conditions is described in Table 4.12. The values in the bracket contains standard deviation of the bias. The mean values and their respective standard deviation values were considered small.

A between-subjects ANOVA with sample size ($N$), items per factor ($J$), item location range ($D$) and factor correlations ($\rho$) as the independent variables (IVs) and absolute $\rho$ parameter bias as the dependent variable was conducted. Based on $\eta^2_p$, all the interaction and main effects of the IVs on parameter bias were considered as small in magnitude (the largest $\eta^2_p$ is 0.006). We can conclude that sample size, test length, item location ranges,
### Table 4.11: Study 1b Successful Replications

<table>
<thead>
<tr>
<th>ρ</th>
<th>J</th>
<th>D</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±2.5</td>
</tr>
<tr>
<td>ρ: 0.3</td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>300</td>
<td>199/1/0</td>
<td>200/0/0</td>
<td>200/0/0</td>
<td>200/0/0</td>
<td>199/0/1</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>199/1/0</td>
<td>200/0/0</td>
<td>199/1/0</td>
<td>200/0/0</td>
<td>200/0/0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ρ: 0.5</th>
<th>J</th>
<th>D</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±2.5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>300</td>
<td>200/0/0</td>
<td>199/0/1</td>
<td>200/0/0</td>
<td>199/0/1</td>
<td>200/0/0</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>200/0/0</td>
<td>200/0/0</td>
<td>200/0/0</td>
<td>200/0/0</td>
<td>200/0/0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ρ: 0.7</th>
<th>J</th>
<th>D</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±2.5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>300</td>
<td>200/0/0</td>
<td>199/0/1</td>
<td>199/1/0</td>
<td>199/0/1</td>
<td>200/0/0</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>200/0/0</td>
<td>200/0/0</td>
<td>199/1/0</td>
<td>200/0/0</td>
<td>200/0/0</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges; ρ = Correlation between factors. The first value represents the number of successful convergence. The second value represents the number of non-convergence and the third value represents the number of convergence but with error.

and factor correlation magnitude have little effect on the recovery of the correlation itself. The ANOVA results are presented in Table 4.13.

### Table 4.12: Mean and Standard Deviation of ρ Parameter Bias

<table>
<thead>
<tr>
<th>ρ: 0.3</th>
<th>J</th>
<th>D</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±2.5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>300</td>
<td>-0.001 (0.056)</td>
<td>0.008 (0.064)</td>
<td>0.003 (0.055)</td>
<td>0.000 (0.057)</td>
<td>0.001 (0.054)</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>-0.035 (0.498)</td>
<td>-0.001 (0.041)</td>
<td>0.000 (0.037)</td>
<td>0.004 (0.038)</td>
<td>0.003 (0.035)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ρ: 0.5</th>
<th>J</th>
<th>D</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±2.5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>300</td>
<td>0.002 (0.044)</td>
<td>0.003 (0.046)</td>
<td>0.004 (0.047)</td>
<td>0.001 (0.048)</td>
<td>0.006 (0.045)</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>-0.001 (0.034)</td>
<td>-0.003 (0.031)</td>
<td>0.001 (0.035)</td>
<td>0.000 (0.030)</td>
<td>-0.002 (0.029)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ρ: 0.7</th>
<th>J</th>
<th>D</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±2.5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>300</td>
<td>0.003 (0.035)</td>
<td>0.002 (0.033)</td>
<td>0.003 (0.032)</td>
<td>-0.016 (0.215)</td>
<td>0.006 (0.030)</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>-0.001 (0.023)</td>
<td>-0.001 (0.023)</td>
<td>0.003 (0.022)</td>
<td>0.002 (0.020)</td>
<td>-0.001 (0.021)</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges; ρ = Correlation between factors. Values represent the mean of the parameter bias and values in bracket represent the standard deviation of the parameter bias.

### Standard Error Bias

The mean of the ρ standard error bias across the different conditions is presented in Table 4.14. The values in the bracket contain standard deviation of the bias. Similar to
Table 4.13: Effect of Independent Variables on Absolute $\rho$ Parameter Bias

<table>
<thead>
<tr>
<th></th>
<th>Degree of freedom</th>
<th>Sum of Squares</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1</td>
<td>0.192</td>
<td>0.003</td>
</tr>
<tr>
<td>J</td>
<td>2</td>
<td>0.038</td>
<td>0.001</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>$\rho$</td>
<td>2</td>
<td>0.390</td>
<td>0.006</td>
</tr>
<tr>
<td>N*J</td>
<td>2</td>
<td>0.020</td>
<td>0.000</td>
</tr>
<tr>
<td>N*D</td>
<td>1</td>
<td>0.022</td>
<td>0.000</td>
</tr>
<tr>
<td>N*$\rho$</td>
<td>2</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>J*D</td>
<td>2</td>
<td>0.013</td>
<td>0.000</td>
</tr>
<tr>
<td>J*$\rho$</td>
<td>4</td>
<td>0.059</td>
<td>0.001</td>
</tr>
<tr>
<td>D*$\rho$</td>
<td>2</td>
<td>0.010</td>
<td>0.000</td>
</tr>
<tr>
<td>N<em>J</em>D</td>
<td>2</td>
<td>0.022</td>
<td>0.000</td>
</tr>
<tr>
<td>N<em>J</em>$\rho$</td>
<td>4</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>N<em>D</em>$\rho$</td>
<td>2</td>
<td>0.012</td>
<td>0.000</td>
</tr>
<tr>
<td>J<em>D</em>$\rho$</td>
<td>4</td>
<td>0.019</td>
<td>0.000</td>
</tr>
<tr>
<td>N<em>J</em>D*$\rho$</td>
<td>4</td>
<td>0.055</td>
<td>0.001</td>
</tr>
<tr>
<td>Error</td>
<td>7152</td>
<td>61.133</td>
<td></td>
</tr>
</tbody>
</table>

$N =$ Sample size; $J =$ Number of item per factor; $D =$ Item location ranges; $\rho =$ Correlation between factors. Interaction effects are denoted by the asterisk symbol (*), for example, the interaction between N and J is denoted as N*J.
the parameter bias in Table 4.12, the mean standard error bias and its associated standard deviation were relatively small.

A between-subjects ANOVA with sample size (N), items per factor (J), item location range (D) and factor correlations (ρ) as the independent variables and absolute ρ standard error bias as the dependent variable was conducted. The ANOVA results were partially reported in Table 4.15. The largest η² effect size was 0.120 (the main effect of N) which is still considered as a small effect according to Cohen’s (1988) rule of thumb. Hence, it could be concluded that the standard error of the correlation is not affected by four independent variables (sample size, test length, item location range, and factor correlations) considered in the ANOVA model.

Table 4.14: Mean and Standard Deviation of ρ Standard Error Bias of the Independent Clusters Model

<table>
<thead>
<tr>
<th></th>
<th>ρ: 0.3</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
<td>±1.5</td>
<td>±2.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.000 (0.004)</td>
<td>-0.001 (0.004)</td>
<td>-0.001 (0.004)</td>
<td>-0.002 (0.004)</td>
<td>-0.000 (0.004)</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>0.000 (0.002)</td>
<td>-0.001 (0.002)</td>
<td>0.001 (0.002)</td>
<td>0.000 (0.002)</td>
<td>0.003 (0.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>0.003 (0.004)</td>
<td>0.002 (0.006)</td>
<td>-0.003 (0.004)</td>
<td>-0.003 (0.004)</td>
<td>-0.001 (0.004)</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>-0.001 (0.002)</td>
<td>0.002 (0.002)</td>
<td>-0.003 (0.002)</td>
<td>-0.001 (0.003)</td>
<td>0.002 (0.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>-0.003 (0.004)</td>
<td>0.000 (0.005)</td>
<td>-0.001 (0.003)</td>
<td>0.001 (0.003)</td>
<td>0.000 (0.003)</td>
</tr>
<tr>
<td></td>
<td>600</td>
<td>0.000 (0.002)</td>
<td>0.001 (0.002)</td>
<td>0.000 (0.001)</td>
<td>0.001 (0.002)</td>
<td>0.000 (0.001)</td>
</tr>
</tbody>
</table>

Note. N = Sample size; J = Number of item per factor; D = Item location ranges; ρ = Correlation between factors. Values represent the mean of the standard error bias and values in bracket represent the standard deviation of the standard error bias.

4.5.3 Summary

Table 4.11 provided us with confidence in Mplus in successfully fitting the OCRUM with more than one dimension in most cases. Furthermore, regardless of the conditions considered in this simulation study (i.e. sample size, test length, item location ranges, and correlation between factors), factor correlation ρ was successfully recovered with little bias. This has shown that Mplus can accurately estimate the correlation between the latent factors in the OCRUM with LMS method for a independent clusters model.

On top of that, the standard error of ρ was also estimated with small to no bias across
Table 4.15: Effect of Independent Variables on Absolute $\rho$
Standard Error Bias of the Independent Clusters Model

<table>
<thead>
<tr>
<th>Degree of freedom</th>
<th>Sum of Squares</th>
<th>$\eta^2_{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1</td>
<td>0.005</td>
</tr>
<tr>
<td>J</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>D</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>$\rho$</td>
<td>2</td>
<td>0.001</td>
</tr>
<tr>
<td>N*J</td>
<td>2</td>
<td>0.001</td>
</tr>
<tr>
<td>N*D</td>
<td>1</td>
<td>0.000</td>
</tr>
<tr>
<td>N*$\rho$</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>J*D</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>J*$\rho$</td>
<td>4</td>
<td>0.000</td>
</tr>
<tr>
<td>D*$\rho$</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>N<em>J</em>D</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>N*$J^\ast$D</td>
<td>4</td>
<td>0.000</td>
</tr>
<tr>
<td>N<em>D</em>$\rho$</td>
<td>2</td>
<td>0.001</td>
</tr>
<tr>
<td>J<em>D</em>$\rho$</td>
<td>4</td>
<td>0.001</td>
</tr>
<tr>
<td>N<em>J</em>D*$\rho$</td>
<td>4</td>
<td>0.000</td>
</tr>
<tr>
<td>Error</td>
<td>7152</td>
<td>0.034</td>
</tr>
</tbody>
</table>

N = Sample size; J = Number of item per factor; D = Item location ranges; $\rho$ = Correlation between factors. Interaction effects are denoted by the asterisk symbol (*), for example, the interaction between N and J is denoted as N*J.
all conditions in this study. An accurate estimation of the standard error is important as it affects the inferential statistics (i.e. t-test). An over-estimation of the standard error of a statistics would result in higher Type-II error and an under-estimation of the standard error would result in a higher Type-I error. In this study, we can conclude that there is no standard error bias and thus inferences based on t-test for the \( \hat{\rho} \) should be accurate within the conditions of the study.

This study has provided us with the evidence that the correlation coefficient between factors and its standard error could be accurately estimated in Mplus, further showcasing the plausibility of modelling the OCRUM in Mplus.
Chapter 5

Study 2: Re-analysis of Empirical Data

The aim of this study was to evaluate the use of the proposed model (Ordered Categorical Response Unfolding Model; OCRUM) with two sets of empirical data and also to compare the performance of the OCRUM with the Generalised Graded Unfolding Model (GGUM) in terms of the estimations of the locations for both the items ($\delta_j$) and person ($\theta_i$). The GGUM was chosen as a comparison model because GGUM is one of the commonly used models in modelling unfolding data in the current literature (e.g. Polak et al., 2009; Tay et al., 2009; Tay, 2011; Weekers & Meijer, 2008). Published results regarding the item locations were also included in the comparison.

Such comparison is necessary when a new model is proposed, according to Andrich (1988), results obtained from a new proposed model should be accompanied by a comparison of results from traditional methods in the development of any new models to investigate the similarities and differences in analysis outcomes.

5.1 General Method

5.1.1 Empirical Datasets

Two empirical datasets were re-analysed in this study: polytomous responses of Thurstone’s (1932) 24-item Attitude Toward Capital Punishment Scale collected by Roberts (1995), and the Ideal Point IRT Order Scale (hereafter referred to as the Order
Scale) to measure the Order facet of the Conscientiousness personality domain (Chernyshenko et al., 2007). Description of the two scales were given in section 5.2 and section 5.3 respectively.

5.1.2 Starting Values in OCRUM

As recalled from previous chapter, in order to conform to the single-peaked response function in equation 3.2, the \( t_j \) parameter has to be a negative value. We chose to providing a starting value of -1 for all the \( t_j \) parameters.

Starting values were also provided for the item location \( \delta_j \) parameters. In order to obtain a reasonable starting value, a linear measurement model (see equation 2.7) was fitted to the data. The \( \hat{\lambda}_j \) was then used as the starting value of the \( \delta_j \) parameters.  

5.1.3 GGUM2004

The generalised graded unfolding model (GGUM) parameters were estimated using the GGUM2004 software (Roberts et al., 2000, 2006). The software estimated the item parameters with the marginal maximum likelihood (MML; or referred to as the Full Information Maximum Likelihood (FIML) in the SEM literature; Forero & Maydeu-Olivares, 2009), while the person location was estimated with the Expected a Posteriori (EAP) procedure (Roberts et al., 2006). The software defaults were used for estimation as accord to the recommendation documented in the GGUM2004 user manual (Roberts et al., 2006).

5.1.4 Analysis

Pearson correlation (\( r_p \)) and Spearman rank correlation (\( r_s \)) were used to estimate the relationship between the estimated item location (and also between person location) between the estimates from OCRUM (from Mplus) and GGUM (from GGUM2004). \( r_p \)

\[ 1 \]

\[ \text{While the relationship between } \lambda_j \text{ (see equation 2.7 and } \delta_j \text{ have not been shown mathematically in any part of this research, based on observations from previous simulations studies, the sign of } \hat{\lambda}_j \text{ corresponded to the item locations, where items with negative } \lambda_j \text{ were located at one end of the latent continuum, and items with positive } \lambda_j \text{ at the opposite end, similar to the interpretation of the signs of } \delta_j. \]

75
was used as an indicator of the similarities of both spacing and ordering of the estimated item (and person) locations, while $r_s$ was used as an indicator of the similarity of the ordering of the estimated item (and person) locations (Polak, 2011).

### 5.2 Empirical Dataset 1: Attitude Toward Capital Punishment Scale

Polytomous response data to Thurstone (1932) 24-item Attitude Toward Capital Punishment Scale was collected by Roberts (1995) from 245 University of South Carolina undergraduates. The order of the items were presented randomly to the respondents and participants responded using a six-point rating scale (1 = strongly disagree, 2 = disagree, 3 = slightly disagree, 4 = slightly agree, 5 = agree, and 6 = strongly agree).

The data has been analysed by Roberts and Laughlin (1996) using the Graded Unfolding Model (GUM; i.e. a "rating scale" version of GGUM, with the $\tau_{jk}$ constrained to be equal across items and $\alpha_j$ fixed to be 1 across items). On the other hand, Polak (2011) used the Correspondence Analysis (CA) to analyse the same set of data.

Before fitting the data to GUM, Roberts and Laughlin (1996) performed a PCA on the data and deleted seven items which were likely to be not unidimensional. An addition of five item from the remaining 17 items were further removed based on Wright and Masters’ (1982) item-fit $t$-statistics obtained from the initial model fitting procedure. Roberts and Laughlin (1996) procedure concluded with a 12-item attitude scale.

On the other hand, Polak (2011) used a iterative procedure in identifying "deviant" items: items with extreme location values based on the CA solutions were first eliminated followed by an inspection of the response frequency of the "strongly agree" option. Items with low endorsement on this option were removed. The responses to the remaining items were subjected to the CA and the procedure was repeated until no item was removed. Polak’s (2011) procedure eliminated three items and concluded with a 21-item attitude scale.

In this study, we analysed both the 12-item and 21-item Attitude toward Capital
Punishment scale. It is noteworthy to point out that the three "deviant" items identified in Polak (2011) were also removed in Roberts and Laughlin (1996). The items were listed in Table 5.1. There were no missing data in the dataset.

5.2.1 Results

In this section, the results of the analysis of the 12-item and 21-item Attitude Towards Capital Punishment Scale Thurstone (1932) described above were reported.

Table 5.1 (see page 78) contained the original Thurstone scale value (i.e. item locations; \( T_j \)), item location estimates from the OCRUM (\( \hat{\delta}_{12}^{OCRUM} \) and \( \hat{\delta}_{21}^{OCRUM} \)), and estimates from GGUM (\( \hat{\delta}_{12}^{GGUM} \) and \( \hat{\delta}_{21}^{GGUM} \)). The superscript 12 and 21 indicated the 12-item scale and the 21-item scale respectively. The values in brackets are the standard error of the estimations.

Comparison of Item Location (\( \hat{\delta}_j \))

Four different comparisons were conducted. Firstly, the original Thurstone scale value with each analysis (i.e. \( \hat{\delta}_j \) of the 12-item scale and 21-item scale of the OCRUM and GGUM) was compared. Secondly, \( \hat{\delta}_{12}^{OCRUM} \) and \( \hat{\delta}_{12}^{GGUM} \) were compared. Thirdly, comparison between the \( \hat{\delta}_j \) of the 21-item scale obtained from the OCRUM and GGUM were done. Lastly, the relationship of the item location estimates of the 12-item scale and the item locations of the respective 12 items from the 21-item scale was examined.

The entries above the diagonal in Table 5.2 are the Spearman rank correlation coefficients while the entries below the diagonal are the Pearson correlation coefficients. The first column (and row) of Table 5.2 showed that \( \hat{\delta}_{12}^{OCRUM} \), \( \hat{\delta}_{12}^{GGUM} \), \( \hat{\delta}_{21}^{OCRUM} \), and \( \hat{\delta}_{21}^{GGUM} \) were highly correlated with the scale values derived by Thurstone (\(|r_p|\) ranged between .79 and .92; \(|r_s|\) ranged between .82 and .90), indicating the resulting order of the items were highly similar, albeit different direction, across different methods (i.e. Thurstone, OCRUM, and GGUM) and across different scale length (i.e. 12-item and 21-item scale).

The absolute value of the correlation coefficient was interpreted instead of the original estimated values because inter-item distance would not be affect by the sign (i.e. negative
Table 5.1: Thurstone’s (1932) Attitude Toward Capital Punishment

<table>
<thead>
<tr>
<th>Item</th>
<th>Statement</th>
<th>Tj</th>
<th>( \hat{\delta}_{ij}^{OCRUM} )</th>
<th>( \hat{\delta}_{ij}^{GGUM} )</th>
<th>( \hat{\delta}_{ij}^{OCRUM} )</th>
<th>( \hat{\delta}_{ij}^{GGUM} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP02</td>
<td>Capital punishment is absolutely never justified.</td>
<td>0.6</td>
<td>10.860 (1.565)</td>
<td>3.823 (9.654)</td>
<td>10.846 (2.002)</td>
<td>3.886 (10.298)</td>
</tr>
<tr>
<td>CP12</td>
<td>I do not believe in capital punishment under any circumstances.</td>
<td>0.1</td>
<td>2.209 (0.678)</td>
<td>1.933 (0.079)</td>
<td>2.717 (1.601)</td>
<td>3.443 (8.103)</td>
</tr>
<tr>
<td>CP19</td>
<td>Capital punishment is the most hideous practice of our time.</td>
<td>0.6</td>
<td>a</td>
<td>a</td>
<td>2.604 (0.805)</td>
<td>2.196 (0.242)</td>
</tr>
<tr>
<td>CP16</td>
<td>Execution of criminals is a disgrace to civilized society.</td>
<td>0.9</td>
<td>3.563 (1.036)</td>
<td>2.896 (0.721)</td>
<td>4.127 (1.825)</td>
<td>3.835 (9.818)</td>
</tr>
<tr>
<td>CP14</td>
<td>We can’t call ourselves civilized as long as we have capital punishment.</td>
<td>1.5</td>
<td>9.516 (1.821)</td>
<td>3.39 (5.700)</td>
<td>8.273 (2.173)</td>
<td>2.593 (0.241)</td>
</tr>
<tr>
<td>CP21</td>
<td>The state cannot teach the sacredness of human life by destroying it.</td>
<td>2.0</td>
<td>a</td>
<td>a</td>
<td>9.584 (1.615)</td>
<td>3.398 (6.074)</td>
</tr>
<tr>
<td>CP05</td>
<td>Capital punishment cannot be regarded as a sane method of dealing with crime.</td>
<td>2.4</td>
<td>a</td>
<td>a</td>
<td>7.837 (3.643)</td>
<td>2.272 (0.131)</td>
</tr>
<tr>
<td>CP08</td>
<td>Capital punishment has never been effective in in preventing crime.</td>
<td>2.7</td>
<td>a</td>
<td>a</td>
<td>8.777 (2.275)</td>
<td>5.054 (22.309)</td>
</tr>
<tr>
<td>CP13</td>
<td>Capital punishment is not necessary in modern civilization.</td>
<td>3.0</td>
<td>9.628 (1.428)</td>
<td>3.739 (8.475)</td>
<td>7.748 (2.425)</td>
<td>2.623 (0.766)</td>
</tr>
<tr>
<td>CP09</td>
<td>I don’t believe in capital punishment but I’m not sure it isn’t necessary.</td>
<td>3.4</td>
<td>2.018 (0.868)</td>
<td>1.278 (0.156)</td>
<td>1.716 (0.731)</td>
<td>1.269 (0.169)</td>
</tr>
<tr>
<td>CP15</td>
<td>Life imprisonment is more effective than capital punishment.</td>
<td>3.4</td>
<td>4.914 (3.940)</td>
<td>2.130 (0.317)</td>
<td>3.761 (2.365)</td>
<td>2.064 (0.219)</td>
</tr>
<tr>
<td>CP11</td>
<td>I think the return of the whipping post would be more effective than capital punishment.</td>
<td>3.9</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>CP22</td>
<td>It doesn’t make any difference to me whether we have capital punishment or not.</td>
<td>5.5</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>CP18</td>
<td>I do not believe in capital punishment but I’m not sure it is necessary to abolish it.</td>
<td>5.8</td>
<td>11.0 (0.327)</td>
<td>10.2 (2.124)</td>
<td>8.8 (1.800)</td>
<td>0.9 (0.182)</td>
</tr>
<tr>
<td>CP01</td>
<td>Capital punishment may be necessary but I wish it were not.</td>
<td>7.2</td>
<td>a</td>
<td>a</td>
<td>5.7 (1.411)</td>
<td>3.6 (0.237)</td>
</tr>
<tr>
<td>CP23</td>
<td>Capital punishment is the most hideous practice of our time.</td>
<td>7.8</td>
<td>18.0 (3.827)</td>
<td>10.3 (2.761)</td>
<td>7.5 (1.761)</td>
<td>0.9 (0.182)</td>
</tr>
<tr>
<td>CP04</td>
<td>Any person, man or woman, young or old, who commits murder, should pay with his own life.</td>
<td>9.5</td>
<td>a</td>
<td>a</td>
<td>b</td>
<td>b</td>
</tr>
<tr>
<td>CP07</td>
<td>Capital punishment is the most hideous practice of our time.</td>
<td>9.5</td>
<td>11.5 (0.317)</td>
<td>10.0 (2.124)</td>
<td>8.8 (1.800)</td>
<td>0.9 (0.182)</td>
</tr>
<tr>
<td>CP20</td>
<td>Capital punishment gives the criminals what they deserve.</td>
<td>9.8</td>
<td>18.0 (3.827)</td>
<td>10.3 (2.761)</td>
<td>7.5 (1.761)</td>
<td>0.9 (0.182)</td>
</tr>
<tr>
<td>CP17</td>
<td>Capital punishment is just and necessary.</td>
<td>9.6</td>
<td>18.0 (3.827)</td>
<td>10.3 (2.761)</td>
<td>7.5 (1.761)</td>
<td>0.9 (0.182)</td>
</tr>
<tr>
<td>CP03</td>
<td>Capital punishment is the most hideous practice of our time.</td>
<td>10.5</td>
<td>a</td>
<td>a</td>
<td>5.7 (1.411)</td>
<td>3.6 (0.237)</td>
</tr>
</tbody>
</table>

Note: 'a' refers to item not included in Roberts and Laughlin (1996); 'b' refers to item not included in Polak (2011).

Tj = original Thurstone scale value; \( \hat{\delta}_{ij}^{OCRUM} \) = item location estimates from 12-item OCRUM; \( \hat{\delta}_{ij}^{GGUM} \) = item location estimates from 12-item GGUM; \( \hat{\delta}_{ij}^{OCRUM} \) = item location estimates from 12-item OCRUM; \( \hat{\delta}_{ij}^{GGUM} \) = item location estimates from 21-item GGUM.
and positive) of the item location \( \hat{\delta}_j \). The sign only affects the interpretation of the two ends of the scale. In the case of the Thurstone’s (1932) item scale values, a lower scale value indicates a negative attitude towards capital punishment while a higher scale value indicates a positive attitude towards capital punishment. In contrast, a lower OCRUM and GGUM \( \delta_j \) value indicates a positive attitude towards capital punishment and a higher \( \delta_j \) is interpreted as a negative attitude towards capital punishment.

Furthermore, both \( r_p \) and \( r_s \) between \( \hat{\delta}_{12}^{OCRUM} \) and \( \hat{\delta}_{12}^{GGUM} \) were larger than .90 (\( r_p = .91 \) and \( r_s = .97 \) respectively). Similarly, the correlation of the \( \hat{\delta}_{21} \) between the two models were larger than or equal to .90 as well (\( r_p = .90 \) and \( r_s = .94 \) respectively). All correlation coefficients were statistically significant at the \( \alpha = .05 \) level.

Table 5.2: Pairwise correlation of the item locations for the Attitude Toward Capital Punishment scale

<table>
<thead>
<tr>
<th></th>
<th>( T_j )</th>
<th>( \hat{\delta}_{12}^{OCRUM} )</th>
<th>( \hat{\delta}_{12}^{GGUM} )</th>
<th>( \hat{\delta}_{12}^{OCRUM} )</th>
<th>( \hat{\delta}_{21}^{GGUM} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_j )</td>
<td>-</td>
<td>-.809</td>
<td>-.869</td>
<td>-.841</td>
<td>-.904</td>
</tr>
<tr>
<td>( \hat{\delta}_{OCRUM} )</td>
<td>-</td>
<td>-</td>
<td>.972</td>
<td>.986</td>
<td>a</td>
</tr>
<tr>
<td>( \hat{\delta}_{GGUM} )</td>
<td>-.918</td>
<td>.912</td>
<td>-</td>
<td>a</td>
<td>.902</td>
</tr>
<tr>
<td>( \hat{\delta}_{OCRUM} )</td>
<td>-.794</td>
<td>.985</td>
<td>a</td>
<td>-</td>
<td>.940</td>
</tr>
<tr>
<td>( \hat{\delta}_{GGUM} )</td>
<td>-.916</td>
<td>a</td>
<td>.927</td>
<td>.900</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. ‘a’ are coefficients that are not of interest of this comparison. \( T_j \) = original Thurstone scale value; \( \hat{\delta}_{OCRUM} \) = item location estimates from 12-item OCRUM; \( \hat{\delta}_{GGUM} \) = item location estimates from 12-item GGUM; \( \hat{\delta}_{OGUM} \) = item location estimates from 12-item GGUM; \( \hat{\delta}_{GGUM} \) = item location estimates from 21-item GGUM.

Regarding the last comparison, the correlation between the item location estimates (\( \hat{\delta}_j \)) of the 12-item scale and the subset of the same 12 items from the 21-item scale were found to be high as well (OCRUM: \( r_p = .99 \), \( r_s = .99 \); GGUM: \( r_p = .93 \), \( r_s = .90 \)). All four correlation coefficients were statistically significant at the \( \alpha = .05 \) level.

**Comparison of Response Function**

Item CP06 from the 21-item scale and CP18 from the 12-item scale served as examples of the theoretical (category) response function for GGUM and OCRUM. As illustrated in
Figure 5.1a and Figure 5.1b, the rate of change of the probability is starkly different, with OCRUM having a higher rate of change, which is characterised by the steepness of each of the category response functions, yet the peak of the middle categories in OCRUM was rounder as compared to GGUM. Similar observations of the response functions can be made between CP18 in Figure 5.2a and Figure 5.2b.

**Comparison of Person Location (θᵢ)**

Similar to the analysis of item locations, the Pearson ($r_p$) and Spearman rank correlation ($r_s$) between the estimated person locations of both models were estimated for the 12-item scale and the 21-item scale separately. For the 12-item scale, the correlations between the $\hat{\theta}_i$ estimated under the GGUM and the OCRUM were high ($r_p = .95$, $p < .001$; $r_s = .99$, $p < .001$). Similarly, the correlation between 21-item $\hat{\theta}_i$ estimated under the GGUM and the OCRUM were equally high ($r_p = .98$, $p < .001$; $r_s = .99$, $p < .001$). The distribution of the $\hat{\theta}_i$ and their scatter plots were displayed in Figure 5.3.

### 5.3 Empirical Dataset 2: Order Scale

The Order Scale was developed by Chernyshenko et al. (2007) to measure the Order facet of the Conscientiousness personality domain based on the unfolding response model. Response data for the 20-item scale was collected from 539 undergraduate students from an American university using a four-point scale (1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree).

In this study, we only used cases with complete data. Approximately 6% of the sample were deleted and the sample size was reduced to N = 505. On top of that, to be in line with the analysis approach used in Chernyshenko et al. (2007), due to infrequent endorsement of the two extreme options (see Table 5.3 for endorsement rates), the polytomous responses were dichotomised before analysis, where strongly disagree and disagree were collapsed and recoded as 0, while agree and strongly agree were collapsed and recoded as 1.
Figure 5.1: Theoretical response function of item CP06 according to (a) GGUM and (b) OCRUM

(a) $\hat{\delta}^{21}_{GGUM} = -0.210$

(b) $\hat{\delta}^{21}_{OCRUM} = -0.170$
Figure 5.2: Theoretical response function of item CP18 according to (a) GGUM and (b) OCRUM

(a) $\hat{\delta}^{12}_{GGUM} = 1.021$

(b) $\hat{\delta}^{12}_{OCRUM} = 1.180$
Figure 5.3: Histograms and scatter plots with linear fit line of the $\hat{\theta}$ obtained from OCRUM (x-axis) and GGUM (y-axis) for the 12-item (a) and the 21-item (b) Thurstone’s (1932) Attitude Toward Capital Punishment Scale.
Table 5.3: Chernyshenko et al.’s (2007) Order Scale Polytomous Response Endorsement Rate

<table>
<thead>
<tr>
<th>Statement</th>
<th>SD</th>
<th>D</th>
<th>A</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORD02 Usually, my notes are so jumbled, even I have a hard time reading them.</td>
<td>177</td>
<td>258</td>
<td>59</td>
<td>11</td>
</tr>
<tr>
<td>ORD06 Most of the time my room is in complete disarray.</td>
<td>178</td>
<td>243</td>
<td>68</td>
<td>16</td>
</tr>
<tr>
<td>ORD10 I frequently forget to put things back in their proper place.</td>
<td>70</td>
<td>268</td>
<td>149</td>
<td>18</td>
</tr>
<tr>
<td>ORD14 I do not like work spaces that are too clean and tidy.</td>
<td>160</td>
<td>279</td>
<td>59</td>
<td>7</td>
</tr>
<tr>
<td>ORD15 For me, being organized is unimportant.</td>
<td>156</td>
<td>270</td>
<td>73</td>
<td>6</td>
</tr>
<tr>
<td>ORD17 Being neat is not exactly my strength.</td>
<td>111</td>
<td>194</td>
<td>156</td>
<td>44</td>
</tr>
<tr>
<td>ORD24 Half of the time I do not put things in their proper place.</td>
<td>86</td>
<td>223</td>
<td>178</td>
<td>18</td>
</tr>
<tr>
<td>ORD30 Although I have a daily organizer, I have hard time keeping it up to date.</td>
<td>117</td>
<td>182</td>
<td>158</td>
<td>48</td>
</tr>
<tr>
<td>ORD20 I do pretty standard maintenance for my property and possessions.</td>
<td>4</td>
<td>45</td>
<td>375</td>
<td>81</td>
</tr>
<tr>
<td>ORD23 My room neatness is about average.</td>
<td>41</td>
<td>109</td>
<td>322</td>
<td>33</td>
</tr>
<tr>
<td>ORD26 My ability to plan is at about average.</td>
<td>48</td>
<td>187</td>
<td>263</td>
<td>7</td>
</tr>
<tr>
<td>ORD29 Although I try to keep everything in its place, it does not always work for me.</td>
<td>12</td>
<td>143</td>
<td>310</td>
<td>40</td>
</tr>
<tr>
<td>ORD37 I prefer to do things in a logical order.</td>
<td>5</td>
<td>52</td>
<td>344</td>
<td>104</td>
</tr>
<tr>
<td>ORD42 I write notes to myself only if I have too many things to do at once.</td>
<td>28</td>
<td>122</td>
<td>283</td>
<td>72</td>
</tr>
<tr>
<td>ORD35 I need a neat environment in order to work well.</td>
<td>27</td>
<td>158</td>
<td>243</td>
<td>77</td>
</tr>
<tr>
<td>ORD38 Organization is a key component of most things I do.</td>
<td>13</td>
<td>142</td>
<td>256</td>
<td>94</td>
</tr>
<tr>
<td>ORD34 I have a daily routine and stick to it.</td>
<td>25</td>
<td>246</td>
<td>210</td>
<td>24</td>
</tr>
<tr>
<td>ORD44 I become annoyed when things around me are disorganized.</td>
<td>16</td>
<td>121</td>
<td>275</td>
<td>93</td>
</tr>
<tr>
<td>ORD46 I keep detailed notes of important meetings and lectures.</td>
<td>25</td>
<td>144</td>
<td>278</td>
<td>58</td>
</tr>
<tr>
<td>ORD50 Every item in my room and on my desk has its own designated place.</td>
<td>62</td>
<td>183</td>
<td>196</td>
<td>64</td>
</tr>
</tbody>
</table>

Note: SD = Strongly Disagree, D = Disagree, A = Agree, and SA = Strongly Agree. The numbers represent the number of sample who endorse the corresponding response option for the items.

5.3.1 Results

In this section, the results of the analysis of the 20-item Order Scale (Chernyshenko et al., 2007) described in the previous section were reported. The analyses were similar to the previous analyses done on the Attitude Toward Capital Punishment Scale (Thurstone, 1932).

Table 5.4 contained the averaged item location ratings ($R_j$) across two raters as reported in Chernyshenko et al. (2007), item location estimates from the OCRUM ($\hat{\delta}_{OCRUM,j}$) and GGUM ($\hat{\delta}_{GGUM,j}$). The values in brackets are the standard error of the estimations.

Comparison of Item Location $\hat{\delta}_j$

Similar to the analysis of Thurstone’s (1932) Attitude Toward Capital Punishment Scale in the previous section, two different comparisons were conducted. Firstly, the raters’ averaged rating ($R_j$) with each analysis (i.e. $\hat{\delta}_j$ from OCRUM and GGUM) was compared. Secondly, the $\hat{\delta}_j$ of scale obtained from the OCRUM and GGUM was compared.

Table 5.5 contained the correlation coefficients of the item locations estimated from
Table 5.4: Chernyshenko et al.’s (2007) Order Scale

<table>
<thead>
<tr>
<th>Statement</th>
<th>( R_j )</th>
<th>( \hat{\delta}_{\text{OCRUM}} )</th>
<th>( \hat{\delta}_{\text{GGUM}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORD02 Usually, my notes are so jumbled, even I have a hard time reading them.</td>
<td>1.0</td>
<td>7.566 (0.946)</td>
<td>3.430 (5.342)</td>
</tr>
<tr>
<td>ORD06 Most of the time my room is in complete disarray.</td>
<td>1.5</td>
<td>2.628 (1.175)</td>
<td>1.895 (0.184)</td>
</tr>
<tr>
<td>ORD10 I frequently forget to put things back in their proper place.</td>
<td>2.5</td>
<td>1.452 (0.193)</td>
<td>1.293 (0.084)</td>
</tr>
<tr>
<td>ORD14 I do not like work spaces that are too clean and tidy.</td>
<td>2.5</td>
<td>3.386 (2.544)</td>
<td>5.171 (36.637)</td>
</tr>
<tr>
<td>ORD15 For me, being organized is unimportant.</td>
<td>2.5</td>
<td>5.752 (2.323)</td>
<td>2.305 (0.399)</td>
</tr>
<tr>
<td>ORD17 Being neat is not exactly my strength.</td>
<td>3.0</td>
<td>1.887 (0.423)</td>
<td>1.623 (0.176)</td>
</tr>
<tr>
<td>ORD24 Half of the time I do not put things in their proper place.</td>
<td>3.5</td>
<td>2.278 (0.641)</td>
<td>1.531 (0.126)</td>
</tr>
<tr>
<td>ORD30 Although I have a daily organizer, I have hard time keeping it up to date.</td>
<td>3.5</td>
<td>2.239 (1.425)</td>
<td>1.493 (0.299)</td>
</tr>
<tr>
<td>ORD20 I do pretty standard maintenance for my property and possessions.</td>
<td>4.0</td>
<td>-0.207 (0.251)</td>
<td>-0.232 (0.212)</td>
</tr>
<tr>
<td>ORD23 My room neatness is about average.</td>
<td>4.0</td>
<td>-0.192 (0.097)</td>
<td>-0.192 (0.079)</td>
</tr>
<tr>
<td>ORD26 My ability to plan is at about average.</td>
<td>4.0</td>
<td>1.188 (0.387)</td>
<td>1.190 (0.289)</td>
</tr>
<tr>
<td>ORD29 Although I try to keep everything in its place, it does not always work for me.</td>
<td>4.5</td>
<td>0.632 (0.109)</td>
<td>0.670 (0.089)</td>
</tr>
<tr>
<td>ORD37 I prefer to do things in a logical order.</td>
<td>5.5</td>
<td>-3.342 (1.808)</td>
<td>-3.460 (23.910)</td>
</tr>
<tr>
<td>ORD42 I write notes to myself only if I have too many things to do at once.</td>
<td>5.5</td>
<td>-0.239 (0.213)</td>
<td>-0.247 (0.242)</td>
</tr>
<tr>
<td>ORD35 I need a neat environment in order to work well.</td>
<td>6.0</td>
<td>-1.700 (0.415)</td>
<td>-1.535 (0.335)</td>
</tr>
<tr>
<td>ORD38 Organization is a key component of most things I do.</td>
<td>6.0</td>
<td>-6.757 (0.629)</td>
<td>-2.203 (12.442)</td>
</tr>
<tr>
<td>ORD34 I have a daily routine and stick to it.</td>
<td>6.5</td>
<td>-6.211 (1.098)</td>
<td>-2.502 (1.601)</td>
</tr>
<tr>
<td>ORD44 I become annoyed when things around me are disorganized.</td>
<td>6.5</td>
<td>-2.984 (1.822)</td>
<td>-2.693 (19.608)</td>
</tr>
<tr>
<td>ORD46 I keep detailed notes of important meetings and lectures.</td>
<td>6.5</td>
<td>-4.138 (2.369)</td>
<td>-3.481 (0.734)</td>
</tr>
<tr>
<td>ORD50 Every item in my room and on my desk has its own designated place.</td>
<td>7.0</td>
<td>-6.541 (0.839)</td>
<td>-2.571 (8.799)</td>
</tr>
</tbody>
</table>

Note. \( R_j \) = raters’ averaged ratings; \( \hat{\delta}_{\text{OCRUM}} \) = item location estimates from OCRUM; \( \hat{\delta}_{\text{GGUM}} \) = item location estimates from GGUM. Values in brackets were the estimated standard error.

the three different methods (i.e. raters, OCRUM, and GGUM). The entries above the diagonal were the Spearman rank correlation coefficients \( (r_s) \) while the entries below the diagonal were the Pearson correlation coefficients \( (r_p) \). The first column (and row) of Table 5.5 showed that the \( \hat{\delta}_j \) of the OCRUM and GGUM were highly correlated with the raters’ ratings \(|r_p|\) were .90 and .89 respectively; \(|r_s|\) were .93 and .92 respectively; Refer to the previous section for rationale of interpreting the absolute value instead of the original estimated parameter values).

On the comparison of the item location estimates between the OCRUM and GGUM, the \( r_p = .89 \) and \( r_s = .95 \). All correlation coefficients were significant at the \( \alpha = .05 \) level. Taken together, even though the exact value of the item location were not the same, the resulted ordering of the three methods were highly similar.

### Comparison of Response Function

Two items (ORD23 and ORD18) were selected to compare the response function as characterised by GGUM and OCRUM based on their item location estimates. Both Figure 5.4a and Figure 5.4b display the single-peak function with the highest probability of endorsement at the item location. However, it is clear that, as compared to GGUM, OCRUM has a flatter tail and rounder peak: The probability of endorsing this item
### Table 5.5: Pairwise correlation of the item locations for the Order scale

<table>
<thead>
<tr>
<th></th>
<th>(R_j)</th>
<th>(\hat{\delta}_{OCRUM,j})</th>
<th>(\hat{\delta}_{GGUM,j})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(R_j)</td>
<td>-</td>
<td>-.926</td>
<td>-.918</td>
</tr>
<tr>
<td>(\hat{\delta}_{OCRUM,j})</td>
<td>-.908</td>
<td>-</td>
<td>.946</td>
</tr>
<tr>
<td>(\hat{\delta}_{GGUM,j})</td>
<td>-.888</td>
<td>.890</td>
<td>-</td>
</tr>
</tbody>
</table>

**Note.** \(R_j\) = raters’ averaged ratings; \(\hat{\delta}_{OCRUM,j}\) = item location estimates from OCRUM; \(\hat{\delta}_{GGUM,j}\) = item location estimates from GGUM.

reaches 0 at a smaller absolute value of \(\theta\) in OCRUM than in GGUM. The highest probability of endorsement is also lower in OCRUM (lesser than .90) than in GGUM (approximate .90). Similar observations can also be made in Figure 5.5a and Figure 5.5b, with the exception of a single-peak function due to extreme item location.

**Comparison of Person Location \(\theta_i\)**

Pearson and Spearman rank correlation between the estimated person locations of both OCRUM and GGUM were estimated and analysed. The \(\hat{\theta}_i\) were found to be highly correlated: \(r_p = .992, p < .001\), and \(r_s = .996, p < .001\). Similarly, the distribution of the \(\hat{\theta}_i\) and their scatter plots were displayed in figure 5.6.

### 5.4 Discussion

The aim of this study was to evaluate the use of proposed model (Ordered Categorical Response Unfolding Model; OCRUM) on the empirical responses on Attitude toward Capital Punishment scale and the Order Scale and also to compare the performance of the OCRUM with the Generalised Graded Unfolding Model (GGUM) in terms of the estimations of the item locations (\(\delta_j\)) and person locations (\(\theta_i\)). On top of that, we were interested in comparing the estimates with published results of item scale values and rated item locations.

Firstly, it was noteworthy to emphasise that Mplus managed to fit OCRUM to the
Figure 5.4: Theoretical response function of item ORD23 according to (a) GGUM and (b) OCRUM

(a) $\hat{\delta}_{GGUM} = -0.192$

(b) $\hat{\delta}_{OCRUM} = -0.192$
Figure 5.5: Theoretical response function of item ORD37 according to (a) GGUM and (b) OCRUM

(a) $\hat{\delta}_{GGUM} = -3.460$

(b) $\hat{\delta}_{OCRUM} = -3.342$
Figure 5.6: Histograms and scatter plot with linear fit line of the $\hat{\theta}$ obtained from OCRUM (x-axis) and GGUM (y-axis) for 20-item Ideal Point IRT Order Scale (Chernyshenko, Stark, Drasgow, & Roberts, 2007).
data without any error, hence partially support the feasibility of fitting the model (i.e. OCRUM) with empirical polytomous and dichotomous data.

Pearson correlation and Spearman rank correlation were used to quantify the similarities between the item location estimates ($\hat{\delta}_j$) with a higher coefficient represent higher similarity. For both set of data, item ordering based on the parameter estimates ($\hat{\delta}_j$) were highly similar as compared to the scale values ($T_j$) and rated item location ($R_j$). If we were to use $T_j$ and $R_j$ as the benchmarks for the "true" item location, OCRUM performed as good as GGUM in reproducing the item locations.

On top of that, as observed from the high correlation between the $\hat{\delta}_{OCRUM}$ and $\hat{\delta}_{GGUM}$ on both scales, this had provided further evidence that the performance of OCRUM in estimating the item locations may be comparable to GGUM. Furthermore, the correlations between the estimated person locations ($\theta_i$) obtained from OCRUM and GGUM were also fairly high, indicating that OCRUM and GGUM produced comparable person locations estimates.

Furthermore, visual comparisons of the response functions of OCRUM and GGUM revealed that the functions are similar, with some stark difference especially on the "steepness" of the functions and the probability of the lowest level of agreement (i.e. Strongly Disagree).

Secondly, referring to the findings of the high correlation between $\hat{\delta}_{12,j}$ and the corresponding item location estimates of the 12-items from the 21-item Attitude to Capital Punishment scale indicated that omission of items in a scale would not affect the estimation of the item locations for OCRUM and GGUM. This finding is important as it showed that estimations of the OCRUM was not affected by the items included in the analysis and that the $\hat{\delta}_j$ estimation was relatively stable.

In brief, through this study, we have showed the feasibility of OCRUM in modelling empirical response data in Mplus. Indications of its comparability to the GGUM, in terms of its performance in estimating the $\hat{\delta}_j$ and $\theta_i$, were also found, suggesting OCRUM as a plausible alternative to the GGUM to be used in the structural equation modelling framework. However, systematic investigation (i.e. simulation studies) is required to
thoroughly compare the two models.
Chapter 6

Conclusion

This thesis has introduced the OCRUM, an unfolding response model, which

1. can be fitted within the Structural Equation Modelling (SEM) framework,

2. models the non-monotonic (single-peaked) relationship between the endorsement
(observed response) and the latent trait (as assumed by the unfolding response
process), and

3. estimation of both the item locations and person locations.

Two response mechanisms, the dominance and the unfolding response mechanism,
have been proposed in the literature to describe how individuals respond to questionnaire
items. However, it is our contention that existing methods of analysing the response data
resulted from the unfolding response mechanism are less acquainted by social scientists
in general and some of the methods have limited model expansion flexibility.

To answer the limitation of the current literature, we proposed a new method, the
Ordered Categorical Response Unfolding Model (OCRUM) and showcased its implementation
in Mplus. The feasibility and performance of the proposed method was investigated
through two simulation studies and two sets of empirical response data.

The two simulation studies looked at the effects of sample size, test length, item
location ranges, and response options on the parameter recovery and model-data fit. The
analysis of empirical response data aimed to illustrate the application of the proposed
method (OCRUM) and also to compare the performance of OCRUM with the Generalised Graded Unfolding Model (GGUM).

In general, we have showed that the person, item locations, and factor correlation specified in OCRUM could be estimated with little bias in Mplus, especially if the dataset consists of polytomous response data, with at least a moderate sample size (N > 300) and moderate test length (J > 10). OCRUM was also shown to produce item location $\hat{\delta}$ and person location $\hat{\theta}$ which are comparable with the GGUM (Generalised Graded Unfolding Model).

As mentioned in previous chapters, the assumption of response mechanism remains an empirical question: does the response data conform to the dominance response mechanism or unfolding response mechanism? This question could be answered from the model comparison approach, where the same dataset is fitted into both dominance model (i.e. linear factor model) and unfolding model (i.e. quadratic factor model) and to compare the fit statistics. From our study, we found that AIC, sBIC, and the Likelihood-ratio test were able to identify the correct model most of the time under the various conditions considered in this study (i.e. sample size, test length, response options, and item location range).

6.1 Limitations of Model

Apart from the evidence that OCRUM may not perform well under small sample size with short dichotomously scored test, one possible limitation of the model is the implementation of the estimation method of person location $\theta$, which assume a normal distribution in person location. This may not be appropriate if the assumption of normality does not hold in the population level, which in reality could be of some distribution other than normal distribution. The effect of violation this normality assumption on the estimation of $\theta$ was not explored in this study and should be pursued in future research.

Another limitation of the model is on the LMS approach of estimating the nonlinear effects. As mentioned in previous chapters, the implementation of LMS in Mplus
requires numerical integration, and this process is computing intensive. Therefore the computing time required to fit the data on OCRUM may get longer when the number of specified factors increases. Furthermore, LMS is currently only available in Mplus, therefore, implementation of OCRUM can only be done in the Mplus software.

Another limitation is the confirmatory nature of the method. Unlike exploratory factor analysis which enable researchers to explore the data structure, OCRUM requires the user to decide on the number of factors \textit{a priori}. However, we believe that researchers can take a model comparison approach, by specifying different number of underlying factors and to search for the best fitting model using the information criteria (e.g. AIC, and sBIC). Nevertheless, as highlighted in the previous paragraph, the computing time required may increase substantially when the number of factors increased.

\section*{6.2 Future Directions}

One of the features of our simulation study was the use of even-spaced item locations $\delta_j$. This is an idealised situation where a scale consists of items which were evenly spaced. However, referring to Study 2 of empirical response data, we acknowledge that it is very likely that the items in a scale would not be evenly spaced, but with gaps in the continuum. We did not consider the latter condition in the simulation. This is worth investigating because Polak (2011) found that in applying the Correspondence Analysis (CA) on response data with uneven-spaced items would result in inaccuracy in the estimation of the person location $\theta$.

On the other hand, the simulated data were either unidimensional or from an independent clusters model, where both refers to the situation that an item only measures one, and only one factor. This may be not a true representation of the empirical data where there may be more than one underlying factors explaining the variance of an item.

In short, we believe future research should look into the effect of violation of the normality assumption of population $\theta$ in the estimation of $\theta$. Furthermore, simulations with uneven-spaced item should be conducted to investigate the effect of such conditions
in the recovery of the item location $\delta$ and person location $\theta$. It is also worth investigating cases where items were multidimensional, instead of unidimensional or of independent clusters pattern in the item parameters.

Future research could also look into integrating OCRUM in the structural model where the person location $\theta$ can be treated as the predictor or the criterion variables where the OCRUM item parameters and the structural coefficients could be estimated simultaneously, which we believe could be implemented in Mplus with ease.
References


97


Reise, S. P. (2010). Thurstone might have been right about attitudes, but Drasgow,
Chernyshenko, and Stark fail to make the case for personality. *Industrial and

doctoral dissertation, University of South Carolina).

model for unfolding unidimensional polytomous responses. *Applied Psychological

program to estimate parameters of the generalized graded unfolding model. *Applied

unfolding responses from a graded disagree-agree response scale. *Applied Psychological

Thurstone approaches to attitude measurement. *Educational and Psychological

Rosander, A. C. & Thurstone, L. L. (1931). Scale of attitude toward censorship: scale
no. 28. In L. L. Thurstone (Ed.), *The measurement of social attitudes*. Chicago,
IL: University of Chicago Press.

Saffir, M. A. (1937). A comparative study of scales constructed by three psychophysical

doi:10.1177/01466216980221001

assumptions about item responding in personality assessment: Should ideal point


Thurstone, L. L. & Chave, E. J. (1929). *The measurement of attitude: A psychophysical method and some experiments with a scale for measuring attitude toward the church*. Chicago, IL: University of Chicago Press.


Appendix A

MATLAB Threshold Generating Script

%% Generating thresholds values

% generate huge data
n = 1000000;
d_m = repmat(0, n, 1); % item location = 0
t_m = repmat(-1, n, 1); % curvature param/loading = -1
a_m = repmat(0, n, 1);
e_m = randn(n, 1)*sqrt(1); % error variance
theta = pearsrnd(0, 1, 0, 3, [n, 1]); % theta~N(0, 1)
theta_m = repmat(theta, 1, 1);
distance = (theta_m - d_m);
data = a_m + t_m .* (distance.^2) + e_m; % data

% get thresholds
a = size(data);
for s = 1:a(2)
    y = data(:, s);
    [f, x] = ecdf(y); % f=cdf, x=data
    t1 = max(x(f <= 0.05));
t2 = max(x(f<=0.23));
t3 = max(x(f<=0.50));
t4 = max(x(f<=0.77));
t5 = max(x(f<=0.95));
threshold(s) = [t1 t2 t3 t4 t5]';
end
ptres = cell2mat(threshold); % polytomous thresholds
csvwrite('polythres.csv', ptres);

for s = 1:a(2)
y = data(:,s);
[f, x] = ecdf(y); %f=cdf, x=data
t1 = max(x(f<=0.50));
dtres = t1; % dichotomous thresholds
end
csvwrite('dichothres.csv', dtres);

% get thresholds
a = size(data);
for s = 1:a(2)
y = data(:,s);
[f, x] = ecdf(y); %f=cdf, x=data
%t1 = latent var at 5% of the distribution,
%t2=at26%,
%t3=at74%,
%t4=at95%
t1 = max(x(f<=0.05));
t2 = max(x(f<=0.26));
t3 = max(x(f<=0.74));
t4 = max(x(f<=0.95));

threshold{s} = [t1, t2, t3, t4]';

end

ptres = cell2mat(threshold);  % polytomous thresholds

csvwrite('polythres5.csv', ptres);
Appendix B

Mplus OCRUM Syntax for Unidimensional Data

This is an example of the Mplus syntax for OCRUM with five items.

TITLE:
Study 1a OCRUM;

DATA:
FILE IS data.csv;

VARIABLE:
NAMES ARE y1-y5;
USEVARIABLES ARE ALL;
CATEGORICAL ARE ALL;

ANALYSIS:
TYPE IS RANDOM;
ESTIMATOR IS MLR;
ALGORITHM IS INTEGRATION;
INTEGRATION IS 50;
MITERATION IS 500;
LINK IS PROBIT;

MODEL:
  f BY y1* (b1)
       y2 (b2)
       y3 (b3)
       y4 (b4)
       y5 (b5);

  f@1;
  [f@0];
  f_sq| f XWITH f;

  y1 ON f_sq* -1 (t1);
  y2 ON f_sq* -1 (t2);
  y3 ON f_sq* -1 (t3);
  y4 ON f_sq* -1 (t4);
  y5 ON f_sq* -1 (t5);

MODEL CONSTRAINT:
  new (d1 d2 d3 d4 d5);

    b1 = -2*d1*t1;
    b2 = -2*d2*t2;
    b3 = -2*d3*t3;
    b4 = -2*d4*t4;
    b5 = -2*d5*t5;
OUTPUT: TECH1 TECH2 TECH8;
SAVEDATA: RESULTS ARE results.csv;
          FILE IS fscores.csv;
          SAVE IS FSCORES;
Appendix C

MATLAB Study 1a Data Simulation Script

This is the MATLAB script used to generate the unidimensional data for the simulation study 1a.

\[
\begin{align*}
N &= \{150; 300; 600\}; \quad \text{% sample size} \\
j &= \{5; 10; 20\}; \quad \text{% test length} \\
sk &= \{0 3\}; \quad \text{% latent trait distribution} \\
delta &= \{-1.5 1.5; -2.5 2.5\}; \quad \text{% item locations} \\
resp &= \{6; 2\}; \quad \text{% response category} \\
seed &= 789132; \quad \text{% set seed} \\
nloop &= 200; \quad \text{% number of replication}
\end{align*}
\]
dtres = csvread('F:\Study 1\Thresholds\dichothres.csv'); % dir for dichothres.csv
ptres = csvread('F:\Study 1\Thresholds\polythres.csv');% dir for polythres.csv

% generate categorical data for each conditions
for iN = 1:length(N)
    for ij = 1:length(j)
        for isk = 1:size(sk, 1)
            for idelta = 1:size(delta, 1)
                for iresp = 1:length(resp)
                    % create folder and cd into folder to save datasets
                    mkdir (datestr(now,'ddmmyy_HHMMSS'));
                    cd(datestr(now,'ddmmyy_HHMMSS'));

                    % save input information
                    header = {'replication', 'N', 'test_length', 'delta_left', 'delta_right', ...
                              'resp','theta_skew', 'theta_kurtosis', 'starting_seed'};
                    info = [nloop N(iN), j(ij) delta(idelta, 1) delta(idelta, 2) resp(iresp), ...
sk(isk,1) sk(isk,2) seed;

input = dataset({info, header{}});

export (input, 'File', 'simulation.csv', 'Delimiter', ','); %export dataset as csv file

% start generating data

d_j = delta(idelta,1):...

(abs(delta(idelta,1)-delta(idelta,2))/(j(ij)-1):...

delta(idelta,2); % equi-distant item locations

for inloop = 1:nloop

rng(seed, 'twister'); % set seed

theta = pearsrnd(0, 1, sk(isk,1), sk(isk,2), N(iN), 1); % person locations
% these are all fixed.

d_m = repmat(d_j, N(iN), 1); % delta

t_m = repmat(-1, N(iN), length(d_j)); % curvature

a_m = repmat(0, N(iN), length(d_j)); % intercept

e_m = randn(N(iN),length(d_j))*sqrt(1); % error

b_j = -2.*d_j.*t_m(1,:); % model beta
\[ u_j = a_m(1,:) + t_m(1,:)*(d_j.^2); \] % model intercept

\[ \theta_m = \text{repmat}(\theta,1,j(ij)); \]
% generate continuous data

distance = (\theta_m-d_m);
data = a_m + t_m.*(distance.^2) + e_m;

\[ \text{if } iresp == 2 \]
% generate dichotomous data
\[ a = \text{size}(\text{data}); \]
\[ \text{for } s = 1:a(2) \text{ } \% a(2) returns ncol \]
\[ y = \text{data}(:,s); \]
\[ t1 = dtres; \text{ } \% \text{use item t threshold} \]

% condition (this must be specified first before recoding).
\[ y1 = y \leq t1; \]
\[ y2 = y > t1; \]
% recoding
y(y1) = 0;
y(y2) = 1;
output{s} = y;
end

else
% generate polytomous (6 category) data
a = size(data);
for s = 1:a(2) %a(2) returns ncol
    y = data(:,s);
t1 = ptres(1);
t2 = ptres(2);
t3 = ptres(3);
t4 = ptres(4);
t5 = ptres(5);
%condition (this must be specified first before recoding).

\[ y_1 = y \leq t_1; \]
\[ y_2 = y > t_1 \& y \leq t_2; \]
\[ y_3 = y > t_2 \& y \leq t_3; \]
\[ y_4 = y > t_3 \& y \leq t_4; \]
\[ y_5 = y > t_4 \& y \leq t_5; \]
\[ y_6 = y > t_5; \]

%recoding

\[ y(y_1) = 0; \]
\[ y(y_2) = 1; \]
\[ y(y_3) = 2; \]
\[ y(y_4) = 3; \]
\[ y(y_5) = 4; \]
\[ y(y_6) = 5; \]

output\{s\} = y;

end
end

% save data

catdata = cell2mat(output);
clearvars output; %clear 'output' to prevent overwriting
csvwrite(["data", num2str(inloop), ".csv"], catdata); % write data
csvwrite(["theta", num2str(inloop), ".csv"], theta); % write theta

seed = seed + 1;

end

xlswrite(’parameters.xls’, d_j, ’delta’);

xlswrite(’parameters.xls’, t_m(1,:), ’curvature’);

xlswrite(’parameters.xls’, b_j, ’beta’);

xlswrite(’parameters.xls’, u_j, ’modelmu’);

cd ..

dl
end
Appendix D

Mplus Linear Factor Model Syntax

This is an example of the Mplus syntax fitting a linear common factor model with five items.

TITLE:

   Study 1a Linear Model;

DATA:

   FILE IS data.csv;

VARIABLE:

   NAMES ARE y1-y5;
   USEVARIABLES ARE ALL;
   CATEGORICAL ARE ALL;

ANALYSIS:

   ESTIMATOR IS MLR;
   LINK IS PROBIT;

MODEL:

   f BY y1* y2-y5;
f@1;

OUTPUT: TECH1 TECH2;

SAVEDATA: RESULTS ARE results.csv;

FILE IS fscores.csv;

SAVE IS FSCORES;
Appendix E

MATLAB Study 1b Data Simulation Script

This is the MATLAB script used to generate the data with independent cluster pattern for the simulation study 1b.

\[ N = [150; 300; 600]; \quad \% \text{sample size} \]
\[ j = [5; 10; 20]; \quad \% \text{test length per factor} \]
\[ \delta = [-1.5 1.5; -2.5 2.5]; \quad \% \text{item locations} \]
\[ \text{resp} = [2; 6]; \quad \% \text{response category} \]
\[ \text{cor} = [0.3; 0.5; 0.7]; \quad \% \text{correlation between factors} \]
\[ \text{sk} = [0 3]; \quad \% \text{latent trait distribution} \]
\[ \text{seed} = 8987658; \quad \% \text{set seed} \]
nloop = 200; \% number of replication

dtres = csvread('F:\Study 1\Thresholds\dichothes.csv'); \% dir for dichothres.csv
ptres = csvread('F:\Study 1\Thresholds\polythres.csv'); \% dir for polythres.csv

for iN = 1:length(N)
    for ij = 1:length(j)
        for isk = 1:size(sk, 1)
            for idelta = 1:size(delta, 1)
                for iresp = 1:length(resp)
                    for icor = 1:length(cor)
                        \% create folder and cd into folder to save datasets
                        newFolder = datestr(now, 'ddmmyy_HHMSS');
                        mkdir(newFolder);
                        cd(newFolder);

                        \% save input information
                        header = {
                            'replication', 'N', 'test_length', 'delta_left', 'delta_right', ...
                        };
        end
    end
end
'resp','theta_skew', 'theta_kurtosis', 'theta_corr', 'starting_seed'}

info = [nloop N(iN) j(ij) delta(idelta, 1) delta(idelta,2) resp(iresp), ...
       sk(isk,1) sk(isk,2) cor(icor) seed];

input = dataset ({info, header{:}});

export (input, 'File', 'simulation.csv', 'Delimiter', ','); % export dataset as csv file.

% start generating data

d_j = delta(idelta,1): (...)
   (abs(delta(idelta,1)-delta(idelta,2)))/(j(ij)-1): ... 
   delta(idelta,2); % equi-distant item locations

k = []; % start empty matrix to save the empirical correlation

for inloop = 1:nloop
    rng(seed, 'twister'); % set seed
    theta = mvnrnd([0 0], [1 cor(icor); cor(icor) 1], N(iN));

    % these are all fixed.
    d_m = repmat(d_j, N(iN), 1); % delta
\[ t_m = \text{repmat}(-1, N(iN), \text{length}(d_j)); \quad \% \text{curvature} \]
\[ a_m = \text{repmat}(0, N(iN), \text{length}(d_j)); \quad \% \text{intercept} \]
\[ b_j = -2 \cdot d_j \cdot t_m(1,:); \quad \% \text{model beta} \]
\[ u_j = a_m(1,:) + t_m(1,:).*(d_j.^2); \quad \% \text{model intercept} \]
\[ \% \text{rng(log(seed)*1000, 'twister'); } \% \text{set seed} \]
\[ e_m = \text{randn}(N(iN), 2*\text{length}(d_j)) \cdot \text{sqrt}(1); \% \text{error} \]

\% factor 1
\[ \text{theta}_m1 = \text{repmat}(\text{theta}(:, 1), 1, j(ij)); \]
\[ e_m1 = e_m(:, 1: \text{length}(d_j)); \]
\% generate continuous data for factor 1
\[ \text{distance} = (\text{theta}_m1 - d_m); \]
\[ \text{data1} = a_m + t_m.\ast(\text{distance}.^2) + e_m1; \]

\% categorise
\[ \text{if resp(iresp) == 2} \]
\[ \quad \% \text{generate dichotomous data} \]
a = size(data1);
for s = 1:a(2) %a(2) returns ncol
    y = data1(:,s);
    t1 = dtres; %use item t threshold
    y1 = y<=t1;
    y2 = y>t1;
    %recoding
    y(y1) = 0;
    y(y2) = 1;
    output{s} = y;
end
else
    % generate polytomous (6 category) data
a = size(data1);

for s = 1:a(2)  %a(2) returns ncol
    y = data1(:,s);
    t1 = ptres(1);
    t2 = ptres(2);
    t3 = ptres(3);
    t4 = ptres(4);
    t5 = ptres(5);

% condition (this must be specified first before recoding).
    y1 = y<=t1;
    y2 = y>t1 & y<=t2;
    y3 = y>t2 & y<=t3;
    y4 = y>t3 & y<=t4;
    y5 = y>t4 & y<=t5;
    y6 = y>t5;
% recoding
y(y1) = 0;
y(y2) = 1;
y(y3) = 2;
y(y4) = 3;
y(y5) = 4;
y(y6) = 5;
output{s} = y;
end
end
catdata1 = cell2mat(output);
clearvars output; % clear 'output' to prevent overwriting

% factor 2
e_m2 = e_m(:, length(d_j)+1:end); % error
theta_m2 = repmat(theta(:, 2), 1, j(ij));
% generate continuous data for factor 2

distance = (theta_m2 - d_m);
data2 = a_m + t_m.*(distance.^2) + e_m2;

% cateogrise

if resp(iresp) == 2

    % generate dichotomous data
    a = size(data2);

    for s = 1:a(2) %a(2) returns ncol
        y = data2(:,s);
        t1 = dtres; %use item t threshold

        %condition (this must be specified first before recoding).
        y1 = y<=t1;
        y2 = y>t1;

        %recoding
y(y1) = 0;
y(y2) = 1;
output{s} = y;
end
else

% generate polytomous (6 category) data
a = size(data2);
for s = 1:a(2) %a(2) returns ncol
    y = data2(:,s);
    t1 = ptres(1);
    t2 = ptres(2);
    t3 = ptres(3);
    t4 = ptres(4);
    t5 = ptres(5);

    % condition (this must be specified first before recoding).
y1 = y<=t1;
y2 = y>t1 & y<=t2;
y3 = y>t2 & y<=t3;
y4 = y>t3 & y<=t4;
y5 = y>t4 & y<=t5;
y6 = y>t5;

% recoding
y(y1) = 0;
y(y2) = 1;
y(y3) = 2;
y(y4) = 3;
y(y5) = 4;
y(y6) = 5;
output{s} = y;

end

end
% save data
catdata2 = cell2mat(output);
clearvars output; %clear 'output' to prevent overwriting

catdata = [catdata1 catdata2]; % 1st factor + 2nd factor
csvwrite(['data', num2str(inloop), '.csv'], catdata); % write data
csvwrite(['theta', num2str(inloop), '.csv'], theta); % write theta

seed = seed + 1;

% save empirical correlation
tmp = corr(theta);
k = [k; tmp(1, 2)];
end

% save parameters
xlswrite('parameters.xls', d_j, 'delta');
xlswrite('parameters.xls', t_m(1,:), 'curvature');
xlswrite('parameters.xls', b_j, 'beta');
xlswrite('parameters.xls', u_j, 'modelmu');

% save empirical correlation
xlswrite('parameters.xls', k, 'EmpCorr');
cd ..
end
Appendix F

Mplus OCRUM Syntax for
Independent Cluster Pattern Data

This is an example of the Mplus syntax for OCRUM two factors with five items each.

TITLE:
Study1b OCRUM.

DATA:
FILE IS data.csv;

VARIABLE:
NAMES ARE y1-y10;
USEVARIABLES ARE ALL;
CATEGORICAL ARE ALL;

ANALYSIS:
TYPE IS RANDOM;
ESTIMATOR IS MLR;
ALGORITHM IS INTEGRATION;
INTEGRATION IS 50;
MITERATION IS 500;
LINK IS PROBIT;

MODEL:

! 1st factor
f1 BY y1* (b1)
   y2 (b2)
   y3 (b3)
   y4 (b4)
   y5 (b5);

f1@1;
[f1@0];
f1_sq| f1 XWITH f1;

y1 ON f1_sq^*1 (t1);
y2 ON f1_sq^*1 (t2);
y3 ON f1_sq^*1 (t3);
y4 ON f1_sq^*1 (t4);
y5 ON f1_sq^*1 (t5);

! 2nd factor
f2 BY y6* (b6)
   y7 (b7)
   y8 (b8)
   y9 (b9)
y10 (b10);

f2@1;
[f2@0];
f2_sq1 | f2 XWITH f2;

y6 ON f2_sq*1 (t6);
y7 ON f2_sq*1 (t7);
y8 ON f2_sq*1 (t8);
y9 ON f2_sq*1 (t9);
y10 ON f2_sq*1 (t10);

! factor correlation
f1 WITH f2;

MODEL CONSTRAINT:
new (d1 d2 d3 d4 d5
d6 d7 d8 d9 d10);

b1 = -2*d1*t1;
b2 = -2*d2*t2;
b3 = -2*d3*t3;
b4 = -2*d4*t4;
b5 = -2*d5*t5;
b6 = -2*d6*t6;
b7 = -2*d7*t7;
b8 = -2*d8*t8;
b9 = -2*d9*t9;
b10 = -2*d10*t10;

OUTPUT: TECH1 TECH2 TECH8;
SAVEDATA: RESULTS ARE results.csv;
FILE IS fscores.csv;
SAVE IS FSCORES;