國立臺灣師範大學數學系碩士班碩士論文

指導教授：蔡蓉青 博士

呂翠珊 博士

The Use of Mixture Response Time Model to Account for Careless Respondents in Survey Questionnaires with a Bogus Item

研究生：林克謙

中 華 民 國 104 年 3 月

致謝

一段經過短暫時間的漫長旅程，遇到很多很多的貴人。在此先感謝我的兩位指導教授：蔡蓉青老師與呂翠珊老師。非常感謝她們的嚴格的要求與耐心、細心地指導才得以完成這篇文章。完成的過程中也得到許多人建議與幫助。感謝口試委員林定香老師給予建議，使文章內容更貼近初讀者並一目瞭然。淑貞學姐與育瑋學長的熱情相助，在反覆的討論中建立對基礎反應模型的認識並給予程式語言的指導。系上助教與彰化師範大學的老師們協助蒐集並提供實證資料，讓這篇文章得以短時間完成。另外，家人的支持和朋友的鼓勵與協助，在研究生活的咖啡中加點糖跟奶精。這篇文章絕非只是學術方面的結果，也是我與這些生命中的貴人在這段旅程中的紀錄。如還有未提到而疏忽的貴人，一言難盡，不勝感激。

摘要

一份問卷調查的健全推論建立在有效率的估計之上，但是有效率的估計往往受到不認真的受訪者的影響。問卷調查廣泛應用虛題 (bogus item) 找出不認真的受訪者、發現那些不看題目的人。另一種方法是反應答題的時間，其使用方式已由 Rasch 混合模型結合反應時間的方法確立，並能有效地提升估計的效率。在這篇研究中，我們提出結合虛題與混合反應時間的模型。此模型是利用虛題並考慮反應時間來區分「兩種不同類型的不認真受訪者的回答以及認真受訪者的回覆」。分析時保留虛題做為共變數與時間做為潛在類別的應變數的優點。在模擬實驗中，當樣本數 500 人以上，反應時間有降低估計值的標準誤的趨勢。更進一步，我們應用該模型分析真實資料時，其中有高達 22% 的不認真受訪者。結果顯示：我們未受限制的模型表現得比受限制的模型好。

關鍵字：虛題、反應時間、潛在類別、Rasch 混合模型結合反應時間

Abstract

A valid inference for a survey is always based on efficient estimation. The efficient estimation is threatened and distorted by the careless response. Bogus item is commonly used in survey questionnaire to detect the careless respondents and powerfully unveils those responding without reading the items. Response time is another way to identify the careless responses and its use of enhancing item parameter estimation was verified by the mixture Rasch model with response time component (MRM-RT). Therefore, we propose a mixture response time model to analyze survey questionnaire with a bogus item. The goal of our proposed model is to classify two behaviors of the careless responding in addition to the attentive respondents, using the bogus item and taking account into response time. We attempt to converse the benefit of the bogus item used as a covariate and the response time used as the latent class indicators in analysis. In the simulation studies, we find that the standard errors of the model with response time decrease substantially as the sample size is equal or larger than 500. Further, we also apply our proposed model to a real data set and with a larger proportion of careless responding (22%), the results show that the unrestricted models perform better than the restricted models.

Keywords: bogus item, response time, latent class, MRM-RT

Contents

List of Tables

List of Figures

1 Introduction

In a research based on survey responses, careless or random responses are often concerned and become increasingly vital. It had been concluded that the careless responses may distort inference of the survey. The distorted inference results from the erroneous estimation, and the careless or random responses lead to the loss in estimation efficiency. For example, Wood (2006) examined five indices: Comparative fit index (CFI), Tucker–Lewis incremental fit index (TLI), root mean square error of approximation (RMSEA), standardized root mean-square residual (SRMR), weighted root mean square residual (WRMR) and found that, with random responses, the two-factor model was often mistakenly concluded to fit better than the true one-factor model. That is, the researchers might reject a single-factor confirmatory factor model because the model with about 10 percent of the respondents carelessly responding to reverse-worded items. Credé (2010) found that even low base rates of random responding can significantly affect the observed correlations. With a 5 percent of random response rate, the observed squared correlation can be as little as $\frac{1}{44}$ of the squared correlation that would be observed without random responding. Hence, to avoid the improper inference, a solid and efficient model which can classify the survey behavior and simultaneously improve the accuracy of estimation is needed.

Learning from a similiar situation of the achivement test is random guessing. The random guessing also distorts inference due to the inefficient estimation. Wise and DeMar (2006) used an effort-moderated model to enhance item parameters by classifying response behavior. When excluding the random-behavior examinees, the bias and root mean-square error (RMSE) of their model are both smaller than the standard 3-parameter logistic (3PL) model. Meade and Craig (2012) described that there were at least 7 kinds of indicators for the careless respondents in a questionnaire. Meade and Craig had applied a factor mixture model to identify careless responding. The model is displayed in Figure 1. Both the effort-moderated model and the mixture factor model rely on the covariate (the position of the careless response indicators of Figure 1) to classify the respondents' latent classes. In the following, we will review a useful covariate employed in a survey: Bogus Item.

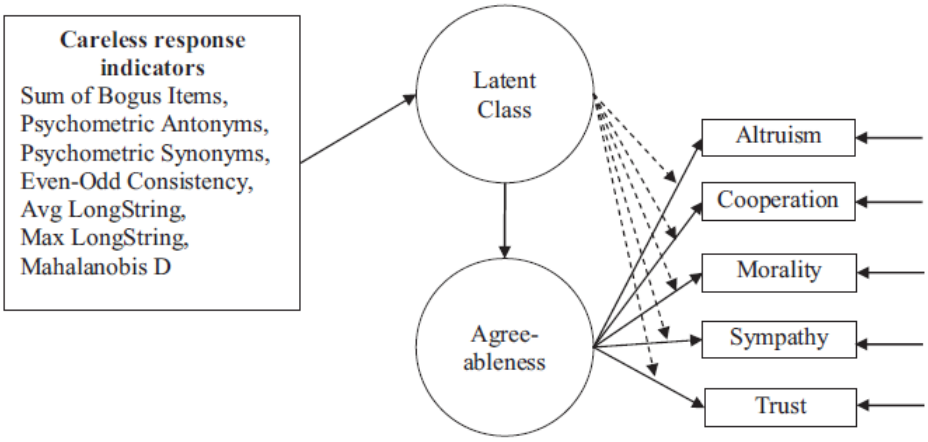
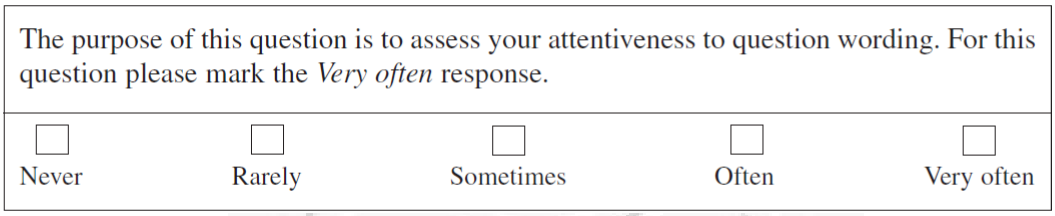


Figure 1: The factor mixture model used in the Meade and Craig's study

The bogus item design is useful and common in a survey. An advantage of an unambiguous bogus item is that if a careless person responds incorrectly, there is little doubt that he or she is responding carelessly or dishonestly (Meade and Craig, 2012). Beach (1988) was aware of the careless responses and illustrated the bogus item like "I was born on Feburary 30^{th}" to detect careless respondents. This example was used in the "Ture or False" survey. In fact, Beach had a good explanation to the bogus item: such a scale is made up of the questions that the subjects who read and understand the question can only answer in one way. Hargittai (2008) gave an example of such a bogus item, as illustrated in Figure 2. With certain response to the bogus items, the careless respondents can be easily identified. Other careless response indicators may be used as well, but it is sometimes harder to directly link certain response pattern to careless or random responding. For example, the researchers so far have no consensus in the criteria for long string and Mahalanobis D. Consistency like antonyms and synonyms may create other negative factors (Schmitt and Stults, 1985). Wood's finding (2006) agreed with their results. Therefore, comparing with other indicators, a bogus item design can be more straightforward and easily implemented in a study.


The purpose of this question is to assess your attentiveness to question wording. For this question please mark the *Very often* response.

Never	Rarely	Sometimes	Often	Very often
☐	☐	☐	☐	☐

Figure 2: An exact example of the bogus item in Hargittai's study without misunderstanding.

The length of response time (RT) can be another method to identify careless respondents in the survey. For instance, RT was treated as a continuous covariate and used in a personality test to detect those dissimulting or dishonest. Their model (Holden, et al., 1992) predicted that dissimulators are relatively faster in answering consistent with the dissimulation and relatively slower in responding inconsistently with the dissimulation. Meade and Craig (2012) also considered RT as indicator of careless responding and excluded the repondents whose RT exceeded 15 minutes for their analysis. In the achivement test, Wise and DeMar (2006) used RT to distinguish solution behavior and random guessing. Their effort-moderated model refined the item parameter estimates by classifying response behavior. However, the thresholds of the RT would be subjective when it was treated as a covariate, and these models relied on the empirical RT plot.

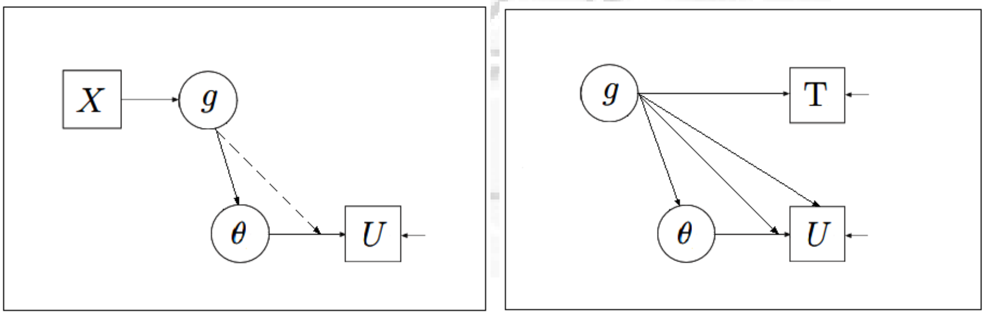
There were other ways to treat RT as a function of the latent trait in the achivement test. One of the models that combined responses and RT is based on speed-distance hypothesis(Ferrando and Lorenzo-Seva, 2007). The hypothesis constructed a measure for the distance between the personal trait and item difficulty, and linked this measure to the logarithmized RT by regression. The measure had described that the RT decreased as the distance of personal trait and item difficulty increased. However, The shorter RT is merely corresponded to the larger distance between the trait and difficulty. It is insufficient to confirm whether the classes are guessing. Another model proposed by Meyer(2010) was based on the assumption that the response and the RT are independent given the latent class. According to Meyer, the item estimates from the mixture Rasch model with response time component (MRM-RT) could be recovered by a Bayesian method. The analysis of extant data showed that a two-class model fit the test data better than the one-class model when 15 percent of examinees engaged in rapid-guessing behavior. Chan, Lu and Tsai (2014) employed the MRM-RT, adjusted it into a structure equation model (SEM) and enhanced the item parameters. According to their study, the mixture SEM was extended to 3 classes which defined random guessing, well-skilled and the ordinary. The mixture SEM used maximum liklihood estimation and it takes much less time than the Bayesian approach. Upon these reaserch, a model treating RT as a latent class indicator can be more precise than a model treating RT as a function of the latent trait. Another consideration for the relation of the RT to a survey is not corresponded to the trait-difficulty distance hypothesis. The distance represent the degree of agreement in a survey. Therefore, we propose a model with RT as a latent class indicator for a survey to avoid the defect of the adjustment in an achievement test.

We will review Chan, Lu and Tsai's model (2014), Meade and Craig's model (2012) and resultantly propose a mixture response time model with a bogus item for the quetionnaire. This model conserves both advantage of the bogus item and RT. In Section 2, we introduce the likelihood of the proposed model and list the identification constaints. In Section 3, the simulation studies are stated and the results are reported and discussed. In Section 4, we apply the model to analyze real data; we will then conclude and discuss our study in Section 5.

2 Model

2.1 The Mixture Response Time Model

According to Meade and Craig (2012), their model structure can be described in Figure 3-(a), where X denotes as the covariates previously shown in Figure 2; g represents the latent classes of the careless and the attentive respondents; θ is the latent trait of interest and U is the response. Meyer (2010) proposed a Bayesian model to enhance the item estimates and later Chan, Lu and Tsai (2014) constructed a structure equation model corresponding to Meyer's. The path diagram is displayed in Figure 3-(b), where T is response time, logarithm-transformed normally distributed and treated as an indicator of the latent classes g; U is the response as before. Note that θ is interpreted as the ability in achievement testing.



(a) (b)

Figure 3: (a) Meade and Craig's model; (b) Chan, Lu and Tsai's model

In Figure 3-(a), the careless respondents were viewed partially careless; therefore, the dotted line represented ”partially” and revealed that some careless responding did not depend on θ. That is, the factor loadings vary across the latent classes.

Figures 3-(a) and 3-(b) attempt to distinguish the careless/random responding from attentive/standard behavior solution. The more accurate the sizes of the latent classes, the more reliable estimates one can obtain. Based on the models above, we propose a mixture model incorporating response time and involving a bogus item; the model is depicted in Figure 4.

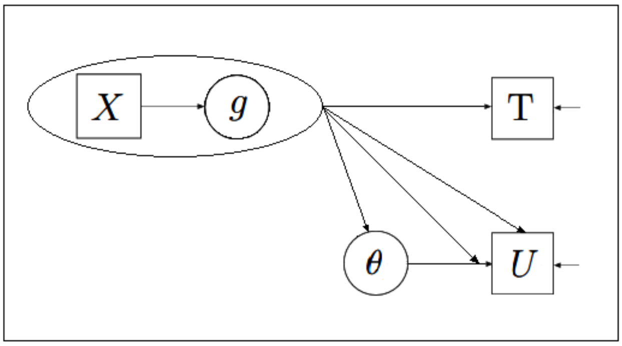


Figure 4: A mixture model incorporating response time and involving a bogus item.

In Figure 4, the ellipse including X and g represents the joint probability. The ellipse substitutes g only since those careless respondents are apt to respond incorrectly on X. That is, we believe those careless without reading in g might still be identified by X. Therefore we suggest using (X, g) instead of g.

According to Meade and Craig (2012), $\mathbf{X} = (1, X_1, X_2, \cdots, X_7)$ representing the covariate vactor in Figure 2 was linked with the latent classes by a logistic regression:

$$P(g_i|\mathbf{X}_i) = \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})} , \quad (1)$$

where g_i is the i^{th} person's latent class and $\mathbf{X}_i$ is the vector of his or her covariates. $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_7)$ are the intercept and the coefficient parameters. In Meyer's justification (2010), the probability of personal responses and response time are independent when the latent class is given. However this justification should be adjusted since now we treat X and g simultaneously. Similarly, the response and response time are independent given $(\mathbf{X}_i, g_i)$, so that we can write

$$P(U_{ij}, T_{ij}|\mathbf{X}_i, g_i) = P(U_{ij}|\mathbf{X}_i, g_i)P(T_{ij}|\mathbf{X}_i, g_i) , \quad (2)$$

where U_{ij} and T_{ij} are respectively the i^{th} person's response and response time for item j. $P(U_{ij}|\mathbf{X}_i, g_i)$ and $P(T_{ij}|\mathbf{X}_i, g_i)$ will be further discussed next.

To relate a personal latent trait of interest and fitting a polytomous item response model, the graded response model (GRM) (Samejima, 1973) was extended and developed from the continuous response model. Forero and Maydeu-Olivares (2009) not only had a summary of the approximation between logistic GRM and normal ogive GRM, but also derived normal ogive GRM from a factor analytic framework. That is, either logistic or normal GRM can be derived from a latent trait model, and moreover the parameter space can be identified via each model constriants. Based on the latent trait model, suppose that the i^{th} person's latent response to item j is U^*_{ij}:

$$U^*_{ij} = \lambda_{(\mathbf{X},g)j}\theta_i + \varepsilon_{ij} , \quad (3)$$

where $\lambda_{(\mathbf{X},g)j}$ is the factor loading of item j. θ_i is the i^{th} person's latent trait of interest and ε_{ij} follows $N(0, \sigma^2)$. As a result, the observation of the i^{th} person responding to item j is:

$$U_{ij} = k \quad \text{if} \quad \tau_{(\mathbf{X},g)jk} \le U^*_{ij} \le \tau_{(\mathbf{X},g)j(k+1)} , \quad (4)$$

where $\boldsymbol{\tau}_{(\mathbf{X},g)j} = (\tau_{(\mathbf{X},g)j0}, \cdots, \tau_{(\mathbf{X},g)j(k+1)})$, specific to the class $(\mathbf{X}, g)$, is a vector consisting of threshold parameters for the item j. Note that $k = 0, \cdots, K-1$, where K is the number of categories of item j; $\tau_{(\mathbf{X},g)j0} = -\infty$ and $\tau_{(\mathbf{X},g)jK} = \infty$. Hence the response model of U_{ij} can be expressed in the following:

$$
\begin{aligned}
& P(U_{ij} = k|\mathbf{X}_i, g_i, \theta_i; \lambda_{(\mathbf{X},g)j}, \boldsymbol{\tau}_{(\mathbf{X},g)j}) \\
& = \boldsymbol{\Phi}(\frac{\tau_{(\mathbf{X},g)j(k+1)} - \lambda_{(\mathbf{X},g)j}\theta_i}{\sigma}) - \boldsymbol{\Phi}(\frac{\tau_{(\mathbf{X},g)jk} - \lambda_{(\mathbf{X},g)j}\theta_i}{\sigma}) \ .
\end{aligned} \tag{5}
$$

It has been shown that lognormal-distributed response time has a better fit when compared with other distributions (Linden & Krimpen-Stoop, 2003). As a result, T_{ij} will be transformed to logarithm scale in this study. Since Meyer (2010) assigned a distribution to each latent class, the response time model will be similarly designed to adapt to $(\mathbf{X}, g)$. The adjusted response time model now is

$$
P(T_{ij} = t_{ij}|\mathbf{X}_i, g_i; \nu_{(\mathbf{X},g)}, \xi_{(\mathbf{X},g)}) \ = \ \frac{1}{\sqrt{2\pi}\xi_{(\mathbf{X},g)}} \exp(\frac{-(t_{ij} - \nu_{(\mathbf{X},g)})^2}{2\xi^2_{(\mathbf{X},g)}}) \ , \tag{6}
$$

where t_{ij} is the realization of T_{ij}; $(\nu_{(\mathbf{X},g)}, \xi^2_{(\mathbf{X},g)})$ are the mean and the variance of response time for class $(\mathbf{X}, g)$. Under conditional independence of all hte response and response time as well, given the class membership $(\mathbf{X}, g)$, a mixture model incorporating response time and involving bogus items is

$$
\begin{aligned}
&P(\mathbf{U},\mathbf{T}|\mathbf{X};\boldsymbol{\Theta}) \\
&= \prod_{i=1}^{N} P(\mathbf{U}_i,\mathbf{T}_i|\mathbf{X}_i;\boldsymbol{\Theta}) \\
&= \prod_{i=1}^{N}\sum_{g_i=1}^{G}\int_{-\infty}^{\infty} P(\mathbf{U}_i,\mathbf{T}_i|\mathbf{X}_i,g_i,\theta_i;\boldsymbol{\Theta})P(g_i,\theta_i|\mathbf{X}_i;\boldsymbol{\Theta})d\theta_i \\
&= \prod_{i=1}^{N}\sum_{g_i=1}^{G}\int_{-\infty}^{\infty} P(\mathbf{U}_i,\mathbf{T}_i|\mathbf{X}_i,g_i,\theta_i;\boldsymbol{\Theta})P(\theta_i|\mathbf{X}_i,g_i;\boldsymbol{\Theta})P(g_i|\mathbf{X}_i;\boldsymbol{\Theta})d\theta_i \\
&= \prod_{i=1}^{N}\sum_{g_i=1}^{G}\int_{-\infty}^{\infty}\Big[\prod_{j=1}^{J} P(U_{ij}|\mathbf{X}_i,g_i,\theta_i;\boldsymbol{\Theta})P(T_{ij}|\mathbf{X}_i,g_i,\theta_i;\boldsymbol{\Theta})\Big]P(\theta_i|\mathbf{X}_i,g_i;\boldsymbol{\Theta})P(g_i|\mathbf{X}_i;\boldsymbol{\Theta})d\theta_i \ ,
\end{aligned}
\tag{7}
$$

where

$$
\boldsymbol{\Theta} = (\sigma,\eta,\delta,\boldsymbol{\lambda},\boldsymbol{\tau},\boldsymbol{\nu},\boldsymbol{\xi},\boldsymbol{\beta}) \ ,
$$

$$
\begin{aligned}
P(g_i|\mathbf{X}_i;\boldsymbol{\Theta}) &= \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})} \ , \\
P(\theta_i|\mathbf{X}_i,g_i;\boldsymbol{\Theta}) &= \frac{1}{\sqrt{2\pi}\delta}\exp\Big(\frac{-(\theta_i-\eta)^2}{2\delta^2}\Big) \ , \\
P(T_{ij}=t_{ij}|\mathbf{X}_i,g_i,\theta_i;\boldsymbol{\Theta}) &= \frac{1}{\sqrt{2\pi}\xi_{(\mathbf{X},g)}}\exp\Big(\frac{-(t_{ij}-\nu_{(\mathbf{X},g)})^2}{2\xi^2_{(\mathbf{X},g)}}\Big) \ , \\
P(U_{ij}=k|\mathbf{X}_i,g_i,\theta_i;\boldsymbol{\Theta}) &= \boldsymbol{\Phi}\Big(\frac{\tau_{(\mathbf{X},g)j(k+1)}-\lambda_{(\mathbf{X},g)j}\theta_i}{\sigma}\Big)-\boldsymbol{\Phi}\Big(\frac{\tau_{(\mathbf{X},g)jk}-\lambda_{(\mathbf{X},g)j}\theta_i}{\sigma}\Big) \ ,
\end{aligned}
$$

and $i=1,\cdots,N$; $j=1,\cdots,J$; $g=1,\cdots,G$, where N, J and G are the total numbers of the respondents, items and latent classes, respectively.

2.2 A Restricted Condition

Without loss of generality, we consider $\mathbf{X}_i$ as a two-dimension vector

$$\mathbf{X}_i \quad = \quad (1, X_i) \; . \tag{8}$$

$X_i = 1$ if person i has an incorrect response on the bogus item, and 0 if correct. Abbreviating the conditional probability $P(g_i|\mathbf{X}_i, \boldsymbol{\Theta})$, we replace it with $P(g_i|X_i)$ as follows. Denote that the latent class $g_i = 1$ if person i is a careless respondent and $g_i = 2$ if he/she is an attentive respondent. The conditional probability can be tabulated as,

	$X_i = 0$	$X_i = 1$
$P(g_i = 1\|X_i)$	$\pi_{1\|0}$	$\pi_{1\|1}$
$P(g_i = 2\|X_i)$	$\pi_{2\|0}$	$\pi_{2\|1}$

where

$$\pi_{1|0} \quad = P(g_i = 1|X_i = 0) \quad = \frac{1}{1 + \exp(\beta_0)} \; , \tag{9}$$

$$\pi_{2|0} \quad = P(g_i = 2|X_i = 0) \quad = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)} \; , \tag{10}$$

$$\pi_{1|1} \quad = P(g_i = 1|X_i = 1) \quad = \frac{1}{1 + \exp(\beta_0 + \beta_1)} \; , \tag{11}$$

$$\pi_{2|1} \quad = P(g_i = 2|X_i = 1) \quad = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)} \; . \tag{12}$$

Meade and Craig did not restrict $\pi_{2|1}$ to be zero when estimating (β_0, β_1). Hence $P(X_i = 1, g_i = 2)$ would not equal 0 in their study. Instead, we believe that it barely happens when an attentive person responds incorrectly, and such that a model constraint will be

$$\beta_1 \to -\infty \quad \Rightarrow \quad \pi_{1|1} = 1, \; \pi_{2|1} = 0 \; . \tag{13}$$

To account for the careless responding, the response distributions of these classes $(X_i = 0, g_i = 1)$ and $(X_i = 1, g_i = 1)$ should be specified. Since these careless respondents do not read the items thoroughly, θ_i of the respondents do not react on $P(U_{ij} = k|\mathbf{X}_i, g_i, \theta_i; \mathbf{\Theta})$. In fact, θ_i of a certain careless person may follow a distribution in his or her class. But regardless of the value of θ_i, the response distribution can be determined once given the class $(\mathbf{X}_i, g_i)$. That is, the careless responding will be simplified to

$$P(U_{ij} = k|\mathbf{X}_i, g_i = 1, \theta_i; \mathbf{\Theta}) = P(U_{ij} = k|\mathbf{X}_i, g_i = 1; \mathbf{\Theta}). \tag{14}$$

To specify such a condition, we can then simply the factor loadings of the careless classes $(X_i = 0, g_i = 1)$ and $(X_i = 1, g_i = 1)$ to zero,

$$\lambda_{(\mathbf{X}, g_i=1)j} = 0 \qquad \forall\, \mathbf{X},\ j\ . \tag{15}$$

Later in the result, we will estimate $\tau_{(\mathbf{X},g)jk}$ under the careless responsees. However it is not allowed to set loadings equal to zeroes in Mplus Version 5. We instead replace it with $\lambda_{(\mathbf{X},g_i=1)j} = 1$ and $\theta_i|_{g_i=1} = 0$. Moreover, we restrict all the response to follow the same distribution within $(X_i = 0, g_i = 1)$ and $(X_i = 1, g_i = 1)$.

2.3 Estimation and Identification

From (7), the likelihood function can be rewritten as:

$$\begin{aligned} &L(\mathbf{\Theta}; \mathbf{U}, \mathbf{T}) \\ &= \prod_{i=1}^{N} \sum_{g_i=1}^{G} \int_{-\infty}^{\infty} \Big[\prod_{j=1}^{J} P(U_{ij}|\mathbf{X}_i, g_i, \theta_i; \mathbf{\Theta}) P(T_{ij}|\mathbf{X}_i, g_i, \theta_i; \mathbf{\Theta}) \Big] P(\theta_i|\mathbf{X}_i, g_i; \mathbf{\Theta}) P(g_i|\mathbf{X}_i; \mathbf{\Theta}) d\theta_i\ . \end{aligned} \tag{16}$$

The estimates are obtained using the maximum likelihood method. In Mplus, the numerical integration is used to approximate the integral of latent response (Muthén and Muthén, 1998-2010).

As we discussed previously, identification becomes an issue in estimating the parameters. For example, taking a look at (5), it is obvious that the same probabilities can be obtained from different response patterns. For example, $\{\theta = 0, \boldsymbol{\tau}_{(\mathbf{X},g)j} = \boldsymbol{\tau}^*, \sigma = 1\}$ and $\{\theta = 0, \boldsymbol{\tau}_{(\mathbf{X},g)j} = 2\boldsymbol{\tau}^*, \sigma = 2\}$ can produce the same probability. Hence, Forero and Maydeu-Olivares (2009) set constraints for item estimation. To confirm a unique solution set of parameters, the distribution of U^*_{ij} is prerequisite and the variance of ε_{ij} has to be specified. The identification constraints used in our model are

$$\mathrm{E}(\theta_i|\boldsymbol{\Theta}) = 0 , \quad (17)$$

$$\mathrm{Var}(\theta_i|\boldsymbol{\Theta}) = 1 , \quad (18)$$

$$\mathrm{Var}(\varepsilon_{ij}) = 1 . \quad (19)$$

That is, $\theta_i \sim N(\eta = 0, \delta^2 = 1)$ and $\varepsilon_{ij} \sim N(0, \sigma^2 = 1)$.

3 Simulation

In this simulation, we employ the mixture model discussed in Section 2 and compare performances of various model fittings. In Section 3.1, we exhibit the data generation, eight settings of model constraints and the criteria used for model comparison. In Section 3.2, the simulation results are presented and we will discuss the results and findings in Section 3.3.

3.1 Data Generation

To simulate survey items, we apply Fraley, Waller and Brennan's analysis (2000) of self-report measures. There were 18 items and the item parameter estimates were obtained from GRM. Their result is displayed in Table 1-(a). Note that a_j and $\boldsymbol{b}_j$ referred to the slope and intercepts, respectively, in IRT terminology; in SEM terminology, a_j is the factor loading and $\boldsymbol{b}_j$ are the thresholds (Forero and Maydeu-Olivares, 2009). Here we use normal ogive in the simulation. According to Koch (1983), we can interchange these two ogives by using

$$\lambda_{(\mathbf{x},g)j} = \frac{a_j}{1.7}, \tag{20}$$

$$\boldsymbol{\tau}_{(\mathbf{x},g)j} = \lambda_{(\mathbf{x},g)j}\boldsymbol{b}_j . \tag{21}$$

We further collapse these thresholds down to four by eliminating two extreme options in order to avoid extremely large standard errors. Therefore, the factor loadings and thresholds we used in the data generation are presented in Table 1-(b).

Table 1: The parameters transformed from logistic orgive to normal orgive

(a) The item estimates cited from Fraley, Waller and Brennan's analysis

j	a_j	$\boldsymbol{b}_j$					
1	2.79	-1.12	-0.39	0.00	0.45	1.08	1.70
2	2.33	-1.38	-0.52	-0.12	0.41	1.15	1.85
3	2.21	-1.07	-0.21	0.25	0.82	1.53	2.11
4	2.10	-1.64	-0.76	-0.39	0.19	0.93	1.80
5	1.98	-1.32	-0.57	-0.29	0.28	0.86	1.58
6	1.93	-1.71	-0.73	-0.20	0.36	1.00	1.75
7	1.87	-1.36	-0.45	0.04	0.50	1.32	2.05
8	1.74	-1.85	-0.89	-0.40	0.16	0.90	1.80
9	1.50	-1.86	-0.67	-0.06	0.60	1.29	2.26
10	1.49	-0.72	0.35	0.87	1.68	2.43	3.62
11	1.36	-1.69	-0.45	0.11	0.73	1.42	2.22
12	1.36	-1.38	-0.29	0.16	1.07	1.96	2.99
13	1.35	-1.31	-0.18	0.30	1.02	1.73	2.59
14	1.35	-0.90	0.11	0.50	1.09	1.91	2.81
15	1.34	-0.97	0.10	0.52	1.00	1.65	2.61
16	1.32	-1.52	-0.45	0.03	0.79	1.74	2.69
17	1.24	-1.91	-0.71	-0.29	0.33	1.18	2.25
18	1.24	-0.45	0.83	1.40	2.17	2.86	3.53

(b) The item estimates from logistic ogive transforemed to normal ogive

j	$\lambda_{(\mathbf{x},g)j}$	$\boldsymbol{\tau}_{(\mathbf{x},g)j}$			
1	1.64	-0.640	0.000	0.739	1.772
2	1.37	-0.713	-0.164	0.562	1.576
3	1.30	-0.273	0.325	1.066	1.989
4	1.24	-0.939	-0.482	0.235	1.149
5	1.16	-0.664	-0.338	0.326	1.002
6	1.14	-0.829	-0.227	0.409	1.135
7	1.10	-0.495	0.044	0.550	1.452
8	1.02	-0.911	-0.409	0.164	0.921
9	0.88	-0.591	-0.053	0.529	1.138
10	0.88	0.307	0.763	1.472	2.130
11	0.80	-0.360	0.088	0.584	1.136
12	0.80	-0.232	0.128	0.856	1.568
13	0.79	-0.143	0.238	0.810	1.374
14	0.79	-0.087	0.397	0.866	1.517
15	0.79	0.079	0.410	0.788	1.300
16	0.78	-0.349	0.023	0.613	1.350
17	0.73	-0.518	-0.212	0.241	0.861
18	0.73	0.605	1.021	1.583	2.086

3.1.1 Class Size

In Section 2, the respondents were classified according to $(\mathbf{X}, g)$. For better understanding, we denote the classes of $(X_i = 0, g_i = 1)$, $(X_i = 1, g_i = 1)$ and $(X_i = 0, g_i = 2)$ as "the careless 1", "the careless 2" and "the attentive," respectively. Most literature suggested that there were 10 percent of careless responding in a survey, so that we choose 10 percent in our simulation. Furthermore, according to Johnson (2005), there was about 3.5 percent of careless responding without reading the items in the population. Hence we set 3.5, 6.5 and 90 percent for "the careless 1", "the careless 2" and "the attentive," respectively.

3.1.2 Response and Response Time

According to Table 1-(b), the response of "the attentive" can be generated via (5). For the response time of "the attentive," we randomly sampled ten students to answer these questions and recorded their response time. Then the obtained emperical time distribution is $lognormal(1.5, 0.35)$. This time distribution will be used to generate reponse time of "the attentive." The response distributions of "the careless 1" and "the careless 2" depend on their tendency. We considered two conditions, the "equal" and "unequal" cases. In the "equal" case, since the response patterns of the "careless 1" and "careless 2" were set unifromly in most literature, the probability distribution of responses in these two classes is set to be $(0.2, 0.2, 0.2, 0.2, 0.2)$. In the "unequal" case, the "careless 1" and "careless 2" have their own response styles. The response distributions were determined from another empirical data. In the empirical data, the response distribution of the "careless 1" was $(0.05, 0.15, 0.25, 0.45, 0.1)$, and for the "careless 2," it was $(0.05, 0.15, 0.2, 0.4, 0.2)$. The response time distribution of the "careless 1" is $lognormal(-0.8, 0.3)$ (Chan, Lu & Tsai, 2014). The response time of the "careless 2” may be longer than “careless 1” but shorter than “the attentive,” so we assigned $lognormal(1.37, 0.4)$ to the “careless 2.”

3.1.3 Factors and Model Fittings

To evaluate the performance of the proposed model, we consider the following eight fitting models:

- A.1 : Only bogus item as covariate, without response time.
  From (9) to (12), we restrict $P(g_i|\mathbf{X}_i;\mathbf{\Theta})$ with $\beta_1 \to -\infty$ and $\beta_0 \to \infty$. Therefore the model switches into the miltiple-group analysis. From (16), the likelihood is rewritten as
  $$L(\mathbf{\Theta};\mathbf{U}) = \prod_{i=1}^{N}\int_{-\infty}^{\infty}\Big[\prod_{j=1}^{J=18} P(U_{ij}|\mathbf{X}_i,\theta_i;\mathbf{\Theta})\Big]P(\theta_i|\mathbf{X}_i;\mathbf{\Theta})d\theta_i\ .$$
- A.2 : Based on A.1, add the factor of response time. The likelihood is
  $$L(\mathbf{\Theta};\mathbf{U}) = \prod_{i=1}^{N}\int_{-\infty}^{\infty}\Big[\prod_{j=1}^{J=18} P(U_{ij}|\mathbf{X}_i,\theta_i;\mathbf{\Theta})P(T_{ij}|\mathbf{X}_i,\theta_i;\mathbf{\Theta})\Big]P(\theta_i|\mathbf{X}_i;\mathbf{\Theta})d\theta_i\ .$$
- B.1 :Without response time, the response distributions of “careless 1” and “careless 2” are fixed at $(0.2, 0.2, 0.2, 0.2, 0.2)$ when fitting model. That is, for $k = \{0, 1, 2, 3, 4\}$, $P(U_{ij} = k|\mathbf{X}_i = (1, 0), g_i = 1, \theta_i;\mathbf{\Theta}) = P(U_{ij} = k|\mathbf{X}_i = (1, 1), g_i = 1, \theta_i;\mathbf{\Theta}) = 0.2$. The likelihood is rewritten as
  $$L(\mathbf{\Theta};\mathbf{U}) = \prod_{i=1}^{N}\sum_{g_i=1}^{G=2}\int_{-\infty}^{\infty}\Big[\prod_{j=1}^{J=18} P(U_{ij}|\mathbf{X}_i,g_i,\theta_i;\mathbf{\Theta})\Big]P(\theta_i|\mathbf{X}_i,g_i;\mathbf{\Theta})P(g_i|\mathbf{X}_i;\mathbf{\Theta})d\theta_i\ .$$
- B.2 : Based on B.1, add the factor of response time.
- C.1 : Without reponse time, we let the thresholds of “careless 1” and “careless 2” be equal when fitting the model, and we relax the probabilities of the responses from being uniform. That is, $k = \{0, 1, 2, 3, 4\}$, $P(U_{ij} = k|\mathbf{X}_i = (1, 0), g_i = 1, \theta_i;\mathbf{\Theta}) = P(U_{ij} = k|\mathbf{X}_i = (1, 1), g_i = 1, \theta_i;\mathbf{\Theta})$.

- C.2 : Based on C.1, add the factor of response time.

- D.1 : Without reponse time, we relax the restriction in Model C.1 of "careless 1" and "careless 2" when fitting the model. "Careless 1" and "careless 2" can have their response distributions.

- D.2 : Based on D.1, add the factor of response time.

In this simulation, these eight models will be compared using the criteria, SE($\hat{\Theta}$), where $\hat{\Theta}$ is the estimate of Θ. Note that Θ is the parameter.

We also consider different sample sizes, N = 250, 500, and 1000. The replication is 200 times for each condition setting.

3.2 Results

Since we generate data from two different response distributions: “equal case” and “unequal case,” we will present their results seperately. In Section 3.2.1, we display the results of the eight fitted models under the “equal case”. In Section 3.2.2, we show the results for the “unequal case” as described in 3.1.2. In Section 3.2.3, we will discuss the results of both cases and make the conclusions.

3.2.1 Under The Equal Case

Tables 2 to 19 present the item estimates of “the attentive.” In Table 3, when $N = 250$, the models considering the latent classes (B.1, B.2, C.1, C.2 and D. 1, D.2) generally have smaller bias than the models without latent classes (A.1, A.2). The loadings of A.1 and A.2 have larger bias than the other models, and the thresholds are the same. Models B.1 and B.2 are the true models with their thresholds fixed to yeild discrete uniform distribution of the response. Comparing B.1 with B.2, we find that there is no substantial difference between the model with response time and the model without reponse time if just eyeballing the estimates and their standard errors. This finding can also be seen in C.1 versus C.2 and D.1 versus D.2. However the bias and the standard errors are not consistent. Comparing C.1 with D.1 and C.2 with D.2, we can not conclude which of C.1, C.2, D.1 and D.2 is better.

When $N = 500$, the bias of the estimates of A.1 and A.2 are larger than the small sample size $N = 250$. However the other models (B.1, B.2, C.1, C.2 and D.1, D.2) have smaller bias than those at $N = 250$. The standard errors are decreasing if coming with the models for $N = 250$, which is as expected. The bias of the models (B.1, B.2, C.1, C.2 and D.1, D.2) do not have consistent performance, but the standard errors of the models with response time (B.2, C.

2 and D.2) have become smaller than the models without response time (B.1, C. 1, and D.1). For example, λ_1 of Item 1, the estimates of B.2, C.2 and D.2 are smaller than the estimates of the models B.1, C.1 and D.1. It seem that the bias of the models with response time estimates do not have much difference from those without time estimates. When $N = 1000$, these phenomenon become more obviously.

Summary: there are some conclusions in this section as following,

- The models considering the latent classes have smaller bias than those without the latent classes.
- The models with response time have smaller standard errors than those without response time.
- As the sample size increases, we obtain consistent results that both the bias and the standard errors decrease.

Table 2: The estimates of the loading and thresholds for Item 1 (equal)

N		λ_1	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.640	-0.640	0.000	0.739	1.772
250	A.1	1.556(0.162)	-0.635(0.143)	-0.004(0.147)	0.706(0.153)	1.676(0.175)
	A.2	1.556(0.162)	-0.635(0.143)	-0.004(0.147)	0.706(0.153)	1.676(0.175)
	B.1	1.660(0.179)	-0.640(0.160)	-0.001(0.163)	0.744(0.171)	1.780(0.209)
	B.2	1.660(0.179)	-0.641(0.158)	-0.001(0.161)	0.743(0.170)	1.780(0.209)
	C.1	1.662(0.179)	-0.638(0.158)	0.001(0.164)	0.745(0.171)	1.779(0.202)
	C.2	1.661(0.178)	-0.641(0.160)	-0.002(0.165)	0.742(0.170)	1.775(0.203)
	D.1	1.661(0.178)	-0.642(0.157)	-0.002(0.163)	0.741(0.168)	1.775(0.202)
	D.2	1.661(0.178)	-0.638(0.158)	0.001(0.164)	0.745(0.171)	1.779(0.202)
500	A.1	1.543(0.104)	-0.640(0.116)	-0.003(0.110)	0.703(0.116)	1.673(0.132)
	A.2	1.543(0.104)	-0.640(0.116)	-0.003(0.110)	0.703(0.116)	1.673(0.132)
	B.1	1.644(0.121)	-0.647(0.123)	0.001(0.119)	0.737(0.130)	1.773(0.151)
	B.2	1.646(0.117)	-0.642(0.120)	0.006(0.117)	0.741(0.129)	1.778(0.151)
	C.1	1.644(0.122)	-0.646(0.123)	0.003(0.120)	0.739(0.130)	1.775(0.151)
	C.2	1.646(0.117)	-0.643(0.120)	0.005(0.118)	0.741(0.130)	1.778(0.152)
	D.1	1.643(0.123)	-0.647(0.123)	0.001(0.120)	0.737(0.129)	1.772(0.154)
	D.2	1.645(0.117)	-0.644(0.120)	0.004(0.118)	0.739(0.128)	1.776(0.151)
1000	A.1	1.538(0.086)	-0.638(0.072)	-0.006(0.072)	0.710(0.080)	1.685(0.096)
	A.2	1.538(0.086)	-0.638(0.072)	-0.006(0.072)	0.710(0.080)	1.685(0.096)
	B.1	1.638(0.094)	-0.641(0.077)	-0.001(0.083)	0.742(0.089)	1.776(0.106)
	B.2	1.641(0.094)	-0.640(0.077)	0.001(0.079)	0.744(0.088)	1.778(0.104)
	C.1	1.637(0.095)	-0.641(0.079)	-0.002(0.081)	0.741(0.088)	1.775(0.107)
	C.2	1.641(0.094)	-0.639(0.078)	-0.001(0.079)	0.742(0.088)	1.777(0.103)
	D.1	1.637(0.095)	-0.641(0.079)	-0.002(0.081)	0.740(0.087)	1.774(0.106)
	D.2	1.640(0.093)	-0.639(0.078)	0.001(0.079)	0.742(0.087)	1.778(0.103)

Table 3: The estimates of the loading and thresholds for Item 2 (equal)

N		λ_2	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.370	-0.713	-0.164	0.562	1.576
250	A.1	1.283(0.142)	-0.710(0.127)	-0.155(0.133)	0.545(0.122)	1.496(0.153)
	A.2	1.283(0.142)	-0.710(0.127)	-0.155(0.133)	0.545(0.122)	1.496(0.153)
	B.1	1.362(0.158)	-0.720(0.142)	-0.166(0.147)	0.566(0.138)	1.572(0.175)
	B.2	1.364(0.159)	-0.721(0.138)	-0.169(0.145)	0.565(0.137)	1.573(0.176)
	C.1	1.364(0.157)	-0.720(0.143)	-0.166(0.148)	0.567(0.140)	1.574(0.176)
	C.2	1.365(0.158)	-0.721(0.138)	-0.169(0.145)	0.565(0.137)	1.572(0.176)
	D.1	1.364(0.160)	-0.720(0.141)	-0.167(0.148)	0.567(0.139)	1.573(0.180)
	D.2	1.364(0.160)	-0.720(0.141)	-0.167(0.148)	0.567(0.139)	1.573(0.180)
500	A.1	1.291(0.095)	-0.707(0.090)	-0.153(0.094)	0.540(0.095)	1.508(0.111)
	A.2	1.291(0.095)	-0.707(0.090)	-0.153(0.094)	0.540(0.095)	1.508(0.111)
	B.1	1.369(0.108)	-0.718(0.102)	-0.163(0.103)	0.559(0.103)	1.585(0.129)
	B.2	1.371(0.108)	-0.715(0.102)	-0.160(0.101)	0.562(0.102)	1.587(0.126)
	C.1	1.369(0.108)	-0.717(0.102)	-0.161(0.103)	0.561(0.103)	1.587(0.129)
	C.2	1.371(0.108)	-0.716(0.103)	-0.160(0.102)	0.562(0.103)	1.587(0.127)
	D.1	1.370(0.109)	-0.718(0.102)	-0.163(0.102)	0.560(0.102)	1.586(0.129)
	D.2	1.371(0.108)	-0.717(0.102)	-0.162(0.101)	0.560(0.102)	1.585(0.126)
1000	A.1	1.290(0.069)	-0.711(0.063)	-0.161(0.057)	0.533(0.063)	1.507(0.074)
	A.2	1.290(0.069)	-0.711(0.063)	-0.161(0.057)	0.533(0.063)	1.507(0.074)
	B.1	1.368(0.077)	-0.715(0.071)	-0.166(0.064)	0.566(0.070)	1.579(0.081)
	B.2	1.367(0.076)	-0.713(0.070)	-0.164(0.063)	0.565(0.069)	1.579(0.080)
	C.1	1.368(0.076)	-0.716(0.071)	-0.166(0.064)	0.565(0.070)	1.577(0.083)
	C.2	1.366(0.074)	-0.715(0.070)	-0.165(0.063)	0.565(0.068)	1.576(0.082)
	D.1	1.368(0.075)	-0.715(0.070)	-0.165(0.063)	0.564(0.069)	1.577(0.084)
	D.2	1.366(0.074)	-0.715(0.069)	-0.164(0.062)	0.565(0.067)	1.575(0.082)

Table 4: The estimates of the loading and thresholds for Item 3 (equal)

N		λ_3	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.300	-0.273	0.325	1.066	1.989
250	A.1	1.263(0.144)	-0.311(0.127)	0.289(0.133)	1.028(0.140)	1.903(0.189)
	A.2	1.263(0.144)	-0.311(0.127)	0.289(0.133)	1.028(0.140)	1.903(0.189)
	B.1	1.324(0.159)	-0.277(0.136)	0.323(0.146)	1.093(0.157)	2.021(0.216)
	B.2	1.324(0.158)	-0.277(0.136)	0.322(0.142)	1.092(0.154)	2.020(0.213)
	C.1	1.325(0.159)	-0.277(0.136)	0.324(0.148)	1.093(0.159)	2.021(0.217)
	C.2	1.325(0.158)	-0.277(0.135)	0.322(0.141)	1.091(0.153)	2.019(0.213)
	D.1	1.326(0.158)	-0.276(0.137)	0.324(0.143)	1.094(0.156)	2.022(0.215)
	D.2	1.326(0.158)	-0.276(0.137)	0.324(0.143)	1.094(0.156)	2.022(0.215)
500	A.1	1.246(0.109)	-0.315(0.095)	0.281(0.097)	1.000(0.107)	1.885(0.130)
	A.2	1.246(0.109)	-0.315(0.095)	0.281(0.097)	1.000(0.107)	1.885(0.130)
	B.1	1.304(0.117)	-0.283(0.106)	0.316(0.110)	1.066(0.121)	2.003(0.148)
	B.2	1.307(0.116)	-0.281(0.105)	0.319(0.107)	1.071(0.118)	2.008(0.145)
	C.1	1.304(0.118)	-0.281(0.106)	0.318(0.110)	1.069(0.123)	2.006(0.150)
	C.2	1.307(0.116)	-0.281(0.105)	0.319(0.108)	1.071(0.120)	2.008(0.147)
	D.1	1.304(0.118)	-0.283(0.105)	0.316(0.109)	1.067(0.122)	2.005(0.150)
	D.2	1.307(0.116)	-0.282(0.105)	0.318(0.108)	1.070(0.119)	2.007(0.146)
1000	A.1	1.248(0.073)	-0.317(0.066)	0.288(0.065)	0.999(0.072)	1.872(0.096)
	A.2	1.248(0.073)	-0.317(0.066)	0.288(0.065)	0.999(0.072)	1.872(0.096)
	B.1	1.300(0.075)	-0.273(0.071)	0.325(0.070)	1.066(0.082)	1.989(0.109)
	B.2	1.303(0.075)	-0.272(0.071)	0.327(0.070)	1.068(0.080)	1.990(0.110)
	C.1	1.301(0.076)	-0.274(0.070)	0.326(0.070)	1.066(0.081)	1.990(0.112)
	C.2	1.303(0.076)	-0.272(0.070)	0.327(0.070)	1.068(0.080)	1.992(0.112)
	D.1	1.302(0.077)	-0.273(0.071)	0.326(0.070)	1.065(0.082)	1.990(0.111)
	D.2	1.304(0.075)	-0.273(0.070)	0.327(0.070)	1.067(0.081)	1.991(0.111)

Table 5: The estimates of the loading and thresholds for Item 4 (equal)

N		λ_4	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.240	-0.939	-0.482	0.235	1.149
250	A.1	1.177(0.135)	-0.919(0.138)	-0.447(0.127)	0.245(0.114)	1.131(0.128)
	A.2	1.177(0.135)	-0.919(0.138)	-0.447(0.127)	0.245(0.114)	1.131(0.128)
	B.1	1.263(0.151)	-0.955(0.152)	-0.490(0.145)	0.237(0.127)	1.168(0.143)
	B.2	1.262(0.149)	-0.954(0.150)	-0.489(0.142)	0.236(0.126)	1.166(0.142)
	C.1	1.265(0.151)	-0.956(0.154)	-0.491(0.145)	0.236(0.125)	1.167(0.144)
	C.2	1.263(0.149)	-0.955(0.149)	-0.490(0.140)	0.235(0.124)	1.164(0.142)
	D.1	1.262(0.150)	-0.953(0.151)	-0.488(0.142)	0.238(0.125)	1.166(0.143)
	D.2	1.262(0.150)	-0.953(0.151)	-0.488(0.142)	0.238(0.125)	1.166(0.143)
500	A.1	1.159(0.089)	-0.916(0.100)	-0.453(0.093)	0.234(0.087)	1.106(0.091)
	A.2	1.159(0.089)	-0.916(0.100)	-0.453(0.093)	0.234(0.087)	1.106(0.091)
	B.1	1.243(0.098)	-0.952(0.107)	-0.495(0.100)	0.225(0.096)	1.142(0.101)
	B.2	1.244(0.097)	-0.947(0.105)	-0.491(0.098)	0.227(0.093)	1.145(0.100)
	C.1	1.243(0.099)	-0.951(0.107)	-0.494(0.101)	0.226(0.096)	1.144(0.102)
	C.2	1.244(0.097)	-0.947(0.106)	-0.491(0.098)	0.227(0.094)	1.144(0.100)
	D.1	1.243(0.099)	-0.952(0.107)	-0.495(0.100)	0.225(0.094)	1.143(0.102)
	D.2	1.244(0.097)	-0.948(0.105)	-0.492(0.098)	0.226(0.093)	1.144(0.100)
1000	A.1	1.144(0.069)	-0.911(0.067)	-0.451(0.063)	0.244(0.064)	1.110(0.069)
	A.2	1.144(0.069)	-0.911(0.067)	-0.451(0.063)	0.244(0.064)	1.110(0.069)
	B.1	1.238(0.076)	-0.941(0.075)	-0.485(0.069)	0.235(0.068)	1.141(0.073)
	B.2	1.239(0.076)	-0.940(0.073)	-0.486(0.069)	0.235(0.068)	1.144(0.072)
	C.1	1.238(0.076)	-0.941(0.074)	-0.487(0.068)	0.236(0.068)	1.141(0.073)
	C.2	1.238(0.076)	-0.940(0.073)	-0.486(0.067)	0.236(0.067)	1.143(0.072)
	D.1	1.239(0.075)	-0.941(0.074)	-0.485(0.068)	0.236(0.068)	1.141(0.072)
	D.2	1.241(0.074)	-0.939(0.073)	-0.486(0.067)	0.237(0.067)	1.144(0.072)

Table 6: The estimates of the loading and thresholds for Item 5 (equal)

N		λ_5	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.160	-0.664	-0.338	0.326	1.002
	A.1	1.123(0.128)	-0.679(0.121)	-0.321(0.117)	0.321(0.121)	1.004(0.129)
	A.2	1.123(0.128)	-0.679(0.121)	-0.321(0.117)	0.321(0.121)	1.004(0.129)
	B.1	1.187(0.142)	-0.684(0.133)	-0.350(0.134)	0.316(0.136)	1.019(0.144)
	B.2	1.187(0.141)	-0.684(0.135)	-0.349(0.132)	0.318(0.136)	1.019(0.144)
250	C.1	1.188(0.143)	-0.684(0.134)	-0.350(0.135)	0.316(0.137)	1.018(0.145)
	C.2	1.188(0.141)	-0.686(0.133)	-0.350(0.132)	0.317(0.135)	1.019(0.143)
	D.1	1.188(0.140)	-0.684(0.133)	-0.348(0.131)	0.319(0.135)	1.021(0.143)
	D.2	1.188(0.140)	-0.684(0.133)	-0.348(0.131)	0.319(0.135)	1.021(0.143)
	A.1	1.101(0.094)	-0.674(0.098)	-0.318(0.090)	0.326(0.088)	0.990(0.098)
	A.2	1.123(0.128)	-0.679(0.121)	-0.321(0.117)	0.321(0.121)	1.004(0.129)
	B.1	1.162(0.105)	-0.680(0.108)	-0.346(0.101)	0.321(0.096)	1.000(0.107)
	B.2	1.163(0.103)	-0.676(0.107)	-0.344(0.099)	0.322(0.094)	1.003(0.105)
500	C.1	1.162(0.105)	-0.678(0.107)	-0.344(0.100)	0.323(0.096)	1.002(0.107)
	C.2	1.163(0.103)	-0.676(0.107)	-0.344(0.099)	0.322(0.095)	1.003(0.106)
	D.1	1.162(0.105)	-0.679(0.107)	-0.345(0.100)	0.321(0.096)	1.001(0.107)
	D.2	1.163(0.103)	-0.677(0.107)	-0.344(0.099)	0.321(0.094)	1.003(0.105)
	A.1	1.107(0.070)	-0.665(0.065)	-0.319(0.065)	0.324(0.058)	0.995(0.066)
	A.2	1.107(0.070)	-0.665(0.065)	-0.319(0.065)	0.324(0.058)	0.995(0.066)
	B.1	1.159(0.075)	-0.668(0.067)	-0.340(0.070)	0.326(0.064)	1.003(0.074)
	B.2	1.158(0.074)	-0.666(0.067)	-0.339(0.069)	0.326(0.062)	1.003(0.074)
1000	C.1	1.160(0.075)	-0.667(0.066)	-0.340(0.069)	0.325(0.063)	1.000(0.074)
	C.2	1.159(0.075)	-0.667(0.066)	-0.340(0.069)	0.325(0.063)	1.001(0.073)
	D.1	1.160(0.073)	-0.666(0.066)	-0.339(0.068)	0.326(0.060)	1.001(0.074)
	D.2	1.159(0.072)	-0.664(0.065)	-0.338(0.068)	0.325(0.060)	1.002(0.073)

Table 7: The estimates of the loading and thresholds for Item 6 (equal)

N		λ_6	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.140	-0.829	-0.227	0.409	1.135
250	A.1	1.086(0.115)	-0.829(0.115)	-0.220(0.110)	0.408(0.118)	1.125(0.138)
	A.2	1.086(0.115)	-0.829(0.115)	-0.220(0.110)	0.408(0.118)	1.125(0.138)
	B.1	1.143(0.121)	-0.839(0.130)	-0.228(0.121)	0.415(0.130)	1.148(0.148)
	B.2	1.145(0.122)	-0.842(0.129)	-0.231(0.121)	0.412(0.129)	1.147(0.148)
	C.1	1.144(0.120)	-0.840(0.131)	-0.228(0.122)	0.415(0.131)	1.147(0.149)
	C.2	1.146(0.121)	-0.842(0.129)	-0.231(0.120)	0.412(0.128)	1.146(0.147)
	D.1	1.145(0.121)	-0.840(0.130)	-0.229(0.120)	0.413(0.130)	1.147(0.148)
	D.2	1.145(0.121)	-0.840(0.130)	-0.229(0.120)	0.413(0.130)	1.147(0.148)
500	A.1	1.082(0.082)	-0.816(0.095)	-0.216(0.083)	0.408(0.084)	1.111(0.095)
	A.2	1.082(0.082)	-0.816(0.095)	-0.216(0.083)	0.408(0.084)	1.111(0.095)
	B.1	1.141(0.089)	-0.829(0.103)	-0.228(0.093)	0.414(0.092)	1.135(0.103)
	B.2	1.143(0.088)	-0.826(0.103)	-0.227(0.092)	0.417(0.091)	1.139(0.101)
	C.1	1.141(0.089)	-0.827(0.104)	-0.226(0.094)	0.415(0.093)	1.137(0.104)
	C.2	1.143(0.088)	-0.826(0.103)	-0.227(0.092)	0.417(0.091)	1.139(0.102)
	D.1	1.141(0.089)	-0.828(0.102)	-0.228(0.093)	0.414(0.093)	1.136(0.104)
	D.2	1.143(0.088)	-0.827(0.102)	-0.228(0.092)	0.416(0.090)	1.139(0.101)
1000	A.1	1.079(0.064)	-0.820(0.060)	-0.219(0.057)	0.408(0.059)	1.110(0.063)
	A.2	1.079(0.064)	-0.820(0.060)	-0.219(0.057)	0.408(0.059)	1.110(0.063)
	B.1	1.139(0.067)	-0.827(0.067)	-0.227(0.062)	0.414(0.063)	1.136(0.073)
	B.2	1.138(0.067)	-0.827(0.067)	-0.226(0.063)	0.416(0.063)	1.138(0.073)
	C.1	1.139(0.068)	-0.830(0.066)	-0.227(0.062)	0.414(0.062)	1.136(0.073)
	C.2	1.139(0.068)	-0.829(0.065)	-0.225(0.062)	0.414(0.061)	1.137(0.073)
	D.1	1.140(0.067)	-0.830(0.066)	-0.227(0.062)	0.413(0.062)	1.136(0.073)
	D.2	1.139(0.067)	-0.829(0.065)	-0.226(0.062)	0.415(0.061)	1.138(0.072)

Table 8: The estimates of the loading and thresholds for Item 7 (equal)

N		λ_7	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.100	-0.495	0.044	0.550	1.452
	A.1	1.071(0.114)	-0.514(0.119)	0.035(0.115)	0.549(0.109)	1.422(0.139)
	A.2	1.071(0.114)	-0.514(0.119)	0.035(0.115)	0.549(0.109)	1.422(0.139)
	B.1	1.111(0.121)	-0.494(0.129)	0.051(0.123)	0.562(0.119)	1.468(0.157)
	B.2	1.112(0.120)	-0.494(0.128)	0.049(0.123)	0.561(0.119)	1.469(0.154)
250	C.1	1.112(0.121)	-0.493(0.130)	0.051(0.124)	0.562(0.119)	1.468(0.157)
	C.2	1.113(0.120)	-0.494(0.127)	0.049(0.122)	0.561(0.118)	1.469(0.153)
	D.1	1.113(0.120)	-0.493(0.128)	0.050(0.123)	0.563(0.120)	1.470(0.154)
	D.2	1.113(0.120)	-0.493(0.128)	0.050(0.123)	0.563(0.120)	1.470(0.154)
	A.1	1.062(0.086)	-0.514(0.091)	0.035(0.088)	0.540(0.089)	1.412(0.109)
	A.2	1.062(0.086)	-0.514(0.091)	0.035(0.088)	0.540(0.089)	1.412(0.109)
	B.1	1.102(0.093)	-0.496(0.096)	0.049(0.096)	0.555(0.094)	1.459(0.118)
	B.2	1.103(0.093)	-0.493(0.095)	0.052(0.096)	0.557(0.094)	1.462(0.115)
500	C.1	1.102(0.094)	-0.495(0.095)	0.050(0.095)	0.557(0.094)	1.460(0.118)
	C.2	1.104(0.093)	-0.493(0.095)	0.052(0.096)	0.557(0.095)	1.462(0.115)
	D.1	1.103(0.094)	-0.497(0.095)	0.049(0.095)	0.555(0.095)	1.460(0.118)
	D.2	1.103(0.092)	-0.494(0.094)	0.051(0.095)	0.557(0.094)	1.461(0.114)
	A.1	1.058(0.060)	-0.519(0.061)	0.029(0.057)	0.538(0.059)	1.406(0.063)
	A.2	1.058(0.060)	-0.519(0.061)	0.029(0.057)	0.538(0.059)	1.406(0.063)
	B.1	1.098(0.062)	-0.496(0.063)	0.044(0.059)	0.548(0.064)	1.453(0.072)
	B.2	1.097(0.062)	-0.496(0.063)	0.043(0.059)	0.550(0.063)	1.453(0.070)
1000	C.1	1.099(0.064)	-0.497(0.063)	0.044(0.060)	0.549(0.063)	1.452(0.072)
	C.2	1.098(0.063)	-0.496(0.063)	0.044(0.060)	0.549(0.063)	1.453(0.070)
	D.1	1.099(0.064)	-0.495(0.063)	0.044(0.059)	0.548(0.063)	1.452(0.072)
	D.2	1.098(0.063)	-0.496(0.062)	0.043(0.059)	0.550(0.063)	1.453(0.070)

Table 9: The estimates of the loading and thresholds for Item 8 (equal)

N		λ_8	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.020	-0.911	-0.409	0.164	0.921
250	A.1	0.975(0.114)	-0.912(0.122)	-0.389(0.121)	0.184(0.112)	0.920(0.122)
	A.2	0.975(0.114)	-0.912(0.122)	-0.389(0.121)	0.184(0.112)	0.920(0.122)
	B.1	1.037(0.126)	-0.937(0.134)	-0.415(0.134)	0.170(0.125)	0.930(0.133)
	B.2	1.038(0.123)	-0.938(0.132)	-0.417(0.133)	0.170(0.126)	0.930(0.132)
	C.1	1.039(0.126)	-0.939(0.135)	-0.417(0.136)	0.169(0.127)	0.929(0.133)
	C.2	1.039(0.122)	-0.939(0.131)	-0.418(0.132)	0.169(0.125)	0.929(0.131)
	D.1	1.038(0.122)	-0.938(0.133)	-0.416(0.134)	0.171(0.127)	0.931(0.133)
	D.2	1.038(0.122)	-0.938(0.133)	-0.416(0.134)	0.171(0.127)	0.931(0.133)
500	A.1	0.961(0.079)	-0.896(0.089)	-0.387(0.080)	0.170(0.084)	0.919(0.090)
	A.2	0.961(0.079)	-0.896(0.089)	-0.387(0.080)	0.170(0.084)	0.919(0.090)
	B.1	1.020(0.087)	-0.919(0.095)	-0.414(0.089)	0.156(0.093)	0.930(0.099)
	B.2	1.019(0.086)	-0.914(0.094)	-0.411(0.087)	0.159(0.092)	0.932(0.097)
	C.1	1.020(0.087)	-0.918(0.094)	-0.413(0.088)	0.157(0.093)	0.931(0.099)
	C.2	1.020(0.086)	-0.914(0.094)	-0.411(0.088)	0.159(0.092)	0.931(0.098)
	D.1	1.020(0.087)	-0.919(0.094)	-0.414(0.088)	0.157(0.093)	0.931(0.099)
	D.2	1.019(0.086)	-0.914(0.093)	-0.412(0.087)	0.158(0.092)	0.931(0.098)
1000	A.1	0.959(0.052)	-0.896(0.054)	-0.387(0.057)	0.177(0.056)	0.913(0.055)
	A.2	0.959(0.052)	-0.896(0.054)	-0.387(0.057)	0.177(0.056)	0.913(0.055)
	B.1	1.018(0.058)	-0.912(0.062)	-0.410(0.064)	0.165(0.060)	0.923(0.066)
	B.2	1.017(0.058)	-0.912(0.062)	-0.409(0.064)	0.165(0.060)	0.925(0.066)
	C.1	1.018(0.057)	-0.912(0.062)	-0.410(0.065)	0.164(0.061)	0.923(0.066)
	C.2	1.018(0.057)	-0.912(0.062)	-0.409(0.065)	0.164(0.060)	0.924(0.065)
	D.1	1.018(0.058)	-0.913(0.062)	-0.410(0.065)	0.164(0.061)	0.923(0.066)
	D.2	1.017(0.057)	-0.912(0.062)	-0.409(0.064)	0.164(0.060)	0.924(0.065)

Table 10: The estimates of the loading and thresholds for Item 9 (equal)

N		λ_9	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.880	-0.591	-0.053	0.529	1.138
250	A.1	0.858(0.102)	-0.602(0.111)	-0.049(0.105)	0.523(0.114)	1.127(0.115)
	A.2	0.858(0.102)	-0.602(0.111)	-0.049(0.105)	0.523(0.114)	1.127(0.115)
	B.1	0.890(0.110)	-0.590(0.120)	-0.041(0.112)	0.540(0.123)	1.149(0.125)
	B.2	0.890(0.109)	-0.590(0.120)	-0.042(0.112)	0.538(0.122)	1.148(0.123)
	C.1	0.891(0.111)	-0.590(0.120)	-0.041(0.112)	0.539(0.123)	1.149(0.126)
	C.2	0.891(0.109)	-0.591(0.120)	-0.042(0.111)	0.538(0.121)	1.147(0.123)
	D.1	0.891(0.109)	-0.590(0.123)	-0.041(0.113)	0.539(0.123)	1.148(0.123)
	D.2	0.891(0.109)	-0.590(0.123)	-0.041(0.113)	0.539(0.123)	1.148(0.123)
500	A.1	0.847(0.074)	-0.618(0.086)	-0.071(0.075)	0.513(0.082)	1.119(0.091)
	A.2	0.847(0.074)	-0.618(0.086)	-0.071(0.075)	0.513(0.082)	1.119(0.091)
	B.1	0.875(0.076)	-0.605(0.090)	-0.062(0.079)	0.528(0.088)	1.136(0.095)
	B.2	0.876(0.076)	-0.602(0.089)	-0.060(0.078)	0.530(0.088)	1.139(0.094)
	C.1	0.875(0.076)	-0.604(0.090)	-0.061(0.079)	0.529(0.088)	1.137(0.095)
	C.2	0.876(0.076)	-0.603(0.089)	-0.061(0.079)	0.529(0.088)	1.139(0.094)
	D.1	0.875(0.076)	-0.605(0.091)	-0.062(0.079)	0.528(0.088)	1.137(0.095)
	D.2	0.876(0.076)	-0.603(0.090)	-0.061(0.078)	0.529(0.087)	1.138(0.093)
1000	A.1	0.849(0.053)	-0.604(0.055)	-0.062(0.052)	0.513(0.056)	1.119(0.062)
	A.2	0.849(0.053)	-0.604(0.055)	-0.062(0.052)	0.513(0.056)	1.119(0.062)
	B.1	0.878(0.059)	-0.591(0.060)	-0.054(0.056)	0.529(0.061)	1.137(0.067)
	B.2	0.878(0.058)	-0.591(0.059)	-0.053(0.056)	0.529(0.060)	1.138(0.066)
	C.1	0.879(0.059)	-0.591(0.060)	-0.054(0.055)	0.528(0.060)	1.137(0.066)
	C.2	0.879(0.058)	-0.591(0.059)	-0.053(0.055)	0.529(0.059)	1.138(0.066)
	D.1	0.878(0.059)	-0.591(0.060)	-0.054(0.055)	0.529(0.061)	1.136(0.068)
	D.2	0.879(0.057)	-0.591(0.059)	-0.053(0.055)	0.529(0.059)	1.137(0.066)

Table 11: The estimates of the loading and thresholds for Item 10 (equal)

N		λ_{10}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.880	0.307	0.763	1.472	2.130
250	A.1	0.867(0.109)	0.231(0.103)	0.687(0.110)	1.370(0.143)	2.007(0.178)
	A.2	0.867(0.109)	0.231(0.103)	0.687(0.110)	1.370(0.143)	2.007(0.178)
	B.1	0.904(0.120)	0.313(0.118)	0.772(0.130)	1.505(0.178)	2.178(0.218)
	B.2	0.904(0.119)	0.311(0.115)	0.770(0.128)	1.500(0.174)	2.174(0.210)
	C.1	0.905(0.120)	0.314(0.118)	0.773(0.130)	1.506(0.179)	2.179(0.219)
	C.2	0.905(0.118)	0.312(0.115)	0.771(0.127)	1.501(0.175)	2.175(0.211)
	D.1	0.904(0.118)	0.313(0.117)	0.771(0.128)	1.501(0.175)	2.173(0.212)
	D.2	0.904(0.118)	0.313(0.117)	0.771(0.128)	1.501(0.175)	2.173(0.212)
500	A.1	0.848(0.080)	0.223(0.081)	0.683(0.086)	1.350(0.100)	1.985(0.135)
	A.2	0.848(0.080)	0.223(0.081)	0.683(0.086)	1.350(0.100)	1.985(0.135)
	B.1	0.882(0.088)	0.303(0.089)	0.764(0.097)	1.477(0.116)	2.151(0.166)
	B.2	0.883(0.087)	0.304(0.089)	0.766(0.096)	1.480(0.116)	2.154(0.162)
	C.1	0.881(0.088)	0.304(0.089)	0.766(0.098)	1.479(0.118)	2.153(0.169)
	C.2	0.883(0.087)	0.304(0.089)	0.766(0.096)	1.480(0.117)	2.154(0.164)
	D.1	0.882(0.088)	0.303(0.089)	0.764(0.099)	1.476(0.118)	2.152(0.171)
	D.2	0.882(0.087)	0.303(0.088)	0.764(0.096)	1.478(0.117)	2.153(0.164)
1000	A.1	0.848(0.063)	0.226(0.055)	0.683(0.054)	1.349(0.073)	1.990(0.090)
	A.2	0.848(0.063)	0.226(0.055)	0.683(0.054)	1.349(0.073)	1.990(0.090)
	B.1	0.879(0.067)	0.307(0.057)	0.763(0.062)	1.472(0.086)	2.136(0.111)
	B.2	0.880(0.067)	0.308(0.057)	0.764(0.060)	1.472(0.085)	2.138(0.109)
	C.1	0.878(0.067)	0.307(0.057)	0.764(0.060)	1.472(0.085)	2.136(0.111)
	C.2	0.880(0.066)	0.308(0.056)	0.764(0.060)	1.473(0.084)	2.137(0.110)
	D.1	0.879(0.067)	0.307(0.057)	0.763(0.060)	1.473(0.086)	2.137(0.110)
	D.2	0.880(0.066)	0.309(0.056)	0.764(0.060)	1.474(0.084)	2.139(0.110)

Table 12: The estimates of the loading and thresholds for Item 11 (equal)

N		λ_{11}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.800	-0.360	0.088	0.584	1.136
250	A.1	0.788(0.103)	-0.396(0.100)	0.067(0.106)	0.573(0.105)	1.134(0.126)
	A.2	0.788(0.103)	-0.396(0.100)	0.067(0.106)	0.573(0.105)	1.134(0.126)
	B.1	0.806(0.106)	-0.365(0.107)	0.089(0.114)	0.593(0.118)	1.149(0.137)
	B.2	0.806(0.105)	-0.364(0.107)	0.089(0.114)	0.593(0.117)	1.149(0.135)
	C.1	0.807(0.106)	-0.365(0.106)	0.088(0.114)	0.593(0.118)	1.149(0.137)
	C.2	0.806(0.105)	-0.364(0.107)	0.089(0.113)	0.593(0.117)	1.149(0.134)
	D.1	0.806(0.105)	-0.363(0.108)	0.090(0.115)	0.595(0.118)	1.150(0.134)
	D.2	0.806(0.105)	-0.363(0.108)	0.090(0.115)	0.595(0.118)	1.150(0.134)
500	A.1	0.775(0.071)	-0.396(0.068)	0.064(0.070)	0.558(0.073)	1.114(0.085)
	A.2	0.775(0.071)	-0.396(0.068)	0.064(0.070)	0.558(0.073)	1.114(0.085)
	B.1	0.792(0.074)	-0.366(0.073)	0.083(0.074)	0.577(0.078)	1.131(0.090)
	B.2	0.793(0.074)	-0.365(0.072)	0.085(0.073)	0.579(0.078)	1.132(0.090)
	C.1	0.791(0.074)	-0.366(0.073)	0.085(0.074)	0.578(0.078)	1.132(0.090)
	C.2	0.793(0.074)	-0.365(0.073)	0.085(0.074)	0.579(0.078)	1.132(0.091)
	D.1	0.792(0.074)	-0.366(0.073)	0.083(0.074)	0.577(0.078)	1.131(0.090)
	D.2	0.793(0.074)	-0.365(0.073)	0.084(0.074)	0.578(0.078)	1.131(0.090)
1000	A.1	0.774(0.055)	-0.396(0.044)	0.068(0.053)	0.563(0.052)	1.121(0.059)
	A.2	0.774(0.055)	-0.396(0.044)	0.068(0.053)	0.563(0.052)	1.121(0.059)
	B.1	0.796(0.055)	-0.362(0.051)	0.088(0.056)	0.581(0.058)	1.135(0.063)
	B.2	0.797(0.054)	-0.361(0.050)	0.087(0.055)	0.584(0.057)	1.136(0.062)
	C.1	0.798(0.055)	-0.362(0.051)	0.088(0.057)	0.582(0.057)	1.135(0.063)
	C.2	0.798(0.053)	-0.361(0.050)	0.087(0.056)	0.584(0.056)	1.136(0.063)
	D.1	0.797(0.055)	-0.362(0.052)	0.087(0.057)	0.582(0.057)	1.135(0.062)
	D.2	0.797(0.054)	-0.360(0.050)	0.086(0.056)	0.584(0.056)	1.137(0.062)

Table 13: The estimates of the loading and thresholds for Item 12 (equal)

N		λ_{12}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.800	-0.232	0.128	0.856	1.568
	A.1	0.792(0.101)	-0.284(0.101)	0.102(0.105)	0.821(0.121)	1.518(0.142)
	A.2	0.792(0.101)	-0.284(0.101)	0.102(0.105)	0.821(0.121)	1.518(0.142)
	B.1	0.813(0.105)	-0.247(0.108)	0.124(0.113)	0.866(0.131)	1.577(0.154)
	B.2	0.814(0.103)	-0.248(0.105)	0.124(0.110)	0.868(0.128)	1.579(0.155)
250	C.1	0.814(0.104)	-0.246(0.107)	0.124(0.113)	0.866(0.132)	1.577(0.154)
	C.2	0.815(0.103)	-0.248(0.105)	0.124(0.110)	0.868(0.129)	1.580(0.156)
	D.1	0.814(0.102)	-0.247(0.106)	0.126(0.112)	0.869(0.131)	1.581(0.159)
	D.2	0.814(0.102)	-0.247(0.106)	0.126(0.112)	0.869(0.131)	1.581(0.159)
	A.1	0.775(0.064)	-0.267(0.071)	0.114(0.074)	0.817(0.076)	1.525(0.096)
	A.2	0.775(0.064)	-0.267(0.071)	0.114(0.074)	0.817(0.076)	1.525(0.096)
	B.1	0.795(0.067)	-0.230(0.077)	0.134(0.080)	0.857(0.082)	1.585(0.106)
	B.2	0.796(0.067)	-0.228(0.076)	0.135(0.079)	0.858(0.081)	1.585(0.105)
500	C.1	0.794(0.067)	-0.229(0.077)	0.135(0.080)	0.858(0.083)	1.586(0.107)
	C.2	0.796(0.067)	-0.228(0.076)	0.135(0.079)	0.858(0.081)	1.584(0.105)
	D.1	0.795(0.066)	-0.230(0.078)	0.134(0.080)	0.857(0.082)	1.585(0.107)
	D.2	0.795(0.066)	-0.229(0.076)	0.134(0.079)	0.857(0.081)	1.584(0.105)
	A.1	0.774(0.052)	-0.276(0.051)	0.104(0.047)	0.804(0.052)	1.506(0.071)
	A.2	0.774(0.052)	-0.276(0.051)	0.104(0.047)	0.804(0.052)	1.506(0.071)
	B.1	0.799(0.055)	-0.232(0.052)	0.127(0.050)	0.853(0.055)	1.567(0.078)
	B.2	0.798(0.055)	-0.234(0.051)	0.128(0.050)	0.854(0.055)	1.568(0.077)
1000	C.1	0.798(0.056)	-0.236(0.052)	0.128(0.051)	0.854(0.054)	1.568(0.077)
	C.2	0.798(0.055)	-0.234(0.051)	0.128(0.049)	0.854(0.054)	1.568(0.076)
	D.1	0.799(0.055)	-0.236(0.053)	0.127(0.050)	0.855(0.055)	1.567(0.077)
	D.2	0.798(0.054)	-0.235(0.052)	0.128(0.049)	0.854(0.053)	1.567(0.076)

Table 14: The estimates of the loading and thresholds for Item 13 (equal)

N		λ_{13}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.790	-0.143	0.238	0.810	1.374
250	A.1	0.785(0.095)	-0.190(0.099)	0.208(0.092)	0.774(0.103)	1.357(0.120)
	A.2	0.785(0.095)	-0.190(0.099)	0.208(0.092)	0.774(0.103)	1.357(0.120)
	B.1	0.800(0.099)	-0.147(0.108)	0.238(0.101)	0.812(0.114)	1.395(0.133)
	B.2	0.800(0.099)	-0.147(0.107)	0.238(0.102)	0.811(0.115)	1.393(0.131)
	C.1	0.801(0.099)	-0.147(0.109)	0.238(0.103)	0.812(0.116)	1.394(0.135)
	C.2	0.800(0.099)	-0.147(0.106)	0.238(0.101)	0.811(0.116)	1.392(0.132)
	D.1	0.800(0.100)	-0.146(0.107)	0.238(0.101)	0.812(0.116)	1.393(0.132)
	D.2	0.800(0.100)	-0.146(0.107)	0.238(0.101)	0.812(0.116)	1.393(0.132)
500	A.1	0.775(0.072)	-0.187(0.072)	0.210(0.071)	0.774(0.081)	1.343(0.092)
	A.2	0.775(0.072)	-0.187(0.072)	0.210(0.071)	0.774(0.081)	1.343(0.092)
	B.1	0.788(0.075)	-0.143(0.079)	0.240(0.077)	0.811(0.089)	1.377(0.098)
	B.2	0.789(0.075)	-0.141(0.077)	0.242(0.076)	0.813(0.087)	1.378(0.098)
	C.1	0.787(0.074)	-0.142(0.079)	0.241(0.077)	0.812(0.089)	1.378(0.099)
	C.2	0.789(0.074)	-0.141(0.078)	0.242(0.076)	0.813(0.087)	1.378(0.097)
	D.1	0.788(0.075)	-0.142(0.079)	0.240(0.077)	0.811(0.090)	1.378(0.099)
	D.2	0.788(0.074)	-0.141(0.077)	0.242(0.076)	0.812(0.087)	1.377(0.098)
1000	A.1	0.776(0.050)	-0.192(0.051)	0.206(0.052)	0.770(0.052)	1.338(0.060)
	A.2	0.776(0.050)	-0.192(0.051)	0.206(0.052)	0.770(0.052)	1.338(0.060)
	B.1	0.790(0.054)	-0.143(0.055)	0.237(0.055)	0.809(0.059)	1.374(0.067)
	B.2	0.791(0.054)	-0.144(0.054)	0.238(0.054)	0.810(0.058)	1.374(0.066)
	C.1	0.790(0.054)	-0.146(0.056)	0.236(0.055)	0.808(0.059)	1.372(0.067)
	C.2	0.791(0.054)	-0.144(0.054)	0.238(0.053)	0.811(0.057)	1.374(0.066)
	D.1	0.791(0.054)	-0.145(0.054)	0.236(0.055)	0.808(0.058)	1.371(0.066)
	D.2	0.792(0.053)	-0.144(0.052)	0.238(0.053)	0.810(0.057)	1.374(0.065)

Table 15: The estimates of the loading and thresholds for Item 14 (equal)

N		λ_{14}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.790	-0.087	0.397	0.866	1.517
250	A.1	0.780(0.106)	0.032(0.097)	0.362(0.103)	0.829(0.121)	1.472(0.136)
	A.2	0.780(0.106)	0.032(0.097)	0.362(0.103)	0.829(0.121)	1.472(0.136)
	B.1	0.792(0.110)	0.092(0.104)	0.404(0.112)	0.870(0.133)	1.524(0.161)
	B.2	0.792(0.110)	0.092(0.103)	0.403(0.110)	0.869(0.131)	1.523(0.158)
	C.1	0.793(0.111)	0.093(0.104)	0.405(0.113)	0.871(0.134)	1.524(0.162)
	C.2	0.792(0.110)	0.093(0.103)	0.403(0.110)	0.869(0.131)	1.522(0.158)
	D.1	0.792(0.110)	0.094(0.105)	0.405(0.111)	0.870(0.132)	1.523(0.159)
	D.2	0.792(0.110)	0.094(0.105)	0.405(0.111)	0.870(0.132)	1.523(0.159)
500	A.1	0.779(0.070)	0.017(0.074)	0.342(0.075)	0.817(0.080)	1.471(0.107)
	A.2	0.779(0.070)	0.017(0.074)	0.342(0.075)	0.817(0.080)	1.471(0.107)
	B.1	0.790(0.072)	0.076(0.081)	0.383(0.081)	0.856(0.085)	1.517(0.115)
	B.2	0.791(0.072)	0.078(0.079)	0.385(0.078)	0.858(0.083)	1.517(0.109)
	C.1	0.790(0.073)	0.077(0.081)	0.383(0.081)	0.858(0.086)	1.518(0.115)
	C.2	0.791(0.072)	0.078(0.079)	0.385(0.079)	0.858(0.083)	1.516(0.110)
	D.1	0.790(0.073)	0.075(0.081)	0.382(0.081)	0.856(0.086)	1.517(0.115)
	D.2	0.791(0.072)	0.077(0.079)	0.384(0.078)	0.857(0.083)	1.516(0.109)
1000	A.1	0.775(0.051)	0.027(0.053)	0.356(0.055)	0.830(0.057)	1.478(0.066)
	A.2	0.775(0.051)	0.027(0.053)	0.356(0.055)	0.830(0.057)	1.478(0.066)
	B.1	0.789(0.054)	0.088(0.056)	0.397(0.058)	0.870(0.060)	1.520(0.073)
	B.2	0.789(0.052)	0.087(0.055)	0.397(0.058)	0.872(0.059)	1.520(0.073)
	C.1	0.789(0.053)	0.088(0.056)	0.397(0.058)	0.870(0.060)	1.519(0.073)
	C.2	0.789(0.052)	0.087(0.054)	0.397(0.057)	0.870(0.059)	1.517(0.072)
	D.1	0.788(0.054)	0.087(0.054)	0.397(0.058)	0.871(0.059)	1.519(0.073)
	D.2	0.789(0.053)	0.087(0.054)	0.397(0.057)	0.872(0.059)	1.517(0.072)

Table 16: The estimates of the loading and thresholds for Item 15 (equal)

N		λ_{15}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.790	0.079	0.410	0.788	1.300
250	A.1	0.794(0.117)	0.015(0.103)	0.367(0.104)	0.759(0.116)	1.293(0.126)
	A.2	0.794(0.117)	0.015(0.103)	0.367(0.104)	0.759(0.116)	1.293(0.126)
	B.1	0.804(0.121)	0.073(0.114)	0.410(0.114)	0.790(0.125)	1.316(0.135)
	B.2	0.805(0.122)	0.072(0.110)	0.409(0.112)	0.789(0.123)	1.315(0.135)
	C.1	0.804(0.121)	0.074(0.115)	0.411(0.114)	0.790(0.124)	1.316(0.135)
	C.2	0.805(0.121)	0.073(0.110)	0.410(0.112)	0.789(0.122)	1.314(0.134)
	D.1	0.804(0.122)	0.073(0.111)	0.410(0.113)	0.790(0.123)	1.315(0.135)
	D.2	0.804(0.122)	0.073(0.111)	0.410(0.113)	0.790(0.123)	1.315(0.135)
500	A.1	0.783(0.076)	0.018(0.078)	0.369(0.078)	0.754(0.078)	1.278(0.091)
	A.2	0.783(0.076)	0.018(0.078)	0.369(0.078)	0.754(0.078)	1.278(0.091)
	B.1	0.790(0.079)	0.076(0.085)	0.412(0.086)	0.784(0.086)	1.300(0.098)
	B.2	0.792(0.078)	0.078(0.083)	0.413(0.085)	0.786(0.084)	1.302(0.099)
	C.1	0.790(0.079)	0.077(0.085)	0.413(0.086)	0.785(0.085)	1.301(0.098)
	C.2	0.792(0.078)	0.077(0.083)	0.413(0.085)	0.786(0.084)	1.302(0.098)
	D.1	0.790(0.079)	0.076(0.086)	0.412(0.086)	0.784(0.085)	1.300(0.098)
	D.2	0.791(0.078)	0.077(0.083)	0.413(0.085)	0.785(0.084)	1.301(0.098)
1000	A.1	0.783(0.050)	0.016(0.050)	0.363(0.050)	0.757(0.053)	1.280(0.063)
	A.2	0.783(0.050)	0.016(0.050)	0.363(0.050)	0.757(0.053)	1.280(0.063)
	B.1	0.788(0.051)	0.077(0.054)	0.406(0.054)	0.788(0.059)	1.301(0.067)
	B.2	0.790(0.050)	0.077(0.054)	0.406(0.053)	0.789(0.059)	1.301(0.066)
	C.1	0.787(0.051)	0.078(0.052)	0.408(0.054)	0.788(0.059)	1.302(0.068)
	C.2	0.790(0.050)	0.077(0.052)	0.407(0.054)	0.789(0.058)	1.301(0.066)
	D.1	0.788(0.051)	0.078(0.053)	0.408(0.054)	0.788(0.059)	1.301(0.066)
	D.2	0.790(0.051)	0.078(0.052)	0.407(0.053)	0.788(0.058)	1.302(0.064)

Table 17: The estimates of the loading and thresholds for Item 16 (equal)

N		λ_{16}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.780	-0.349	0.023	0.613	1.350
250	A.1	0.762(0.108)	-0.381(0.105)	0.009(0.108)	0.598(0.103)	1.331(0.125)
	A.2	0.762(0.108)	-0.381(0.105)	0.009(0.108)	0.598(0.103)	1.331(0.125)
	B.1	0.782(0.111)	-0.352(0.114)	0.024(0.113)	0.622(0.108)	1.372(0.140)
	B.2	0.783(0.112)	-0.353(0.112)	0.022(0.112)	0.620(0.108)	1.369(0.138)
	C.1	0.782(0.111)	-0.351(0.114)	0.024(0.113)	0.621(0.107)	1.371(0.139)
	C.2	0.783(0.112)	-0.353(0.112)	0.022(0.111)	0.620(0.108)	1.369(0.138)
	D.1	0.783(0.112)	-0.353(0.113)	0.023(0.112)	0.621(0.109)	1.369(0.139)
	D.2	0.783(0.112)	-0.353(0.113)	0.023(0.112)	0.621(0.109)	1.369(0.139)
500	A.1	0.753(0.074)	-0.390(0.078)	0.003(0.077)	0.593(0.089)	1.311(0.097)
	A.2	0.753(0.074)	-0.390(0.078)	0.003(0.077)	0.593(0.089)	1.311(0.097)
	B.1	0.772(0.076)	-0.361(0.084)	0.015(0.084)	0.613(0.096)	1.345(0.109)
	B.2	0.773(0.076)	-0.358(0.083)	0.017(0.082)	0.615(0.094)	1.347(0.106)
	C.1	0.772(0.076)	-0.360(0.084)	0.016(0.084)	0.614(0.096)	1.346(0.110)
	C.2	0.773(0.075)	-0.359(0.083)	0.017(0.083)	0.615(0.094)	1.347(0.106)
	D.1	0.772(0.075)	-0.361(0.084)	0.015(0.083)	0.613(0.095)	1.345(0.109)
	D.2	0.773(0.075)	-0.359(0.082)	0.016(0.081)	0.614(0.093)	1.346(0.106)
1000	A.1	0.766(0.049)	-0.383(0.052)	0.005(0.054)	0.585(0.049)	1.318(0.060)
	A.2	0.766(0.049)	-0.383(0.052)	0.005(0.054)	0.585(0.049)	1.318(0.060)
	B.1	0.780(0.053)	-0.350(0.055)	0.024(0.056)	0.610(0.053)	1.351(0.067)
	B.2	0.782(0.053)	-0.349(0.054)	0.021(0.056)	0.609(0.053)	1.351(0.066)
	C.1	0.780(0.052)	-0.348(0.054)	0.022(0.055)	0.611(0.053)	1.353(0.066)
	C.2	0.782(0.052)	-0.349(0.055)	0.021(0.053)	0.611(0.052)	1.352(0.065)
	D.1	0.781(0.053)	-0.350(0.054)	0.022(0.055)	0.610(0.053)	1.353(0.066)
	D.2	0.781(0.052)	-0.350(0.053)	0.020(0.054)	0.611(0.052)	1.351(0.064)

Table 18: The estimates of the loading and thresholds for Item 17 (equal)

N		λ_{17}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.730	-0.518	-0.212	0.241	0.861
250	A.1	0.709(0.098)	-0.540(0.112)	-0.207(0.103)	0.248(0.103)	0.867(0.113)
	A.2	0.709(0.098)	-0.540(0.112)	-0.207(0.103)	0.248(0.103)	0.867(0.113)
	B.1	0.733(0.103)	-0.523(0.121)	-0.210(0.110)	0.241(0.110)	0.866(0.119)
	B.2	0.734(0.103)	-0.524(0.119)	-0.211(0.110)	0.241(0.109)	0.865(0.117)
	C.1	0.733(0.103)	-0.523(0.121)	-0.211(0.111)	0.241(0.111)	0.865(0.118)
	C.2	0.734(0.103)	-0.524(0.119)	-0.212(0.110)	0.240(0.110)	0.864(0.116)
	D.1	0.733(0.103)	-0.523(0.119)	-0.211(0.110)	0.241(0.111)	0.865(0.117)
	D.2	0.733(0.103)	-0.523(0.119)	-0.211(0.110)	0.241(0.111)	0.865(0.117)
500	A.1	0.713(0.068)	-0.540(0.073)	-0.212(0.073)	0.244(0.073)	0.866(0.080)
	A.2	0.713(0.068)	-0.540(0.073)	-0.212(0.073)	0.244(0.073)	0.866(0.080)
	B.1	0.736(0.072)	-0.523(0.079)	-0.215(0.078)	0.237(0.079)	0.864(0.087)
	B.2	0.737(0.073)	-0.521(0.078)	-0.213(0.077)	0.239(0.078)	0.867(0.087)
	C.1	0.736(0.072)	-0.522(0.078)	-0.214(0.079)	0.238(0.080)	0.865(0.087)
	C.2	0.737(0.073)	-0.521(0.078)	-0.213(0.077)	0.239(0.078)	0.866(0.087)
	D.1	0.736(0.072)	-0.523(0.078)	-0.214(0.078)	0.238(0.080)	0.865(0.087)
	D.2	0.736(0.072)	-0.521(0.078)	-0.214(0.077)	0.239(0.078)	0.866(0.087)
1000	A.1	0.710(0.042)	-0.539(0.044)	-0.213(0.046)	0.239(0.042)	0.862(0.050)
	A.2	0.710(0.042)	-0.539(0.044)	-0.213(0.046)	0.239(0.042)	0.862(0.050)
	B.1	0.729(0.047)	-0.518(0.052)	-0.213(0.050)	0.238(0.048)	0.861(0.055)
	B.2	0.730(0.048)	-0.519(0.053)	-0.213(0.050)	0.238(0.048)	0.860(0.056)
	C.1	0.728(0.048)	-0.518(0.052)	-0.212(0.049)	0.237(0.048)	0.858(0.055)
	C.2	0.730(0.047)	-0.519(0.053)	-0.212(0.048)	0.237(0.048)	0.858(0.056)
	D.1	0.730(0.048)	-0.519(0.052)	-0.213(0.050)	0.238(0.048)	0.860(0.055)
	D.2	0.730(0.047)	-0.518(0.051)	-0.212(0.049)	0.237(0.047)	0.860(0.055)

Table 19: The estimates of the loading and thresholds for Item 18 (equal)

N		λ_{18}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.730	0.605	1.021	1.583	2.086
250	A.1	0.707(0.105)	0.498(0.108)	0.918(0.119)	1.457(0.142)	1.956(0.170)
	A.2	0.707(0.105)	0.498(0.108)	0.918(0.119)	1.457(0.142)	1.956(0.170)
	B.1	0.735(0.114)	0.608(0.131)	1.036(0.145)	1.615(0.178)	2.138(0.230)
	B.2	0.735(0.114)	0.607(0.123)	1.034(0.137)	1.611(0.163)	2.130(0.221)
	C.1	0.737(0.114)	0.611(0.132)	1.041(0.148)	1.620(0.185)	2.142(0.234)
	C.2	0.736(0.113)	0.608(0.124)	1.036(0.137)	1.613(0.165)	2.131(0.221)
	D.1	0.736(0.114)	0.609(0.126)	1.037(0.139)	1.615(0.171)	2.130(0.223)
	D.2	0.736(0.114)	0.609(0.126)	1.037(0.139)	1.615(0.171)	2.130(0.223)
500	A.1	0.696(0.074)	0.495(0.073)	0.900(0.076)	1.431(0.093)	1.922(0.110)
	A.2	0.696(0.074)	0.495(0.073)	0.900(0.076)	1.431(0.093)	1.922(0.110)
	B.1	0.721(0.081)	0.601(0.083)	1.013(0.088)	1.581(0.109)	2.084(0.141)
	B.2	0.722(0.082)	0.603(0.080)	1.016(0.088)	1.583(0.108)	2.088(0.139)
	C.1	0.721(0.082)	0.602(0.083)	1.014(0.089)	1.583(0.112)	2.085(0.143)
	C.2	0.722(0.082)	0.603(0.080)	1.015(0.088)	1.583(0.109)	2.088(0.139)
	D.1	0.722(0.082)	0.601(0.082)	1.014(0.089)	1.582(0.112)	2.085(0.143)
	D.2	0.722(0.082)	0.602(0.080)	1.015(0.088)	1.582(0.109)	2.087(0.140)
1000	A.1	0.709(0.050)	0.495(0.051)	0.920(0.053)	1.437(0.060)	1.936(0.082)
	A.2	0.709(0.050)	0.495(0.051)	0.920(0.053)	1.437(0.060)	1.936(0.082)
	B.1	0.731(0.057)	0.604(0.054)	1.022(0.060)	1.586(0.069)	2.089(0.099)
	B.2	0.730(0.057)	0.605(0.054)	1.023(0.058)	1.586(0.070)	2.090(0.097)
	C.1	0.731(0.057)	0.605(0.054)	1.022(0.059)	1.586(0.070)	2.091(0.099)
	C.2	0.730(0.057)	0.605(0.053)	1.023(0.058)	1.587(0.070)	2.091(0.096)
	D.1	0.730(0.057)	0.605(0.054)	1.023(0.060)	1.586(0.071)	2.091(0.097)
	D.2	0.730(0.056)	0.605(0.052)	1.023(0.058)	1.587(0.069)	2.090(0.095)

3.2.2 Under The Unequal Case

Under the "unequal case," the results are presented form Tables 20 to 37. In Table 20, when $N = 250$, we have the same conclusion about the models considering the latent classes as described in 3.2.1. The models with latent classes (B.1, B. 2, C.1, C.2 and D.1, D.2) have smaller bias than the models without latentent classes (A.1 and A.2). An unrestricted model is better than a restricted model. For instance, the bias of the estimates of B.2 are smaller than B.1. Both of B.1 and B.2 response distribution are not exact, but the cooperating with response time model (B.2) can minimize the damage of improper response distribution model (B.1). Therefore, the models C.1, C.2 and D.1, D.2 have smaller bias of the estimates because their response distributions are relaxed. For example, the τ_1 of Item 1, the bias of C.1 and C.2 are smaller than B.1 and B.2. It is similiar when we compare the models B.1 nad B.2 versus D.1 and D.2. However, the bias are not consistent as described in Section 3.2.1. The bias of the models (C. 1, C.2 and D.1, D.2) have larger bias than B.1 and B.2. But it becomes better while the sample size is large enough. The comparison of the models C.1 and C. 2 versus D.1 and D.2 are stated: the bias of D.1 and D.2 are larger than C.1 and C.2. Though D.1 and D.2 are the true models, the models C.1 and C.2 which constrain the carelss 1 and careless 2 to have the similiar response distribution with smaller bias than models D.1 and D.2.

When $N = 500$, there are two same findings as Section 3.2.1. First, the bias and the standard errors become smaller than the size 250 as expected. Second, the effect of the response time to the standard errors has become obviously. The models with response time (B.2, C.2 and D.2) have smaller standard errors than the models without time (B.1, C.1 and D.1).

Summary: there are some conclusions in this section as following,

- The unrestricted models (B.2, C.1, C.2, and D.1, D.2) have smaller bias than the restricted model (B.1).
- The models with response time have smaller standard errors than those without response time, but there are no consistency upon the bias.

Table 20: The estimates of the loading and thresholds for Item 1 (unequal)

N		λ_1	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.640	-0.640	0.000	0.739	1.772
	A.1	1.611(0.167)	-0.724(0.154)	-0.074(0.159)	0.681(0.158)	1.736(0.196)
	A.2	1.611(0.167)	-0.724(0.154)	-0.074(0.159)	0.681(0.158)	1.736(0.196)
	B.1	1.629(0.175)	-0.726(0.158)	-0.075(0.162)	0.686(0.162)	1.752(0.204)
	B.2	1.649(0.178)	-0.712(0.160)	-0.059(0.162)	0.701(0.164)	1.773(0.210)
250	C.1	1.668(0.174)	-0.659(0.159)	-0.006(0.162)	0.748(0.167)	1.794(0.213)
	C.2	1.673(0.175)	-0.662(0.160)	-0.008(0.163)	0.744(0.169)	1.793(0.212)
	D.1	1.675(0.175)	-0.651(0.159)	0.002(0.164)	0.755(0.170)	1.807(0.218)
	D.2	1.675(0.174)	-0.653(0.162)	0.002(0.167)	0.753(0.173)	1.805(0.215)
	A.1	1.587(0.122)	-0.710(0.117)	-0.074(0.105)	0.669(0.100)	1.720(0.124)
	A.2	1.587(0.122)	-0.710(0.117)	-0.074(0.105)	0.669(0.100)	1.720(0.124)
	B.1	1.602(0.126)	-0.710(0.117)	-0.073(0.105)	0.675(0.101)	1.735(0.125)
	B.2	1.623(0.128)	-0.692(0.117)	-0.054(0.107)	0.696(0.103)	1.763(0.131)
500	C.1	1.641(0.130)	-0.639(0.120)	0.001(0.109)	0.741(0.104)	1.786(0.133)
	C.2	1.646(0.130)	-0.638(0.119)	0.002(0.109)	0.740(0.105)	1.788(0.133)
	D.1	1.648(0.132)	-0.632(0.121)	0.010(0.110)	0.748(0.106)	1.800(0.135)
	D.2	1.649(0.130)	-0.634(0.120)	0.006(0.109)	0.744(0.105)	1.795(0.134)
	A.1	1.584(0.086)	-0.710(0.074)	-0.070(0.072)	0.673(0.081)	1.720(0.099)
	A.2	1.584(0.086)	-0.710(0.074)	-0.070(0.072)	0.673(0.081)	1.720(0.099)
	B.1	1.593(0.085)	-0.711(0.076)	-0.064(0.074)	0.677(0.085)	1.730(0.102)
	B.2	1.620(0.083)	-0.690(0.079)	-0.044(0.076)	0.692(0.088)	1.758(0.105)
1000	C.1	1.642(0.088)	-0.639(0.080)	0.002(0.077)	0.743(0.091)	1.779(0.108)
	C.2	1.641(0.086)	-0.638(0.080)	0.002(0.078)	0.741(0.090)	1.781(0.107)
	D.1	1.643(0.088)	-0.637(0.080)	0.001(0.079)	0.744(0.091)	1.785(0.108)
	D.2	1.642(0.086)	-0.638(0.081)	0.000(0.078)	0.742(0.090)	1.783(0.107)

Table 21: The estimates of the loading and thresholds for Item 2 (unequal)

N		λ_2	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.370	-0.713	-0.164	0.562	1.576
250	A.1	1.325(0.138)	-0.771(0.143)	-0.229(0.132)	0.507(0.141)	1.535(0.168)
	A.2	1.325(0.138)	-0.771(0.143)	-0.229(0.132)	0.507(0.141)	1.535(0.168)
	B.1	1.338(0.141)	-0.774(0.147)	-0.232(0.136)	0.509(0.144)	1.546(0.171)
	B.2	1.354(0.144)	-0.762(0.147)	-0.220(0.135)	0.519(0.145)	1.562(0.176)
	C.1	1.367(0.146)	-0.719(0.149)	-0.177(0.138)	0.554(0.150)	1.576(0.178)
	C.2	1.373(0.149)	-0.722(0.151)	-0.181(0.138)	0.550(0.149)	1.574(0.179)
	D.1	1.373(0.149)	-0.712(0.152)	-0.171(0.139)	0.560(0.151)	1.586(0.180)
	D.2	1.374(0.151)	-0.714(0.152)	-0.173(0.139)	0.557(0.150)	1.584(0.181)
500	A.1	1.325(0.100)	-0.775(0.093)	-0.222(0.092)	0.508(0.090)	1.533(0.116)
	A.2	1.325(0.100)	-0.775(0.093)	-0.222(0.092)	0.508(0.090)	1.533(0.116)
	B.1	1.335(0.102)	-0.774(0.094)	-0.222(0.093)	0.511(0.092)	1.543(0.119)
	B.2	1.351(0.104)	-0.759(0.095)	-0.208(0.094)	0.527(0.094)	1.562(0.121)
	C.1	1.364(0.107)	-0.714(0.097)	-0.163(0.095)	0.563(0.098)	1.579(0.125)
	C.2	1.368(0.108)	-0.713(0.098)	-0.162(0.095)	0.563(0.097)	1.579(0.125)
	D.1	1.368(0.109)	-0.707(0.098)	-0.156(0.096)	0.569(0.099)	1.588(0.127)
	D.2	1.370(0.108)	-0.710(0.098)	-0.159(0.095)	0.565(0.097)	1.584(0.123)
1000	A.1	1.332(0.070)	-0.781(0.075)	-0.222(0.068)	0.505(0.069)	1.543(0.082)
	A.2	1.332(0.070)	-0.781(0.075)	-0.222(0.068)	0.505(0.069)	1.543(0.082)
	B.1	1.349(0.073)	-0.781(0.078)	-0.219(0.069)	0.514(0.071)	1.547(0.086)
	B.2	1.361(0.076)	-0.766(0.078)	-0.182(0.069)	0.526(0.070)	1.570(0.087)
	C.1	1.370(0.078)	-0.714(0.079)	-0.167(0.077)	0.560(0.073)	1.576(0.090)
	C.2	1.372(0.076)	-0.714(0.078)	-0.166(0.076)	0.560(0.072)	1.578(0.090)
	D.1	1.374(0.078)	-0.713(0.079)	-0.165(0.077)	0.561(0.070)	1.576(0.090)
	D.2	1.373(0.077)	-0.712(0.078)	-0.166(0.076)	0.562(0.070)	1.575(0.091)

Table 22: The estimates of the loading and thresholds for Item 3 (unequal)

N		λ_3	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.300	-0.273	0.325	1.066	1.989
250	A.1	1.311(0.140)	-0.378(0.129)	0.227(0.132)	0.966(0.149)	1.935(0.193)
	A.2	1.311(0.140)	-0.378(0.129)	0.227(0.132)	0.966(0.149)	1.935(0.193)
	B.1	1.319(0.142)	-0.372(0.132)	0.233(0.136)	0.977(0.154)	1.951(0.201)
	B.2	1.321(0.143)	-0.351(0.132)	0.255(0.136)	1.000(0.153)	1.977(0.205)
	C.1	1.304(0.142)	-0.292(0.129)	0.314(0.139)	1.054(0.159)	1.994(0.211)
	C.2	1.306(0.143)	-0.293(0.131)	0.313(0.140)	1.053(0.158)	1.993(0.211)
	D.1	1.305(0.143)	-0.284(0.131)	0.323(0.141)	1.062(0.161)	2.004(0.214)
	D.2	1.305(0.142)	-0.285(0.132)	0.322(0.143)	1.061(0.160)	2.003(0.214)
500	A.1	1.309(0.101)	-0.361(0.089)	0.232(0.091)	0.980(0.109)	1.944(0.135)
	A.2	1.309(0.101)	-0.361(0.089)	0.232(0.091)	0.980(0.109)	1.944(0.135)
	B.1	1.315(0.102)	-0.355(0.089)	0.240(0.092)	0.990(0.111)	1.961(0.139)
	B.2	1.315(0.102)	-0.327(0.090)	0.267(0.092)	1.018(0.113)	1.983(0.139)
	C.1	1.302(0.103)	-0.270(0.089)	0.324(0.090)	1.070(0.113)	1.999(0.139)
	C.2	1.304(0.105)	-0.269(0.089)	0.326(0.092)	1.071(0.113)	2.002(0.138)
	D.1	1.302(0.105)	-0.261(0.092)	0.333(0.093)	1.077(0.114)	2.009(0.140)
	D.2	1.304(0.105)	-0.264(0.090)	0.331(0.093)	1.075(0.113)	2.007(0.139)
1000	A.1	1.307(0.068)	-0.373(0.069)	0.232(0.073)	0.970(0.074)	1.929(0.090)
	A.2	1.307(0.068)	-0.373(0.069)	0.232(0.073)	0.970(0.074)	1.929(0.090)
	B.1	1.321(0.074)	-0.325(0.079)	0.236(0.072)	1.198(0.071)	1.961(0.079)
	B.2	1.311(0.073)	-0.300(0.070)	0.284(0.072)	1.004(0.073)	1.883(0.079)
	C.1	1.300(0.073)	-0.270(0.079)	0.324(0.070)	1.066(0.073)	1.999(0.079)
	C.2	1.302(0.075)	-0.272(0.079)	0.326(0.072)	1.065(0.072)	1.999(0.077)
	D.1	1.301(0.075)	-0.271(0.072)	0.323(0.073)	1.067(0.074)	1.998(0.075)
	D.2	1.301(0.075)	-0.274(0.070)	0.324(0.073)	1.067(0.073)	1.997(0.075)

Table 23: The estimates of the loading and thresholds for Item 4 (unequal)

N		λ_4	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.240	-0.939	-0.482	0.235	1.149
250	A.1	1.171(0.128)	-1.001(0.142)	-0.532(0.130)	0.201(0.126)	1.128(0.142)
	A.2	1.171(0.128)	-1.001(0.142)	-0.532(0.130)	0.201(0.126)	1.128(0.142)
	B.1	1.186(0.132)	-1.007(0.145)	-0.539(0.133)	0.198(0.129)	1.133(0.142)
	B.2	1.204(0.135)	-1.001(0.147)	-0.535(0.136)	0.199(0.132)	1.138(0.147)
	C.1	1.230(0.139)	-0.971(0.147)	-0.504(0.135)	0.226(0.138)	1.146(0.152)
	C.2	1.236(0.140)	-0.974(0.147)	-0.508(0.136)	0.221(0.137)	1.143(0.150)
	D.1	1.237(0.141)	-0.967(0.149)	-0.500(0.136)	0.229(0.139)	1.152(0.152)
	D.2	1.238(0.141)	-0.969(0.149)	-0.502(0.138)	0.226(0.139)	1.150(0.153)
500	A.1	1.174(0.082)	-0.975(0.093)	-0.509(0.086)	0.207(0.085)	1.136(0.092)
	A.2	1.174(0.082)	-0.975(0.093)	-0.509(0.086)	0.207(0.085)	1.136(0.092)
	B.1	1.186(0.085)	-0.978(0.095)	-0.514(0.087)	0.206(0.085)	1.141(0.094)
	B.2	1.206(0.087)	-0.970(0.095)	-0.508(0.087)	0.213(0.087)	1.152(0.096)
	C.1	1.236(0.090)	-0.941(0.097)	-0.480(0.089)	0.237(0.090)	1.160(0.098)
	C.2	1.239(0.091)	-0.941(0.098)	-0.480(0.089)	0.236(0.090)	1.106(0.099)
	D.1	1.240(0.093)	-0.934(0.099)	-0.473(0.089)	0.242(0.090)	1.168(0.099)
	D.2	1.242(0.091)	-0.938(0.098)	-0.477(0.088)	0.238(0.089)	1.164(0.098)
1000	A.1	1.179(0.065)	-0.977(0.063)	-0.520(0.060)	0.210(0.060)	1.132(0.066)
	A.2	1.179(0.065)	-0.977(0.063)	-0.520(0.060)	0.210(0.060)	1.132(0.066)
	B.1	1.190(0.069)	-0.975(0.064)	-0.518(0.059)	0.201(0.059)	1.136(0.068)
	B.2	1.211(0.070)	-0.967(0.066)	-0.512(0.062)	0.209(0.062)	1.140(0.070)
	C.1	1.233(0.070)	-0.937(0.066)	-0.483(0.061)	0.235(0.063)	1.151(0.069)
	C.2	1.236(0.071)	-0.937(0.066)	-0.483(0.063)	0.234(0.064)	1.150(0.070)
	D.1	1.239(0.071)	-0.937(0.067)	-0.482(0.064)	0.237(0.063)	1.148(0.070)
	D.2	1.239(0.071)	-0.936(0.066)	-0.483(0.063)	0.236(0.063)	1.150(0.070)

Table 24: The estimates of the loading and thresholds for Item 5 (unequal)

N		λ_5	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.160	-0.664	-0.338	0.326	1.002
250	A.1	1.129(0.128)	-0.729(0.131)	-0.392(0.129)	0.284(0.129)	1.004(0.140)
	A.2	1.129(0.128)	-0.729(0.131)	-0.392(0.129)	0.284(0.129)	1.004(0.140)
	B.1	1.139(0.132)	-0.729(0.134)	-0.395(0.131)	0.283(0.130)	1.003(0.142)
	B.2	1.155(0.135)	-0.719(0.135)	-0.390(0.133)	0.287(0.132)	1.000(0.143)
	C.1	1.173(0.135)	-0.683(0.140)	-0.357(0.137)	0.316(0.138)	1.001(0.147)
	C.2	1.179(0.136)	-0.685(0.137)	-0.360(0.134)	0.311(0.137)	0.996(0.147)
	D.1	1.178(0.136)	-0.678(0.141)	-0.353(0.137)	0.319(0.140)	1.005(0.149)
	D.2	1.180(0.137)	-0.679(0.139)	-0.354(0.137)	0.316(0.139)	1.003(0.149)
500	A.1	1.113(0.092)	-0.713(0.085)	-0.374(0.086)	0.298(0.081)	1.016(0.104)
	A.2	1.113(0.092)	-0.713(0.085)	-0.374(0.086)	0.298(0.081)	1.016(0.104)
	B.1	1.120(0.093)	-0.711(0.087)	-0.376(0.086)	0.299(0.082)	1.017(0.105)
	B.2	1.133(0.095)	-0.696(0.086)	-0.365(0.085)	0.308(0.083)	1.018(0.107)
	C.1	1.151(0.097)	-0.658(0.087)	-0.331(0.087)	0.334(0.087)	1.016(0.107)
	C.2	1.155(0.098)	-0.657(0.087)	-0.330(0.088)	0.334(0.089)	1.015(0.108)
	D.1	1.155(0.098)	-0.651(0.088)	-0.324(0.088)	0.340(0.088)	1.022(0.108)
	D.2	1.157(0.098)	-0.654(0.087)	-0.328(0.088)	0.336(0.089)	1.017(0.107)
1000	A.1	1.117(0.058)	-0.711(0.066)	-0.375(0.064)	0.299(0.062)	1.014(0.067)
	A.2	1.117(0.058)	-0.711(0.066)	-0.375(0.064)	0.299(0.062)	1.014(0.067)
	B.1	1.124(0.059)	-0.711(0.067)	-0.378(0.065)	0.298(0.063)	1.015(0.068)
	B.2	1.138(0.059)	-0.695(0.068)	-0.366(0.067)	0.308(0.064)	1.016(0.069)
	C.1	1.156(0.061)	-0.657(0.069)	-0.331(0.068)	0.336(0.065)	1.014(0.071)
	C.2	1.159(0.061)	-0.656(0.070)	-0.330(0.069)	0.336(0.065)	1.013(0.071)
	D.1	1.159(0.062)	-0.651(0.070)	-0.326(0.068)	0.340(0.065)	1.018(0.071)
	D.2	1.160(0.061)	-0.653(0.070)	-0.328(0.068)	0.338(0.065)	1.016(0.071)

Table 25: The estimates of the loading and thresholds for Item 6 (unequal)

N		λ_6	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.140	-0.829	-0.227	0.409	1.135
250	A.1	1.110(0.121)	-0.880(0.124)	-0.287(0.124)	0.364(0.121)	1.115(0.143)
	A.2	1.110(0.121)	-0.880(0.124)	-0.287(0.124)	0.364(0.121)	1.115(0.143)
	B.1	1.119(0.124)	-0.881(0.126)	-0.287(0.127)	0.365(0.123)	1.117(0.144)
	B.2	1.130(0.127)	-0.871(0.128)	-0.278(0.129)	0.374(0.124)	1.118(0.147)
	C.1	1.142(0.130)	-0.833(0.133)	-0.239(0.132)	0.404(0.126)	1.119(0.149)
	C.2	1.146(0.129)	-0.836(0.132)	-0.242(0.131)	0.401(0.126)	1.115(0.147)
	D.1	1.146(0.130)	-0.828(0.133)	-0.233(0.133)	0.408(0.128)	1.124(0.150)
	D.2	1.147(0.130)	-0.830(0.133)	-0.235(0.132)	0.407(0.127)	1.122(0.150)
500	A.1	1.107(0.086)	-0.886(0.091)	-0.284(0.092)	0.372(0.088)	1.136(0.092)
	A.2	1.107(0.086)	-0.886(0.091)	-0.284(0.092)	0.372(0.088)	1.136(0.092)
	B.1	1.114(0.086)	-0.886(0.091)	-0.283(0.092)	0.374(0.088)	1.139(0.093)
	B.2	1.125(0.087)	-0.871(0.092)	-0.269(0.093)	0.384(0.089)	1.145(0.094)
	C.1	1.137(0.090)	-0.833(0.092)	-0.230(0.093)	0.414(0.091)	1.143(0.099)
	C.2	1.139(0.090)	-0.832(0.093)	-0.229(0.095)	0.414(0.091)	1.143(0.098)
	D.1	1.140(0.091)	-0.827(0.092)	-0.223(0.094)	0.419(0.091)	1.150(0.100)
	D.2	1.141(0.091)	-0.829(0.093)	-0.226(0.095)	0.416(0.091)	1.146(0.098)
1000	A.1	1.114(0.053)	-0.899(0.063)	-0.288(0.059)	0.364(0.061)	1.128(0.070)
	A.2	1.114(0.053)	-0.899(0.063)	-0.288(0.059)	0.364(0.061)	1.128(0.070)
	B.1	1.117(0.056)	-0.893(0.062)	-0.285(0.058)	0.363(0.065)	1.130(0.073)
	B.2	1.131(0.059)	-0.879(0.064)	-0.277(0.059)	0.378(0.066)	1.138(0.075)
	C.1	1.140(0.058)	-0.822(0.065)	-0.233(0.059)	0.408(0.066)	1.137(0.076)
	C.2	1.140(0.059)	-0.824(0.064)	-0.230(0.058)	0.409(0.067)	1.137(0.076)
	D.1	1.140(0.059)	-0.825(0.066)	-0.229(0.059)	0.409(0.067)	1.133(0.076)
	D.2	1.141(0.059)	-0.824(0.064)	-0.229(0.058)	0.409(0.066)	1.132(0.076)

Table 26: The estimates of the loading and thresholds for Item 7 (unequal)

N		λ_7	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.100	-0.495	0.044	0.550	1.452
250	A.1	1.111(0.132)	-0.586(0.125)	-0.046(0.118)	0.489(0.134)	1.442(0.158)
	A.2	1.111(0.132)	-0.586(0.125)	-0.046(0.118)	0.489(0.134)	1.442(0.158)
	B.1	1.117(0.134)	-0.581(0.128)	-0.043(0.119)	0.491(0.136)	1.448(0.161)
	B.2	1.121(0.134)	-0.565(0.130)	-0.026(0.122)	0.502(0.139)	1.455(0.164)
	C.1	1.117(0.137)	-0.518(0.126)	0.022(0.121)	0.537(0.141)	1.464(0.168)
	C.2	1.120(0.138)	-0.519(0.128)	0.021(0.120)	0.534(0.141)	1.461(0.169)
	D.1	1.119(0.139)	-0.510(0.128)	0.028(0.121)	0.542(0.142)	1.471(0.171)
	D.2	1.119(0.138)	-0.512(0.128)	0.028(0.121)	0.541(0.142)	1.469(0.173)
500	A.1	1.101(0.086)	-0.572(0.077)	-0.028(0.075)	0.513(0.083)	1.446(0.099)
	A.2	1.101(0.086)	-0.572(0.077)	-0.028(0.075)	0.513(0.083)	1.446(0.099)
	B.1	1.105(0.086)	-0.568(0.077)	-0.024(0.075)	0.515(0.082)	1.452(0.100)
	B.2	1.108(0.088)	-0.546(0.079)	-0.003(0.077)	0.528(0.083)	1.465(0.103)
	C.1	1.106(0.090)	-0.500(0.079)	0.044(0.078)	0.562(0.085)	1.472(0.103)
	C.2	1.108(0.091)	-0.498(0.080)	0.045(0.079)	0.562(0.086)	1.473(0.105)
	D.1	1.107(0.092)	-0.492(0.079)	0.051(0.079)	0.567(0.086)	1.480(0.105)
	D.2	1.109(0.092)	-0.495(0.080)	0.049(0.079)	0.564(0.086)	1.478(0.106)
1000	A.1	1.098(0.059)	-0.562(0.061)	-0.030(0.059)	0.506(0.060)	1.426(0.073)
	A.2	1.098(0.059)	-0.562(0.061)	-0.030(0.059)	0.506(0.060)	1.426(0.073)
	B.1	1.096(0.061)	-0.558(0.063)	-0.027(0.058)	0.500(0.063)	1.437(0.075)
	B.2	1.100(0.061)	-0.543(0.064)	-0.006(0.059)	0.518(0.065)	1.440(0.076)
	C.1	1.097(0.062)	-0.494(0.063)	0.043(0.058)	0.550(0.067)	1.450(0.080)
	C.2	1.098(0.062)	-0.493(0.062)	0.044(0.059)	0.551(0.067)	1.450(0.079)
	D.1	1.098(0.062)	-0.493(0.064)	0.043(0.058)	0.552(0.067)	1.452(0.080)
	D.2	1.099(0.062)	-0.494(0.063)	0.045(0.058)	0.551(0.066)	1.453(0.079)

Table 27: The estimates of the loading and thresholds for Item 8 (unequal)

N		λ_8	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	1.020	-0.911	-0.409	0.164	0.921
250	A.1	0.979(0.112)	-0.952(0.121)	-0.453(0.108)	0.137(0.110)	0.917(0.128)
	A.2	0.979(0.112)	-0.952(0.121)	-0.453(0.108)	0.137(0.110)	0.917(0.128)
	B.1	0.989(0.116)	-0.954(0.124)	-0.456(0.112)	0.134(0.111)	0.917(0.130)
	B.2	1.002(0.118)	-0.948(0.126)	-0.450(0.114)	0.135(0.112)	0.916(0.133)
	C.1	1.020(0.123)	-0.918(0.126)	-0.421(0.113)	0.155(0.114)	0.915(0.135)
	C.2	1.024(0.123)	-0.922(0.128)	-0.424(0.115)	0.151(0.115)	0.911(0.136)
	D.1	1.024(0.124)	-0.915(0.128)	-0.417(0.115)	0.157(0.115)	0.919(0.136)
	D.2	1.025(0.123)	-0.917(0.127)	-0.418(0.115)	0.155(0.116)	0.917(0.137)
500	A.1	0.981(0.079)	-0.961(0.073)	-0.454(0.077)	0.138(0.075)	0.924(0.091)
	A.2	0.981(0.079)	-0.961(0.073)	-0.454(0.077)	0.138(0.075)	0.924(0.091)
	B.1	0.989(0.079)	-0.963(0.073)	-0.457(0.077)	0.136(0.075)	0.925(0.092)
	B.2	1.002(0.081)	-0.952(0.073)	-0.449(0.081)	0.141(0.078)	0.926(0.094)
	C.1	1.021(0.082)	-0.923(0.074)	-0.419(0.083)	0.161(0.078)	0.924(0.098)
	C.2	1.024(0.083)	-0.923(0.075)	-0.419(0.084)	0.160(0.080)	0.923(0.098)
	D.1	1.024(0.083)	-0.918(0.075)	-0.414(0.084)	0.164(0.079)	0.929(0.099)
	D.2	1.026(0.083)	-0.920(0.075)	-0.417(0.084)	0.161(0.079)	0.925(0.098)
1000	A.1	0.966(0.059)	-0.950(0.057)	-0.445(0.056)	0.149(0.055)	0.936(0.059)
	A.2	0.966(0.059)	-0.950(0.057)	-0.445(0.056)	0.149(0.055)	0.936(0.059)
	B.1	0.974(0.060)	-0.952(0.056)	-0.447(0.057)	0.146(0.056)	0.920(0.061)
	B.2	0.996(0.060)	-0.941(0.059)	-0.433(0.060)	0.152(0.056)	0.922(0.063)
	C.1	1.015(0.062)	-0.914(0.058)	-0.406(0.060)	0.164(0.061)	0.921(0.063)
	C.2	1.018(0.062)	-0.914(0.059)	-0.406(0.060)	0.163(0.060)	0.921(0.063)
	D.1	1.016(0.063)	-0.910(0.058)	-0.407(0.060)	0.166(0.060)	0.925(0.063)
	D.2	1.017(0.062)	-0.909(0.059)	-0.406(0.060)	0.164(0.059)	0.923(0.063)

Table 28: The estimates of the loading and thresholds for Item 9 (unequal)

N		λ_9	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.880	-0.591	-0.053	0.529	1.138
250	A.1	0.900(0.102)	-0.667(0.113)	-0.125(0.108)	0.476(0.108)	1.146(0.128)
	A.2	0.900(0.102)	-0.667(0.113)	-0.125(0.108)	0.476(0.108)	1.146(0.128)
	B.1	0.904(0.102)	-0.663(0.115)	-0.122(0.109)	0.479(0.109)	1.148(0.131)
	B.2	0.904(0.103)	-0.647(0.114)	-0.107(0.108)	0.491(0.109)	1.147(0.132)
	C.1	0.898(0.105)	-0.605(0.118)	-0.064(0.111)	0.526(0.114)	1.145(0.132)
	C.2	0.899(0.105)	-0.606(0.116)	-0.064(0.110)	0.525(0.114)	1.143(0.132)
	D.1	0.899(0.106)	-0.600(0.118)	-0.057(0.111)	0.530(0.114)	1.148(0.132)
	D.2	0.899(0.105)	-0.600(0.118)	-0.058(0.112)	0.530(0.115)	1.148(0.133)
500	A.1	0.880(0.076)	-0.660(0.083)	-0.115(0.078)	0.481(0.076)	1.139(0.086)
	A.2	0.880(0.076)	-0.660(0.083)	-0.115(0.078)	0.481(0.076)	1.139(0.086)
	B.1	0.883(0.077)	-0.656(0.084)	-0.112(0.078)	0.485(0.076)	1.142(0.087)
	B.2	0.884(0.078)	-0.638(0.086)	-0.094(0.081)	0.500(0.079)	1.143(0.089)
	C.1	0.880(0.078)	-0.597(0.086)	-0.052(0.080)	0.532(0.078)	1.139(0.088)
	C.2	0.881(0.079)	-0.596(0.087)	-0.051(0.081)	0.533(0.080)	1.140(0.089)
	D.1	0.880(0.079)	-0.591(0.086)	-0.046(0.081)	0.537(0.079)	1.142(0.089)
	D.2	0.881(0.079)	-0.593(0.086)	-0.048(0.081)	0.535(0.080)	1.142(0.090)
1000	A.1	0.886(0.053)	-0.665(0.057)	-0.121(0.058)	0.377(0.055)	1.130(0.062)
	A.2	0.886(0.053)	-0.665(0.057)	-0.121(0.058)	0.377(0.055)	1.130(0.062)
	B.1	0.886(0.053)	-0.660(0.056)	-0.118(0.057)	0.379(0.057)	1.137(0.065)
	B.2	0.888(0.055)	-0.637(0.057)	-0.008(0.059)	0.495(0.058)	1.136(0.064)
	C.1	0.880(0.057)	-0.594(0.058)	-0.055(0.059)	0.528(0.059)	1.136(0.067)
	C.2	0.880(0.057)	-0.594(0.058)	-0.054(0.058)	0.528(0.059)	1.136(0.066)
	D.1	0.881(0.057)	-0.591(0.057)	-0.052(0.058)	0.530(0.059)	1.136(0.067)
	D.2	0.881(0.056)	-0.591(0.056)	-0.053(0.058)	0.530(0.058)	1.135(0.066)

Table 29: The estimates of the loading and thresholds for Item 10 (unequal)

N		λ_{10}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.880	0.307	0.763	1.472	2.130
250	A.1	0.929(0.121)	0.188(0.110)	0.640(0.120)	1.321(0.141)	2.059(0.194)
	A.2	0.929(0.121)	0.188(0.110)	0.640(0.120)	1.321(0.141)	2.059(0.194)
	B.1	0.932(0.122)	0.201(0.116)	0.653(0.127)	1.342(0.148)	2.085(0.207)
	B.2	0.926(0.125)	0.230(0.115)	0.686(0.129)	1.387(0.154)	2.124(0.217)
	C.1	0.897(0.125)	0.288(0.112)	0.752(0.127)	1.461(0.155)	2.148(0.222)
	C.2	0.896(0.126)	0.289(0.111)	0.753(0.127)	1.462(0.156)	2.151(0.223)
	D.1	0.894(0.126)	0.298(0.115)	0.763(0.131)	1.472(0.160)	2.163(0.227)
	D.2	0.894(0.127)	0.297(0.112)	0.761(0.129)	1.470(0.157)	2.161(0.227)
500	A.1	0.922(0.084)	0.193(0.077)	0.644(0.084)	1.332(0.103)	2.054(0.123)
	A.2	0.922(0.084)	0.193(0.077)	0.644(0.084)	1.332(0.103)	2.054(0.123)
	B.1	0.925(0.085)	0.205(0.079)	0.656(0.086)	1.351(0.109)	2.078(0.135)
	B.2	0.917(0.086)	0.237(0.080)	0.693(0.088)	1.399(0.112)	2.113(0.135)
	C.1	0.889(0.087)	0.296(0.079)	0.760(0.090)	1.482(0.115)	2.146(0.141)
	C.2	0.887(0.088)	0.299(0.080)	0.764(0.090)	1.486(0.113)	2.150(0.139)
	D.1	0.886(0.088)	0.307(0.080)	0.772(0.091)	1.495(0.115)	2.160(0.142)
	D.2	0.885(0.088)	0.305(0.080)	0.769(0.090)	1.491(0.114)	2.156(0.141)
1000	A.1	0.920(0.050)	0.223(0.054)	0.640(0.055)	1.335(0.062)	2.003(0.086)
	A.2	0.920(0.050)	0.223(0.054)	0.640(0.055)	1.335(0.062)	2.003(0.086)
	B.1	0.909(0.055)	0.237(0.060)	0.660(0.057)	1.353(0.068)	2.055(0.092)
	B.2	0.904(0.054)	0.249(0.060)	0.701(0.058)	1.399(0.068)	2.098(0.094)
	C.1	0.884(0.055)	0.311(0.059)	0.764(0.057)	1.469(0.068)	2.136(0.094)
	C.2	0.883(0.055)	0.310(0.059)	0.765(0.057)	1.470(0.068)	2.133(0.094)
	D.1	0.880(0.056)	0.308(0.059)	0.766(0.057)	1.474(0.068)	2.134(0.093)
	D.2	0.879(0.055)	0.309(0.058)	0.765(0.057)	1.475(0.066)	2.134(0.092)

Table 30: The estimates of the loading and thresholds for Item 11 (unequal)

N		λ_{11}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.800	-0.360	0.088	0.584	1.136
250	A.1	0.820(0.100)	-0.436(0.110)	0.016(0.110)	0.548(0.103)	1.159(0.130)
	A.2	0.820(0.100)	-0.436(0.110)	0.016(0.110)	0.548(0.103)	1.159(0.130)
	B.1	0.822(0.101)	-0.430(0.111)	0.021(0.111)	0.552(0.106)	1.161(0.134)
	B.2	0.820(0.101)	-0.412(0.112)	0.038(0.112)	0.564(0.108)	1.160(0.137)
	C.1	0.810(0.101)	-0.369(0.112)	0.082(0.112)	0.595(0.108)	1.153(0.134)
	C.2	0.811(0.101)	-0.369(0.111)	0.081(0.112)	0.594(0.110)	1.151(0.135)
	D.1	0.810(0.101)	-0.362(0.113)	0.088(0.114)	0.599(0.109)	1.156(0.135)
	D.2	0.811(0.101)	-0.363(0.113)	0.087(0.113)	0.599(0.111)	1.155(0.136)
500	A.1	0.805(0.074)	-0.428(0.082)	0.025(0.080)	0.535(0.078)	1.138(0.084)
	A.2	0.805(0.074)	-0.428(0.082)	0.025(0.080)	0.535(0.078)	1.138(0.084)
	B.1	0.806(0.073)	-0.422(0.082)	0.030(0.080)	0.539(0.078)	1.139(0.085)
	B.2	0.804(0.074)	-0.401(0.081)	0.050(0.080)	0.554(0.079)	1.141(0.085)
	C.1	0.794(0.076)	-0.357(0.083)	0.095(0.082)	0.587(0.081)	1.137(0.086)
	C.2	0.794(0.076)	-0.355(0.084)	0.097(0.082)	0.588(0.081)	1.138(0.086)
	D.1	0.793(0.076)	-0.350(0.084)	0.102(0.083)	0.592(0.081)	1.142(0.087)
	D.2	0.793(0.076)	-0.351(0.084)	0.100(0.082)	0.590(0.080)	1.140(0.086)
1000	A.1	0.817(0.054)	-0.429(0.052)	0.019(0.053)	0.527(0.057)	1.150(0.060)
	A.2	0.817(0.054)	-0.425(0.052)	0.019(0.053)	0.527(0.057)	1.150(0.060)
	B.1	0.808(0.055)	-0.420(0.056)	0.036(0.056)	0.541(0.059)	1.144(0.064)
	B.2	0.810(0.055)	-0.396(0.056)	0.066(0.056)	0.560(0.059)	1.142(0.064)
	C.1	0.802(0.055)	-0.363(0.056)	0.090(0.056)	0.582(0.060)	1.137(0.065)
	C.2	0.801(0.055)	-0.360(0.056)	0.091(0.055)	0.583(0.059)	1.137(0.065)
	D.1	0.798(0.055)	-0.358(0.057)	0.093(0.056)	0.586(0.059)	1.138(0.066)
	D.2	0.797(0.055)	-0.359(0.057)	0.095(0.055)	0.585(0.059)	1.138(0.065)

Table 31: The estimates of the loading and thresholds for Item 12 (unequal)

N		λ_{12}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.800	-0.232	0.128	856	1.568
250	A.1	0.834(0.104)	-0.309(0.100)	0.055(0.108)	0.789(0.126)	1.556(0.151)
	A.2	0.834(0.104)	-0.309(0.100)	0.055(0.108)	0.789(0.126)	1.556(0.151)
	B.1	0.836(0.104)	-0.302(0.102)	0.060(0.110)	0.798(0.129)	1.566(0.156)
	B.2	0.831(0.105)	-0.281(0.100)	0.078(0.111)	0.819(0.130)	1.578(0.157)
	C.1	0.814(0.106)	-0.235(0.099)	0.124(0.112)	0.862(0.129)	1.584(0.155)
	C.2	0.814(0.107)	-0.234(0.099)	0.125(0.111)	0.862(0.129)	1.584(0.155)
	D.1	0.814(0.107)	-0.228(0.100)	0.131(0.112)	0.868(0.131)	1.591(0.157)
	D.2	0.813(0.107)	-0.228(0.100)	0.130(0.112)	0.868(0.131)	1.590(0.156)
500	A.1	0.824(0.073)	-0.308(0.077)	0.059(0.071)	0.784(0.078)	1.541(0.105)
	A.2	0.824(0.073)	-0.308(0.077)	0.059(0.071)	0.784(0.078)	1.541(0.105)
	B.1	0.825(0.074)	-0.302(0.078)	0.064(0.072)	0.791(0.080)	1.550(0.107)
	B.2	0.820(0.074)	-0.278(0.078)	0.086(0.073)	0.815(0.081)	1.563(0.107)
	C.1	0.803(0.075)	-0.231(0.078)	0.132(0.072)	0.860(0.084)	1.574(0.110)
	C.2	0.802(0.075)	-0.228(0.079)	0.134(0.072)	0.862(0.083)	1.575(0.111)
	D.1	0.801(0.075)	-0.224(0.079)	0.138(0.072)	0.866(0.084)	1.581(0.111)
	D.2	0.801(0.075)	-0.224(0.080)	0.138(0.073)	0.865(0.084)	1.579(0.111)
1000	A.1	0.829(0.050)	-0.308(0.050)	0.062(0.053)	0.790(0.056)	1.550(0.068)
	A.2	0.829(0.050)	-0.308(0.050)	0.062(0.053)	0.790(0.056)	1.550(0.068)
	B.1	0.818(0.053)	-0.300(0.055)	0.054(0.056)	0.782(0.059)	1.544(0.070)
	B.2	0.810(0.053)	-0.290(0.055)	0.091(0.057)	0.833(0.059)	1.559(0.070)
	C.1	0.805(0.054)	-0.236(0.053)	0.124(0.056)	0.859(0.061)	1.572(0.071)
	C.2	0.805(0.053)	-0.235(0.053)	0.126(0.056)	0.860(0.061)	1.570(0.071)
	D.1	0.799(0.054)	-0.231(0.053)	0.130(0.056)	0.858(0.061)	1.573(0.071)
	D.2	0.798(0.053)	-0.232(0.052)	0.129(0.056)	0.859(0.061)	1.572(0.070)

Table 32: The estimates of the loading and thresholds for Item 13 (unequal)

N		λ_{13}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.790	-0.143	0.238	0.810	1.374
250	A.1	0.822(0.100)	-0.224(0.110)	0.163(0.103)	0.749(0.107)	1.378(0.124)
	A.2	0.822(0.100)	-0.224(0.110)	0.163(0.103)	0.749(0.107)	1.378(0.124)
	B.1	0.824(0.102)	-0.216(0.111)	0.168(0.105)	0.755(0.110)	1.382(0.127)
	B.2	0.818(0.102)	-0.194(0.110)	0.189(0.106)	0.773(0.111)	1.384(0.125)
	C.1	0.800(0.102)	-0.147(0.111)	0.237(0.110)	0.813(0.111)	1.385(0.126)
	C.2	0.801(0.103)	-0.147(0.111)	0.237(0.110)	0.812(0.113)	1.382(0.126)
	D.1	0.799(0.103)	-0.140(0.111)	0.245(0.111)	0.819(0.113)	1.390(0.128)
	D.2	0.799(0.103)	-0.141(0.112)	0.244(0.111)	0.818(0.114)	1.387(0.127)
500	A.1	0.821(0.072)	-0.225(0.076)	0.157(0.075)	0.739(0.076)	1.375(0.091)
	A.2	0.821(0.072)	-0.225(0.076)	0.157(0.075)	0.739(0.076)	1.375(0.091)
	B.1	0.821(0.072)	-0.217(0.077)	0.164(0.076)	0.746(0.078)	1.379(0.092)
	B.2	0.816(0.072)	-0.192(0.078)	0.187(0.077)	0.767(0.080)	1.385(0.092)
	C.1	0.798(0.072)	-0.144(0.079)	0.236(0.078)	0.810(0.080)	1.388(0.092)
	C.2	0.797(0.072)	-0.141(0.079)	0.238(0.077)	0.811(0.081)	1.388(0.092)
	D.1	0.795(0.072)	-0.136(0.080)	0.243(0.079)	0.815(0.080)	1.392(0.091)
	D.2	0.796(0.072)	-0.137(0.078)	0.242(0.077)	0.814(0.081)	1.391(0.092)
1000	A.1	0.816(0.046)	-0.230(0.050)	0.166(0.053)	0.741(0.057)	1.359(0.062)
	A.2	0.816(0.046)	-0.230(0.050)	0.166(0.053)	0.741(0.057)	1.359(0.062)
	B.1	0.810(0.052)	-0.226(0.054)	0.171(0.056)	0.779(0.060)	1.363(0.067)
	B.2	0.804(0.052)	-0.198(0.054)	0.191(0.056)	0.782(0.060)	1.380(0.067)
	C.1	0.793(0.052)	-0.144(0.054)	0.239(0.055)	0.805(0.060)	1.379(0.068)
	C.2	0.792(0.051)	-0.144(0.054)	0.238(0.055)	0.804(0.059)	1.374(0.068)
	D.1	0.793(0.052)	-0.143(0.054)	0.240(0.055)	0.805(0.060)	1.373(0.068)
	D.2	0.793(0.051)	-0.143(0.053)	0.239(0.054)	0.805(0.059)	1.373(0.067)

Table 33: The estimates of the loading and thresholds for Item 14 (unequal)

N		λ_{14}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.790	-0.087	0.397	0.866	1.517
250	A.1	0.839(0.114)	-0.011(0.116)	0.296(0.110)	0.796(0.113)	1.510(0.141)
	A.2	0.839(0.114)	-0.011(0.116)	0.296(0.110)	0.796(0.113)	1.510(0.141)
	B.1	0.840(0.115)	-0.001(0.121)	0.304(0.113)	0.803(0.115)	1.516(0.143)
	B.2	0.833(0.115)	0.024(0.119)	0.326(0.112)	0.818(0.113)	1.520(0.141)
	C.1	0.811(0.116)	0.075(0.117)	0.378(0.113)	0.862(0.110)	1.525(0.142)
	C.2	0.810(0.116)	0.075(0.117)	0.379(0.114)	0.860(0.111)	1.522(0.142)
	D.1	0.809(0.116)	0.083(0.117)	0.386(0.114)	0.868(0.112)	1.530(0.144)
	D.2	0.808(0.117)	0.082(0.118)	0.385(0.114)	0.865(0.111)	1.527(0.143)
500	A.1	0.828(0.077)	-0.001(0.078)	0.313(0.075)	0.805(0.081)	1.509(0.091)
	A.2	0.828(0.077)	-0.001(0.078)	0.313(0.075)	0.805(0.081)	1.509(0.091)
	B.1	0.828(0.078)	0.008(0.080)	0.321(0.077)	0.812(0.084)	1.516(0.093)
	B.2	0.820(0.077)	0.036(0.080)	0.346(0.078)	0.832(0.085)	1.524(0.095)
	C.1	0.799(0.079)	0.088(0.081)	0.397(0.077)	0.874(0.085)	1.527(0.095)
	C.2	0.798(0.078)	0.090(0.081)	0.400(0.078)	0.876(0.086)	1.529(0.095)
	D.1	0.796(0.080)	0.097(0.082)	0.405(0.078)	0.881(0.086)	1.533(0.097)
	D.2	0.797(0.078)	0.095(0.081)	0.404(0.079)	0.879(0.086)	1.532(0.096)
1000	A.1	0.817(0.055)	-0.008(0.053)	0.318(0.055)	0.803(0.058)	1.500(0.072)
	A.2	0.817(0.055)	-0.008(0.053)	0.318(0.055)	0.803(0.058)	1.500(0.072)
	B.1	0.810(0.058)	0.010(0.057)	0.324(0.058)	0.810(0.060)	1.512(0.076)
	B.2	0.806(0.058)	0.048(0.057)	0.347(0.058)	0.841(0.060)	1.519(0.076)
	C.1	0.792(0.058)	0.089(0.058)	0.399(0.059)	0.870(0.060)	1.520(0.076)
	C.2	0.792(0.058)	0.087(0.057)	0.397(0.058)	0.869(0.060)	1.519(0.075)
	D.1	0.793(0.058)	0.088(0.057)	0.400(0.058)	0.869(0.060)	1.520(0.076)
	D.2	0.792(0.057)	0.088(0.057)	0.401(0.058)	0.868(0.060)	1.519(0.075)

Table 34: The estimates of the loading and thresholds for Item 15 (unequal)

N		λ_{15}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.790	0.079	0.410	0.788	1.300
250	A.1	0.828(0.108)	-0.024(0.113)	0.320(0.108)	0.727(0.109)	1.315(0.133)
	A.2	0.828(0.108)	-0.024(0.113)	0.320(0.108)	0.727(0.109)	1.315(0.133)
	B.1	0.828(0.109)	-0.013(0.118)	0.328(0.112)	0.732(0.111)	1.317(0.135)
	B.2	0.821(0.108)	0.010(0.118)	0.351(0.113)	0.750(0.113)	1.319(0.137)
	C.1	0.800(0.109)	0.062(0.114)	0.404(0.113)	0.792(0.117)	1.317(0.139)
	C.2	0.799(0.110)	0.062(0.114)	0.404(0.113)	0.791(0.117)	1.315(0.140)
	D.1	0.798(0.110)	0.070(0.116)	0.412(0.114)	0.797(0.118)	1.322(0.141)
	D.2	0.798(0.111)	0.069(0.115)	0.411(0.114)	0.796(0.118)	1.320(0.142)
500	A.1	0.824(0.075)	-0.014(0.072)	0.323(0.074)	0.724(0.078)	1.302(0.090)
	A.2	0.824(0.075)	-0.014(0.072)	0.323(0.074)	0.724(0.078)	1.302(0.090)
	B.1	0.824(0.075)	-0.005(0.073)	0.331(0.075)	0.730(0.080)	1.305(0.092)
	B.2	0.816(0.075)	0.024(0.075)	0.359(0.076)	0.749(0.081)	1.307(0.093)
	C.1	0.796(0.076)	0.075(0.073)	0.411(0.075)	0.789(0.081)	1.302(0.092)
	C.2	0.795(0.077)	0.078(0.074)	0.413(0.076)	0.790(0.082)	1.302(0.093)
	D.1	0.793(0.077)	0.083(0.075)	0.419(0.077)	0.795(0.082)	1.304(0.093)
	D.2	0.794(0.077)	0.083(0.074)	0.418(0.076)	0.792(0.081)	1.304(0.093)
1000	A.1	0.820(0.060)	-0.024(0.055)	0.323(0.057)	0.730(0.060)	1.306(0.066)
	A.2	0.820(0.060)	-0.024(0.055)	0.323(0.057)	0.730(0.060)	1.306(0.066)
	B.1	0.822(0.061)	-0.009(0.058)	0.331(0.060)	0.733(0.062)	1.304(0.069)
	B.2	0.817(0.060)	0.024(0.058)	0.364(0.060)	0.755(0.062)	1.306(0.069)
	C.1	0.791(0.060)	0.073(0.059)	0.408(0.060)	0.788(0.062)	1.302(0.069)
	C.2	0.792(0.061)	0.074(0.058)	0.409(0.059)	0.789(0.063)	1.301(0.069)
	D.1	0.792(0.060)	0.077(0.059)	0.410(0.060)	0.790(0.062)	1.299(0.069)
	D.2	0.793(0.060)	0.077(0.058)	0.412(0.059)	0.791(0.062)	1.298(0.068)

Table 35: The estimates of the loading and thresholds for Item 16 (unequal)

N		λ_{16}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.780	-0.349	0.023	0.613	1.350
250	A.1	0.796(0.099)	-0.434(0.107)	-0.052(0.098)	0.550(0.097)	1.343(0.126)
	A.2	0.796(0.099)	-0.434(0.107)	-0.052(0.098)	0.550(0.097)	1.343(0.126)
	B.1	0.799(0.100)	-0.429(0.108)	-0.050(0.098)	0.554(0.098)	1.349(0.129)
	B.2	0.796(0.100)	-0.410(0.108)	-0.034(0.099)	0.566(0.098)	1.354(0.130)
	C.1	0.784(0.100)	-0.367(0.109)	0.008(0.100)	0.602(0.098)	1.358(0.132)
	C.2	0.785(0.102)	-0.367(0.111)	0.009(0.102)	0.600(0.098)	1.357(0.131)
	D.1	0.784(0.101)	-0.361(0.111)	0.014(0.102)	0.606(0.100)	1.364(0.135)
	D.2	0.784(0.102)	-0.361(0.112)	0.014(0.103)	0.605(0.100)	1.362(0.134)
500	A.1	0.802(0.075)	-0.420(0.076)	-0.044(0.071)	0.559(0.075)	1.344(0.100)
	A.2	0.802(0.075)	-0.420(0.076)	-0.044(0.071)	0.559(0.075)	1.344(0.100)
	B.1	0.804(0.076)	-0.414(0.077)	-0.040(0.072)	0.563(0.076)	1.349(0.102)
	B.2	0.801(0.076)	-0.392(0.077)	-0.020(0.071)	0.578(0.075)	1.355(0.101)
	C.1	0.790(0.077)	-0.349(0.076)	0.021(0.072)	0.613(0.078)	1.359(0.104)
	C.2	0.790(0.077)	-0.347(0.077)	0.024(0.072)	0.614(0.078)	1.359(0.103)
	D.1	0.790(0.077)	-0.343(0.076)	0.028(0.072)	0.618(0.080)	1.365(0.105)
	D.2	0.790(0.077)	-0.343(0.077)	0.027(0.072)	0.616(0.079)	1.362(0.103)
1000	A.1	0.796(0.053)	-0.430(0.055)	-0.044(0.056)	0.549(0.059)	1.338(0.070)
	A.2	0.796(0.053)	-0.430(0.055)	-0.044(0.056)	0.549(0.059)	1.338(0.070)
	B.1	0.791(0.054)	-0.427(0.056)	-0.038(0.057)	0.565(0.060)	1.345(0.072)
	B.2	0.790(0.054)	-0.406(0.056)	-0.029(0.057)	0.581(0.060)	1.355(0.072)
	C.1	0.781(0.055)	-0.353(0.056)	0.020(0.058)	0.612(0.061)	1.353(0.073)
	C.2	0.780(0.055)	-0.352(0.056)	0.021(0.058)	0.613(0.061)	1.352(0.072)
	D.1	0.779(0.055)	-0.353(0.056)	0.024(0.057)	0.614(0.061)	1.357(0.072)
	D.2	0.779(0.054)	-0.352(0.056)	0.022(0.057)	0.613(0.060)	1.358(0.072)

Table 36: The estimates of the loading and thresholds for Item 17 (unequal)

N		λ_{17}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.730	-0.518	-0.212	0.241	0.861
250	A.1	0.737(0.095)	-0.587(0.114)	-0.267(0.112)	0.206(0.098)	0.867(0.112)
	A.2	0.737(0.095)	-0.587(0.114)	-0.267(0.112)	0.206(0.098)	0.867(0.112)
	B.1	0.741(0.095)	-0.583(0.116)	-0.266(0.114)	0.205(0.100)	0.865(0.113)
	B.2	0.743(0.096)	-0.568(0.116)	-0.255(0.113)	0.210(0.101)	0.861(0.116)
	C.1	0.744(0.097)	-0.532(0.116)	-0.221(0.113)	0.231(0.101)	0.854(0.117)
	C.2	0.745(0.099)	-0.532(0.117)	-0.221(0.113)	0.231(0.100)	0.852(0.117)
	D.1	0.745(0.098)	-0.527(0.117)	-0.217(0.115)	0.233(0.101)	0.857(0.119)
	D.2	0.744(0.099)	-0.527(0.118)	-0.217(0.115)	0.235(0.102)	0.856(0.118)
500	A.1	0.725(0.064)	-0.584(0.069)	-0.265(0.066)	0.208(0.069)	0.873(0.077)
	A.2	0.725(0.064)	-0.584(0.069)	-0.265(0.066)	0.208(0.069)	0.873(0.077)
	B.1	0.728(0.064)	-0.580(0.069)	-0.264(0.067)	0.207(0.069)	0.872(0.078)
	B.2	0.730(0.064)	-0.564(0.070)	-0.25(0.068)	0.215(0.071)	0.870(0.079)
	C.1	0.731(0.066)	-0.527(0.070)	-0.215(0.070)	0.237(0.072)	0.862(0.080)
	C.2	0.731(0.066)	-0.526(0.070)	-0.214(0.070)	0.238(0.072)	0.861(0.081)
	D.1	0.731(0.066)	-0.522(0.071)	-0.210(0.071)	0.241(0.072)	0.864(0.081)
	D.2	0.731(0.067)	-0.523(0.069)	-0.212(0.069)	0.239(0.072)	0.863(0.080)
1000	A.1	0.727(0.044)	-0.580(0.050)	-0.263(0.053)	0.222(0.056)	0.872(0.059)
	A.2	0.727(0.044)	-0.580(0.050)	-0.263(0.053)	0.222(0.056)	0.872(0.059)
	B.1	0.729(0.046)	-0.581(0.052)	-0.258(0.056)	0.227(0.058)	0.870(0.060)
	B.2	0.729(0.046)	-0.559(0.052)	-0.248(0.056)	0.231(0.058)	0.871(0.060)
	C.1	0.731(0.047)	-0.520(0.054)	-0.211(0.057)	0.244(0.059)	0.864(0.061)
	C.2	0.731(0.047)	-0.521(0.054)	-0.211(0.056)	0.245(0.059)	0.865(0.061)
	D.1	0.731(0.046)	-0.519(0.054)	-0.214(0.057)	0.242(0.058)	0.862(0.061)
	D.2	0.730(0.046)	-0.519(0.053)	-0.213(0.056)	0.240(0.058)	0.866(0.061)

Table 37: The estimates of the loading and thresholds for Item 18 (unequal)

N		λ_{18}	τ_1	τ_2	τ_3	τ_4
Sample Size	Model	0.730	0.605	1.021	1.583	2.086
250	A.1	0.796(0.111)	0.483(0.111)	0.883(0.112)	1.418(0.141)	2.036(0.174)
	A.2	0.796(0.111)	0.483(0.111)	0.883(0.112)	1.418(0.141)	2.036(0.174)
	B.1	0.799(0.113)	0.500(0.116)	0.901(0.118)	1.442(0.150)	2.062(0.187)
	B.2	0.787(0.114)	0.535(0.117)	0.944(0.119)	1.492(0.146)	2.094(0.192)
	C.1	0.753(0.115)	0.600(0.117)	1.024(0.123)	1.594(0.160)	2.125(0.198)
	C.2	0.750(0.115)	0.603(0.116)	1.029(0.122)	1.597(0.150)	2.126(0.194)
	D.1	0.748(0.116)	0.611(0.118)	1.037(0.124)	1.607(0.163)	2.137(0.201)
	D.2	0.742(0.116)	0.611(0.117)	1.037(0.124)	1.604(0.155)	2.133(0.200)
500	A.1	0.790(0.081)	0.486(0.077)	0.887(0.078)	1.425(0.095)	2.011(0.131)
	A.2	0.790(0.081)	0.486(0.077)	0.887(0.078)	1.425(0.095)	2.011(0.131)
	B.1	0.791(0.082)	0.501(0.079)	0.904(0.081)	1.447(0.100)	2.031(0.134)
	B.2	0.781(0.083)	0.540(0.079)	0.950(0.081)	1.502(0.099)	2.070(0.136)
	C.1	0.745(0.084)	0.606(0.081)	1.030(0.084)	1.599(0.105)	2.100(0.141)
	C.2	0.743(0.084)	0.610(0.080)	1.036(0.083)	1.604(0.104)	2.105(0.139)
	D.1	0.740(0.085)	0.618(0.081)	1.043(0.085)	1.613(0.107)	2.115(0.142)
	D.2	0.732(0.083)	0.615(0.082)	1.041(0.086)	1.608(0.111)	2.107(0.148)
1000	A.1	0.774(0.052)	0.477(0.057)	0.881(0.058)	1.403(0.066)	2.003(0.085)
	A.2	0.774(0.052)	0.477(0.057)	0.881(0.058)	1.403(0.066)	2.003(0.085)
	B.1	0.770(0.055)	0.490(0.059)	0.893(0.060)	1.434(0.073)	2.022(0.089)
	B.2	0.764(0.056)	0.548(0.059)	0.940(0.060)	1.489(0.073)	2.058(0.089)
	C.1	0.733(0.057)	0.607(0.059)	1.017(0.063)	1.584(0.074)	2.088(0.091)
	C.2	0.732(0.056)	0.606(0.059)	1.020(0.062)	1.584(0.074)	2.089(0.090)
	D.1	0.731(0.057)	0.607(0.059)	1.023(0.062)	1.582(0.073)	2.087(0.091)
	D.2	0.731(0.056)	0.608(0.059)	1.022(0.062)	1.582(0.072)	2.088(0.090)

3.3 Discussion

From the simulstion results, there are some conclusions to discuss. First in the equal case, it is a good news that the proposed Models D.1 and D.2 have close estimates as the true Models B.1 and B.2. That is, the mixture response model with a bogus item can provide nearly the same estimates as the model without time, and moreover it has more smaller standard errors than the model without time. Although Models C.1 and C.2 provide the similiar results as Model D. 1 and D.2, these results can not show that we have enough evidence to recommend Models D.1 and D.2. We suspect the uniformly distributed response of the careless leads to the results. Therefore we designed the unequal case which assigns the careless 1 and careless 2 different response distributions to highlight the advantage of Models D.1 and D.2.

Second in the unequal case, flexible models perform better in estimation. As expected, D.1 and D.2 have much more unbiased estimates than B.1 and B.2 because the fittings of the careless response distributions for D.1 and D.2 are unrestricted. The consideration of the joint probability of response and response time has an advantage. From the comparison of B.1 and B.2, if the fit of response distribution fails, the response time can help improving the estimation. However in this design, we still fail to prove D.1 and D.2 are better than C.1 and C.2. In fact, C.1 and C.2 present similar results as well as D.1 and D.2. The consideration of the different careless response distributions to the careless 1 and careless 2 has no benefit to estimation in such situation, but it has no damage either.

To sum up, we currently can not conclude which model of C.1, C.2, D.1 and D.2 is the best because further investigation is needed. It seems that we divide the general condition about the careless response distributions into the equal and unequal case, but denoting that "both discrete uniformly distributed" and "neither discrete uniformly distributed" will be more precise. The concepts of "the same distributions" and "both discrete uniform distributions" are confounded. Fortunately, in the unequal case, the careless response distributions are different and neither discrete uniform. That is, based on this design, all the four models (C.1, C.2, D.1 and D.2) were set on the same baseline and they provided similiar estimates. The keys of the contest are the setting of each careless response distribution and the size of each careless group. In a word, we suspect our setting of the careless reponse distributions and size of the careless lead to the similiar results from the four models.

4 Data Analysis

We apply our proposed model to the real data and examine some model fit indices in this chapter. We collect and analyze data of the questionnaire answered by a group of junior high school students. The questionnaire was designed by Cheng, Hung, and Huang (2013), and the purpose was to investigate the multiple teaching stragies arousing the junior high school students' environmental interests. The content of the items in this questionnaire were revised from Linnenbrink-Garcia's scale of Measuring Situational Interest in Academic Domains (2010) and Mitchell's scale of Person Interest (1993). The validity of the questionnaire design has been verified by George and Mallery (2003). A sample of the quetionnaire is included in the Appendix. There were 20 items with 5 options (Extremly agree, Slightly agree, Neither agree nor disagree, Slightly disagree and Extremly disagree). We added a bogus item to the quetionnaire as the 15^{th} item (Figure 5). We had 575 junior high students to answer the quetionnaire with 21 items in total. We then use Akaike information criterion (AIC), Bayesian information criterion (BIC), and Sample-Size Adjusted BIC to compare eight fitted models and present the results in Section 4.2.

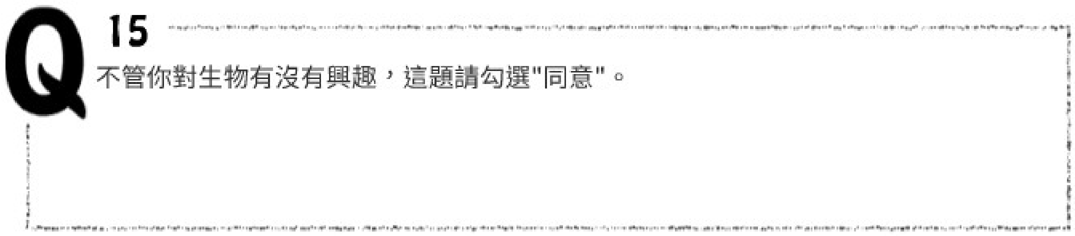


Figure 5: The added bogus item in the questionnaire.

4.1 Data

The data were collected from 575 junior high school students through administering the survey on the computer. The response time was recorded for each item, so that each student had his or her own response time. The quetionnaire is used to investigate if the teaching stragies arouse the students' interests to biology. The students were anonymous and unobliged. The resulting sample size used for the analysis was 569 because some students did not finish or they had missing response time for some items. Note that there was 22 percent of the students who answered the bogus item incorrectly. In addition, because the students rarely endorsed the last two options (Slightly disagree and Extremly disagree), we combined the last two options to avoid sparseness that could cause estimation problems

4.2 Results

The AICs, BICs and adjusted BICs of the eight models are summarized in Table 38. Model D.1 and D.2 have the smallest values among the models in each condition (without and with time). It suggested that the unrestricted models (D. 1 and D.2) are better than the restricted models (B.1 & C.1 and B.2 & C.2). The item estimates of D.1 and D.2 are reported in Table 38.

Table 39 displays the item estimates of the attentive. Comparing D.1 with D.2, the estimates of the loadings and thresholds are close. The standard errors of D.1 are slightly smaller than D.2. Table 40 reports the careless respondents' response distributions. The response distribution of Careless 1 is $(0.55, 0.13, 0.26, 0.06)$ after calculating from the estimated thresholds $(0.114, 0.465, 1.573)$. The response distribution of Careless 2 is $(0.01, 0.99, 0, 0)$ estimated from D.1; from D.2, the response distribution is $(0.04, 0.96, 0, 0)$. It is clear that the response pattern of

Careless 2 is long string in this case. Thus, we find that the proposed model can also identify the long string when it occurs in the extreme condition (i.e., some respondents frequently endose the same options). Table 42 presents the estimates of the response time of the three classes and we draw their distributions in Figure 6. In fact, we expected the reponse time of Careless 2 should be longer than Careless 1. Contrarily, we obtained that the mean of the response time of Careless 2 is substantially smaller than Careless 1. We might reconsider to place the answer of the bogus item in the extreme options in the future. The size of the three classes (Careless 1, Careless 2 and Attentive, respectively) is estimated $(0.22, 0.14, 0.64)$ from both D.1 and D.2.

Table 38: The model fit indices of each model setting

	Without Time				With Time		
	AIC	BIC	Adjusted BIC		AIC	BIC	Adjusted BIC
A.1	23839.173	24190.885	23933.747	A.2	57463.207	57923.658	57587.156
B.1	23585.607	23941.661	23681.349	B.2	55607.119	55989.226	55709.866
C.1	21142.314	21511.394	21241.558	C.2	52117.985	52513.118	52224.235
D.1	20263.529	20645.636	20366.276	D.2	49888.001	50296.160	49997.754

Table 39: The item estimates for Model D.1 and Model D.2

	Model D.1				Model D.2			
j	λ_j	τ_1	τ_2	τ_3	λ_j	τ_1	τ_2	τ_3
1	1.510(0.153)	-0.998(0.127)	0.747(0.120)	2.842(0.205)	1.509(0.154)	-1.025(0.133)	0.748(0.126)	2.835(0.209)
2	1.563(0.149)	-1.055(0.128)	0.621(0.120)	2.487(0.190)	1.558(0.149)	-1.067(0.133)	0.613(0.129)	2.477(0.195)
3	1.284(0.130)	-0.798(0.111)	0.574(0.107)	2.521(0.200)	1.285(0.131)	-0.823(0.114)	0.579(0.116)	2.518(0.203)
4	1.549(0.138)	-1.293(0.130)	0.084(0.116)	2.089(0.157)	1.539(0.138)	-1.324(0.135)	0.092(0.127)	2.079(0.163)
5	1.064(0.116)	-0.674(0.099)	0.642(0.100)	2.185(0.177)	1.066(0.118)	-0.695(0.102)	0.645(0.106)	2.184(0.181)
6	1.371(0.152)	-0.600(0.112)	1.116(0.129)	2.842(0.241)	1.370(0.154)	-0.608(0.115)	1.105(0.136)	2.836(0.245)
7	1.290(0.132)	-0.700(0.108)	0.734(0.110)	2.668(0.206)	1.291(0.133)	-0.702(0.111)	0.737(0.118)	2.665(0.210)
8	0.994(0.104)	-1.335(0.109)	-0.207(0.092)	1.507(0.124)	0.982(0.103)	-1.370(0.113)	-0.221(0.099)	1.499(0.127)
9	1.317(0.155)	-0.316(0.105)	0.951(0.118)	2.580(0.219)	1.326(0.159)	-0.326(0.109)	0.958(0.127)	2.584(0.226)
10	1.475(0.170)	0.680(0.118)	0.467(0.114)	2.643(0.233)	1.476(0.173)	0.690(0.122)	0.471(0.125)	2.639(0.239)
11	1.576(0.167)	-0.836(0.125)	0.474(0.120)	2.595(0.226)	1.578(0.170)	-0.847(0.131)	0.480(0.132)	2.592(0.233)
12	1.860(0.188)	-0.882(0.142)	0.507(0.132)	3.083(0.260)	1.876(0.192)	-0.890(0.153)	0.519(0.147)	3.094(0.269)
13	2.237(0.240)	-1.218(0.174)	0.697(0.158)	3.569(0.299)	2.247(0.244)	-1.244(0.189)	0.707(0.174)	3.572(0.309)
14	1.701(0.171)	-1.226(0.139)	0.443(0.127)	2.700(0.228)	1.701(0.172)	-1.238(0.149)	0.451(0.138)	2.695(0.233)
16	1.637(0.160)	-1.180(0.139)	0.641(0.126)	3.124(0.257)	1.659(0.164)	-1.204(0.146)	0.654(0.138)	3.144(0.267)
17	1.565(0.159)	-1.015(0.128)	0.581(0.122)	3.025(0.291)	1.591(0.165)	-1.044(0.133)	0.595(0.135)	3.047(0.303)
18	1.441(0.147)	-0.867(0.118)	0.904(0.125)	2.871(0.239)	1.465(0.152)	-0.920(0.123)	0.915(0.135)	2.892(0.248)
19	1.295(0.153)	-0.750(0.113)	0.787(0.115)	2.614(0.232)	1.307(0.156)	-0.802(0.119)	0.794(0.124)	2.623(0.238)
20	1.168(0.121)	-0.778(0.104)	0.751(0.107)	2.368(0.181)	1.180(0.123)	-0.813(0.107)	0.758(0.114)	2.375(0.186)
21	1.337(0.138)	-0.705(0.110)	0.856(0.118)	2.643(0.220)	1.353(0.141)	-0.740(0.117)	0.865(0.126)	2.655(0.227)

Table 40: The thresholds of the careless responses estimated from D.1 and D.2

	D.1				D.2		
	τ_1	τ_2	τ_3		τ_1	τ_2	τ_3
Careless 1	0.114	0.465	1.573	Careless 1	0.114	0.465	1.573
Careless 2	-2.500	2.527	3.234	Careless 2	-1.771	2.767	3.235

Table 41: The response time distribution estimated from D.2 (logarithm scale)

	mean	standard error
Attentive	0.914	0.856
Careless 1	0.046	1.255
Careless 2	-0.696	0.916

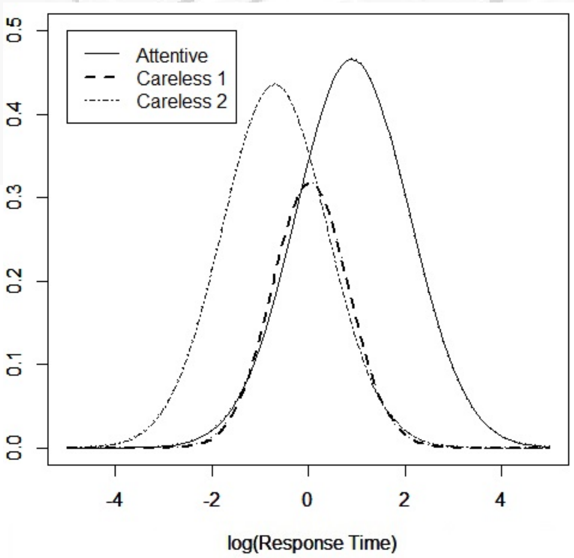


Figure 6: The response time of logarithm scale distribution

5 Conclusion and Future Direction

Learning from the achievement test and the survey, we attempt to converse the advantage of the covariates and the indicators. Hence, we propose the mixture response time model with a bogus item to estimate item parameters of a questionnaire. The flexible model did not perform well in the simulation but outperformed the other models in the real data analysis. There are some critical components in the difference between the generation and the data.

First, the careless response distributions are extremely different in real data. We consider they may have similiar response distribution and then assign the response distributions for Careless 1 and Careless 2 in simulation. However, the response pattern of Careless 2 in reality is an extreme long string. The character of these respondents in experience were regarded as without reading with high possibility. The explanation is these respondents customarily endorse the nonnegative option, and thus our bogus item position failed to identify them. This result has been discussed in Section 4.2. Here we focus on the response distributions. How sensitive of the model to distinguish between the two differnt response distributions can be a direction for future research.

Second, the response time distributions are unexpected but match the experience in real data. We have expected Careless 2 to have a larger mean of response time because they answered correctly to the bogus item. While Figure 6 presents the distributions, the character of Careless 2 becomes clear. The mean of response time of Careless 2 is smaller than Careless 1, and response are long string. Therefore, the character of Careless 2 is reacting without reading and finishing as fast as they can, and fortunately they endorse the correct answer to the bogus item. Based on this case, the benefit of using of response and response time joint probability is strengthened. In addition, the model not only provides the item estimates but also analyzes the character of the latent classes. This is another direction to define the performance of the reponse time distribution given the bogus item.

Third, the sizes of Careless 1 and Careless 2 in real data are larger than in simulation, especially Careless 2. From real data analysis, Careless 2 exists indeed. To treat Careless 1 and Careless 2 responses as uniformly distributed is challenged. Since Careless 1 and Careless 2 respondents have reacted differently on the bogus item, it is inferior to consider their response distribution being the same. However, how sensitive is the benefit of such a model to the size of the careless is yet another direction for future research.

In the end, we conclude the mixture response time model with a bogus item (Model D.2) can provide valid item estimates of a questionnaire. In addition, it can analyze the extreme long string and inspect the response time distributions of each latent class.

Reference

Beach, D. A. (1989). Identifying the random responder. *Journal of Psychology: Interdisciplinary and Applied, 123*, 101-103.

Chan, S. C., Lu, T. S. & Tsai, R. C. (2014). Incorporating Response Time to Analyze Test Data with Mixture Structural Equation Modeling. *Psychological Testing, 61(4)*, 463-488.

Cheng, J. C., Hung, J. F., & Huang, T. C. (2013). Promoting junior high students' situational interests with multiple teaching strategies in informal nanometer-related curricula. *Journal of Educational Practice and Research, 26(2)*, 1-28.

Credé, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement, 70(4)*, 596-612.

Ferrando, P. J. & Lorenzo-Seva, U. (2007). An item response theory model for incorporating response time data in binary personality items. *Applied Psychological Measurement, 31(6)*, 525-543.

Forero, C. G. & Maydeu-Olivares A. (2009). Estimation of IRT graded response models: limited versus full information methods. *Psychological Methods, 14(3)*, 275-299.

Fraley, R. C., Waller, N. G. & Brennan, K. A. (2000). Item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology 78(2)*, 350-365.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston, MA: Allyn & Bacon.

Hargittai, E. (2008). An update on survey measures of web-oriented digital literacy. *Social Science Computer Review 27(1)*, 130-137.

Holden, R. R., Kroner, D. G., Fekken, G. C. & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social*

Psychology 63(2), 272-279.

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality 39*, 103–129.

Meade, A. W. & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17(3)*, 437-455.

Meyer, J. P. (2010). A mixture Rasch model with item response time components, *Applied Psychological Measurement, 34(7)*, 521-538.

Muthén, L. K. & Muthén, B. O. (1998-2010). *Mplus User's Guide.* Los Angeles: Muthén & Muthén.

Koch, W. R. (1983). Likert scaling using the graded response latent trait model. *Applied Psychological Measurement, 7(1)*, 15-32.

Linnenbrink-Garcia, L., Durik, A. M., Conley, A. M., Barron, K. E., Tauer, J. M., Karabenick, S. A., & Harackiewicz, J. M. (2010). Measuring situational interest in academic domains. *Educational and Psychological Measurement, 70(4)*, 647-671.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85*, 424-436.

Samejima, F. (1973). Homogeneous case of the continuous response model. *Psychometrika, 38*, 203-219.

Schmitt, N. & Stults, D. M. (1985). Factors defined by negatively keyed items: the result of careless respondents. *Applied Psychological Measurement, 9(4)*, 367-373.

Wise, S. L. & DeMars, C. E. (2006). An application of item response time: the effort-moderated IRT model. *Journal of Educational Measurement 43(1)*, 19-38.

Woods, C. M. (2006). Careless responding to reverse-worded items: implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment, 28(3)*, 189-194.

Appendix

生物科學習興趣量表

		非常同意	同意	普通	不同意	非常不同意
1	我享受生物科主題。					
2	我喜歡生物科。					
3	我喜歡從事生物科活動。					
4	學習生物科對我而言是令人興奮的。					
5	對我來說生物科是有用的。					
6	生物科在日常生活中對我來說是有幫助的。					
7	對我來說能運用或理解生物科是重要的。					
8	科學性的思考是我擁有的特質。					
9	生物科課程老師上課是有趣的。					
10	生物科課程老師能引起我上課的注意力。					
11	生物科課程經常是好玩的。					
12	生物科課程生動有趣能吸引我的注意力。					
13	我在生物科課程所學習到得事物對我是有吸引力的。					
14	我在生物科課程所學習的事物是我感到興奮的。					
15	我喜歡我在生物科課程所學會的事物。					
16	我在生物科課程所做得事物是有趣的。					
17	生物科課程學習到的事物，可幫助我對其他事物了解。					
18	我在生物科課程學習到的事物，對我是重要的。					
19	我在生物科課程學習到的事物，我可以運用於生活中。					
20	我在生物科課程學習到的事物，對我是有價值的。					