INCORPORATING DIFFERENTIAL SPEED IN COGNITIVE DIAGNOSTIC MODELS WITH POLYTOMOUS ATTRIBUTES

By

Hope Onyinye Akaeze

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Measurement and Quantitative Methods—Doctor of Philosophy

2020

ABSTRACT

INCORPORATING DIFFERENTIAL SPEED IN COGNITIVE DIAGNOSTIC MODELS WITH POLYTOMOUS ATTRIBUTES

By

Hope Onyinye Akaeze

The recent increase in interest for instructional relevance and fine-grained feedback from assessments has led to a unified paradigm of educational measurement, combining cognitive psychology with psychometrics, and thus, cognitive diagnostic assessment or CDA. CDAs are particularly useful for identifying areas of students' needs as well as designing individualized instruction and learning/teaching interventions to meet those needs. However, the typical CDAs assess coarsely defined attributes and lack information on the cognitive processes that underlie test performance.

Cognitive processing takes time. A typical CDA is time-limited and the time an examinee allocates to tasks can provide insight on the cognitive process underlying the response. Response time (RT) has therefore been identified as important collateral information that can be used to account for examinee behavior in cognitive assessment. However, the use of RT in measurement models has, so far, been limited to approaches with the strict assumption that a test taker maintains a constant speed over the test process. In addition, most cognitive diagnostic modeling approaches have been directed towards classification of examinees based on their profiles on dichotomized status on the latent skill. Classifying latent attribute status into mastery and non-mastery not only obscures information but also ignores the fact that learning can be progressive, and respondents in the same category (mastery/non-mastery) may possess the skill to a considerably varying degree. These two concerns are the focus of the current study.

This study aims to develop a more adaptable and informative modeling approach for examining and accounting for the effect of time *speededness* on examinees' cognitive processing behavior and ability in diagnostic models with polytomous attributes, thereby increasing the diagnostic potentials of CDAs. This is achieved by integrating variable working speed and partial mastery (polytomous attributes) into cognitive assessment model. The strengths of the model are assessed and compared to existing models using an empirical data and a simulation study. This new model, where applicable, allows for finer-grained feedback and flexibility in the assumed role of RT in cognitive diagnostic assessment while providing useful supplementary information to better understand testing strategies and behaviors.

To the memory of my beloved parents, Israel and Obioma Ilechukwu

and

To my “Dad” & “Mom”, Ikechukwu and Chibuzor Ilechukwu
Thank you for investing in me unconditionally.
“Your greatest contribution to the universe may not be something you do,
but someone you raise” - Unknown

ACKNOWLEDGMENTS

Many thanks to my advisor and dissertation chair, Dr. Kimberly Kelly. I would not have made it this far without her unwavering support. I thank her for believing in me, even when I doubted myself. I would also like to thank my committee members, Dr. Mark Reckase, Dr. Aline Godfroid, and Dr. Chi Chang. This project was developed and refined by their constructive feedback and insights. I lack the words to adequately describe how blessed I feel to have their unconditional support, even with the inconveniences created by the COVID-19 situation.

My immense gratitude goes to my brother ("Dad"), Ikechukwu Ilechukwu, and his wife ("Mom"), Chibuzor Ilechukwu. I am grateful for their unfailing emotional, financial, and spiritual support; their dedication to my success fuels my dreams. I also want to thank my siblings and siblings-in-law. Their unconditional love and support cannot be measured. They held me through and gave me a reason to persevere to the finish line, especially after my mother's demise. I am eternally grateful to God for them. Special thanks go to my husband, Henry Akaeze, for his sacrifices and patience through this arduous journey. A big thank you to my nieces and nephews for their support and sincere belief in my ability to set a good example for them.

I would also like to thank my mentors and colleagues at the Center for Statistical Training and Consulting (CSTAT). The mentoring and support I received as a member of the CSTAT team, together with the statistical consulting experience, enriched my Ph.D. experience in no small measure. Many thanks also go to the Office of International Students and Scholars (OISS) and the Community Volunteers for International Program (CVIP) at MSU for their unwavering financial support through the most challenging periods of my journey through MSU.

For their prayers and support, I sincerely appreciate Pastor and Mrs. Ozoemena Ani, and terrific friends like Nathalie, the Aboludes, the Nwankwos, the Mwikas, and the Flynns. Time and space would fail me to mention everyone by name. I also want to especially thank my friends and study buddies, Talesha, Tingqiao, Kathy, Lin, and Lora, who held me accountable for every minute of procrastination and helped me stay on track. I cannot forget many other friends who traveled the rough paths with me, cheered me on, and celebrated each accomplishment – the list is long. I am grateful for such good friends.

My special thanks also go to Dr. Peida Zhan of Zhejiang Normal University, Jinhua, China, for providing me with helpful references, reviewing and providing feedback on my R codes, and generously sharing his knowledge about the cognitive diagnostic model with me. I also thank Dr. Tzur Karelitz of The National Institute for Testing & Evaluation, Jerusalem, Israel, for granting me free access to his data.

Finally, and most importantly, “To the only wise God our Saviour, be glory and majesty, dominion and power, both now and ever. Amen.” Jude vs. 25.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES xi

CHAPTER 1: INTRODUCTION 1
1.1 Statement of Problem 3
1.1.1 The Dichotomy Problem 3
1.1.2 The Speededness Effect 4
1.2 Purpose of Study 6
1.3 Research questions 8
1.4 Overview of Chapters 8

CHAPTER 2: LITERATURE REVIEW 10
2.1 Cognitive Diagnostic Modeling 10
2.2 The Q-Matrix 11
2.3 The Nature of Attributes 14
2.3.1 Higher-order Latent Trait Model 16
2.3.2 Attribute Hierarchy Model 17
2.4 Implementation of Cognitive Diagnostic Models 19
2.4.1 Bayesian Estimation Using MCMC 20
2.4.2 Estimating attribute profiles 24
2.4.3 Assessing model fit 26
2.4.4 Response Time Models 27
2.5 Response Time and Response Accuracy 27
2.5.1 Joint Models of Response Time with Accuracy 29
2.6 Response Time in Cognitive Diagnostic Models 29

CHAPTER 3: METHODOLOGY 31
3.1 The log-normal random quadratic variable speed model 31
3.2 Deterministic, Input, Noisy ‘And’ Gate (DINA) Model for Polytomous Attributes 33
3.3 The Joint Differential Speed DINA (JDS-DINA) Model 36
3.3.1 Model Specifications 41
3.3.2 Parameter Estimation 42
3.3.3 Assessment of Model Fit 46
3.4 Real data analysis 47
3.4.1 Model Comparisons 49
3.5 The Simulation Study 51
3.5.1 Simulation design 51
3.5.2 Data Generation 54
3.5.3 Model Evaluation 56

CHAPTER 4: RESULTS 58

4.1 Research Question 1 ... 58
4.1.1 Model fit statistics ... 58
4.1.2 Standard error of item parameter estimates ... 60
4.2 Research Question 2 ... 61
4.2.1 Comparison of model fit ... 61
4.2.2 Classification accuracies ... 62
4.3 Research Question 3 ... 63
4.3.1 Overall parameter recovery of the JDS-DINA model ... 63
4.3.2 Effect of design conditions on parameter recovery ... 67
4.4 Research Question 4 ... 72
4.5 Research question 5 ... 80

CHAPTER 5: DISCUSSION AND CONCLUSION ... 84
5.1 Summary of Findings ... 84
5.1.1 Dichotomization ... 84
5.1.2 Response time ... 84
5.1.3 Variable speed ... 85
5.1.4 Supplementary RT information ... 85
5.2 Limitations and future research ... 87
5.2.1 Test length ... 87
5.2.2 Parameter values ... 88
5.2.3 Simulation conditions ... 88
5.2.4 Computational burden ... 89
5.3 Summary ... 90

APPENDICES ... 92
APPENDIX A: SUPPLEMENTARY MATERIALS FOR RESEARCH QUESTION 4 ... 93
APPENDIX B: SUPPLEMENTARY MATERIALS FOR RESEARCH QUESTION 5 ... 101
APPENDIX C: JAGS CODES FOR STUDY MODELS ... 119

REFERENCES ... 130

LIST OF TABLES

Table 1 Binary Q-matrix 12

Table 2 Polytomous Q-matrix 13

Table 3 Design conditions – person parameters 53

Table 4 Design conditions – structural parameters 54

Table 5 Model fit statistics for the 2012 PISA computer-based mathematics test 59

Table 6 Estimated item parameters for the 2012 PISA computer-based mathematics items 60

Table 7 Model fit statistics for the Language Rule data 61

Table 8 Classification accuracy rates of attributes for the language data 62

Table 9 Bias and RMSE of item and structural parameters of JDS DINA with polytomous attributes 63

Table 10 MANOVA and ANOVA results for item parameters 67

Table 11 MANOVA and ANOVA results for attribute structural parameters 69

Table 12 Bias and RMSE of δ_0 by simulation design conditions – binary attributes 73

Table13 Exact and weighted classification accuracy rates 79

Table 14 Bias and RMSE of δ_0 by simulation design conditions – polytomous attributes 81

Table 15. Exact and weighted classification accuracy rates with polytomous attributes 83

Table 16 Computation times (in minutes) for study models 89

Table A1 Bias and RMSE of λ by simulation design conditions – binary attributes 102

Table A2 Bias and RMSE of δ_1 by simulation design conditions – binary attributes 104

Table A3 Bias and RMSE of γ_{0k2} by simulation design conditions – binary attributes 106

Table B1 Bias and RMSE of λ by simulation design conditions – polytomous attributes 110

Table B2 Bias and RMSE of δ_1 by simulation design conditions – polytomous attributes 112

Table B3 Bias and RMSE of γ_{0k2} by simulation design conditions – polytomous attributes.....114

Table B4 Bias and RMSE of γ_{0k3} by simulation design conditions – polytomous attributes.....116

Table B5 Bias and RMSE of γ_{0k4} by simulation design conditions – polytomous attributes.....118

Table B6 Bias and RMSE of γ_1 by simulation design conditions – polytomous attributes........120

LIST OF FIGURES

Figure 1 Forms of attribute hierarchy 18

Figure 2 Diagrammatic representation of the DINA model 34

Figure 3 Higher order polytomous attributes model 39

Figure 4 K × I Q matrix for binary attributes 52

Figure 5 K × I Q matrix for binary attributes 52

Figure 6 Profile correct classification rates 65

Figure 7 Effect of $\sigma^2_{\tau_0}$ on absolute bias and standard error of λ 68

Figure 8 Effect of $\sigma^2_{\tau_1}$ on absolute bias and standard error of λ 68

Figure 9 Effect of $\sigma^2_{\tau_2}$ on absolute bias and standard error of λ 69

Figure 10 Effect of $\rho_{\theta\tau_0}$ on absolute bias and standard error of $\gamma_{0_{k2}}$ 70

Figure 11 Effect of $\rho_{\theta\tau_0}$ on absolute bias and standard error of $\gamma_{0_{k3}}$ 70

Figure 12 Effect of $\rho_{\theta\tau_0}$ on absolute bias and standard error of $\gamma_{0_{k4}}$ 71

Figure 13 Effect of $\sigma^2_{\tau_0}$ on absolute bias and standard error of γ_1 71

Figure 14 Effect of $\rho_{\theta\tau_0}$ on absolute bias and standard error of γ_1 72

Figure 15 Bias of λ across simulation conditions 76

Figure 16 RMSE of λ across simulation conditions 76

Figure 17 Bias of δ_0 across simulation conditions 77

Figure 18 RMSE of δ_0 across simulation conditions 77

Figure 19 Bias of δ_1 across simulation conditions 78

Figure 20 RMSE of δ_1 across simulation conditions 78

Figure 21 Bias of γ_{0k2} across simulation conditions 79

Figure 22 Bias of δ_0 across simulation conditions – polytomous attribute configuration 80

Figure 23 RMSE of δ_0 across simulation conditions – polytomous attribute configuration.......83

Figure 24 Relationships among person parameters from PISA computer-based Math test 86

Figure A1 RMSE of γ_{0k2} across simulation conditions – polytomous attribute configuration...108

Figure A2 Bias of γ_1 across simulation conditions – polytomous attribute configuration..........108

Figure A3 RMSE of γ_1 across simulation conditions – polytomous attribute configuration......109

Figure B1 Bias of δ_1 across simulation conditions – polytomous attribute configuration..........122

Figure B2 RMSE of δ_1 across simulation conditions – polytomous attribute configuration......122

Figure B3 Bias of γ_{0k2} across simulation conditions – polytomous attribute configuration.......123

Figure B4 RMSE of γ_{0k2} across simulation conditions – polytomous attribute configuration...123

Figure B5 Bias of γ_{0k3} across simulation conditions – polytomous attribute configuration.......124

Figure B6 RMSE of γ_{0k3} across simulation conditions – polytomous attribute configuration...124

Figure B7 Bias of γ_{0k4} across simulation conditions – polytomous attribute configuration.......125

Figure B8 RMSE of γ_{0k4} across simulation conditions – polytomous attribute configuration...125

Figure B9 Bias of γ_1 across simulation conditions – polytomous attribute configuration..........126

Figure B10 RMSE of γ_1 across simulation conditions – polytomous attribute configuration 126

Figure B11 Bias of λ across simulation conditions – polytomous attribute configuration..........127

Figure B12 RMSE of λ across simulation conditions – polytomous attribute configuration......127

CHAPTER 1: INTRODUCTION

Learning, even from the best-designed instruction, can only be verified through assessment. A well-designed assessment provides evidence to validate the expected effect of instruction. It is, therefore, an indispensable tool in any teaching and learning process. Educational assessments are of two broad types – summative assessment and formative assessment – depending on the purpose. Summative assessments are comprehensive assessments administered at the end of the course of study, as a summary evaluation of student learning. On the other hand, formative assessments are designed to evaluate students' learning over the study period, primarily to inform and enhance the teaching and learning process. The feedback from formative assessments is particularly useful for diagnosing the strengths and weaknesses of students and/or instructional materials/approach and for determining the best improvement strategy, when necessary.

Over the past few decades, there has been an increased push for fine-grained feedback from formative assessments; feedback that provide information, not only on examinees' cognitive abilities, but also on their proficiencies in the required processing skills (Leighton, Gierl, & Hunka, 2004; Sessoms & Henson, 2018; Sheehan & Mislevy, 1990) on a test. This interest has led to a unified paradigm of educational measurement, combining cognitive psychology with psychometrics, and thus, cognitive diagnostic assessment or CDA (Leighton & Gierl, 2007). Cognitive diagnostic assessment is an alternative form of assessment that provides formative information on students' cognitive strengths and weaknesses in the targeted skills. Such assessments are particularly useful for identifying areas of students' needs as well as designing individualized instruction and learning/teaching interventions to meet those needs.

Truly, CDAs provide detailed information on cognitive ability or performance level, but the classic CDA is devoid of information regarding the cognitive processes that underlie test performance (De Boeck & Jeon, 2019). The knowledge of whether a student got an item right/wrong, or whether a student has mastered a skill or not, is insufficient to tell the cognitive process that led to the answer (Tatsuoka & Tatsuoka, 1979). Process information provides an answer to the 'why' and 'how' of a task response. Knowledge of this cognitive process has several advantages – detection of aberrant test behaviors, better understanding and interpretation of test scores, better calibration of tests and test items, and richer information for developing remedial interventions.

One approach for developing theoretical information about cognitive process is using expert knowledge of the process domain in assessment development (Rupp, Templin, & Henson, 2010). Another method is to retrospectively or concurrently review the examinees' instructional background (Tatsuoka & Tatsuoka, 1979) , or probe the examinees for self-report of their solution strategies (Rupp et al., 2010); both of which can be daunting with large scale assessments. Alternative evidence for cognitive process comes from eye tracking information (De Boeck & Jeon, 2019; Rupp et al., 2010). Predicated on the fact that the mind follows the eye, this technique tracks eye movement and uses location and duration of fixation to approximate the cognitive processes of examinees. As Rupp et al. (2010) noted, this procedure can be resource-intensive. De Boeck and Jeon (2019) identified an even more sophisticated procedure - evaluating the brain's electrical activity from electroencephalogram (EEG) – as another source of collateral information for process-related measurement.

Cognitive processing, as we know it, takes time; and the time an examinee allocates to tasks can provide insight on the cognitive process underlying the response (De Boeck & Jeon, 2019). The time taken to carry out all the operations required for a task is therefore a useful source of information that can be put to multiple uses, including improved estimation of examinees' performance. Moreover, with the advent of computerized testing, response time data have become accessible for this purpose. Several research efforts have also been directed towards modeling the relationship between response time and test performance, especially within the IRT framework (e.g., Sen, 2012; Tatsuoka & Tatsuoka, 1979); van der Linden, 2007; Verhelst, Verstralen, & Jansen, 1997; Wang & Hanson, 2005). A few others have also been devoted to improving cognitive diagnostic model (CDM) estimation and inferences by integrating response time with responses (e.g. Huang, 2019; Zhan, Jiao & Liao, 2018a; Zhan, Liao, & Bian, 2018b).

1.1 Statement of Problem

1.1.1 The Dichotomy Problem

As noted earlier, cognitive diagnostic assessments are developed to meet the need for finer-grained feedback from educational tests and to make these tests more relevant to classroom instruction. However, most modeling approaches have been directed towards classification of examinees based on their profiles as dichotomized status on the latent skills – mastery/non-mastery. Like any random variable, dichotomization leads to loss of information. Classifying latent attribute status into mastery and non-mastery not only obscures information (Karelitz, 2008), but it also ignores the fact that learning can be progressive (Karelitz, 2004), and respondents in the same category (mastery/non-mastery) may possess the skill to a considerably varying degree (Zhan, Ma, Jiao, & Ding, 2019a). On the other hand, modeling continuous attributes places students on a continuum that is, in most cases, not informative enough for meaningful formative and diagnostic purposes (Karelitz, 2004). To create a middle ground between these two extremes, researchers are calling for cognitive diagnostic modeling with theoretically relevant polytomous attributes that would allow some gradation in skills diagnosis (e.g., Hartz, 2002; Karelitz, 2004, 2008; Zhan, Wang, & Li, 2019b).

1.1.2 The Speededness Effect

Response time (RT), as a measure of cognitive process, has been identified as important collateral information that can be used to account for examinee behavior and improve the estimation of students' proficiency levels in cognitive assessment (Schnipke & Scrams, 2002). A good number of modeling approaches have therefore been proposed to explore the benefits of response time in item response models (e.g., Sen, 2012; Simonetto, 2011; van der Linden, 2007) and cognitive diagnostic models (e.g., Zhan et al., 2018a; Zhan et al., 2018b). These studies have shown that incorporating response time in item response modeling can improve estimation and classification accuracy, detect aberrant test-taking behaviors, and differentiate among different test-taking strategies.

Gulliksen (1950) identified two unique types of tests, with respect to timing – power tests and speed tests. Power tests are designed to measure only the knowledge level of examinees. For these tests, examinees are allowed unlimited time and are scored based on their responses alone. Speed tests are designed to measure cognitive processing speed and are scored based on the time taken to answer a fixed number of items or the number of items completed within a set time interval. Contemporary educational assessments, however, though designed to measure knowledge only, are usually time-limited. This time constraint on a power test introduces construct irrelevant variance into the measurement (Wollack, Cohen, & Wells, 2003; Kahraman, Cuddy, & Clauser, 2013), due to the intricate relationship between response time and accuracy/ability. This relationship is reflected in the phenomena known as the speed-accuracy tradeoff and speed-ability relationship. The speed-accuracy tradeoff defines a within-person negative nonlinear relationship between accuracy and time; the faster an examinee completes a task, the lower his or her level of accuracy on the tasks (van der Linden, 2007). At the between-person level, we can only define a speed-ability relationship whereby examinees with higher ability take less time to complete the test. While changes in speed are sometimes negligible (van der Linden, Breithaupt, Chuah, & Zhang, 2007), more substantial speed changes within an examinee are frequent in time-limited high-stakes tests. These can present unobserved dependencies in the item responses.

In cognitive assessments, RT is sometimes treated as a parallel dependent variable with response accuracy (RA), as a covariate for RA, or as a co-dependent variable with RA to explain local dependency (De Boeck & Jeon, 2019). While all three options for incorporating RT have been explored in the IRT framework (e.g., ; Fox & Marianti, 2016; Molenaar, Tuerlinckx & van der Maas, 2015; van der Linden, 2007) and cognitive diagnosis (e.g., Huang, 2019; Zhan et al., 2018a; Zhan et al., 2018b), the use of RT in measurement models has, so far, been limited to approaches with the assumption of constant speed. These approaches assume that a test taker maintains a constant speed over the test period. Few exceptions to this are the works by Fox & Marianti (2016) and Molenaar, Oberski, Vermunt, & De Boeck (2016), where response time is incorporated into the IRT model with differential speed. To date, no similar work has been recorded in cognitive diagnostic modeling.

Pure cognitive diagnostic models that use only the responses ignore the *speededness* effect – the effect that time constraint has on responses, and subsequent ability or mastery-based inferences. Accounting for response time with a constant speed assumption improves estimation but only captures the between-level relationship between response time and responses. However, this between-level relationship tells nothing about the within-level relationship and may cause the item responses to violate the local independence assumption. In a time-limited test, an examinee may change the speed of response due to fatigue, change in strategy, a reminder of a time limit, or other aberrant behavior like cheating or guessing. Hence, it would be naive to impose a constant speed assumption in modeling students' cognitive speed of performance on a test.

If RTs are to become routinely used as supplementary information in cognitive diagnostic modeling, more flexibility in the framework for incorporating RT is needed. Such modeling framework would provide insight into the effect of test *speededness* and reveal aberrant behaviors that could distort the ability estimates of examinees. It would also allow researchers to integrate changes in working speed and test their specific hypotheses about the role of RT in cognitive diagnostic models.

1.2 Purpose of the Study

The insight for this study is drawn from Fox & Marianti (2016)'s joint model for improving IRT estimation by incorporating RT with varying speed; Zhan et al. (2018a)'s work on incorporating RT in cognitive diagnostic models; and the need to increase the diagnostic potential of CDMs via polytomous attributes (Chen & de la Torre, 2013; Karelitz, 2004, 2008; Zhan et al., 2019b; Hartz, 2002). The current study provides an adaptable and informative modeling framework to examine and account for the effect of time *speededness* on examinees' cognitive ability and processing behavior in diagnostic models. The research goal is to propose a new approach that allows for finer-grained feedback and flexibility in the assumed role of RT in cognitive diagnostic assessment, and to compare the performance of the approach with existing ones, where applicable.

Karelitz, (2004) and Zhan et al. (2019b) developed two modeling frameworks to address diagnostic measurement with polytomous attributes. Karelitz (2004) proposed an ordered category attribute coding (OCAC) framework to model attributes with ordinal levels, each coded from 0 (for the lowest level) to the highest level. However, his approach assumes (1) invariant item parameters, which is unrealistically restrictive, and (2) a saturated latent structural model for the attributes, which can quickly become computationally intensive. To improve on the OCAC framework, Zhan et al. (2019b) proposed the partial mastery, the higher-order latent structural model for polytomous attributes. This model relaxes the OCAC constraint of common *slipping* and *guessing* parameters across items and shows that constraining the structural model with a higher-order latent trait model was equally good at parameter recovery and more parsimonious and time efficient (Zhan et al., 2019b). However, these two approaches, did not recognize the lack of local independence in responses that could be attributed to the underlying cognitive processes students engage in while responding to tasks used in measuring these attributes.

To account for the effect of cognitive processes using response time, Fox & Marianti (2016) proposed a joint model for responses and response times that allows for differential speed via a latent growth model for response times. Such an extension has never been explored with cognitive diagnostic models. On the cognitive diagnostic modeling side, Zhan et al. (2018a) proposed joint modeling of RT and RA, with a correlational structure between the model person and item parameters, following the hierarchical modeling framework of van der Linden (2007). Modeling the correlational structure of parameters, with a constant speed component, accounts for the interdependence between RT and RA but fails to account for the effect of differential test *speededness* on RA.

The current study (1) extends Fox & Marianti's (2016) work to cognitive diagnostic models, (2) generalizes the study by Zhan et al. (2018a) to polytomous attributes and (3) permits variable speed as examinees progress through the tasks on a test. The main objectives of this study are to (1) propose a new flexible model for incorporating response time into cognitive diagnostic models, with polytomous attributes and differential speed across tasks; (2) assess the performance of the new model in terms of parameter recovery for different conditions of sample size, number of items and correlations between RT and ability and (3) compare the performance of the new model with that of existing models on real data, in terms of model fit and precision of parameter estimates.

1.3 Research questions

In line with the objectives listed above, this study seeks to answer the following research questions:

1. How does the new model compare with existing models, in terms of model fit and precision of estimates?
2. How does the dichotomization of polytomous attributes affect correct classification accuracies?
3. How is the recovery of item and person parameter estimates in the new model affected by the variances of RT parameters and the correlation between RT and RA parameters?
4. How well does the new model recover person and item parameter estimates when attributes are dichotomized?
5. How well does the new model for polytomous attributes recover person and item parameter estimates?

1.4 Overview of Chapters

In the following chapters, cognitive diagnostic models, response time models, and their joint models are discussed in greater detail, with emphasis on the aspects that are germane to the objectives of the current study. Chapter 2 presents definitions of relevant concepts and terminologies related to response time and cognitive diagnostic models. In this chapter, the study's rationale is established through a comprehensive literature review of existing studies related to exploiting response time in cognitive diagnostic modeling.

Chapter 3 describes the methods used to address the research questions, including the technical details of the proposed and existing models, model estimation and assessment procedures, description of empirical data, and the design and implementation of the pertinent simulation study. The results from the methods described in Chapter 3 are summarized and presented in Chapter 4. Finally, Chapter 5 provides a discussion of the results and their implication for cognitive diagnostic modeling. This fifth chapter concludes the study with the limitations of the current study and recommendations for future studies.

CHAPTER 2: LITERATURE REVIEW

2.1 Cognitive Diagnostic Modeling

Cognitive diagnostic models or CDMs (de la Torre, 2009; Huebner, 2010) are psychometric models specially designed to assess examinees' proficiency and classify them based on their mastery or non-mastery of postulated attributes. CDMs are also known by several other names in literature, such as restricted latent class models (Haertel, 1989; Xu, 2017), diagnostic classification models (Rupp et al., 2010; Sessoms & Henson, 2018), cognitive psychometric models, multiple classification latent class models or MCLCM (Maris, 1999) structured item response theory (SIRT) models (Rupp & Mislevy, 2007; Leighton et al., 2004), and latent response models (Maris, 1995) (check Rupp & Templin, 2008 for more labels). These variants differ in functional forms, assumptions, complexities, and areas of emphasis, but they all provide nuanced information on examinees' skills, which are then used to classify them based on their score profiles on the skills.

The major difference between CDMs and traditional multidimensional item response theory (IRT) or CTT models is that the former is concerned with a binary (mastery or non-mastery) or polytomous latent variable for diagnostic and criterion-referenced purposes while the latter provide scores on a continuously valued latent variable, mostly for norm-referenced interpretations. The categorical nature of the latent variable is one similarity between CDMs and conventional latent class models. However, in latent class models, subjects are classified into one of many possible categories of a single latent variable based on their observed response pattern, but CDM classifies subjects based on their membership to latent categories of many latent variables or attributes (Maris, 1999) and classification is restricted by the assumed form of interaction among the measured attributes.

Like every other modeling tool, CDM has had its fair share of criticisms. Most of the criticisms of CDM applications are concerned with the lack of evidence for reliability, validity, distinctiveness of attributes, measurement invariance, and informed practical decision-making (e.g., Sinharay & Haberman, 2009; Bradshaw, Izsák, Templin & Jacobson, 2014; Chen & de la Torre, 2014; Henson, 2009; Jurich & Bradshaw, 2014; Ravand, 2016; Rupp & Templin, 2008). A more overarching problem, which seems to be the source of all the other limitations, is that most existing educational assessments are designed to align to content instead of attributes. As a result, CDM applications employ the retrofitting procedure, with items that are coded for attributes after the test had been developed and administered. This qualifies the utility of inference generated from CDMs since the items were not originally developed to measure these micro-level attributes (Gierl, Alves, & Majeau, 2010). Details of these limitations are discussed elsewhere in Roussos, Templin & Henson (2007).

2.2 The Q-Matrix

In the cognitive diagnostic assessment framework, any skill or specific knowledge that a student requires to perform a task is generally referred to as an attribute. For any item/task, the combination of attributes needed for a correct response to it is known as its attribute profile. Every examinee is also characterized and classified by attribute profile, which is his/her mastery levels on the vector of attributes being measured. The complete list of all possible combinations of attributes assessed in a test is called the latent attribute space (Tatsuoka, 1990). The Q matrix is a matrix representation of the relationship between test items and the attributes of interest in a diagnostic assessment. It is a I by K matrix with elements q_{ik} indicating the mastery level of attribute k required to answer item i correctly, where I is the number of items, and K is the number of attributes (Karelitz, 2008; Zhan et al., 2019). The columns of the Q matrix represent the attributes and the rows represent the items so that $q_{jk} = l\text{-}1$ ($l = 1 \ldots L_k$) if item i requires examinees to possess level l of attribute k ($k = 1 \ldots K$) for a correct response, where L_k is the number of mastery levels measured for attribute k. The first mastery level of every attribute is set to 0 so that $q_{jk} = 0$ if item i requires the lowest mastery level of attribute k for a correct response.

The Q matrix for a test with binary attributes is a special case of that described above. If all attributes are measured at only two levels, mastery/non-mastery, then the Q matrix reduces to a binary matrix of zeros and ones but with the same dimension. As an example, Table 1 provides the Q-matrix for the attributes in the *Numbers* content domain of TIMSS 2011 eighth-grade mathematics, adapted from Table 2 of Terzi & Sen (2019).

Table 1

Binary Q-matrix

Item j	Attribute k		
	α_1	α_2	α_3
1	1	1	0
2	0	1	0
4	1	0	0
5	1	1	1
15	1	1	0
16	1	1	0
18	0	0	1
30	0	0	1
31	0	0	1

Note. α_1 – Possesses understanding of fraction equivalence and ordering; uses equivalent fractions as a strategy to add and subtract fractions. α_2 – Understands decimal notation for fractions and compares decimal fractions; performs operations with decimals. α_3 – Understands ratio concepts and uses ratio reasoning to solve problems; finds a percent of a quantity as a rate per 100.

Table 2 represents the same information on Table 1 but, for illustrative purposes only, hypothetical entries have been assigned in the matrix to represent a hypothetical testing scenario for polytomous attributes. Here, the first attribute is measured at two levels and the last two at three mastery levels (non-mastery, intermediate, and mastery).

Table 2

Polytomous Q-matrix

Item *j*	Attribute *k*		
	α_1	α_2	α_3
1	0	0	0
2	1	0	0
4	1	0	2
5	1	1	2
15	0	2	0
16	1	2	0
18	0	1	0
30	0	1	1
31	0	0	2

The Q-matrix is very similar to the loading matrix of a factor analysis model. However, unlike the loading matrix of FA, the Q-matrix is a key input for CDMs, and its correct specification is essential for a valid CDM-based assessment. An incorrect specification can lead to wrong parameter estimation and incorrect classification of examinees into proficiency groups (Chiu, Douglas, & Li, 2009; Köhn & Chiu, 2018a). For a Q-matrix to be valid, it must be complete, which means it should allow for all the possible proficiency profiles of examinees to be identified (e.g., Chiu et al., 2009).

Construction of Q-matrix is usually done by a panel of subject-matter experts, item developers, or teachers. The results of such combined qualitative inputs can be very subjective. Given the crucial role of Q-matrix in CDM, the subjectivity in its construction can pose severe problems in parameter estimation and model validation. As such, several studies in the literature have been devoted to studying the completeness, validation, and impact of misspecification of the Q-matrix on CDM and CDM-based inferences (e.g., Kunina-Habenicht, Rupp & Wilhelm, 2012;

Chiu, 2013; Terzi & Sen 2019; DeCarlo, 2011; de la Torre, 2008; de la Torre & Chiu, 2016; Liu et al., 2012; Terzi & de la Torre, 2018).

2.3 The Nature of Attributes

Latent traits, skills, attributes, latent characteristics, and elements of processes are all different labels used in literature for the categorical latent variables assessed in CDMs. The choice of label depends on the theoretical interpretations and inferences to be made about them. The degree of detail desired in the resulting inference determines the definitional specificity or definitional grain size of the attributes (Rupp, et al., 2010; Hong, Wang, Lim, & Douglas, 2015). A task with broad scope would often be operationalized with coarse-grained attributes to keep the dimension of the corresponding CDM practicable. The more finely-defined the attributes are, the higher the number of mastery levels for the attributes, and the more unmanageable the CDM becomes. To overcome this limitation, most CDMs are implemented with coarsely defined attributes for tasks that are broad in scope and finely defined attributes for tasks with smaller scope (Rupp et al., 2010). The selection, labeling, definition, and coding instructions for attributes must be done carefully to sufficiently represent the theoretical basis and intended use of the diagnostic assessment and, at the same time, prevent ambiguity and high inter-rater disagreement (Rupp, et al., 2010). For a meaningful diagnosis, the definition of grain size for an attribute must have theoretical support for its existence and developmental levels (Karelitz, 2008)

In practice, the skills or attributes measured in a test are conceptually related. Such relationships need to be accounted for in cognitive assessment modeling (de la Torre & Douglas, 2004) to reflect the way the attributes interact in the response process. The assumed nature of this relationship among attributes at the item level leads to the categorization of CDM into compensatory and non-compensatory models (Roussos, Templin, & Henson, 2007).

Non-compensatory models assume that the solution to an item depends on a combination of attributes. An examinee must possess of all these attributes for a correct response on that task. Examples of such models include the DINA (deterministic input noisy and) model (Haertel, 1989); NIDA model (Junker & Sijtsma, 2001); HYBRID model (Gitomer & Yamamoto, 1991); unified model (UM) (DiBello, Stout, & Roussos, 1995); re-parameterized unified model (RUM) (Hartz, 2002); and the conjunctive MCLCM" (Roussos et al., 2007). These models assume that an examinee who lacks any one of the required attributes cannot provide a correct response to the task unless by guessing. An examinee who possess all the required attributes cannot get it wrong unless by slipping.

Compensatory models, on the other hand, assume that the correct response to a task can be achieved if an examinee has mastered at least one of the attributes required to perform the task. Such models are particularly applicable in psychiatric and other medical diagnoses, where a disease can be considered present if at least one of its symptoms is present (Roussos et al., 2007) or in tasks for which multiple appropriate strategies (each requiring different skills) could be applied to arrive at the correct answer. Some examples of compensatory cognitive diagnostic models are the disjunctive MCLCM and compensatory MCLCM of Maris (1999), the DINO (deterministic input noisy or) model of Templin and Henson (2006), and the NIDO (noisy input deterministic or) model of Templin, Henson, and Douglas (2006) (as cited in Roussos et al., 2007).The form of attributes' interaction assumed in a model depends on the purpose of the assessment and the definition of the attributes.

Whether compensatory or non-compensatory, CDMs make the fundamental assumption of conditional independence of response vectors, like the traditional IRT and LCA models. The item responses are independent, given an examinee's mastery profile on the latent attributes. The latent attributes, however, do not influence the responses in isolation. If the k polytomous attributes had L_k levels, then, without any relationship or constraints imposed on the polytomous attributes, the maximum number of possible latent profiles is $C = \prod_{k=1}^{K} L_k$ where L_k is the number of mastery levels for attribute k, and K is the number of the attributes measured by the assessment. In the actual implementation of CDM, only $\prod_{k=1}^{K} L_k - 1$ parameters are estimated since the profiles are mutually exclusive and collectively exhaustive. The model with all parameters estimated is the unstructured or saturated model. For the constrained CDM, there are several approaches for accounting for the correlation among the attributes – the higher-order (HO) latent trait model of de la Torre & Douglas (2004), the attribute hierarchy model (AHM) of Leighton et al. (2004), and the hierarchical diagnostic classification model (HDCM) of Templin & Bradshaw (2014). There has been only one attempt, by Zhan, Ma, Jiao, & Ding (2019), to combine two of these approaches in one model.

2.3.1 Higher-order Latent Trait Model

The higher-order latent trait model assumes that a continuously valued latent variable or general ability underlies the binary latent attributes, such that a two-parameter logistic model defines each attribute as a function of the underlying latent trait (de la Torre & Douglas, 2004):

$$P(\alpha_{jk} = 1|\theta_j) = \frac{e^{\gamma_{0k}+\gamma_{1k}\theta_j}}{1 + e^{\gamma_{0k}+\gamma_{1k}\theta_j}} \qquad (1)$$

Where θ_j is the trait level of examinee *j* and is assumed to follow the standard normal distribution, γ_{0k} is the intercept or location parameter and γ_{1k} is the slope or discrimination parameter for attribute *k*. The reasoning behind this approach is that an examinee with a higher value on the latent trait is more likely to demonstrate mastery of an attribute (de la Torre & Douglas, 2004). When the higher-order approach is incorporated into a CDM, it reduces the dimension of the model parameter space to from $2^K - 1$ to 2K and provides summative information on the latent trait, in addition to the mastery levels of the latent attributes.

2.3.2 Attribute Hierarchy Model

The attribute hierarchy model of Leighton et al. (2004) is based on the assumption that the attributes assessed on a test are the basic cognitive processes necessary to solve the task correctly, and the performance on the test is based on a set of skills that are hierarchically organized such that mastery of the lower-level attributes in the hierarchy is prerequisite to mastery of the higher-level ones. A fundamental premise for the application of this framework is that the attribute hierarchy must be determined before the test "because the hierarchical organization of attributes must guide the development of test items" (Leighton et al., 2004). Leighton et al. (2004) identified four distinct forms of attribute hierarchy – the linear, convergent, divergent, and unstructured hierarchies. Two or more of these hierarchies can be combined to form a complex hierarchy, where the complexity varies with the cognitive load of the task (Kim, 2001).

The linear attribute hierarchy organizes attributes in increasing order of cognitive load. It requires that all attributes be mastered sequentially such that mastering attribute 1 is a prerequisite for mastering 2; attribute 3 cannot be mastered without 2, and so on. With the linear hierarchy, there is only one attribute at the top of the hierarchy and only one path to get from the lowest to highest skill level. The convergent hierarchy also has only one attribute at the top of the hierarchy, but an examinee can attain the highest skill level through multiple different paths. For instance, attribute 1 is a prerequisite for 2 and 3; attribute 4 can be mastered if an examinee has mastered 2 or 3, and attribute 4 is a prerequisite for 5. With this structure, an examinee can attain the skill level of 5 by mastering 1, 2, and 4 or 1, 3, and 4. The divergent and unstructured hierarchies also start with a single prerequisite attribute, but end with multiple skills at the highest level. With the divergent hierarchy, we have intermediate prerequisites before reaching any of the top skills. For the unstructured, all attributes except the single prerequisite, are at the top level so that there are no intermediaries. Figure 1 (Leighton et al., 2004) represents a diagrammatic representation of these hierarchy forms using six attributes.

Specifying a hierarchical relationship among attributes reduces the number of plausible attribute profiles. For instance, suppose that a ‘1’ denotes mastery, and ‘0’ indicates non-mastery of an attribute, then, for the linear hierarchy, it is not possible to have the profile 101111 because skill 3 cannot be mastered without skill 2. With the linear form of hierarchy and six attributes, the number of possible profiles reduces from $2^6 = 64$ to $6+1 = 7$. The extent of reduction depends on the number of attributes and the form of hierarchy stipulated. Implementation of the attribute hierarchy approach requires proper identification of the plausible attribute profiles that map onto the specified kind of hierarchy and guarantees a complete Q-matrix (Köhn & Chiu, 2018a; Templin, & Bradshaw, 2014).

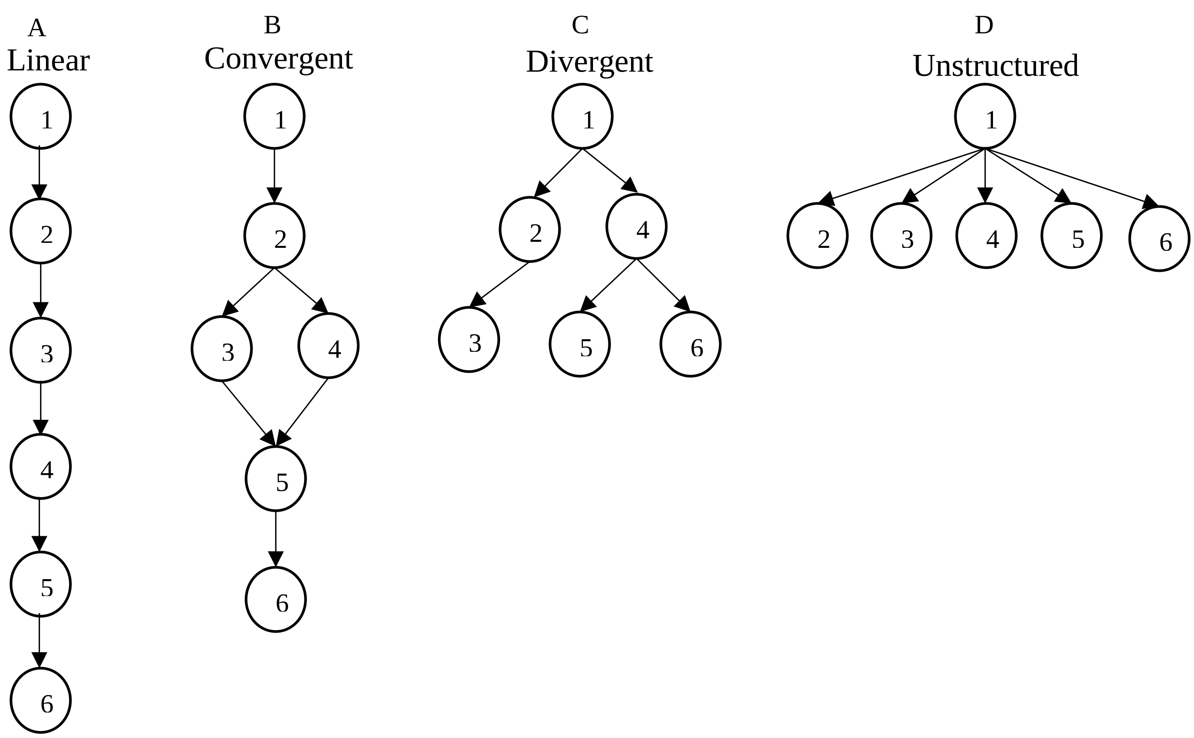


Figure 1 Forms of attribute hierarchy

2.4 Implementation of Cognitive Diagnostic Models

CDM estimation entails estimating the item parameters (as defined by the model of choice), structural parameters, and attribute profiles (respondent parameters). These sets of parameters may be estimated simultaneously by joint maximum likelihood estimation (ML) or marginalized maximum likelihood (MML) estimation using the expectation-maximization (EM) algorithms (de la Torre, 2009; von Davier, 2005; Rupp et al., 2010). ML estimates can become computationally complex when complex constraints are imposed on them (Rupp et al., 2010). In the MML, a population distribution is assumed for the structural parameters, which are then treated as 'known' and marginalized out for the estimation of the item parameters. The estimated item parameters are then treated as 'known' for the estimation of attribute profiles, and the entire process is repeated until the stopping criterion is satisfied. Implementation of the MML for LCDM in the M*plus* (Muthen & Muthen, 2018) statistical software is outlined in Rupp et al. (2010). The MML procedure is computationally expensive because it requires integration across the distribution of the latent variables for each examinee variable. The EM algorithm increases in computational intensity with increase in number of latent classes. It also requires the specification of starting values to initialize the algorithm convergence of the algorithm. Convergence may take longer or never be attained if these values are far from the true unknown values (Rupp et al., 2010).

Alternatively, item parameters and attribute profiles may be obtained simultaneously in the Bayesian estimation context using a Markov Chain Monte Carlo (MCMC) estimation (e.g., de la Torre & Douglas, 2004). This approach is especially useful when dealing with complicated likelihood functions, for which optimization with EM algorithm is not feasible. It focuses on determining the posterior distribution for each parameter from which specific estimates are obtained as a summary statistic like the mean, mode, or percentile of the distribution. While this approach provides an alternative for cases where EM estimation is not feasible, its application is impeded by technical details such as choice of prior, burn-in length, and occasional convergence issues (Templin, 2004; Rupp et al., 2010).

2.4.1 Bayesian Estimation Using MCMC

In Bayesian inference, the uncertainty about parameter estimates is expressed in terms of probability models (distributions), implying that parameters are random instead of fixed. It is grounded in the Bayes' theorem, whereby one's prior knowledge or belief about the unknown parameter is combined with the data at hand to derive an updated or posterior knowledge about the parameter. The initial knowledge is commonly specified in the form of a probability density or the prior distribution. Information from data at hand is defined in terms of its likelihood, and the resulting updated distribution is the posterior distribution. The algebraic expression of this process is expressed with the Bayes' theorem as follows:

$$P(\boldsymbol{\Omega}|\boldsymbol{Y}) = \frac{P(Y|\boldsymbol{\Omega}) \times P(\boldsymbol{\Omega})}{P(\boldsymbol{Y})} \qquad (2)$$

Where $\boldsymbol{\Omega}$ is a set of unknown parameters that are of interest in the estimation, $P(\boldsymbol{\Omega})$ is the prior distribution of $\boldsymbol{\Omega}$, $P(Y|\boldsymbol{\Omega})$ is the likelihood of the data given $\boldsymbol{\Omega}$, and $P(\boldsymbol{Y})$ represents the marginal likelihood of the data. Since the observed data $\boldsymbol{Y}$ is considered as fixed, $P(\boldsymbol{Y})$ is simply a normalizing constant to ensure that $P(Y|\boldsymbol{\theta}) \times P(\boldsymbol{\Omega})$ is a true density, and can be dropped from equation (2) to yield:

$$P(\boldsymbol{\Omega}|\boldsymbol{Y}) \propto P(Y|\boldsymbol{\Omega}) \times P(\boldsymbol{\Omega}) \qquad (3)$$

The left-hand side of equation (3) is the posterior distribution, obtained by modifying the prior knowledge about the parameter $P(\boldsymbol{\Omega})$ by the likelihood of the observed data $P(Y|\boldsymbol{\Omega})$. Bayesian estimates of the unknown parameters are obtained as descriptive measure of the corresponding posterior distribution – mean, median, mode, credible interval, etc.

With simple models, the parameter estimates can be obtained algebraically from the posterior distribution but, when models are considerably complex, determining the exact solution from closed form of the posterior is often impossible (Kruschke, 2014; Robert & Casella, 1999). For instance, with the DINA model, the posterior distribution would be a complex joint distribution of attribute profiles for all examinees as well as guessing and slipping parameters for all items. This limitation with complex models had restricted earlier implementation of Bayesian estimation to the use of conjugate priors – priors that are in the same distributional family as their resulting posteriors (Robert & Casella, 1999). When making draws from posterior distributions proves difficult, simulation methods like the Markov Chain Monte Carlo Methods (MCMC) are used to obtain and characterize the posterior distribution.

Monte Carlo integration is a method for drawing independent samples from a required distribution and using the sample averages in approximating the expectation of the distribution. A Markov chain is a sequence of random variables X_t with the property that the state *t* of the variable only depends on the state *t-1* of the random variable generated just previously. This property ensures that estimates based on any of the MCMC methods at each iteration depend only the iteration just preceding it (Roberts, 1996). In the MCMC estimation procedure, a Markov chain is constructed by generating samples from the posterior distribution. This begins with some trial initial values, followed by a series of random draws. These steps are run many times until a stationary distribution is reached. For X_t to attain a stationary distribution, the chain must be irreducible, aperiodic, and positive recurrent. See Roberts (1996) for more details. The stationary distribution for each parameter represents its posterior distribution (Roberts, 1996; Rupp, et al., 2010). In the MCMC process, there are several techniques for drawing random samples from the posterior distribution. These include the slice sampling, the Metropolis-Hastings algorithm, and the Gibbs sampling, among others. These sampling techniques differ in terms of the proposal distribution chosen to construct the chain and the probability of moving between states.

The Metropolis-Hastings (or M-H) algorithm begins the sampling procedure by drawing from a *proposal* distribution that depends on the current state of the Markov chain and computes an acceptance probability to decide whether to retain the sample and move to the next state or not. The proposal distribution is defined by a step-size, which must be adjusted (or tuned) at each step. If a Gaussian distribution is chosen, as is commonly the case, the variance parameter is used as the step size (Dittmar, 2013). The typical M-H algorithm proceeds as follows (Hastings, 1970):

1. Choose a random starting value X_t and an arbitrary proposal distribution $g(y|x_t)$ from which the next sample value y is drawn given the previous value, x_t. The density $g(.|x_t)$ must be a symmetric distribution so that $g(y|x_t) = g(x_t|y)$.
2. For each iteration, draw a candidate value y from $g(.|x_t)$ and accept y with probability:

$$\alpha(x_t, y) = min\left(1, \frac{g(x_t|y)\pi(y)}{g(y|x_t)\pi(x_t)}\right)$$

3. If y is accepted, set $x_{t+1} = y$; otherwise, set $x_{t+1} = x_t$

Unlike the M-H, the Gibbs sampling algorithm uses the conditional posterior distribution as the proposal distribution, and acceptance probability is set to 1, making Gibbs sampling a special case of M-H algorithm. This technique requires the conditional posterior distribution for each parameter, given the other parameters and the data, to be fully specified (Geman & Geman, 1984). It also assumes that, if the regularity conditions are met, the joint posterior distribution is determined by all the full conditional posterior distributions (Geman & Geman, 1984; Casella & George, 1992). The Gibbs sampling proceeds as follows:

1. Choose random starting values for all P unknown parameters $\theta_1^0, \theta_2^0 \cdots \theta_P^0$

2. Given the starting values $\theta_1^{(0)}, \theta_2^{(0)} \cdots \theta_P^{(0)}$, sample θ_p^{t+1} from the conditional posterior distribution of θ_k given the data and presumably known values of the other parameters, $P\left(\theta_p | Y, \theta_1^{(t+1)}, \theta_2^{(t+1)} \cdots \theta_{p-1}^{(t+1)}, \theta_{p+1}^{(t)}, \cdots \theta_P^{(t)}\right)$. This step is looped through all the P parameters in the set, replacing the already sampled parameters $\theta_1, \theta_2, \cdots \theta_{p-1}$ with their sampled values $\theta_1^{(t+1)}, \theta_2^{(t+1)} \cdots \theta_{p-1}^{(t+1)}$ in the conditional posterior distribution.

The Gibbs sampling algorithm is feasible and straightforward when the joint distribution is not explicitly known or is difficult to sample from directly, but the conditional distribution of each parameter is known and is easy to sample from. In more complicated models, where the conditional distribution cannot be directly sampled from, alternative MCMC algorithms like the M-H algorithm or the slice sampling may be adopted for this step.

The slice sampling (Neal, 2003) is not as popular as the M-H or Gibbs algorithm but circumvents the problem of tuning proposal distributions by adaptively adjusting the step-size to match the local properties of the density function. The basic idea is that one can sample from a distribution by sampling uniformly from the region under the density plot. The slice sampling algorithm may be summarized as follows (Neal, 1997):

1. Choose a starting value x_0 for which $f(x_0) > 0$, where $f(x)$ is proportional to the posterior density of interest
2. Draw a $y \sim (0, f(x_0)$
3. Slice $f(x)$ horizontally at y
4. Sample a point (x, y) from the line segment
5. Repeat steps 2 through 4 using the new x.

For the multivariate posterior distribution, the univariate algorithm above can be used to sample and update each parameter in turn (Neal, 2003).

Irrespective of the algorithm used to draw samples, the sequence of values produced in the MCMC process are dependent since every new state depends on the previous one, making the first set of values unrepresentative of the posterior distribution being sampled (Kim & Bolt, 2007). As a result, the initial set of values, also known as *burn-in*, are discarded before assessing the posterior distribution. There are, therefore, some practical concerns in the implementation of MCMC estimation. These include: (1) The length of chain; (2) The *burn-in* length; (3) The choice of starting values, which may affect convergence; (4) The number of chains needed to attain stationary posterior distributions for each parameter. While there are suggestions in literature for each of these concerns (e.g., Roberts, 1996; Gelfand & Smith, 1990; Gelman & Rubin, 1992; Raftery & Lewis 1992a), diagnostic checks of the Markov chain can be used to evaluate the performance of the process before parameter estimates are extracted.

For convergence check Gelman & Rubin (1992) proposed the $\hat{R}$ statistic that is based on a comparison of the pooled between chain variances and within chain variances for each parameter. Stability is indicated by a ratio that is close to 1. Other diagnostic approaches have also been recommended (e.g., Geweke, 1992; Raftery & Lewis 1992b), but there is no consensus on which is optimal. Knowledge about whether the Markov chain has converged via convergence diagnostics or examination of trace plots can be used to inform the decision about the required burn-in.

2.4.2 Estimating attribute profiles

Given the class membership and class-specific response probabilities, examinees can be scored and classified to classes based on their mastery level of the attribute vector using one of three common approaches - via maximum likelihood estimation (MLE), maximum a posteriori

(MAP), or expected a posterior (EAP) (Huebner & Wang, 2011). The MLE approach assigns an examinee to the attribute pattern $\hat{\alpha}_{MLE}$ that maximizes the likelihood of the responses:

$$L(\boldsymbol{X}_j|\boldsymbol{\alpha}) = \prod_{i=1}^{I} P(X_{ij}|\boldsymbol{\alpha}) \qquad (2)$$

Sometimes, prior information is available on the proportion of examinees expected in each skill pattern, and this can be incorporated into the likelihood with the maximum a posterior (MAP) approach. With a non-informative prior, the MLE and MAP yield the same results (Huebner & Wang, 2011). Given C skill patterns, the prior probability can be denoted as $P(\boldsymbol{\alpha}_c)$ such that $\sum_{c=1}^{C} P(\boldsymbol{\alpha}_c) = 1$ then, each examinee is classified into the skill pattern that maximizes the posterior probability. The posterior probability is defined by Huebner & Wang (2011), from Bayes theorem, as:

$$P(\boldsymbol{\alpha}_c|\boldsymbol{X}_j) = \frac{L(\boldsymbol{X}_j|\boldsymbol{\alpha}_c)P(\boldsymbol{\alpha}_c)}{\sum_{m=1}^{C} L(\boldsymbol{X}_j|\boldsymbol{\alpha}_m)P(\boldsymbol{\alpha}_m)} \qquad (3)$$

Though statistically straightforward, the MLE and MAP results may be hard to interpret because they do not provide separate probability estimates for each attribute. Expected a posterior (EAP) approach, on the other hand, provides probability estimates for each of the attributes for all response patterns by taking the aggregate of probabilities across all latent classes where the specific attribute has been mastered. EAP calculates the probabilities of mastery for each attribute and sets up a cutoff probability at (usually) 0.5 to determine if the attribute has been mastered or not for each examinee (Huebner & Wang, 2011; Rupp et al., 2010; Embretson & Reise, 2000).

Huebner & Wang (2011) compared all three approaches in a simulation study. Their results show that, across all the varied conditions, MLE/MAP had a higher proportion of correctly classified examinees on all skills, but the EAP presented a higher proportion correct classification on total skills. They conclude that none of the methods can be judged as better than the others;

rather, preference for classification method should be guided by the purpose of the diagnostic assessment.

2.4.3 Assessing model fit

As with all statistical models, the results of CDMs are meaningless if the model fit is unacceptable. Assessment of CDM fit could be in terms of absolute or relative fit. Measures of absolute model fit include the absolute value of the deviations of Fisher-transformed correlations and the limited information RMSEA (Houts & Cai, 2013) or RMSEA2 (Hu, Miller, Huggins-Manley, & Chen, 2016). RMSEA2 values of <0.089 indicate adequate fit, while values <0.05 are indicative of a close fit for multidimensional IRT (Maydeu-Olivares & Joe, 2014), and these values have been adopted for CDMs as well (e.g., Hu et al., 2016).

To evaluate relative model fit, CDM researchers use relative fit indices like the Akaike Information Criterion (AIC; Akaike, Parzen, Tanabe & Kitagawa, 1998) and Bayesian Information Criterion (BIC; Schwarz, 1978) (Sessoms & Henson, 2018). These fit statistics are used to compare fit among multiple competing models (e.g., de la Torre and Douglas, 2008). Kunina-Habenicht et al. (2012) and Hu et al. (2016) compared the performances of relative fit indices in DCMs. Kunina-Habenicht et al., (2012) studied the performance of AIC, BIC, and SABIC for the LCDM and multidimensional IRT models under varying levels of item quality, and base rate of attribute mastery. They found that all the indices performed well with strong-quality items but extremely poor with medium to low-quality items. Hu et al. (2016) implemented a similar study but focused more on the misspecification of the Q-matrix with the G-DINA model. Their study found that AIC and BIC are sensitive to over- and under-specification of the Q-matrix. Both studies recommend that model fit assessment should rely on multiple sources of evidence to select among non-nested models.

2.4.4 Response Time Models

The RT models are focused on describing the non-negative positively skewed distribution of RT. Of these RT models, the log-normal model by van der Linden (2006) is the most popular. This model can handle the skewness of RTs while allowing the benefits of the statistical properties for normal distribution for the log-transformed RT. Schnipke & Scrams (1997). Fox, Klein Entink, and van der Linden (2007) and Klein Entink, Fox, & van der Linden (2009a) extended the log-normal model to include a slope parameter for the person speed parameter that characterizes the differential effects of items on the speed of examinees. Klein Entink, van der Linden & Fox (2009b) also acknowledged additional limitation of the log-normal model in handling skewness of RT and proposed the Box-Cox normal model to provide more flexibility in characterizing RT data. However, van der Linden, Scrams, & Schnipke (1999) have demonstrated a good model fit for RT using the log-normal distribution. More details on this model follow in the next chapter.

Other models have also been used to characterize RT. These include the Weibull (Loeys et al., 2011), inverse Gaussian (Lo & Andrews, 2015), Gamma (Maris, 1993), Ex-Gaussian (Ratcliff & McKoon, 2008) and the shifted Wald (Anders, Alario, & Van Maanen, 2016). For an overview of these distributions, see De Boeck & Jeon (2019).

2.5 Response Time and Response Accuracy

Response time (RT), in educational testing, refers to the amount of time an examinee takes to provide a response (correct or incorrect) to a task (item or test). Prior to the advent of computers in educational testing, it was difficult to record RT on tests. Hence, research on RT in educational testing gained interest only recently. The spike in interest is predicated on its relationship with response accuracy (RA) and the need to understand examinee test-taking behaviors and the cognitive processes that lead to correct or incorrect responses. Ignoring these behaviors can lead to a violation of the local independence assumption of the popular IRT models and compromised test validity (Wang & Xu, 2015). For instance, an examinee may speed through the questions in a high stakes test to answer all the questions. Such rapid response may contaminate the estimate of examinee's true ability with construct irrelevant variance, posing a severe threat to score interpretation (Lu & Sireci, 2007).

Response time provides important person and item-level information that researchers can use to improve the design, administration, and quality control of a test. At the person level, RT provides insight on the working speed of the examinee and, on the item level, it gives information on the time intensity of the item (Zhan et al., 2018b). Rapid guessing in a high-stakes test could indicate test *speededness*, and rapid response in a low-stake test may suggest a lack of motivation (Lee & Chen, 2011). Either of these examinee behaviors introduces construct irrelevant variance and harms validity and score interpretation. Response time information on item time-intensities can be used to improve item calibration, selection and assembly in adaptive tests and educational tests in general (Kahraman et al., 2013; Lee & Chen, 2011; van der Linden, 2007).

Successful incorporation of RT into measurement requires an appropriate statistical model for their distribution, and several models have been proposed to this end. De Boeck & Jeon (2019) classified these models into four broad categories – distributional RT models (e.g., Maris, 1993; Loeys, Rosseel, & Baten, 2011; and van der Linden, 2006), joint models of RT with other dependent variables like accuracy (e.g., van der Linden, 2007; and Zhan et al., 2018a), local dependency models with RT as one of two or more correlated dependent variables (e.g., Partchev & De Boeck, 2012; Wang & Xu, 2015), and covariate models with RT as an explanatory variable (e.g., Sen, 2012; Naumann & Goldhammer, 2017). Each of these models differs in functional form, flexibility, and the assumption it makes about the response process. For a review of these other forms of RT models, see De Boeck & Jeon (2019) and Schnipke & Scrams (2002).

2.5.1 Joint Models of Response Time with Accuracy

As the name implies, these models take a multivariate approach to simultaneously analyze response and response time, usually to improve the parameter estimates of the response model. Several frameworks have been proposed in literature for this purpose. van der Linden (2007) proposed a hierarchical framework for modeling speed and accuracy on test items where the response times and responses are modeled separately at the first level. The dependency among their respective parameters is modeled at a higher level. This framework is one of the most popular tools to explain the relationship between response speed and accuracy and has been adapted to several combinations of response and response time models (e.g., Klein Entink et al., 2009a; Wang & Xu, 2015; Fox & Marianti, 2016).

2.6 Response Time in Cognitive Diagnostic Models

The need to account for cognitive process using response time has also been emphasized and addressed in cognitive diagnostic models. For instance, Zhan et al. (2018b) proposed a joint model for RT and the attributes in cognitive diagnosis. Their study integrated the lognormal model for RT and the DINA model for latent attributes using the hierarchical modeling framework of van der Linden (2007). The proposed joint model was assessed using simulated data and the PISA 2012 computer-based mathematics data. Their results showed that incorporating RT into the DINA model improved the precision of model parameters and the classification rates of attributes and profiles. Zhan et al. (2018b) extended the work by Zhan et al. (2018a) to a joint-testlet model to address the issue of paired local item dependence due to testlet effects from response and response times. This study used simulated data and the 2015 PISA computer-based mathematics data to demonstrate the utility and application of this extension.

The adaptive testing procedure is not left out. Huang (2019) explored a model for improving item calibration in cognitive diagnostic computerized adaptive testing (CD-CAT) with higher-order DINA. The study used a modified posterior-weighted Kullback-Leibler (PWKL) method that maximizes the item information per time unit and a shadow-test method that assembles a provisional test subject to a specified time constraint were developed. The results showed that the incorporation of RT is associated with a lower risk of running out of time while ensuring acceptable latent trait and speed parameter estimates.

The importance of cognitive diagnostic modeling to educational assessment is not contestable; neither is the relevance of response time in assessing the cognitive processes that underlie test responses. However, cognitive diagnostic modeling has been unnecessarily limited to the simplistic configuration of attributes into mastery and non-mastery, even when the available models can do much better. The works by Karelitz (2004) and Zhan et al. (2019a) have shown that there are greater possibilities with cognitive diagnosis. The significance of incorporating cognitive processing in skill diagnosis via response time has also been demonstrated, especially with item response modeling. However, current approaches for exploiting information from response time in cognitive diagnostic modeling have been unduly constrained to the assumption of constant speed. Relaxing these constraints for more informative and instructionally relevant skill diagnosis is the focus of the current study. The method used to achieve this is laid out in the next chapter.

CHAPTER 3: METHODOLOGY

3.1 The log-normal random quadratic variable speed model

Let T_{ik} denote the response time of person *j* (*j* = 1, ..., N) on item *i* (*i* = 1, ..., *I*) and assume that examinee *j* chooses his/her speed of response τ_j at the start of the test and maintains this speed throughout the test, van der Linden (2006) defines the log-normal distribution for RT of person *j* on item *i* as:

$$f\left(t_{ij};\tau_j,\sigma_i,\lambda_i\right)=\frac{\sigma_i}{t_{ij}\sqrt{2\pi}}exp\left\{-\frac{1}{2}\left[\sigma_i\left(lnt_{ij}-\left(\lambda_i-\tau_j\right)\right)\right]^2\right\} \qquad (4)$$

Which implies that the log of RT can be modeled as (van der Linden, 2016):

$$ln\left(t_{ij}\right)=\lambda_i-\tau_j+\varepsilon_{ij}; \qquad \varepsilon_{ij}\sim N(0,1/\sigma_i^2) \qquad (5)$$

Where $\tau_j\epsilon(-\infty,\infty)$ is the speed of the examinee *j* on the test, λ_i is the time intensity or time *consumingness* of item *i* and $\sigma_i\epsilon(0,\infty)$ is the discrimination parameter that captures the contribution of item *i* to the precision of the estimate of examinee's speed (van der Linden, 2016).

To be identified, the constraint $\sum_{j=1}^{n}\tau_j=0$ is imposed on the speed parameters. With the mean of the distribution being $\lambda_i-\tau_j$, this constraint equates the expected log RT over items and persons to the average item difficulty so that person speed parameter values (τ_j) are estimated as deviations from that average (van der Linden, 2006). An examinee with positive (negative) value is working faster (slower) than the average level in the population. Note that, while this model allows variance of log RT to be item dependent, it assumes examinee speed is constant across items, which may not always be true.

λ_i in the log-normal model represents the cognitive load of an item on a time scale. The higher the magnitude of λ_i, the more time-intensive item *i* is. An examinee working at a higher speed would complete the item with a lower response time. However, the change in response time due to change in speed may vary from item to item. To account for this possible variation, Klein Entink et al. (2009a) and Fox (2010) introduced an item discrimination parameter into the log-normal model. This extended the log-normal model to:

$$ln(t_{ij}) = \lambda_i - \phi_i \tau_j + \varepsilon_{ij}; \qquad \varepsilon_{ij} \sim N(0, \sigma_i^2) \qquad (6)$$

Where ϕ_i is the time discrimination of item i and σ_i^2 is the residual variance.

Building on the log-normal RT model of van der Linden (2006, 2016) and its extended version by Fox (2010) and Klein Entink et al. (2009a), Fox & Marianti (2016) proposed a log-normal random quadratic variable model for RTs to account for changes in examinee's speed across the items on a test. To do this, they defined items in a test as the measurement occasions, and the response time as the time between two subsequent items. In particular, $\boldsymbol{X}_{ij} = X_{1j}, X_{2j} \dots X_{Ij}$ is the time variable representing the measurement occasions of items 1 through I for examinee j where $X_{1j} = 0$ so that the speed from first item defines the intercept. The time variable is placed on an arbitrary 0 to 1 scale by defining $X_{ij} = \left(X_{(ij)} - 1\right)/I$ where $X_{(ij)}$ is the order in which item i is completed by examinee j and I is the number of items. By this definition, the time scale for this model is only meaningful if respondents are not allowed to take breaks between items or go back to review previous items. It is expected, given a typical testing situation, that measurement occasions would not be equidistant. To address this, they assume a testing situation where the total test time is small enough that the non-equidistance of time has little to no effect on the results.

With these assumptions in place, Fox & Marianti (2016) defined the log-normal RT model with a linear and quadratic trend for speed using the time variable X as:

$$ln(T_{ij}) = \lambda_i - \phi_i\left(\tau_{j0} + \tau_{j1}X_{ij} + \tau_{j2}X_{ij}^2\right) + \varepsilon_{ij};$$

$$\begin{pmatrix}\tau_{0j}\\ \tau_{1j}\\ \tau_{2j}\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\\0\end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix}\sigma^2_{\tau_0} & \sigma_{\tau_0\tau_1} & \sigma_{\tau_0\tau_2}\\ \sigma_{\tau_0\tau_1} & \sigma^2_{\tau_1} & \sigma_{\tau_1\tau_2}\\ \sigma_{\tau_0\tau_2} & \sigma_{\tau_1\tau_2} & \sigma^2_{\tau_2}\end{pmatrix}\right) \qquad (7)$$

Where λ_i is the usual time intensity of item *i*, τ_{0j} represents the initial value of speed, τ_{1j} is the random slope in speed and τ_{2j} is the random quadratic term to characterize the acceleration or deceleration in speed of examinee *j*. The initial speed as well as the random linear and quadratic slope terms are assumed to follow a normal distribution with mean vector of **0** and covariance matrix **Σ** such that the expected response time for the test is still the average time intensities.

3.2 Deterministic, Input, Noisy 'And' Gate (DINA) Model for Polytomous Attributes

There are numerous CDMs proposed in literature and DINA is one of the most popular choice because of its simplicity, parsimony and ease of interpretation (de la Torre, 2009; Zhang, 2015). The DINA model requires the mastery of all required attributes for an item to solve the item correctly. Suppose a test has been designed to assess examinees on K latent attributes where each attribute has L_k mastery levels ($L_k \geq 2$). Let α_{jk} be the mastery level of examinee *j* on attribute *k*. The lowest mastery level on each attribute is set to 0 so that $q_{ik} = l - 1$ if item *i* requires the l^{th} level of mastery on attribute *k* for a correct response, and $\alpha_{jk} = l - 1$ if examinee *j* has attained the l^{th} mastery level on attribute *k*.

For each item in the DINA model, an examinee is classified into one of two latent classes – those who have attained the required mastery level on the attributes, as required by the item, and those who have not attained the required mastery level on at least one of the attributes. Let η_{ij} denote this latent class variable for the j^{th} examinee on the i^{th} item; then, $\eta_{ij} = 1$ if examinee *j* is at or above the mastery level on the attributes as required by item *i* for a correct response and 0 otherwise (Zhan et al., 2019). This definition assigns the same probability of success to examinees who lack the required mastery level on at least one of the requisite attributes for an item. Given the examinee’s status on η_{ij}, the probability of a correct response for the DINA model is given by:

$$P(Y_{ij} = 1|\eta_{ij}) = (1 - s_i)^{\eta_{ij}} g_i^{(1-\eta_{ij})} \qquad (9)$$

Where Y_{ij} is the response of examinee *j* to item *i*; s_i is the *“slipping”* parameter or the probability that an examinee who has attained the necessary mastery levels on the required attributes for item *i* would answer the item incorrectly by mistake, and g_i is the *“guessing”* parameter or the probability that an examinee who falls short on at least one of the required attributes for an item would answer the item correctly by guessing or using alternative strategies that are not specified by the Q-matrix (de la Torre, 2009). These two parameters, assuming the Q matrix has been correctly specified, incorporate the noise (Karelitz, 2004) – the reasons why an examinee with $\eta_{ij} = 0$ could get an item right and an examinee with $\eta_{ij} = 1$ could answer the item incorrectly.

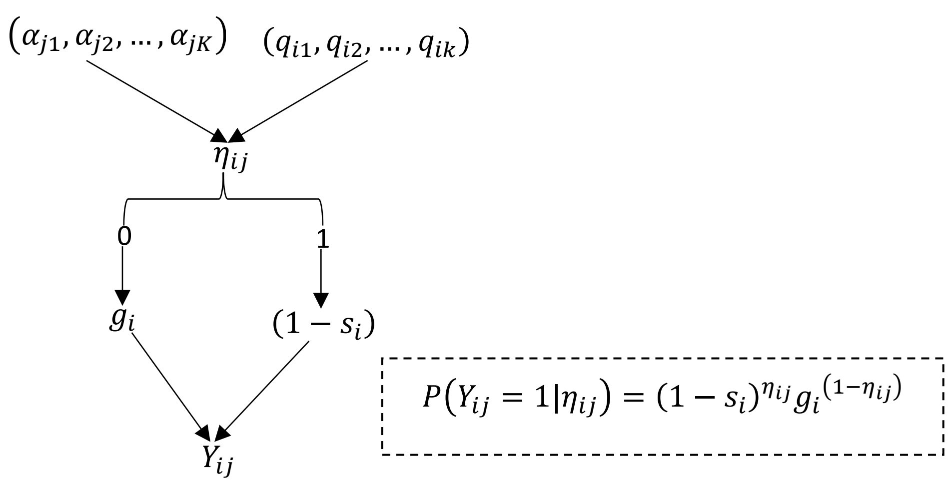


Figure 2 Diagrammatic representation of the DINA model

The two parameters of the DINA model are indexed by item but not attributes; which means the complexity of the DINA model stays the same irrespective of number of attributes considered in the Q-matrix, keeping the model parsimonious (Zhang, 2015). The probability of a correct response is $1 - s_i$ if an examinee has attained the requisite mastery levels on all the required attributes (i.e., $\eta_{ij} = 1$) and g_i otherwise (Henson, Templin, & Willse, 2009). The probability function of the DINA model is further constrained by the condition that $(1 - s_i) > g_i$ so that an examinee who has achieved all the required mastery levels would always have a higher probability of correct answer to the item than one who falls short on at least one of the required attributes. As de la Torre (2009) succinctly shows in Figure 2, the DINA model assigns two probability values to examinees, $(1 - s_i)$ for those that have mastered required attributes and g_i to those who lack the required mastery level on at least one attribute. Partial mastery (mastery below the required level or adequate mastery of a subset of the required attributes) is irrelevant to the probability of correct response. This feature, though restrictive, is the reason DINA model is considered easily estimable and interpretable.

The DINA model for binary attributes is a special case of the polytomous DINA, with L_k = 2 for all attributes. With binary attributes, we are concerned with mastery/non-mastery classification so that $q_{ik} = 1$ if item *i* requires attribute *k* for a correct response, and 0 otherwise and $\alpha_{ik} = 0$ if examinee *j* has mastered attribute *i* and 0 otherwise. Also, the latent variable $\eta_{ij} = 1$ if examinee *j* has mastered all the attributes required for a correct response on item *i*. With these modified definitions in place, equation (9) defines the conditional probability that examinee *j* (with attribute profile $\boldsymbol{\alpha}_j$) would provide a correct response to item *i*.

The two parameters of the DINA model are both probabilities and hence, bounded between 0 and 1. To relax these boundaries and ease parameter estimation, DeCarlo (2011) proposed a reparameterization of the DINA model where the slipping and guessing parameters are expressed as functions that yield positive values within the boundaries of 0 and 1. His proposition gave rise to the *reparameterized* DINA model or RDINA which is written as:

$$P(Y_{ij} = 1|\eta_{ij}) = \delta_{0_i} + \delta_{1_i}\eta_{ij} \qquad (10)$$

Where δ_{0_i} and δ_{1_i} are the intercept and interaction parameters, respectively; and η_{ij} is as defined in equation (9). δ_{1_i} is referred to as interaction because it reflects the difference between those who possess the required mastery levels ($\eta_{ij} = 1$) for an item and those who do not ($\eta_{ij} = 0$). The guessing and slip δ_{0_i} and δ_{1_i} are both logit functions of the guessing and slipping parameters and can be used to recover g_i and s_i with the following conversion formulas:

$$\delta_{0_i} = logit(g_i)$$

$$\delta_{1_i} = logit(1 - s_i) - logit(g_i)$$

3.3 The Joint Differential Speed DINA (JDS-DINA) Model

Data on response time (RT) provides an additional source of information besides the task responses on the test. Both sources of information result from the interaction of item and person characteristics. Task responses are determined by examinee ability (mastery) level and the item characteristics (e.g., difficulty and discrimination) and response times are determined by examinee speed parameters and item parameters (time intensities and discriminations). So far, two separate models have been defined for these two sources of information, but it is important to also model the relationships between the two models to exploit the benefits of response time information in the estimation of item and person parameters.

From van der Linden (2016), it is understood that the primary reason for the introduction of time discrimination parameter into the log-normal model was to achieve similarity between the log-normal model for time and the 2PL IRT model for item responses. This similarity in model specifications is not relevant to the objective of this study. He further argued that the additional parameter is unnecessary and could lead to overparameterization, since the variation in the effect of speed on response time is already captured by the residual variance parameter of the log-normal model such that $\phi_i = 1/\sigma_i^2$. As a result, this parameter is excluded from the log-normal random quadratic variable speed model so that the model becomes:

$$ln(T_{ij}) = \lambda_i - (\tau_{j0} + \tau_{j1}X_{ij} + \tau_{j2}X_{ij}^2) + \varepsilon_{ij}; \qquad (11)$$

$$\varepsilon_{ij} \sim N(0, \phi_i) \; and \; \begin{pmatrix} \tau_{0j} \\ \tau_{1j} \\ \tau_{2j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_{speed} = \begin{pmatrix} \sigma_{\tau_0}^2 & \sigma_{\tau_0\tau_1} & \sigma_{\tau_0\tau_2} \\ \sigma_{\tau_0\tau_1} & \sigma_{\tau_1}^2 & \sigma_{\tau_1\tau_2} \\ \sigma_{\tau_0\tau_2} & \sigma_{\tau_1\tau_2} & \sigma_{\tau_2}^2 \end{pmatrix} \right)$$

To account for possible correlation or hierarchical structure in the skills, the higher order latent trait approach was used, where it is assumed that a continuously valued general ability variable θ_j, underlies the attributes, such that attributes of examinee *j* are independent conditional on θ_j. Conventionally, θ_j is assumed to follow a standard normal distribution for model identification (Zhan et al., 2019a). The adjacent category logit model of (Zhan et al., 2019a), is used to probability of mastery to the underlying latent trait. This model defines the relationship between θ_j and the polytomous attributes as:

$$P_{jkl} = P(\alpha_{jk} = l - 1|\theta_j) = \frac{exp\left(\sum_{u=1}^{l} \gamma_{1_k}\theta_j - \gamma_{0_{ku}}\right)}{\sum_{v=1}^{L_k} exp\left(\sum_{u=1}^{v} \gamma_{1_k}\theta_j - \gamma_{0_{ku}}\right)} \qquad (12)$$

Where $P_{jkl}$is the probability that examinee *j* attains mastery level *l* on attribute *k;* θ_j is the continuously valued latent variable for examinee *j*, γ_{1_k} is the slope parameter for attribute *k* and $\gamma_{0_{ku}}$ is the intercept or location parameter for the *l*th level of attribute *k*, with $\gamma_{0_{k1}} \equiv 0$ (Zhan et al., 2019a). If θ_j truly underlies the latent attributes then, given θ_j, the P_{jkl} are conditionally independent so that the probability of an attribute profile *c* (*c* = 1 … C) for examinee *j* is a product of the probabilities for the corresponding levels on individual attributes:

$$\pi_{jc} = P(\boldsymbol{\alpha_j} = \boldsymbol{\alpha}_c|\theta_j) = \prod_{k=1}^{K}\prod_{l=1}^{L_k} {P_{jkl}}^{I(\alpha_{ck}=l-1)} \qquad (13)$$

Where C is the number of permissible attribute profiles on the test; π_{jc} is the probability that examinee *j* has attribute profile *c*, α_{ck} is the entry in the C by K matrix of permissible profiles; and $I(\alpha_{ck} = l - 1) = 1$ if the k^{th} attribute in $\boldsymbol{\alpha}_c$ is at the l^{th} level and 0 otherwise.

When $L_k = 2$ for all the attributes on a test, equation (12) reduces to:

$$P_{jk} = P(\alpha_{jk} = 1|\theta_j) = \frac{exp(\gamma_{1_k}\theta_j - \gamma_{0_k})}{1 + exp(\gamma_{1_k}\theta_j - \gamma_{0_k})} \qquad (14)$$

for the relationship between the binary attributes and the higher order latent trait, θ_j; and the probability of a response pattern in equation (13) becomes:

$$\pi_{jc} = P(\boldsymbol{\alpha}_j = \boldsymbol{\alpha}_c|\theta_j) = \prod_{k=1}^{K} {P_{jk}}^{\boldsymbol{\alpha}_{ck}}(1 - P_{jk})^{1-\boldsymbol{\alpha}_{ck}} \qquad (15)$$

As noted previously, the higher-order approach to modeling correlation among attributes provides us with summative information on the underlying latent ability for each examinee, in addition to the mastery level for each attribute. In addition, it reduces the number of parameters required to estimate the mastery levels from $(\prod_{k=1}^{K} L_k) - 1$ to $K + \sum_{k=1}^{K}(L_k - 1)$ and thereby reduces the complexity of the estimation process. For instance, a test with 3 attributes, each measured at 3 mastery levels, would require 26 structural parameters but the higher order model reduces that number to only 9. Figure 3 depicts an example of higher order model with three attributes measured by a five-binary-item test. Here, the first attribute (Att1) is measured at three mastery levels by all five items, the second attribute at two levels by the last four items and the third attribute at two levels by the last two items. The horizontal lines indicate latent thresholds representing marginal percentage correct for items or marginal percent mastery at each level for attributes (Rupp et al., 2010).

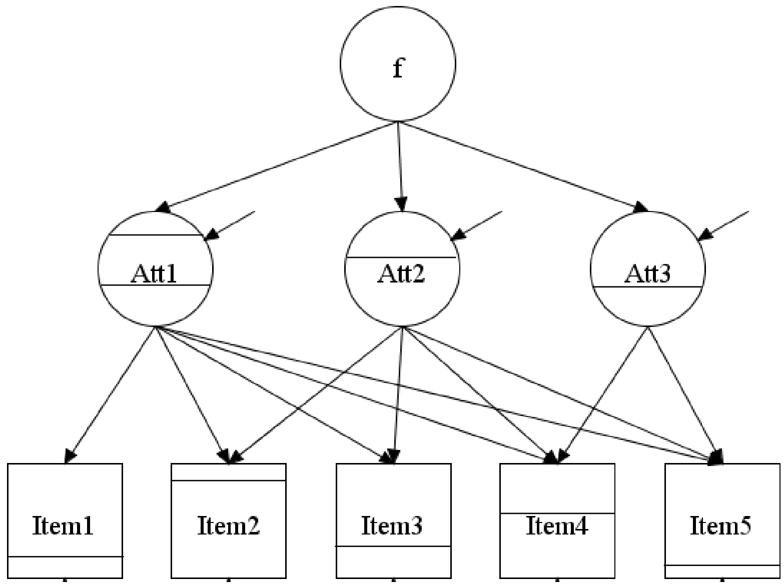


Figure 3 Higher order polytomous attributes model

The unequal positioning of the horizontal lines is to indicate that thresholds can vary across items and latent attributes. When the higher order model depicted in Figure 3 is incorporated for the structural parameters of the DINA model, we obtain the higher-order DINA or HO-DINA.

Following the hierarchical modelling framework of (van der Linden, 2007), equations (10) through (13) make up the separate models for response time and item response at the first level. At the second level, two variance-covariance structures are defined to model the dependencies among item parameters and person parameters, respectively. To define the joint model, some local independence assumptions are required:

a. Given the person speed parameters, the log response times $ln(T_{ij})$ are conditionally independent
b. Given the latent ability, θ_j,the latent attributes, α_{jk}, are conditionally independent

c. The responses, Y_{ij}, are conditionally independent, given the latent examinees attribute profile, $\boldsymbol{\alpha}_j$

d. Given the ability and speed parameters, the responses, Y_{ij} and log response times, $ln(T_{ij})$ are conditionally independent.

With these local independence assumptions in place, the joint differential speed DINA model for polytomous item is defined as:

Level 1:

a. Measurement part:

$$ln(T_{ij}) = \lambda_i - \left(\tau_{0_j} + \tau_{1_j}X_{ij} + \tau_{2_j}X_{ij}^2\right) + \varepsilon_{ij};\ \varepsilon_{ij} \sim N(0, 1/\sigma_i^2)$$

$$P(Y_{ij} = 1|\eta_{ij}) = \delta_{0_i} + \delta_{1_i}\eta_{ij}$$

b. Structural part:

$$P_{jkl} = P(\alpha_{jk} = l - 1|\theta_j) = \frac{exp\left(\sum_{u=1}^{l}\gamma_{1_k}\theta_j - \gamma_{0_{ku}}\right)}{\sum_{v=1}^{L_k} exp\left(\sum_{u=1}^{v}\gamma_{1_k}\theta_j - \gamma_{0_{ku}}\right)}$$

$$\pi_{jc} = P(\boldsymbol{\alpha_j} = \boldsymbol{\alpha}_c|\theta_j) = \prod_{k=1}^{K}\prod_{l=1}^{L_k} {P_{jkl}}^{I(\boldsymbol{\alpha}_{ck} = l-1)}$$

Level 2:

a. Person parameters:

$$\begin{pmatrix} \theta_j \\ \tau_{0_j} \\ \tau_{1_j} \\ \tau_{2_j} \end{pmatrix} \sim N\left(\boldsymbol{\mu}_{person}, \boldsymbol{\Sigma}_{person} = \begin{pmatrix} 1 & \sigma_{\theta\tau_0} & \sigma_{\theta\tau_1} & \sigma_{\theta\tau_2} \\ \sigma_{\theta\tau_0} & \sigma_{\tau_0}^2 & \sigma_{\tau_0\tau_1} & \sigma_{\tau_0\tau_2} \\ \sigma_{\theta\tau_1} & \sigma_{\tau_0\tau_1} & \sigma_{\tau_1}^2 & \sigma_{\tau_1\tau_2} \\ \sigma_{\theta\tau_2} & \sigma_{\tau_0\tau_2} & \sigma_{\tau_1\tau_2} & \sigma_{\tau_2}^2 \end{pmatrix}\right)$$

b. Item parameters:

$$\begin{pmatrix} \lambda_i \\ \delta_{0_i} \\ \delta_{1_i} \end{pmatrix} \sim N\left(\boldsymbol{\mu}_{item}, \boldsymbol{\Sigma}_{item} = \begin{pmatrix} \sigma_{\lambda}^2 & \sigma_{\lambda\delta_0} & \sigma_{\lambda\delta_1} \\ \sigma_{\lambda\delta_0} & \sigma_{\delta_0}^2 & \sigma_{\delta_0\delta_1} \\ \sigma_{\lambda\delta_1} & \sigma_{\delta_0\delta_1} & \sigma_{\delta_1}^2 \end{pmatrix} \right) \qquad (16)$$

Where the person parameters are assumed to follow a multivariate normal with mean vector $\boldsymbol{\mu}_{person}$ and variance-covariance matrix. The items are also assumed to follow a tri-variate normal distribution with mean $\boldsymbol{\mu}_{item}$ and variance-covariance matrix $\boldsymbol{\Sigma}_{item}$. Other terms in the model are as previously defined in equations (9) through (13).

3.3.1 Model Specifications

Rupp et al. (2010) have shown that the probability of an observed response pattern in cognitive diagnostic model, much like the latent class analysis model, is a structural equation model with a structural component and a measurement component. Incorporation of response time therefore entails an extension of the measurement model for relating observed response times to latent speed variables as well as additional structural relationships between latent speed factors to the latent attributes.

The person and item covariance matrices capture the relationship among person and item parameters, respectively. The residual error variance, $\sigma_i^2$is assumed to be independently distributed and therefore, not modeled with the item parameters at the second level. For the identifiability between θ_j and λ_i, the mean and variance of θ_j are set to 0 and 1 respectively, which also follows from the higher order latent trait model of equation (12). To identify the scale of the speed parameters, the mean vector of the speed parameters is fixed to **0**, which means that the average time intensity parameter represents the population average time it takes to complete the item – when the speed parameters are all zeros.

As with regular growth parameters, some form of association is expected among the growth parameters. Fox & Marianti (2016) suggested a negative relationship, such that examinees with high initial speed tend to slow down towards the end of the test while those with low initial speed tend to increase their speed later to finish the test within the time limit. However, for model identifiability, they restricted the covariances among the growth parameters to zeros.

The higher order latent trait of de la Torre & Douglas (2004) is based on the idea that student with higher values on the underlying latent ability should have a higher probability of attaining a higher mastery level on an attribute. To ensure that this is the case, the slope parameter is constrained to be positive, $\gamma_{1_k} > 0$. Finally, the *guessing* parameter is constrained as $g_i < (1 - s_i)$ to guarantee that an examinee who has achieved all the required mastery levels would always have a higher probability of correct answer to the item than one who falls short on at least one of the required attributes.

3.3.2 Parameter Estimation

Estimation of the JDS-DINA model can be achieved through the fully Bayesian estimation with MCMC procedure for higher-order DINA with polytomous items, as proposed by (Zhan et al., 2019). In the present study, parameters of interest include item intercept and interaction parameters, time intercept, interaction, time intensity and item time discrimination parameters ($\delta_{0_i}, \delta_{1_i} \lambda_i, \sigma_i^2$) and person attribute and speed parameters ($\alpha_j, \tau_{0_j}, \tau_{1_j}, \tau_{2_j}$). Given the local independence assumptions *a* through *d* of the JDS-DINA above and a random sample from the population of examinees, the joint likelihood of the observed response and response times is given as:

$$L(\boldsymbol{Y}, \boldsymbol{ln}(\boldsymbol{T})|\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta_0}, \boldsymbol{\delta_1}, \boldsymbol{\tau_0}, \boldsymbol{\tau_1}, \boldsymbol{\tau_2}, \boldsymbol{\sigma^2})$$

$$= \prod_{j=1}^{n} \prod_{i=1}^{I} P\left(Y_{ij}|\alpha_j, \delta_{0_i}, \delta_{1_i}\right) f\left(ln\left(T_{ij}\right)|\lambda_i, \tau_{0_j}, \tau_{1_j}, \tau_{2_j}, \sigma_i^2\right) \qquad (17)$$

Where

$$f\left(ln\left(T_{ij}\right)|\lambda_i, \tau_{0_j}, \tau_{1_j}, \tau_{2_j}, \sigma_i^2\right)$$

$$= \frac{\sigma_i}{t_{ij}\sqrt{2\pi}} exp\left\{-\frac{1}{2}\left[\sigma_i\left(lnt_{ij} - \left(\lambda_i - \tau_{0_j} - \tau_{1_j} - \tau_{2_j}\right)\right)\right]^2\right\}$$

3.3.2.1 Prior distributions

The posterior distribution of the parameter space is proportional to the product of the likelihood in (17) and all the prior distributions of the parameters. The posterior distributions are then derived by drawing samples from the prior distributions and updating the likelihood of the observed response times and responses. Hence, the choice of prior distributions is important for ensure model convergence. Given the relationship among item parameters and person parameters of the model, the following prior distributions are adopted for person and item parameters prior distributions for the JDS-DINA.

The prior for response time residual variance, σ_i^2 is chosen as $InvGamma(1,1)$. Following Gelman, Carlin, Stern, Dunson, Vehtari, & Rubin (2013), the joint prior distribution for the person parameters is set as multivariate normal

$$\begin{pmatrix} \theta_j \\ \tau_{0_j} \\ \tau_{1_j} \\ \tau_{2_j} \end{pmatrix} \sim N\left(\boldsymbol{\mu}_{person} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_{person} = \begin{pmatrix} 1 & \sigma_{\theta\tau_0} & \sigma_{\theta\tau_1} & \sigma_{\theta\tau_2} \\ \sigma_{\theta\tau_0} & \sigma_{\tau_0}^2 & 0 & 0 \\ \sigma_{\theta\tau_1} & 0 & \sigma_{\tau_1}^2 & 0 \\ \sigma_{\theta\tau_2} & 0 & 0 & \sigma_{\tau_2}^2 \end{pmatrix}\right)$$

The normality assumption and mean vector of $\mathbf{0}$ for the person parameters follow from the identifiability conditions of the higher order latent trait model of (de la Torre & Douglas,

2004) and Fox & Marianti (2016)'s variable speed quadratic model for response time. Going by the model identifiability constraints, the variance covariance matrix of the growth parameters is tridimensional matrix with zeros on the off diagonals. Hence, the only covariance terms of interest are those of ability with each of the growth parameters.

Completely specifying the covariance matrix entries would be unrealistic. Rather, the covariance matrix is defined by hyper-priors with hyperparameters. Since some of the entries of $\boldsymbol{\Sigma}_{person}$ are fixed, the inverse-Wishart distribution is not applicable (Zhan et al., 2019). Following the example of (Zhan et al., 2018a), $\boldsymbol{\Sigma}_{person}$ is first re-parameterized in terms of its Cholesky decomposition and priors are then placed on the entries of the resulting lower triangular matrix. The Cholesky decomposition of $\boldsymbol{\Sigma}_{person}$ is given as:

$$\boldsymbol{\Sigma}_{person} = \Delta_{person}\Delta'_{person} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \sigma_{\theta\tau_0} & \omega_0 & 0 & 0 \\ \sigma_{\theta\tau_1} & \varphi_{01} & \omega_1 & 0 \\ \sigma_{\theta\tau_2} & \varphi_{02} & \varphi_{12} & \omega_2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ \sigma_{\theta\tau_0} & \omega_0 & 0 & 0 \\ \sigma_{\theta\tau_1} & \varphi_{01} & \omega_1 & 0 \\ \sigma_{\theta\tau_2} & \varphi_{02} & \varphi_{12} & \omega_2 \end{pmatrix}'$$

Where

$$\varphi_{01} = \frac{-\sigma_{\theta\tau_0}\sigma_{\theta\tau_1}}{\sqrt{\sigma^2_{\tau_0} - \sigma^2_{\theta\tau_0}}};\ \varphi_{02} = \frac{-\sigma_{\theta\tau_0}\sigma_{\theta\tau_2}}{\sqrt{\sigma^2_{\tau_0} - \sigma^2_{\theta\tau_0}}};\ \varphi_{12} = \frac{-\sigma_{\theta\tau_1}\sigma_{\theta\tau_2}\sigma^2_{\tau_0}}{\sqrt{\left(\sigma^2_{\tau_0} - \sigma^2_{\theta\tau_0}\right)\left(\sigma^2_{\tau_1}\sigma^2_{\tau_0} - \sigma^2_{\tau_1}\sigma^2_{\theta\tau_0} - \sigma^2_{\theta\tau_1}\sigma^2_{\tau_0}\right)}}$$

$$\omega_0 = \sqrt{\sigma^2_{\tau_0} - \sigma^2_{\theta\tau_0}};\ \omega_1 = \sqrt{\sigma^2_{\tau_1} - \sigma^2_{\theta\tau_1} - \varphi^2_{01}};\ \text{and}\ \omega_2 = \sqrt{\sigma^2_{\tau_2} - \sigma^2_{\theta\tau_2} - \varphi^2_{02} - \varphi^2_{12}}$$

The priors for the elements of Δ_{person} are set such that φ_{01}, φ_{02}, and φ_{12} are each assumed to follow $N(0,1)$ while ω_0, ω_1, and ω_2 are each $Gamma(1,1)$ (Zhan et al., 2018a). Similarly, following Zhan et al. (2018a), the joint prior distribution adopted for item parameters is the multivariate normal distribution:

$$\begin{pmatrix} \lambda_i \\ \delta_{0i} \\ \delta_{1i} \end{pmatrix} \sim N\left(\begin{bmatrix} \mu_\lambda \\ \mu_{\delta_0} \\ \mu_{\delta_1} \end{bmatrix}, \boldsymbol{\Sigma}_{item}\right)$$

Again, the parameters for this tri-variate normal distribution were drawn from the following hyper-priors:

$$\mu_{\lambda} \sim N(3, 2)$$

$$\mu_{\delta_0} \sim N(-2.197, 2)$$

$$\mu_{\delta_1} \sim N(4.394, 2) I(\mu_{\delta_1} > 0)$$

$$\mathbf{\Sigma}_{item} \sim invWishart(\boldsymbol{R}, 3)$$

where **R** is a tridimensional identity matrix and $I(\mu_{\delta_1} > 0)$ puts a truncation on μ_{δ_1} to satisfy the DINA model constraint of $g_i < (1 - s_i)$. The parameters of these priors are also drawn from (Zhan et al., 2018a).

For the higher-order latent structural model, the structural parameters are assumed to be independently distributed so that their individual priors specified as:

$$\gamma_1 \sim N(0, \sigma_{\gamma_1}^2) I(\gamma_1 > 0)$$

$$\gamma_{0k} \sim N(0, \sigma_{\gamma_0}^2)$$

Where $I(\gamma_1 > 0)$ indicates that the distribution is truncated from below at zero, to align with the belief that higher level of the trait is associated with higher probability of possessing a higher mastery level on attribute *k*. Also, for the outcome variables in the model: $Y_{ij} \sim Bernoulli\left(P(Y_{ij} = 1 | \eta_{ij})\right)$ since it is a binary outcome and $ln(T_{ij}) \sim N\left(\left(\lambda_i - \tau_{0_j} - \tau_{1_j} - \tau_{2_j}\right), 1/\sigma_i^2\right)$. The latent attribute variable is a categorical variable and therefore, follows a categorical distribution: $\alpha_{jk} \sim Categorical(\boldsymbol{P}_{jk})$ where $\boldsymbol{P}_{jk}$ is a 1 by L_k vector of probabilities for the L_k levels of attribute *k*.

Applying the Bayes' theorem to the likelihood and priors defined above, the joint posterior probability distribution for the JDS-DINA model is given as:

$$P\big(\boldsymbol{\Omega}|\boldsymbol{Y}, ln(\boldsymbol{T})\big) \propto L(\boldsymbol{Y}, \boldsymbol{ln}(\boldsymbol{T})|\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta_0}, \boldsymbol{\delta_1}, \boldsymbol{\tau_0}, \boldsymbol{\tau_1}, \boldsymbol{\tau_2}, \boldsymbol{\sigma^2}) \times P(\boldsymbol{\alpha}|\boldsymbol{\Omega}, \boldsymbol{\gamma_0}, \boldsymbol{\gamma_1}) \times P(\boldsymbol{\gamma_0}) \times P(\boldsymbol{\gamma_1})$$

$$\times P\big(\boldsymbol{\Omega}, \boldsymbol{\tau_0}, \boldsymbol{\tau_1}, \boldsymbol{\tau_2}|0, \boldsymbol{\Sigma}_{person}\big) \times P\big(\boldsymbol{\Sigma}_{person}\big) \times P(\boldsymbol{\lambda}, \boldsymbol{\delta_0}, \boldsymbol{\delta_1}|, \boldsymbol{\Sigma}_{item}) \times P(\boldsymbol{\Sigma}_{item})$$

$$\times P(\boldsymbol{\sigma^2})$$

Where $\boldsymbol{\Omega}$ is the set of all the parameters to be estimated in the model.

Given the complex nature of the joint posterior defined above, it would be impossible to sample directly from it. Hence, one of the MCMC algorithm described previously was employed. Once convergence is achieved, with stationary posterior distributions for all the parameters, the mode of the posterior distribution is treated as estimate for α_{jk}. For the rest of the parameters, the means of their posterior distributions was used.

3.3.3 Assessment of Model Fit

Beyond the usual model checks in the Bayesian estimation framework, additional model assessment specific to joint cognitive diagnostic and response time models have not been adequately studied. Hence, given convergence model checks are satisfied, further evaluation of model fit for the JDS-DINA model would follow the example of Zhan et al. (2018a). Model fit was evaluated separately for the response and response time outcomes.

For the responses, Zhan et al. (2018a) implemented the posterior predictive model check of (Gelman et al., 2014) using the sum of the squared Pearson residuals (Yan, Mislevy, & Almond, 2003) to assess the fit for item responses. This discrepancy measure is defined in (Zhan et al., 2018a) as:

$$D\big(Y_{ij}; \alpha_j, \delta_{0i}, \delta_{1i}\big) = \sum_{j=1}^{n}\sum_{i=1}^{I}\left(\frac{Y_{ij} - P_{ij}}{\sqrt{P_{ij}\big(1 - P_{ij}\big)}}\right)^2 \qquad (18)$$

Where $P_{ij} = P\left(Y_{ij} = 1|\eta_{ij}\right)$ is as defined in equations (9) and (15). Values close to 0.5 are indicative of adequate model fit (Zhan et al., 2018a).

Similarly, the fit for response times is assessed using the sum of the standardized error function of $ln\left(T_{ij}\right)$, appropriately modified from (Zhan et al., 2018a) to reflect the variable speed quadratic model for response time. This modified statistic is defined as:

$$D\left(ln\left(T_{ij}\right); \lambda_i, \tau_{0_j}, \tau_{1_j}, \tau_{2_j}, \sigma_i^2\right) = \sum_{j=1}^{n}\sum_{i=1}^{I}\left(\frac{ln\left(T_{ij}\right) - E\left(ln\left(T_{ij}\right)\right)}{\sigma_i}\right)^2 \qquad (19)$$

Where $E\left(ln\left(T_{ij}\right)\right) = \left(\lambda_i - \tau_{0_j} - \tau_{1_j} - \tau_{2_j}\right)$ is the conditional mean of $ln\left(T_{ij}\right)$. Values close to 0.5 are also indicative of adequate model fit for the response times (Zhan et al., 2018a).

3.4 Real data analysis

For this section of the study, two datasets are available to address the study objectives that are related to real data applications. One is the 2012 computer-based PISA mathematics data and the other is dataset created and kindly shared by Karelitz (2004).

The 2012 computer-based PISA mathematics test was developed to assess domain-specific knowledge and skills in mathematics of students from 65 EU, OECD and OECD-partner countries. Test areas include Mathematics, Reading, Science, and financial literacy OECD (2014). However, only the computer-based Math test provides the response time data that is relevant for this study. To keep this study comparable to that of Zhan et al. (2018a), only the data for 1,584 students from Brazil (BRA), Germany (DEU), Shanghai-China (QCN), and the United States of America (USA) were used. Also, for the same reason, only ten of the Mathematics test questions were considered. These 10 questions were designed to assess seven attributes – change and relationships (α_1), quantity (α_2), space and shape (α_3), uncertainty and data (α_4), occupational (α_5), societal (α_6), and scientific (α_7) (OECD, 2014; Zhan et al., 2018a). For more details, see OECD (2014).

The second dataset is a 40-item test designed and administered to 200 University of Illinois undergraduates by Karelitz (2004), henceforth referred to as the *language rule* data. The test was designed to assess their general language proficiency using fictional grammatical rules. The rules tested were grouped into three skills (attributes) with three to four mastery levels each. Participants were first randomly assigned to 10 groups and taught rules, where groups differ by the type of rule they were taught. Thereafter, questions testing all rules were developed and administered to all participants. Responses were then scored as right or wrong. More details on this experiment can be found in Karelitz (2004).

For model comparisons, the HO-DINA differs from the JDS-DINA and JRT-DINA because it excludes response time. However, an additional difference is that, in both JDS-DINA and JRT-DINA, the item parameters are also modeled at the second level to account for inter-relationships among them. Since the objective of this study is to highlight the difference that response time makes, it is important that every comparison model should only differ in terms of how the response time variable is handled. To ensure this is so for comparisons that are made with HO-DINA, a modified HO-DINA (MHO-DINA) introduced by Zhan et al. (2018a) was used. MHO-DINA is essentially HO-DINA that includes a higher-level model for the relationship among item parameters. The MHO-DINA model is defined as follow:

Level 1:

a. *Measurement part:* $P\left(Y_{ij}=1|\eta_{ij}\right)=\delta_{0_i}+\delta_{1_i}\eta_{ij}$

b. *Structural part:* $P_{jkl}=P\left(\alpha_{jk}=l-1|\theta_j\right)=\frac{exp\left(\sum_{u=1}^{l}\gamma_{1k}\theta_j-\gamma_{0_{ku}}\right)}{\sum_{v=1}^{L_k}exp\left(\sum_{u=1}^{v}\gamma_{1k}\theta_j-\gamma_{0_{ku}}\right)}$

$$\pi_{jc} = P(\boldsymbol{\alpha_j} = \boldsymbol{\alpha}_c | \theta_j) = \prod_{k=1}^{K} \prod_{l=1}^{L_k} P_{jkl}{}^{I(\boldsymbol{\alpha}_{ck}=l-1)}$$

Level 2:

c. *Person parameters:* $\theta_j \sim N(0,1)$

d. *Item parameters:* $\begin{pmatrix} \delta_{0_i} \\ \delta_{1_i} \end{pmatrix} \sim N\left(\boldsymbol{\mu}_{item}, \boldsymbol{\Sigma}_{item} = \begin{pmatrix} \sigma_{\delta_0}^2 & \sigma_{\delta_0\delta_1} \\ \sigma_{\delta_0\delta_1} & \sigma_{\delta_1}^2 \end{pmatrix}\right)$ (20)

3.4.1 Model Comparisons

This section describes how the models, estimation method, and datasets discussed so far are employed to address the research objectives of this study that are related to real data analysis.

The availability of response time together with item responses in the PISA data makes it appropriate for comparison of models that differ in terms of response time. Unfortunately, the Q-matrix for this test defines only binary attributes. Hence, the first two research questions are restricted to comparisons involving the JDS-DINA model for binary attributes.

On the other hand, the data from Karelitz (2004) concerns attributes with qualitatively ordered mastery levels, which would have been more appropriate for the application of JDS-DINA model for polytomous attributes. However, this data lacks information on response time. Hence, the fourth research question in this section is a comparison between polytomous and binary attribute specifications using the MHO-DINA model.

3.4.1.1 Research Question 1: How do the JDS-DINA, JRT-DINA, and MHO-DINA models compare in terms of model fit?

This comparison would serve to verify the advantage of a differential speed response time as opposed to constant speed model. All three models, JRT-DINA and MHO-DINA and JDS-DINA with binary attributes, were compared based on model fit statistics – the deviance information criterion (DIC), and the posterior predictive probability (PPP) defined in equation (18). The standard deviation of the posterior distributions would also be used to assess and compare precision across the three models of interest.

It is expected that the MHO-DINA would present the worst performance among the three models. The difference between the JRT-DINA and the JDS-DINA would depend largely on what structure of response time model is most appropriate for the data at hand. If there is no significant change in speed among the examinees, then the JRT-DINA may present relatively smaller standard errors because of its parsimony. The reverse would apply if there is a significant change in speed of response among the examinees.

3.4.1.2 Research Question 2: How does dichotomization of polytomous attributes affect person correct classification accuracy?

The aim of this question is to verify, in the absence of response time information, that dichotomization of attributes leads to poorer model results, as suggested by previous studies (Karelitz, 2004; Karelitz, 2008; Zhan et al., 2019; Chen & de la Torre, 2013). This comparison would employ the *language rule* data by Karelitz (2004). To create the binary attributes from the Q-matrix of this data, the lowest mastery level for each attribute was coded *non-mastery* (0) and all mastery levels beyond the lowest level were coded as *mastery* (i.e., 1).

The appeal of Karelitz (2004)'s data is that, even though it is empirical, the truth about examinees mastery levels is known. Hence, besides comparison of model fit indices, as in research question 1a, the models with polytomous and binary attributes would also be compared in terms of classification accuracies of attributes and attribute patterns. These two quantities are defined as follows (Zhan et al., 2019; Chen & de la Torre, 2013):

$$ACCR_k = \frac{\sum_{j=1}^{n} W_{jk}}{n} \qquad (21)$$

$$PCCR_c = \frac{\sum_{r=1}^{R}\sum_{j=1}^{n}\prod_{k=1}^{K} W_{jk}}{n} \qquad (22)$$

Where n is the number of examinees or sample size, R is the number of replications for the pattern of interest, and $W_{jk} = 1$ if $\alpha_{jk} = \hat{\sigma}_{jk}$ and 0 otherwise. For polytomous attributes, misclassification could be of varying degrees – classification into adjacent or non-adjacent mastery levels. To account for the degree of misclassification, Chen & de la Torre (2013) recommend the use of weighted classification accuracy, where $W_{jk} = 1/2^{|\alpha_{jk} - \hat{\sigma}_{jk}|}$ if $|\alpha_{jk} - \hat{\sigma}_{jk}| < L_k - 1$ and 0 otherwise.

3.5 The Simulation Study

For this section of the study, data were generated according to the JDS-DINA model for polytomous attributes and analyzed with true model and the alternative models. Alternative models to be considered are the JRT-DINA and the MHO-DINA of Zhan et al. (2018a). The JRT-DINA accounts for response time in the model but assumes constant speed across the test period. The MHO-DINA excludes response time completely, ignoring the effect of time. The reason for using MHO-DINA instead of HO-DINA is to ensure that the item parameters are modeled similarly across all models so that differences in model performances across these models can only be attributed to either the grain size of the attributes, the treatment of response time in the model, or both. Each of these models was considered with polytomous and binary items to further assess the effect of dichotomizing attributes with respect to each model. In summary, data were generated with one model, but six models were estimated and compared using the simulated data.

3.5.1 Simulation design

In examining and comparing cognitive diagnostic models, previous researchers have shown significant variation in model performance due to number of items or test length, number of examinees or sample size, misspecification of the Q-matrix, number of attributes, among others. However, for the purpose of this study, these factors were held constant, to keep the study design manageable. Sample size of 200 was considered. For the Q matrix, number of attributes, and number of items, the specification provided in Zhan et al. (2019) was adopted. This means the current study considered a 30-item test measuring four attributes with four mastery levels each and a Q matrix as defined in Zhan et al. (2019) and shown below.

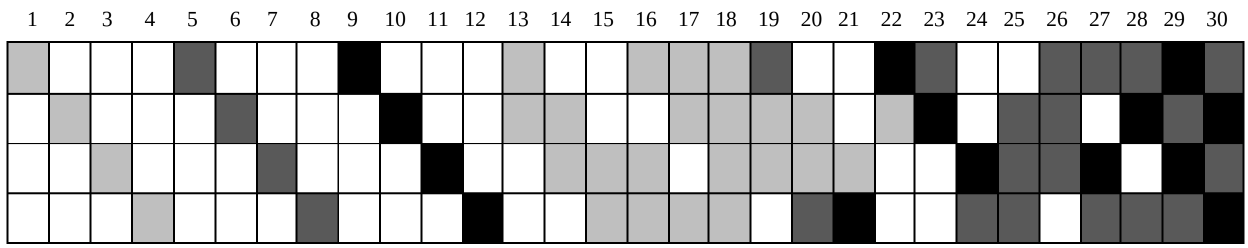


Figure 4 K × *I* Q matrix for binary attributes.
Blank means “0,” light gray means “1”, dark gray means “2” and black means “3”

To dichotomize the attributes, the Q matrix in Figure 4 is revised such that all levels above the first are categorized as mastery (1) and the first level is classified as non-mastery (0). This gives rise to the Q matrix in Figure 5 below for the models with binary attributes.

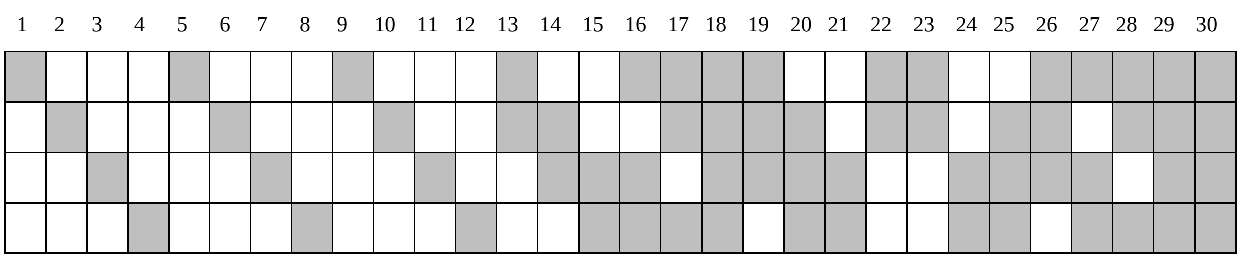


Figure 5 K × *I* Q matrix for binary attributes. Blank means “0,” light gray means “1”

The factors to be manipulated in the study are variances of speed components, and correlation between examinee ability and the speed components. This is to evaluate the performance of the JDS-DINA model under varying conditions of assumed variability in speed. If speed remains constant throughout the test then, only the variance of the initial speed component would be significantly different from zero. The variances of the speed components and their covariances with the ability parameter were considered at two levels each.

There are only a few studies that have considered the variable speed model in combination with item response models. In most cases, the item response model used is an IRT model. However, estimates from these studies provide suggestions on plausible values for the correlation between ability and the speed components. Zhan et al. (2018a), working with the JRT-DINA model, found a theoretically contradictory negative correlation of -0.57 between the initial speed component and ability. Their model did not consider the variable speed components and was based on a 10-item test for seven attributes, which is not likely to produce trustworthy estimates. Fox & Marianti (2016), on the other hand, reported a positive correlation of 0.72 between initial speed and ability and negative correlations of -0.02 and -0.09 for the linear and quadratic components, respectively.

Following the estimates reported in Fox & Marianti (2016), the parameters for the speed components are set as 0.1 and 0.5 for low and high variances respectively for each of the speed components. Also, the correlation parameters are set to (0.3, -0.05, -0.1) and (0.7, -0.1, -0.3) as low and high correlations between ability and the initial speed, linear trend, and quadratic components, respectively. Table 3 below shows these values.

Table 3

Design conditions – person parameters

Level	$\boldsymbol{\sigma^2_{\tau_0}}$	$\boldsymbol{\sigma^2_{\tau_1}}$	$\boldsymbol{\sigma^2_{\tau_0}}$	$\boldsymbol{\sigma_{\theta\tau_0}}$	$\boldsymbol{\sigma_{\theta\tau_1}}$	$\boldsymbol{\sigma_{\theta\tau_2}}$
Low	0.1	0.1	0.1	0.3	-0.05	-0.1
High	0.5	0.5	0.5	0.7	-0.1	-0.3

The covariances among the parameters were determined by the correlations and variances specified above. Unlike the person parameters, the variance of time residuals is fixed at 0.25

and the true values of the higher-order latent trait model parameter, γ_{1_k} and $\gamma_{0_{kl}}$ were specified like in Zhan et al. (2019), as shown in Table 4 below:

Table 4

Design conditions – structural parameters

Attribute	$\boldsymbol{\gamma_{1_k}}$	$\boldsymbol{\gamma_{0_{kl}}}$		
		l=2	*l=3*	*l=4*
a_1	1.5	-1.00	-0.50	0.00
a_2	1.5	-0.50	-0.25	0.25
a_3	1.5	-0.25	0.25	0.50
a_4	1.5	0.00	0.50	1.00

The values for the item parameters were motivated from results of real data analysis reported by Zhan et al. (2018a) and set to

$$\boldsymbol{\Sigma}_{item} = \begin{array}{c} \\ \lambda \\ \delta_0 \\ \delta_1 \end{array} \begin{array}{c} \begin{array}{ccc} \lambda & \delta_0 & \delta_1 \end{array} \\ \begin{pmatrix} 0.24 & -0.44 & 0.25 \\ -0.44 & 3.86 & -2.45 \\ 0.25 & -2.45 & 2.50 \end{pmatrix} \end{array}$$

$$\boldsymbol{\mu}_{item} = \begin{pmatrix} 4.30 \\ -2.31 \\ 3.25 \end{pmatrix}$$

The choices of variances and covariance values make a total of 64 simulation conditions – two variances each of the speed components by two covariances each of speed components with ability. For each condition, 60 samples were generated.

3.5.2 Data Generation

To generate the data, person and item parameters were randomly sampled from the following multivariate normal distributions:

$$\begin{pmatrix}\theta_j \\ \tau_{0_j} \\ \tau_{1_j} \\ \tau_{2_j}\end{pmatrix} \sim N\left(\begin{bmatrix}0\\0\\0\\0\end{bmatrix}, \boldsymbol{\Sigma}_{person} = \begin{pmatrix}1 & \sigma_{\theta\tau_0} & \sigma_{\theta\tau_1} & \sigma_{\theta\tau_2} \\ \sigma_{\theta\tau_0} & \sigma_{\tau_0}^2 & 0 & 0 \\ \sigma_{\theta\tau_1} & 0 & \sigma_{\tau_1}^2 & 0 \\ \sigma_{\theta\tau_2} & 0 & 0 & \sigma_{\tau_2}^2\end{pmatrix}\right)$$

$$\begin{pmatrix}\lambda_i \\ \delta_{0_i} \\ \delta_{1_i}\end{pmatrix} \sim N\left(\begin{bmatrix}4.30\\-2.31\\3.25\end{bmatrix}, \boldsymbol{\Sigma}_{item} = \begin{pmatrix}0.24 & -0.44 & 0.25 \\ -0.44 & 2.86 & -2.45 \\ 0.25 & -2.45 & 2.5\end{pmatrix}\right)$$

Where the unknown entries of $\boldsymbol{\Sigma}_{person}$ are replaced by the values specified in Table 3, and higher-order latent trait model parameters are drawn from Table 4.

Once the parameters have been generated, the attribute mastery level on each attribute for each examinee, α_{jk}, would be generated from a categorical distribution with probabilities $(P_{jk1}, P_{jk2}, P_{jk3}, P_{jk4})$, where P_{jkl} is as defined in equation (12). The log response time would then be randomly drawn according to equation (11), and the binary response variable is generated with equation (10), where η_{ik} is determined from the examinee mastery status α_{jk} and the Q matrix in Figure 3.

Data generation and model estimations were carried out using JAGS, automated within R (Plummer, 2012). JAGS implements the Bayesian MCMC estimation procedure using an adaptive sampling scheme. In other words, JAGS searches through the catalog of samplers and chooses the sampling algorithm most appropriate for the conditional posterior distribution for each parameter (Plummer, 2012). For each of the 60 replications, two Markov chains were generated to improve the precision of parameter estimates (Brooks & Gelman, 1998), with 10,000 iterations per chain. Based on inspection of the trace plot, burn-in was set at 5,000. Random starting values were used for all model parameters. Model convergence was assessed using trace plots and the Gelman–Rubin potential scale reduction factor $\hat{R}$, where $\hat{R} < 1.2$ indicates approximate convergence (Brooks & Gelman, 1998).

3.5.3 Model Evaluation

To evaluate parameter recovery, the absolute bias (AB) and the root mean square error (RMSE) were computed. These two quantities are defined as follows:

$$AB(\hat{\nu}) = |\hat{\nu}_r - \nu| \qquad (23)$$

$$RMSE(\hat{\nu}) = \sqrt{\frac{\sum_{r=1}^{R}(\hat{\nu}_r - \nu)^2}{R}} \qquad (24)$$

Where R is the number of replications (60 in this study), ν is the true value of the parameter of interest and $\hat{\nu}_r$ is its estimated value at the r^{th} replication. The commonly used relative bias was not used in this study because some of the parameters have zeros as true value.

To evaluate item parameter recovery, the bias was averaged across all items; thus the mean absolute bias was reported to avoid the cancellation of positive and negative bias. For the classification accuracy, this study calculated and compared the attribute correct classification rate (ACCR) and the pattern correct classification rate (PCCR), as defined in equations 21 and 22.

3.5.3.1 Research Question 3: How is the recovery of item and person parameter estimates in JDS-DINA affected by variance of speed components and their correlations with person ability?

This part of the analysis was done using multivariate analysis of variance (MANOVA) model with eight factors. The factors to be considered are (1) variance of each initial speed component with two levels each, (2) variance of each linear trend component with two levels, (3) variance of each quadratic speed component with two levels, (4) covariance between initial speed components and ability with two levels, (5) covariance between linear trend component and ability with two levels, and (6) covariance between quadratic speed components and ability with two levels.

The MANOVA was carried out separately for each parameter. Outcome variables for each parameter were the AB and standard errors of estimates. Significant effects from the MANOVA results were probed further with univariate ANOVA and graphically.

3.5.3.2 Research Question 4: How well does the JDS-DINA for dichotomous attributes recover person and item parameter estimates (as reflected in the bias of estimates)?

For this research question, the bias, RMSE, ACCR, and PCCR were compared across all three models but with only the binary configuration of the attributes. The aim of this is to see which of these models gets the closest to the truth in the presence of information loss due to categorization of attributes. Only the model estimates with the binary attribute configuration were compared.

The comparison between JDS-DINA and JRT-DINA should reveal the effect of ignoring differential *speededness* on the accuracy of parameter estimates. The comparison between MHO-DINA with JDS-DINA and JRT-DINA would highlight the effect of ignoring time in CDM estimation.

3.5.3.3 Research Question 5: How well does the JDS-DINA for polytomous attributes recover person and item parameter estimates (as reflected in the bias of estimates)?

This last question is like the previous, but with the polytomous configuration of the attributes. All three models were compared as well. The idea is to determine how much loss, if at all, is incurred when we fail to account for the *speededness* effect in modeling data that came from a population of students with differential test speed. The model with the minimum average bias and RMSE and the maximum ACCR and PCCR is the preferred model.

CHAPTER 4: RESULTS

This chapter presents the results of the analyses, organized into five sections corresponding to the five research questions posited in Chapter 1 and described in Chapter 3. The first section summarizes the results of empirical data analysis that uses the PISA data to determine the best-fitting model, while the second section investigates the effect imposing a binary-attribute model on a polytomous attribute data using the higher other DINA model. Sections 3 through 5 are based on the simulation study, beginning with model evaluation, to assess how well the proposed differential speed DINA model for polytomous attributes, JDSP for short, recovers parameter estimates and attribute profiles under varying data conditions. Sections 4 and 5 are concerned with model comparisons, to examine the effect of ignoring the *speededness* effect as well as the effect of wrong specification of attribute categories, as implied by the model choice.

4.1 Research Question 1

In this first study, data from the 2012 computer-based PISA mathematics test was used to compare the three models of interest in this study – the modified higher-order (MHO) DINA, the joint response time (JRT) DINA models and the proposed joint differential speed (JDS) DINA. The adequacy of these three models for the PISA Mathematics data was assessed via relative model fit statistics as well as standard error of estimates. The following two subsections summarize the results of this study.

4.1.1 Model fit statistics

The DIC and BIC were used to assess the relative adequacy of the three models. The model with the smallest DIC and BIC is preferred. Posterior predictive probability (PPP) was also used to assess the model fit for responses (PPP-Score) and response times (PPP-Time). PPP values range from zero to one, where values close to 0 or 1 mean that observed discrepancies are extreme values and are suggestive of model-data misfit (Almond, Mislevy, Steinberg, and Williamson, 2015). The model with PPP value closest to 0.5 is preferred.

Table 5

Model fit statistics for the 2012 PISA computer-based mathematics test

Parameter	**MHO-DINA**	**JRT-DINA**	**JDS-DINA**
DIC	67812.87	52301.180	53142.52
BIC	416972.2	150292.3	156290.4
PPP-Score	0.544	0.580	0.599
PPP-Time	---	0.591	0.608
Posterior SD			
δ_0	0.306	0.265	0.234
δ_1	0.369	0.324	0.304
λ	1.578	0.018	0.017
θ	0.339	0.629	0.630

Table 5 presents the summary information about overall model fit statistics. The deviance-based statistics, DIC and BIC, point to the JRT-DINA model as preferred, but the PPP value chooses the MHO-DINA. However, standard error of posterior distribution for the item and person parameters are relatively high with MHO-DINA model, suggesting high instability in these estimates. The JDS-DINA model, on the other hand, shows greater stability in parameter estimates.

The small item pool in this test may have favored the parsimony of the MHO-DINA, resulting in the relatively good PPP-Score. Every other fit statistic rejects MHO-DINA in favor of the models that account for response time. Almond et al. (2015) also noted that the PPP value can be too conservative, failing to reject model-data misfits. Standard errors of parameter estimates are similar for the item time intensity and higher-order ability estimates, but not for the item intercept and slope parameters. These values were obtained by aggregating across the ten items. The next section takes a closer look at these items, to evaluate model fit with respect to individual items.

4.1.2 Standard error of item parameter estimates

Given the true values of these item parameters are unknown, it is impossible to tell which of these models provide the true estimates for these items but, the standard deviations provide some information on the reliability of these estimates. The item level estimates in Table 6 show similarities in the parameter estimates provided by the three models, particularly between the JRT and JDS model. Overall, the JDS-DINA provides the smallest standard errors for the item parameter estimates.

Table 6

Estimated item parameters for the 2012 PISA computer-based mathematics items

		δ_0			δ_1	
Item	**MHO**	**JRT**	**JDS**	**MHO**	**JRT**	**JDS**
Item1	-1.102(0.831)	-0.614(0.267)	-.583(0.225)	4.400(0.883)	3.823(0.435)	3.894(.457)
Item2	-5.912(0.952)	-5.592(0.703)	-5.446(0.225)	5.933(0.963)	5.567(0.703)	5.432(.656)
Item3	-4.001(0.418)	-4.269(0.652)	-4.089(0.648)	5.059(0.448)	5.292(0.658)	5.108(.494)
Item4	-3.400(0.209)	-3.439(0.220)	-3.452(0.221)	3.224(0.235)	3.166(0.246)	3.165(.251)
Item5	-0.586(0.067)	-0.601(0.069)	-.606(0.068)	2.029(0.186)	1.924(0.175)	1.919(.180)
Item6	-2.225(0.133)	-2.355(0.150)	-2.377(.153)	3.088(0.195)	3.177(0.201)	3.188(.200)
Item7	-0.913(0.076)	-0.957(0.078)	-.959(0.080)	2.344(0.176)	2.326(0.174)	2.311(.173)
Item8	0.410(0.067)	0.375(0.071)	.375(0.071)	0.895(0.174)	0.853(0.157)	.863(.161)
Item9	-1.934(0.123)	-2.217(0.172)	-2.205(.158)	2.068(0.179)	2.297(0.207)	2.302(.199)
Item10	-2.497(0.192)	-2.812(0.265)	-2.766(.243)	3.041(0.250)	3.115(0.283)	3.077(.265)

The results of this study clearly exclude the MHO-DINA as a plausible model for the PISA Mathematics data. This suggests that, subject to the sample of models considered here, models that account for response time provide a better fit for this data. The relative model fit statistics favor the choice of JRT DINA, but item level assessment suggests that the JDS provides better local fit for the items on the test. The results of this study may have been limited by the number of items, relative to the number of attributes – ten to seven. The effect of this limitation in item pool may vary across these models and affect their results differently.

4.2 Research Question 2

The aim of this question is to verify that, in the absence of response time information, imposing a dichotomous-attribute model on data obtained from polytomous attributes leads to poorer model results, as suggested by previous studies (Karelitz, 2004; Karelitz, 2008; Zhan et al., 2019; Chen & de la Torre, 2013). This comparison used the language rule data by Karelitz (2004), to compare two models – the higher-order DINA (HO-DINA) model and the reparametrized partial mastery higher-order DINA (RPa-DINA) model of Zhan et al. (2019). HO-DINA fits a binary-attribute model while the RPa-DINA models the ordered categories of the attribute using the adjacent category logit model for the structural parameters.

4.2.1 Comparison of model fit

Table 7

Model fit statistics for the Language Rule data

Parameter	HO-DINA	RPa-DINA
DIC	8176.217	7753.088
Deviance	7839.987	7020.388
BIC	9621.282	10902.46
PPP Score	0.497	0.665

The deviance and DIC statistics selected the RPa-DINA over the HO-DINA model. The BIC and PPP, however, did not favor the RPa-DINA model. It has been noted earlier that the PPP value can be conservative in rejecting wrong models. Also, for Bayesian estimation of cognitive diagnostic models, the DIC is preferred over the BIC (personal communication with Dr. Peida Zhan). Hence, judging from the DIC and the deviance statistics, the RPa-DINA model is preferred to the HO-DINA. In other words, imposing a binary-attribute model on polytomous-attribute data leads to poorer model fit. The next subsection further examines the effect of this model-data mismatch on the correct classification rates.

4.2.2 Classification accuracies

The language rule data comes with true mastery level data for all 200 participants, which means that the attribute correct classification rate (ACCR) and person correct classification rates (PCCR) can be computed and compared for these models. Tables 8 shows the relevant results for this comparison.

Table 8

Classification accuracy rates of attributes for the language data

Model	Exact				Weighted			
	ACCR			PCCR	ACCR			PCCR
	α_1	α_3	α_3		α_1	α_3	α_3	
HO-DINA	0.020	0.185	0.675	0.000	0.426	0.507	0.838	0.169
RPa-DINA	0.660	0.400	0.660	0.205	0.889	0.789	0.830	0.579

The exact classification rate considers the exact match between estimated and true mastery status, ignoring the degree of adjacency in the mismatch. The weighted classification rate formula adjusts for the degree of adjacency between estimates and true values and is therefore recommended for models with categorical attributes. As expected, the exact classification rates are generally low for both models. From the weighted classification rates, we see additional evidence in favor of the RPa-DINA model, especially with the first two attributes.

The PCCRs are generally low, though relatively higher with the RPA-DINA model. The aim of this study was to assess the effect of dichotomizing polytomous attributes by imposing a binary-attribute model on data. The DINA model has been used for this purpose, but there are many other options that could have been considered. These low values may stem from the fact that the assumptions of the DINA model may not have aligned well with the data, to begin with. Nonetheless, the results of this section show that, keeping the base model constant (DINA), dichotomizing polytomous attributes leads to a considerable loss in model fit and accuracy of parameter estimates.

4.3 Research Question 3

In this section, a simulation study was conducted to evaluate the parameter recovery of the proposed model, the joint differential speed (JDS) DINA, and to assess the effect of select design conditions on the model. The independent variables manipulated for this simulation were the variance of each of the speed components (two levels each) and the correlation of each speed component with the higher-order ability, θ, also at two levels each. See Table 3 for the specific values of these manipulated factors. Data were simulated using the JDS model with polytomous configuration for the attributes, where each attribute had four category levels.

4.3.1 Overall parameter recovery of the JDS-DINA model

Table 9

Bias and RMSE of item and structural parameters of JDS DINA with polytomous attributes

Attributes	Bias				RMSE			
	B_k	d_{kl}			B_k	d_{kl}		
		l=2	l=3	l=4		l=2	l=3	l=4
A1	0.119	0.100	-0.054	-0.091	0.390	0.578	0.565	0.545
A2	0.164	0.020	-0.025	-0.038	0.417	0.502	0.558	0.502
A3	0.138	0.053	-0.028	-0.033	0.406	0.470	0.541	0.529
A4	0.071	0.021	0.048	-0.113	0.365	0.420	0.476	0.539
Item parameters								
$\boldsymbol{\lambda}$			0.000				0.069	
$\boldsymbol{\delta_0}$			-0.004				0.689	
$\boldsymbol{\delta_1}$			-0.001				0.786	

Table 9 presents the bias and RMSE of item and structural parameters, averaged across all simulation conditions. In terms of bias and RMSE, the recovery of item and structural parameters was quite good, with similar recovery of structural parameters across attributes. ACCR ranged from 85% to 89%. The bias and RMSE values are similar to those reported in previous studies for the same data (e.g., Zhan et al., 2018a)

The weighted classification accuracies are generally low, but Figures 6 provides an explanation for these low values – small sample size. The proportions within the bars represent the average recovery rate for that profile group, while the numbers above the bars represent the average sample size (across all conditions) for that profile group. Profiles with higher number of test-takers are associated with higher recovery rates. A sample size of 200 test-takers implies that some of the 256 profiles that result from the four attributes would be empty. Larger sample size is required and, since profiles would be generated and assigned at random, sample size should be sufficiently greater than 256 to ensure that every profile group is assigned, at least, one test taker.

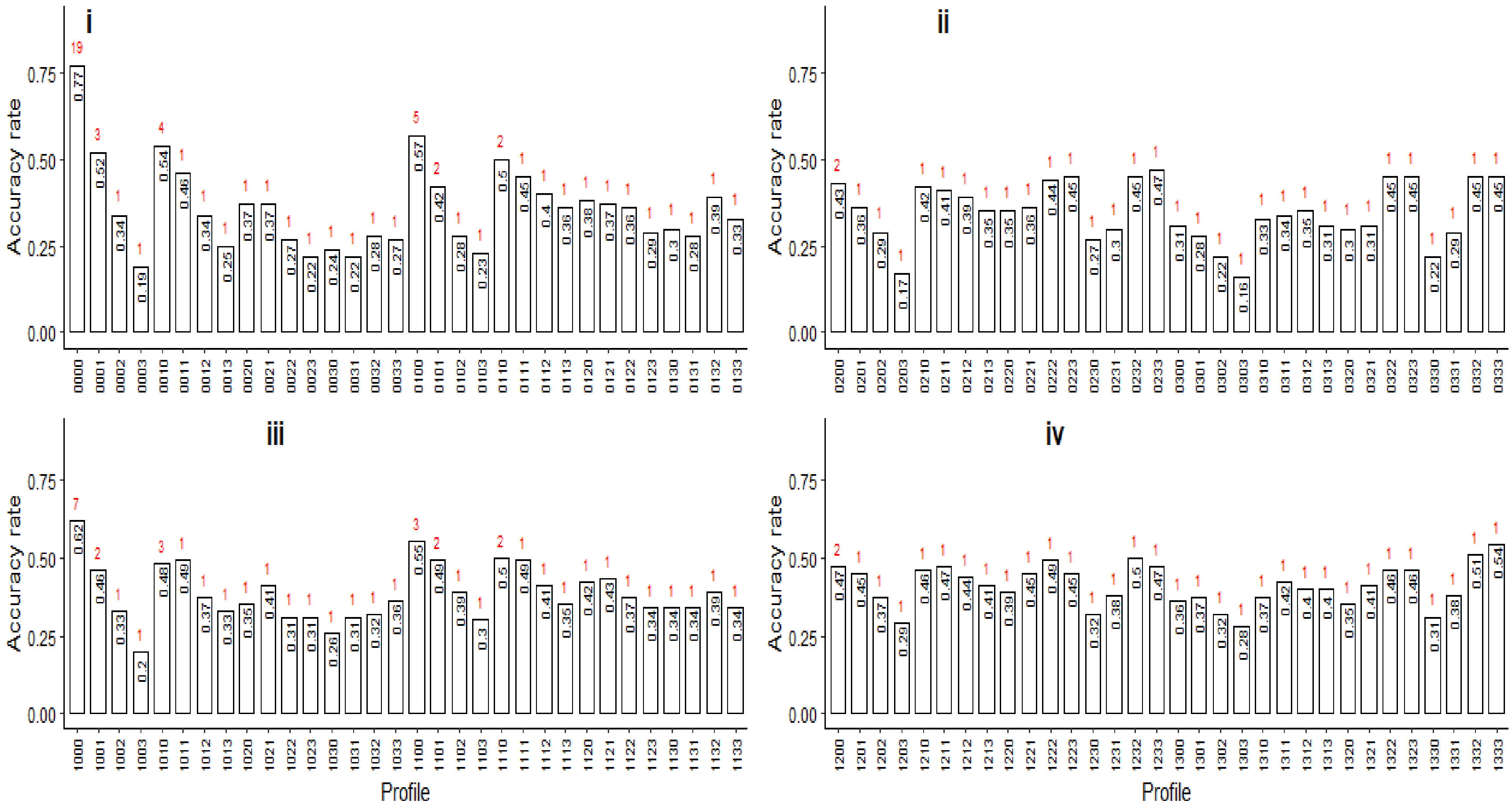
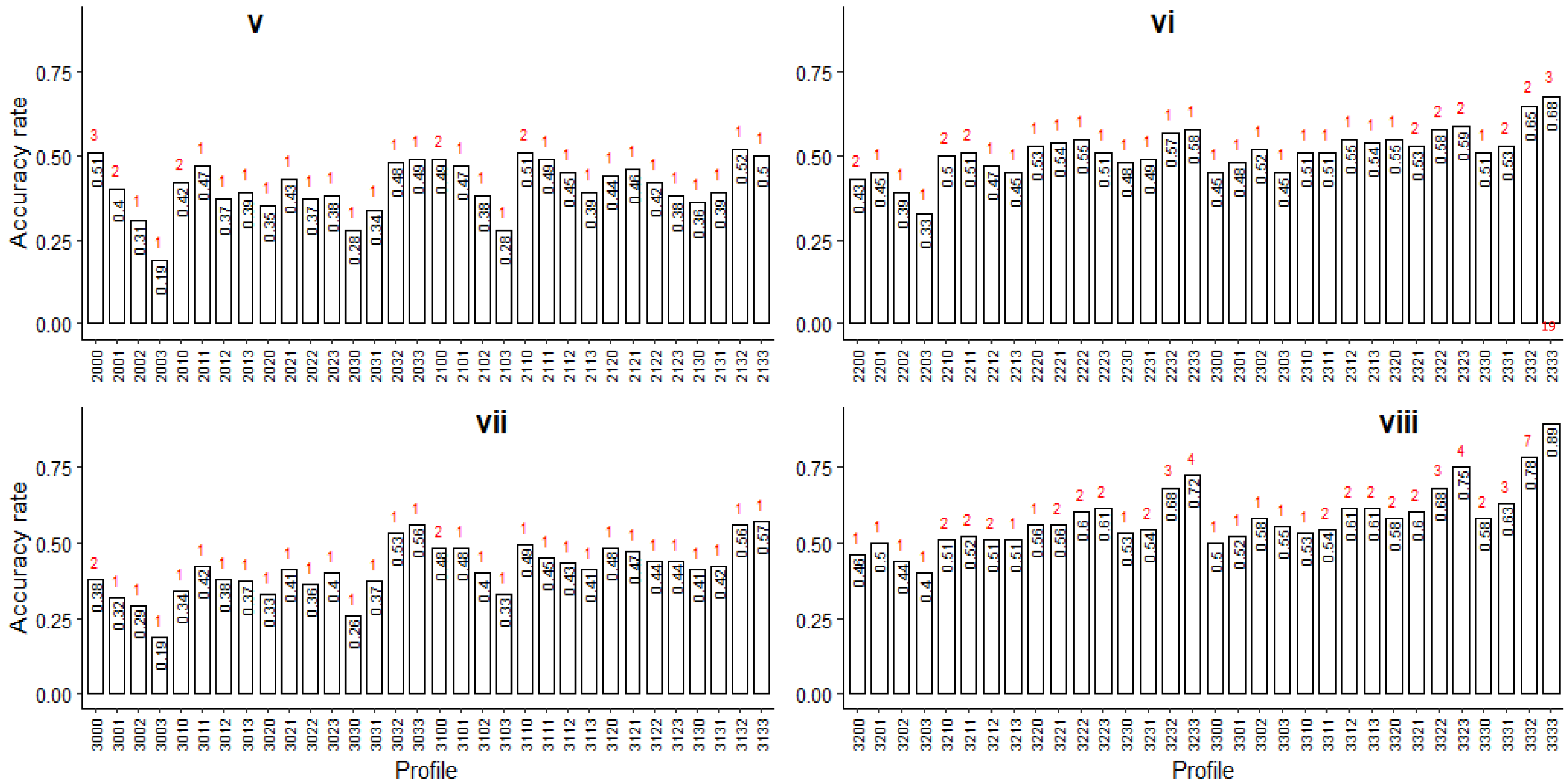


Figure 6 Profile correct classification rates

Figure 6 (cont'd)



4.3.2 *Effect of design conditions on parameter recovery*

To assess the effect of simulation design variables on parameter recovery for the JDS-DINA model, multivariate analysis of variance (MANOVA) was used to analyze the absolute bias and standard error for each of the parameters. For results that were significant, graphical analysis and univariate analysis of variance (ANOVA) were used to determine which of the variables was significantly affected. Tables 10 and 11 summarize the MANOVA and ANOVA results for the item and structural parameters, respectively. Test statistics and p-values are reported for the MANOVA, but only p-values are reported for ANOVA.

Table 10

MANOVA and ANOVA results for item parameters

Model	**Parameter**	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_1}$	$\rho_{\theta\tau_2}$
				λ			
MANOVA	Pillai	0.693	0.209	0.075	0.001	<0.001	0.001
	p-value	<0.001	<0.001	<0.001	0.331	0.760	0.304
ANOVA	Bias	<0.001	<0.001	0.001	0.320	0.712	0.416
(*p*-values)	SD	<0.001	<0.001	<0.001	0.272	0.518	0.187
				δ_0			
MANOVA	Pillai	<0.001	0.001	0.001	<0.001	<0.001	<0.001
	p-value	0.459	0.363	0.185	0.446	0.553	0.613
				δ_1			
MANOVA	Pillai	0.001	0.001	0.001	<0.001	<0.001	<0.001
	p-value	0.207	0.308	0.321	0.515	0.618	0.567

From the MANOVA results in Table 10, none of the correlation variables had effect on the item parameters. The variance of speed components had effect on the recovery of the item time intensity parameter, but not on the item slope and the intercept. The follow-up ANOVA result shows that the variance components affect both the bias and standard error of λ. These results were further examined graphically, in Figures 7 through 9.

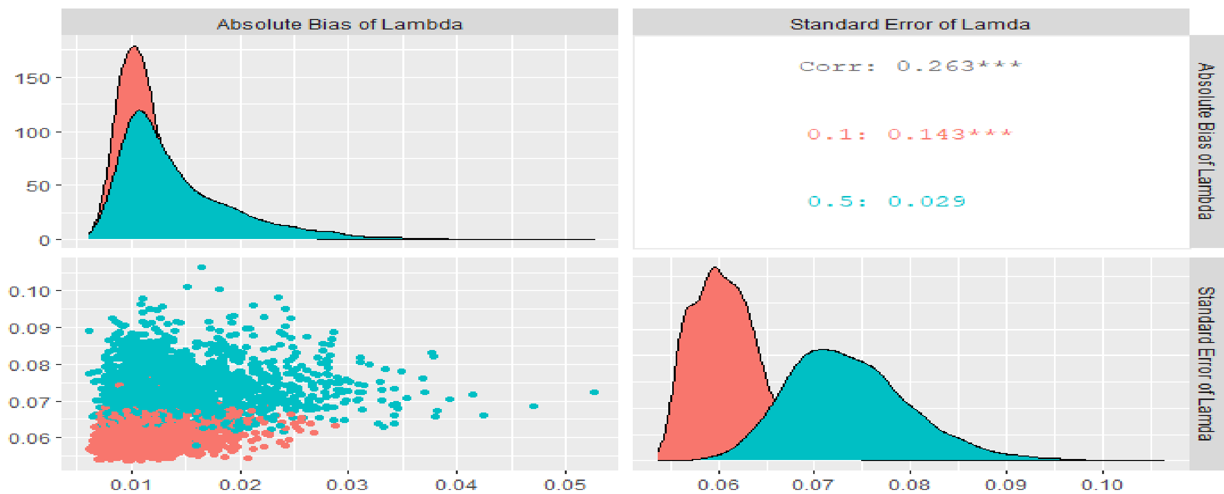


Figure 7 Effect of $\sigma^2_{\tau_0}$ on absolute bias and standard error of λ

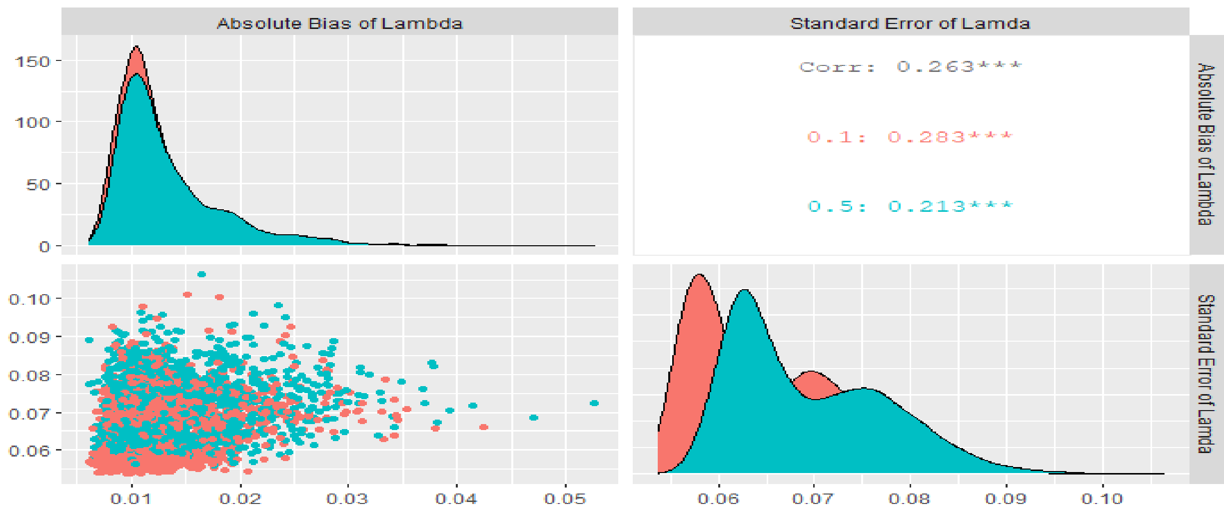


Figure 8 Effect of $\sigma^2_{\tau_1}$ on absolute bias and standard error of λ

A review of the graphs shows that the effect of variance of speed components is greater on the standard error of λ, and that low values of bias and standard error of λ were obtained when variance of the variances of speed components were smaller. The graphs also show that the variance of the initial speed component τ_0 has the strongest effect on parameter recovery. This implies that the JDS-DINA performs better when the variability in speed components is low.

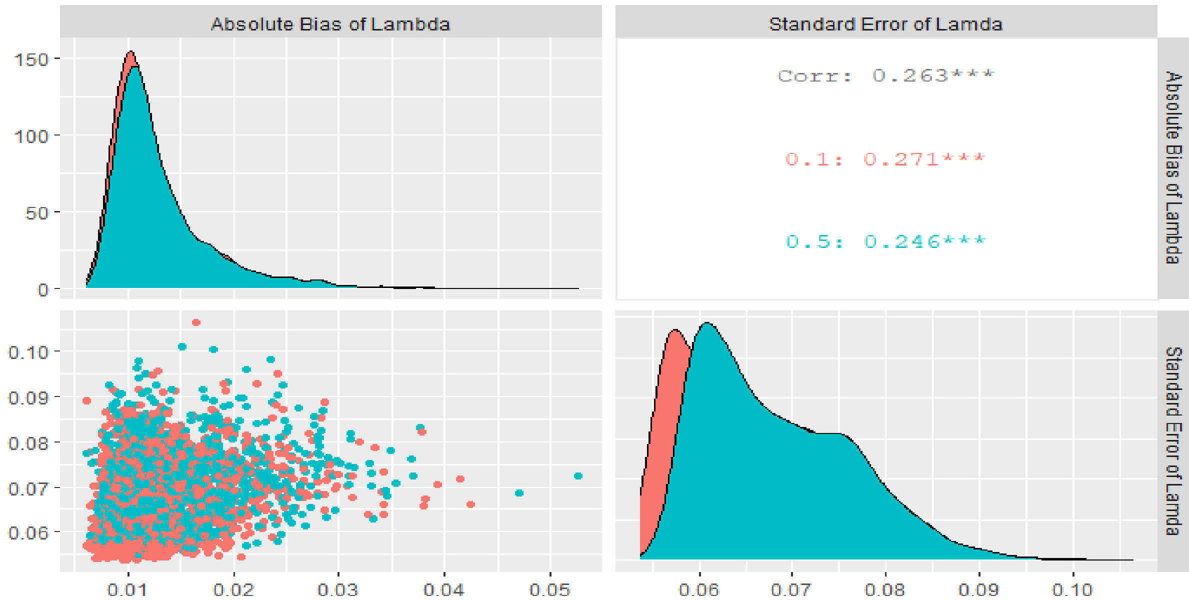


Figure 9 Effect of $\sigma^2_{\tau_2}$ on absolute bias and standard error of λ

Table 11

MANOVA and ANOVA results for attribute structural parameters

Model	**Parameter**	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_1}$	$\rho_{\theta\tau_2}$
				γ_{0k2}			
MANOVA	Pillai	<0.001	<0.001	0.001	0.004	0.001	0.001
	p-value	0.663	0.945	0.354	<0.001	0.102	0.200
ANOVA	Bias	0.867	0.849	0.206	0.879	0.308	0.107
	SD	0.367	0.750	0.313	<0.001	0.126	0.231
				γ_{0k3}			
MANOVA	Pillai	<0.001	<0.001	0.001	0.002	<0.001	<0.001
	p-value	0.684	0.498	0.065	0.011	0.919	0.811
ANOVA	Bias	0.395	0.764	0.088	0.039	0.798	0.740
(*p*-values)	SD	0.687	0.305	0.268	0.008	0.806	0.652
				γ_{0k4}			
MANOVA	Pillai	<0.001	<0.001	<0.001	0.004	0.001	<0.001
	p-value	0.570	0.979	0.787	<0.001	0.197	0.794
ANOVA	Bias	0.649	0.995	0.702	0.340	0.173	0.745
	SD	0.426	0.846	0.653	<0.001	0.447	0.507
				γ_1			
MANOVA	Pillai	0.003	0.001	<0.001	0.027	<0.001	0.001
	p-value	0.004	0.099	0.433	<0.001	0.959	0.227
ANOVA	Bias	0.707	0.311	0.756	0.729	0.817	0.467
(*p*-values)	SD	0.005	0.397	0.432	<0.001	0.991	0.446

For the structural parameters, Table 11, only the correlation between initial speed component and higher-order ability $\rho_{\theta\tau_0}$, had a significant effect on the recovery of all the attribute threshold parameters. The follow-up univariate ANOVA results show that $\rho_{\theta\tau_0}$ affects the standard errors of these parameters but not their biases. Figures 10 through 14 provide a graphical display for these significant results. The figures show that these effects, though statistically significant, may not be practically meaningful. Further investigation is required to obtain a more conclusive evidence for these effects.

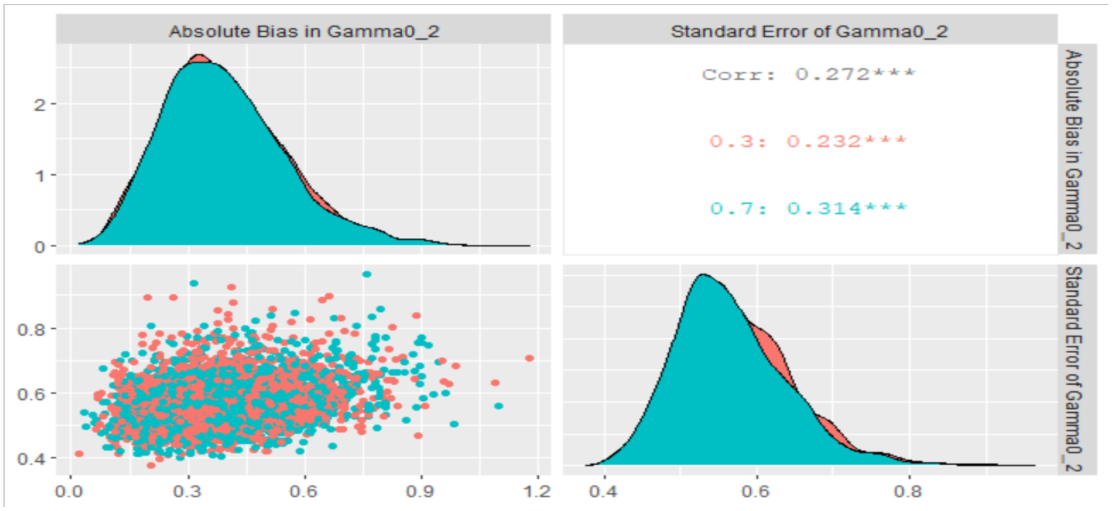


Figure 10 Effect of $\rho_{\theta\tau_0}$ on absolute bias and standard error of $\gamma_{0_{k2}}$

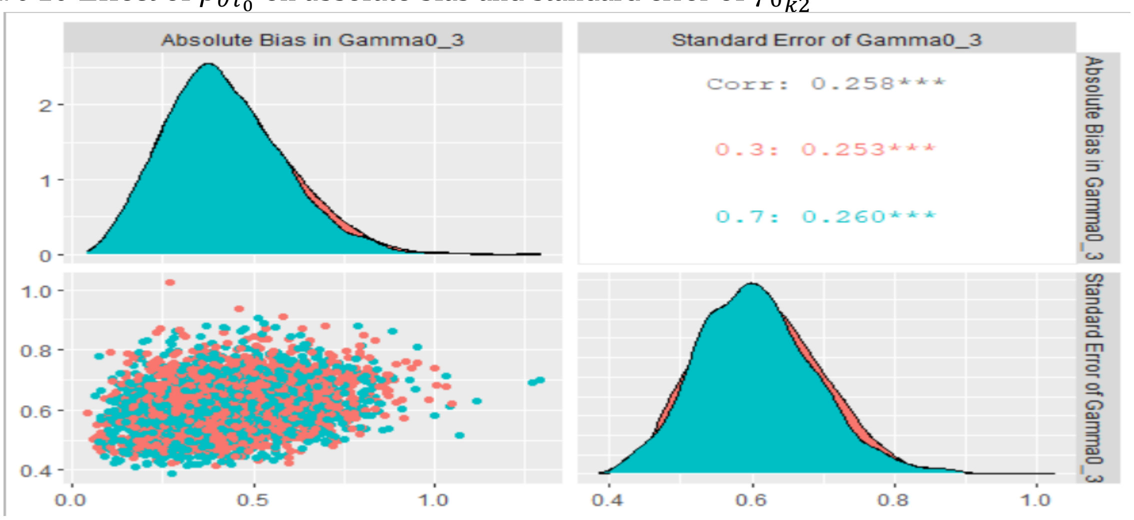


Figure 11 Effect of $\rho_{\theta\tau_0}$ on absolute bias and standard error of $\gamma_{0_{k3}}$

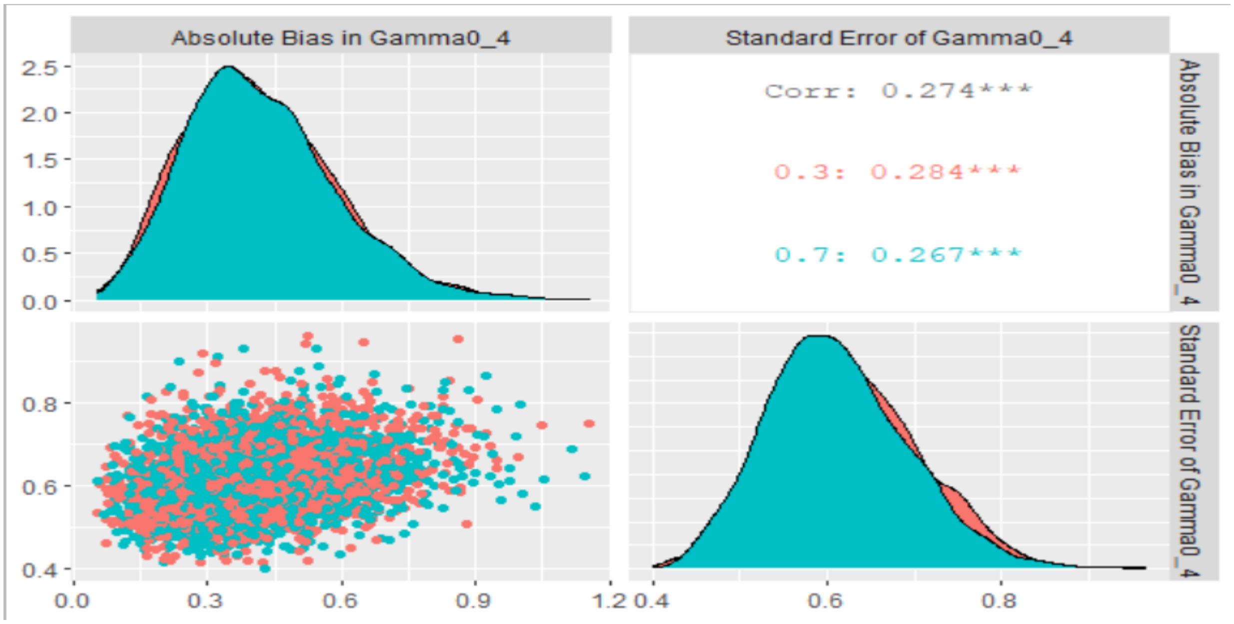


Figure 12 Effect of $\rho_{\theta\tau_0}$ on absolute bias and standard error of $\gamma_{0_{k4}}$

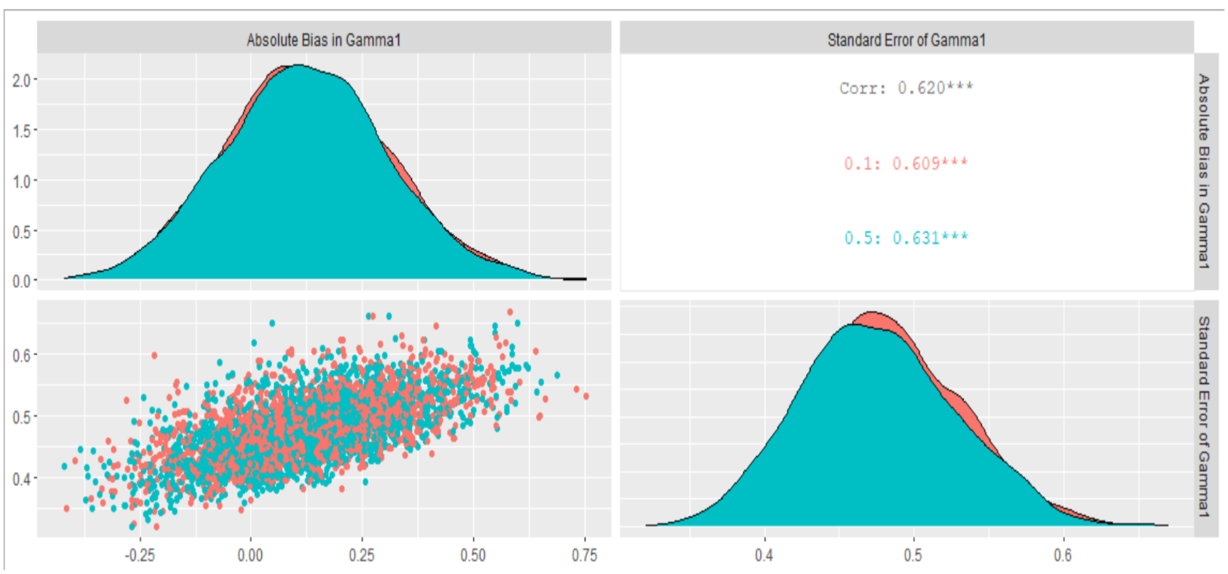


Figure 13 Effect of $\sigma^2_{\tau_0}$ on absolute bias and standard error of γ_1

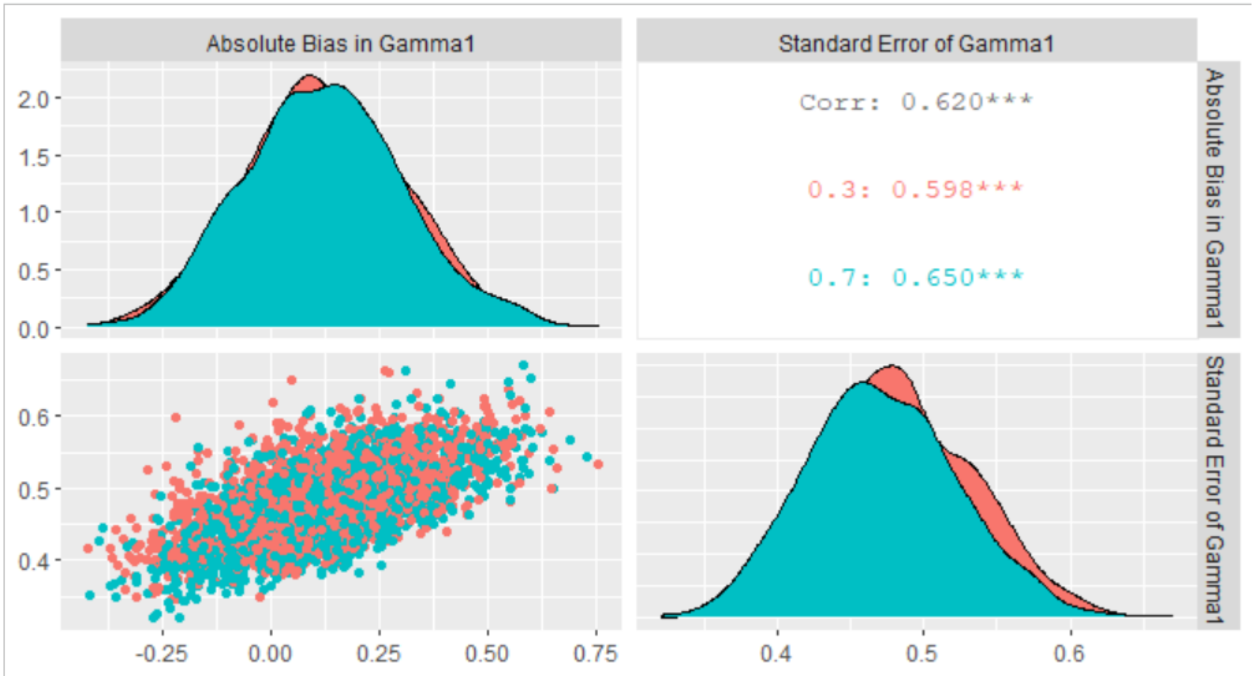


Figure 14 Effect of $\rho_{\theta\tau_0}$ on absolute bias and standard error of γ_1

4.4 Research Question 4

This fourth study was focused on comparing the JDS-DINA model to existing models, specifically, the MHO-DINA and JRT-DINA model. MHO-DINA ignores response time in modeling test responses, while JRT-DINA accounts for response time with the assumption of constant speed. The importance of this study is to highlight the importance of response time information in cognitive diagnostic model estimation as well as the effect of imposing a wrong attribute configuration on polytomous attribute data. Data for this comparison were generated using the JDS DINA model with polytomous attribute configuration, but all the models used here were fit with the binary attribute configuration. The idea is to understand which of these models comes closest to the truth, given that the wrong attribute configuration has been imposed on the data by the model choice

Table 12

Bias and RMSE of δ_0 by simulation design conditions – binary attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOB	Bias JRTB	Bias JDSB	RMSE MHOB	RMSE JRTB	RMSE JDSB
1	0.1	0.1	0.1	0.3	-0.05	-0.1	0.335	0.347	0.344	1.041	1.030	1.030
2						-0.3	0.348	0.356	0.359	1.005	1.001	1.005
3					-0.10	-0.1	0.330	0.341	0.337	0.989	0.984	0.983
4						-0.3	0.354	0.362	0.358	1.021	1.006	1.005
5				0.7	-0.05	-0.1	0.323	0.335	0.335	0.977	0.968	0.968
6						-0.3	0.329	0.344	0.343	0.989	0.988	0.988
7					-0.10	-0.1	0.353	0.361	0.360	1.003	0.997	0.996
8						-0.3	0.366	0.376	0.371	1.055	1.045	1.046
9			0.5	0.3	-0.05	-0.1	0.342	0.347	0.346	0.966	0.952	0.952
10						-0.3	0.337	0.349	0.346	1.013	1.006	1.006
11					-0.10	-0.1	0.316	0.329	0.328	0.996	0.987	0.986
12						-0.3	0.368	0.374	0.370	1.039	1.031	1.030
13				0.7	-0.05	-0.1	0.353	0.372	0.370	1.062	1.056	1.053
14						-0.3	0.342	0.357	0.355	0.988	0.982	0.984
15					-0.10	-0.1	0.332	0.350	0.350	0.987	0.981	0.981
16						-0.3	0.346	0.361	0.360	1.005	1.003	1.004
17		0.5	0.1	0.3	-0.05	-0.1	0.339	0.349	0.347	1.020	1.010	1.012
18						-0.3	0.338	0.343	0.346	0.981	0.967	0.969
19					-0.10	-0.1	0.359	0.366	0.366	1.054	1.045	1.046
20						-0.3	0.331	0.342	0.341	0.990	0.985	0.985
21				0.7	-0.05	-0.1	0.341	0.357	0.355	1.002	0.990	0.989
22						-0.3	0.343	0.356	0.356	0.997	0.984	0.986
23					-0.10	-0.1	0.325	0.334	0.333	1.032	1.018	1.015
24						-0.3	0.399	0.406	0.404	1.058	1.049	1.048
25			0.5	0.3	-0.05	-0.1	0.327	0.338	0.338	1.007	1.001	1.002
26						-0.3	0.368	0.379	0.376	1.020	1.013	1.011
27					-0.10	-0.1	0.379	0.386	0.383	1.048	1.042	1.039
28						-0.3	0.357	0.369	0.369	1.011	1.006	1.007
29				0.7	-0.05	-0.1	0.311	0.324	0.323	0.981	0.974	0.976
30						-0.3	0.340	0.350	0.350	1.028	1.017	1.016
31					-0.10	-0.1	0.352	0.361	0.360	1.004	1.001	0.999
32						-0.3	0.321	0.331	0.330	1.007	0.995	0.996

Table 12 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias			RMSE		
							MHOP	JRTP	JDSP	MHOP	JRTP	JDSP
33	0.5	0.1	0.1	0.3	-0.05	-0.1	0.348	0.358	0.358	1.061	1.051	1.051
34						-0.3	0.324	0.328	0.326	1.006	0.992	0.995
35					-0.10	-0.1	0.320	0.335	0.335	1.025	1.014	1.016
36						-0.3	0.305	0.320	0.318	0.998	0.982	0.980
37				0.7	-0.05	-0.1	0.351	0.382	0.385	1.022	1.016	1.019
38						-0.3	0.344	0.358	0.359	1.046	1.040	1.041
39					-0.10	-0.1	0.377	0.386	0.387	1.046	1.037	1.039
40						-0.3	0.347	0.369	0.370	1.014	1.023	1.022
41			0.5	0.3	-0.05	-0.1	0.316	0.329	0.325	0.978	0.967	0.968
42						-0.3	0.303	0.318	0.321	1.028	1.026	1.026
43					-0.10	-0.1	0.328	0.335	0.335	0.987	0.981	0.980
44						-0.3	0.346	0.357	0.356	1.027	1.021	1.019
45				0.7	-0.05	-0.1	0.331	0.350	0.349	1.014	1.003	1.003
46						-0.3	0.358	0.373	0.370	1.021	1.012	1.010
47					-0.10	-0.1	0.341	0.352	0.352	1.010	0.996	0.997
48						-0.3	0.336	0.345	0.342	0.998	0.992	0.989
49		0.5	0.1	0.3	-0.05	-0.1	0.316	0.334	0.332	0.987	0.969	0.965
50						-0.3	0.363	0.371	0.369	1.061	1.053	1.052
51					-0.10	-0.1	0.317	0.330	0.332	0.987	0.980	0.980
52						-0.3	0.364	0.370	0.374	1.080	1.073	1.075
53				0.7	-0.05	-0.1	0.350	0.375	0.374	1.041	1.036	1.035
54						-0.3	0.337	0.360	0.358	1.010	1.006	1.006
55					-0.10	-0.1	0.320	0.343	0.346	0.996	0.990	0.992
56						-0.3	0.320	0.332	0.333	0.955	0.943	0.942
57			0.5	0.3	-0.05	-0.1	0.316	0.323	0.320	1.024	1.007	1.007
58						-0.3	0.349	0.358	0.357	1.021	1.011	1.011
59					-0.10	-0.1	0.360	0.372	0.373	1.056	1.051	1.052
60						-0.3	0.345	0.351	0.352	1.019	1.010	1.010
61				0.7	-0.05	-0.1	0.347	0.366	0.366	1.025	1.016	1.016
62						-0.3	0.347	0.363	0.366	1.035	1.019	1.024
63					-0.10	-0.1	0.330	0.355	0.355	1.053	1.051	1.054
64						-0.3	0.298	0.308	0.304	1.003	0.986	0.985
						Mean	**0.340**	**0.352**	**0.352**	**1.016**	**1.007**	**1.007**

Table 12 shows the bias and RMSE for δ_0. The design conditions are numbered, 1 through 64 in the first column. The next six columns define these conditions. For instance, condition 1 has $\sigma^2_{\tau_0} = 0.5$, $\sigma^2_{\tau_1} = 0.1$, $\sigma^2_{\tau_2} = 0.1$, $\rho_{\theta\tau_0} = 0.3$, $\rho_{\theta\tau_2} = -0.05$ and $\rho_{\theta\tau_2} = -0.1$, and so on. For each row, the bias and RMSE were averaged across the 60 replications for each of the comparison models.

The results show that MHO-DINA has the worst performance, which is not surprising since data were generated from the JDS-DINA with polytomous attributes. However, JDS-DINA and JRT-DINA have very similar results. The reason may be that, given the values that have been chosen for the speed components, the relatively low parsimony of the JDS model trumps its ability to extract additional information from the differential speed component of the model. Similar result tables are available for all the item and structural parameters in the appendix. All parameters show similar results as with δ_0 shown here.

Graphical analysis was used to further examine and compare the results from these three models. Figures 15 through 20 display the results for the item parameters. Each of these graphs showed evidence that the JRT-DINA and JDS-DINA are indistinguishable in performance, but the MHO -DINA model was consistently poor performing.

Of note is the step or shift observed in the graphs for the item time intensity parameter, λ. This occurs right after condition 32. The first 32 conditions all have one thing in common - $\sigma^2_{\tau_0} = 0.1$. This suggests that the JRT and JDS DINA models perform better (lower bias and RMSE) at $\sigma^2_{\tau_0} = 0.1$, compared to $\sigma^2_{\tau_0} = 0.5$. In other words, high variability in initial speed is associated with poorer parameter recovery. This further highlights the need to account for this variance in cognitive diagnostic model estimation.

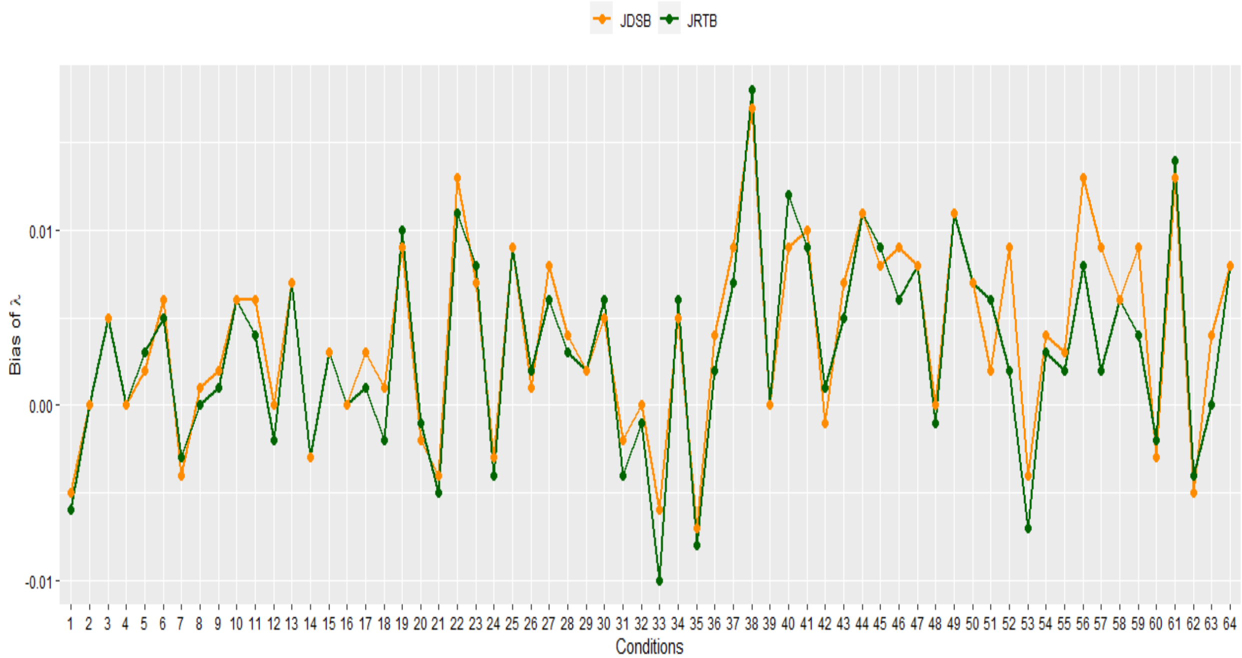


Figure 15 Bias of λ across simulation conditions

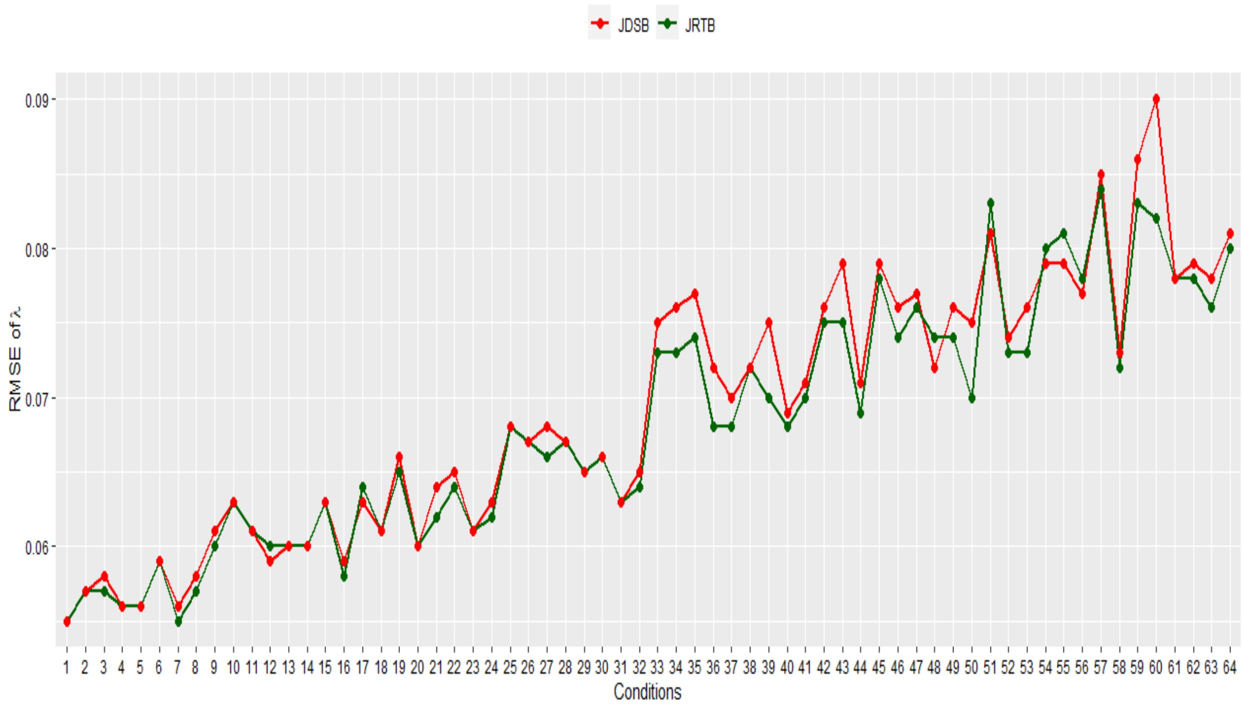


Figure 16 RMSE of λ across simulation conditions

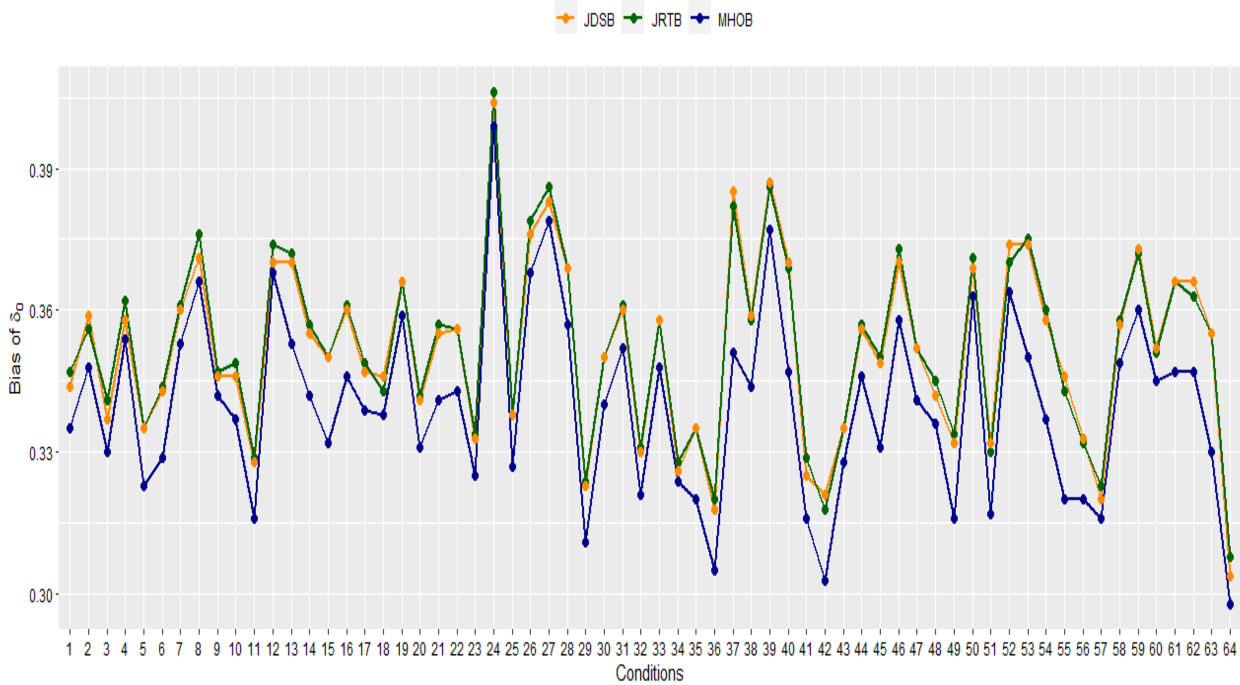


Figure 17 Bias of δ_0 across simulation conditions

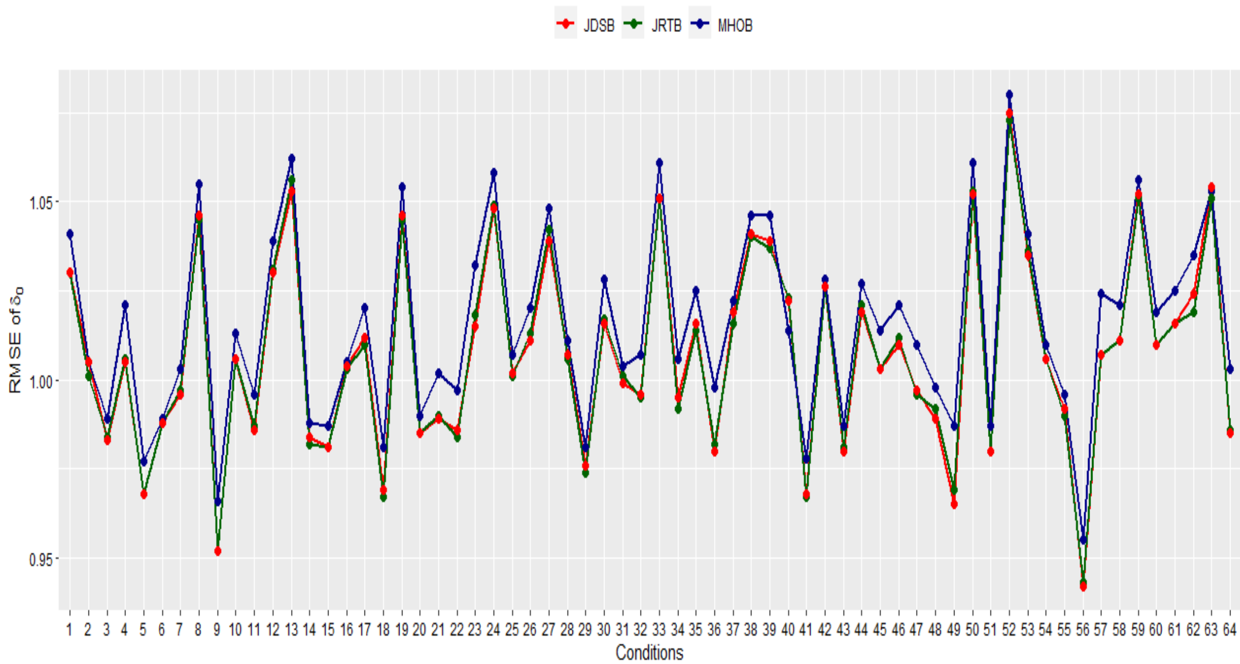


Figure 18 RMSE of δ_0 across simulation conditions

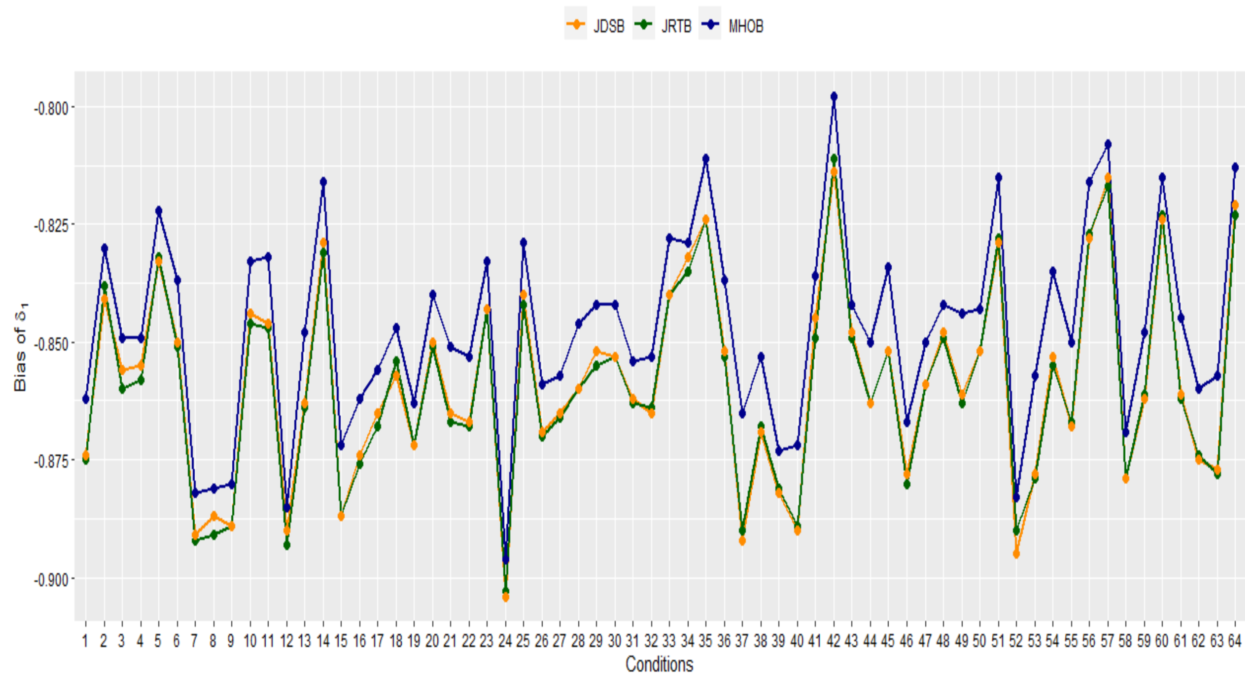


Figure 19 Bias of δ_1 across simulation conditions

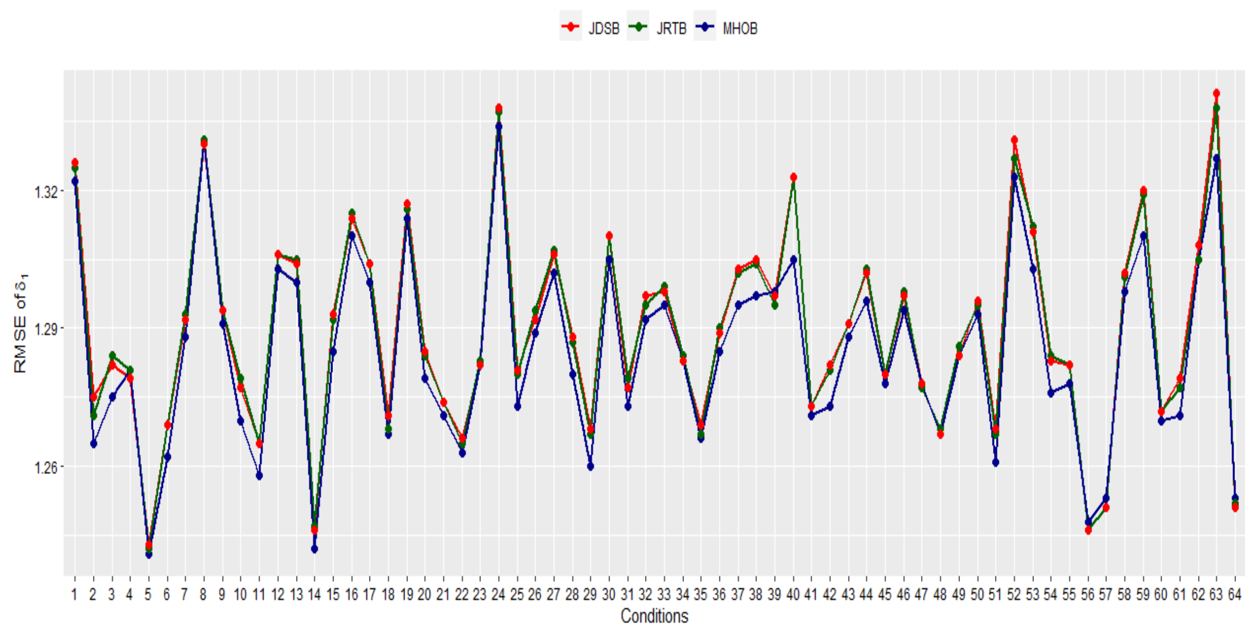


Figure 20 RMSE of δ_1 across simulation conditions

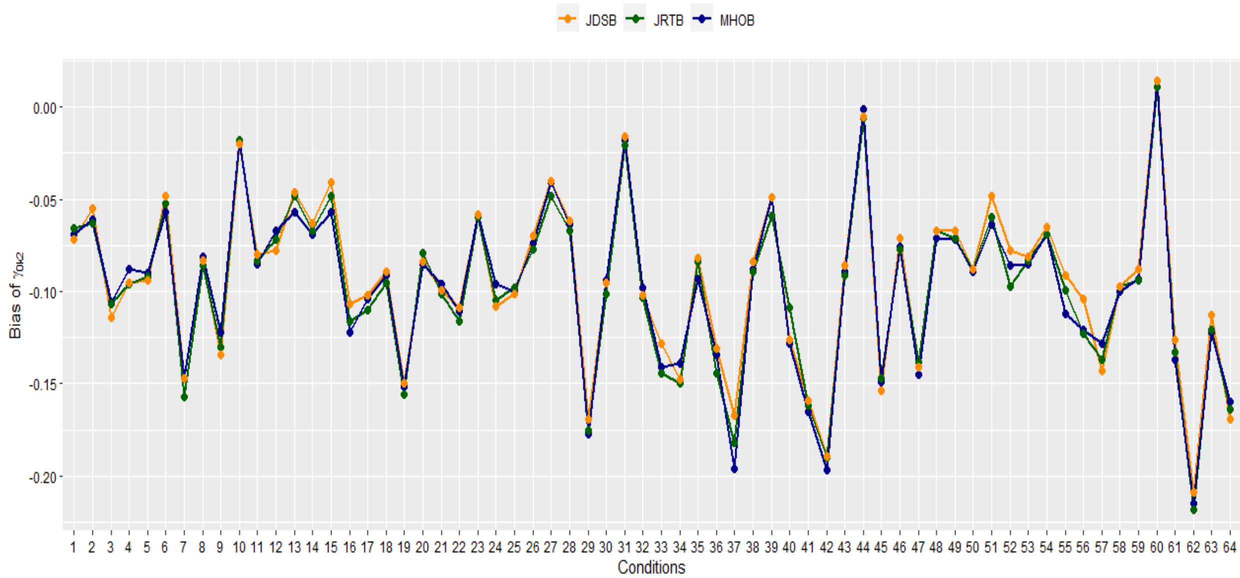


Figure 21 Bias of γ_{0k2} across simulation conditions

The bias and RMSE values for structural parameters were similar across models. While this would suggest that the most parsimonious model, MHO-DINA is appropriate, it is also important to remember that the results here are subject to the values that have been chosen for the speed components in the simulation. These results may be different for other values of speed components. Moreover, the primary aim of a cognitive diagnostic test is to estimate attribute profile, and the importance of model performance with respect to classification cannot be overemphasized. Table 13 below compares the attribute and person classification accuracies across these three models.

Table13

Exact and weighted classification accuracy rates

Model	Exact					Weighted				
	ACCR				PCCR	ACCR				PCCR
	α_1	α_2	α_3	α_4		α_1	α_2	α_3	α_4	
MHOB	0.192	0.220	0.268	0.252	0.025	0.438	0.469	0.510	0.489	0.107
JRTB	0.253	0.348	0.425	0.468	0.094	0.511	0.578	0.634	0.673	0.214
JDSB	0.253	0.348	0.425	0.468	0.094	0.511	0.578	0.634	0.673	0.214

The classification accuracies are generally low across models. This is expected since all three models were estimated with binary attribute configuration, but data was generated with polytomous configuration. Given the wrong attribute configuration, the relatively higher accuracy rates for the JDS-DINA and JRT-DINA underscores the importance of response time.

Taken together, the results from this study show that, while JDS and JRT DINA outperform MHO-DINA and are similar in item and structural parameters, the additional complexity introduced by JDS-DINA in accounting for *speededness* has no added benefit for correct classification accuracy rates. In other words, accounting for *speededness* makes no difference if the wrong attribute configuration has been imposed by the model.

4.5 Research question 5

This last study is like the previous, but with the correct attribute configuration. Data was generated with the JDS-DINA for polytomous attributes, and all three models are once again estimated and compared as before. For all three models, the results with this polytomous attribute configuration are better than with binary configurations. However, the patterns observed in the estimates and among models remain the same. The JRT and JDS DINA models remain considerably indistinguishable, while MHO-DINA still shows poor performance, as expected.

Table 14 shows the results for δ_0. The tables and figures for other item and structural parameters are also available in the appendix.

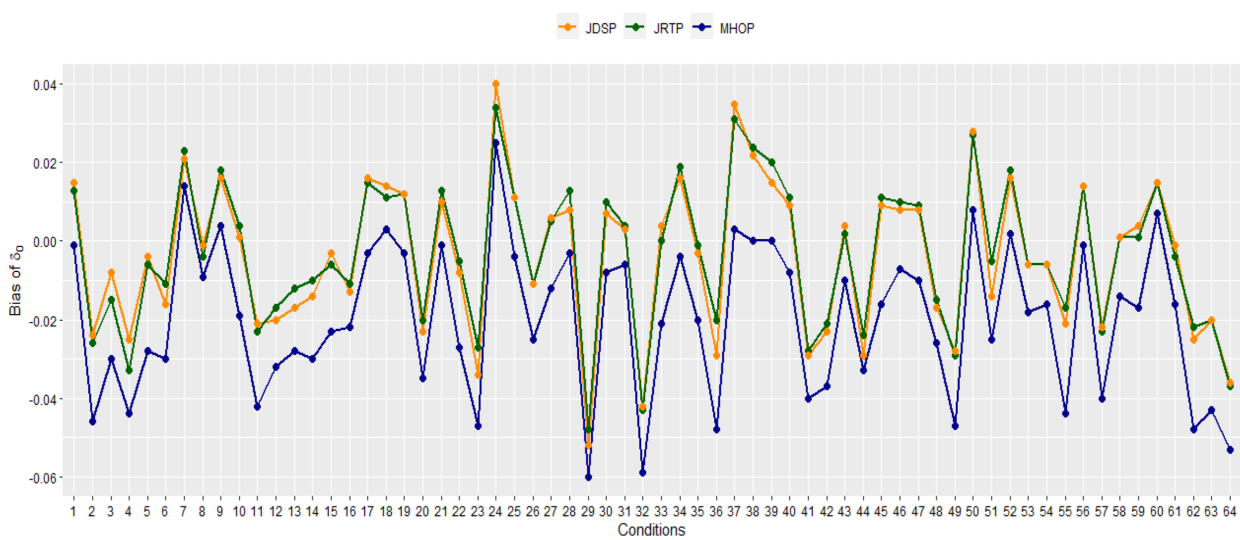


Figure 22 Bias of δ_0 across simulation conditions – polytomous attribute configuration

Table 14

Bias and RMSE of δ_0 by simulation design conditions – polytomous attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias			RMSE		
							MHOP	JRTP	JDSP	MHOP	JRTP	JDSP
1	0.1	0.1	0.1	0.3	-0.05	-0.1	-0.001	0.013	0.015	0.746	0.735	0.732
2						-0.3	-0.046	-0.026	-0.024	0.695	0.677	0.680
3					-0.10	-0.1	-0.030	-0.015	-0.008	0.719	0.708	0.702
4						-0.3	-0.044	-0.033	-0.025	0.685	0.670	0.665
5				0.7	-0.05	-0.1	-0.028	-0.006	-0.004	0.660	0.645	0.643
6						-0.3	-0.030	-0.011	-0.016	0.712	0.695	0.697
7					-0.10	-0.1	0.014	0.023	0.021	0.676	0.668	0.667
8						-0.3	-0.009	-0.004	-0.001	0.721	0.708	0.709
9			0.5	0.3	-0.05	-0.1	0.004	0.018	0.016	0.662	0.644	0.650
10						-0.3	-0.019	0.004	0.001	0.708	0.691	0.694
11					-0.10	-0.1	-0.042	-0.023	-0.021	0.730	0.715	0.715
12						-0.3	-0.032	-0.017	-0.020	0.703	0.683	0.685
13				0.7	-0.05	-0.1	-0.028	-0.012	-0.017	0.700	0.687	0.688
14						-0.3	-0.030	-0.010	-0.014	0.681	0.663	0.664
15					-0.10	-0.1	-0.023	-0.006	-0.003	0.676	0.660	0.664
16						-0.3	-0.022	-0.011	-0.013	0.696	0.683	0.682
17		0.5	0.1	0.3	-0.05	-0.1	-0.003	0.015	0.016	0.688	0.680	0.678
18						-0.3	0.003	0.011	0.014	0.685	0.668	0.668
19					-0.10	-0.1	-0.003	0.012	0.012	0.715	0.699	0.702
20						-0.3	-0.035	-0.020	-0.023	0.697	0.690	0.690
21				0.7	-0.05	-0.1	-0.001	0.013	0.010	0.680	0.670	0.672
22						-0.3	-0.027	-0.005	-0.008	0.702	0.681	0.680
23					-0.10	-0.1	-0.047	-0.027	-0.034	0.737	0.721	0.723
24						-0.3	0.025	0.034	0.040	0.738	0.718	0.716
25			0.5	0.3	-0.05	-0.1	-0.004	0.011	0.011	0.682	0.675	0.676
26						-0.3	-0.025	-0.011	-0.011	0.686	0.679	0.683
27					-0.10	-0.1	-0.012	0.005	0.006	0.713	0.696	0.690
28						-0.3	-0.003	0.013	0.008	0.715	0.714	0.716
29				0.7	-0.05	-0.1	-0.060	-0.048	-0.052	0.686	0.687	0.685
30						-0.3	-0.008	0.010	0.007	0.725	0.709	0.707
31					-0.10	-0.1	-0.006	0.004	0.003	0.682	0.667	0.669
32						-0.3	-0.059	-0.043	-0.042	0.704	0.691	0.688

Table 14 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOP	Bias JRTP	Bias JDSP	RMSE MHOP	RMSE JRTP	RMSE JDSP
33	0.5	0.1	0.1	0.3	-0.05	-0.1	-0.021	0.000	0.004	0.732	0.707	0.706
34						-0.3	-0.004	0.019	0.016	0.688	0.682	0.682
35					-0.10	-0.1	-0.020	-0.001	-0.003	0.714	0.686	0.685
36						-0.3	-0.048	-0.020	-0.029	0.709	0.692	0.694
37				0.7	-0.05	-0.1	0.003	0.031	0.035	0.718	0.681	0.685
38						-0.3	0.000	0.024	0.022	0.718	0.706	0.703
39					-0.10	-0.1	0.000	0.020	0.015	0.708	0.697	0.697
40						-0.3	-0.008	0.011	0.009	0.708	0.706	0.709
41			0.5	0.3	-0.05	-0.1	-0.040	-0.028	-0.029	0.682	0.649	0.653
42						-0.3	-0.037	-0.021	-0.023	0.730	0.705	0.708
43					-0.10	-0.1	-0.010	0.002	0.004	0.679	0.684	0.683
44						-0.3	-0.033	-0.024	-0.029	0.697	0.687	0.690
45				0.7	-0.05	-0.1	-0.016	0.011	0.009	0.712	0.685	0.688
46						-0.3	-0.007	0.010	0.008	0.680	0.660	0.657
47					-0.10	-0.1	-0.010	0.009	0.008	0.717	0.692	0.692
48						-0.3	-0.026	-0.015	-0.017	0.699	0.671	0.671
49		0.5	0.1	0.3	-0.05	-0.1	-0.047	-0.029	-0.028	0.690	0.670	0.672
50						-0.3	0.008	0.027	0.028	0.742	0.728	0.731
51					-0.10	-0.1	-0.025	-0.005	-0.014	0.700	0.691	0.699
52						-0.3	0.002	0.018	0.016	0.714	0.718	0.714
53				0.7	-0.05	-0.1	-0.018	-0.006	-0.006	0.714	0.693	0.696
54						-0.3	-0.016	-0.006	-0.006	0.664	0.650	0.654
55					-0.10	-0.1	-0.044	-0.017	-0.021	0.698	0.667	0.669
56						-0.3	-0.001	0.014	0.014	0.657	0.632	0.630
57			0.5	0.3	-0.05	-0.1	-0.040	-0.023	-0.022	0.713	0.685	0.683
58						-0.3	-0.014	0.001	0.001	0.705	0.689	0.691
59					-0.10	-0.1	-0.017	0.001	0.004	0.730	0.716	0.718
60						-0.3	0.007	0.015	0.015	0.730	0.716	0.718
61				0.7	-0.05	-0.1	-0.016	-0.004	-0.001	0.725	0.705	0.704
62						-0.3	-0.048	-0.022	-0.025	0.707	0.679	0.678
63					-0.10	-0.1	-0.043	-0.020	-0.020	0.768	0.750	0.749
64						-0.3	-0.053	-0.037	-0.036	0.716	0.704	0.705
						Mean	-0.020	-0.003	-0.004	0.704	0.688	0.689

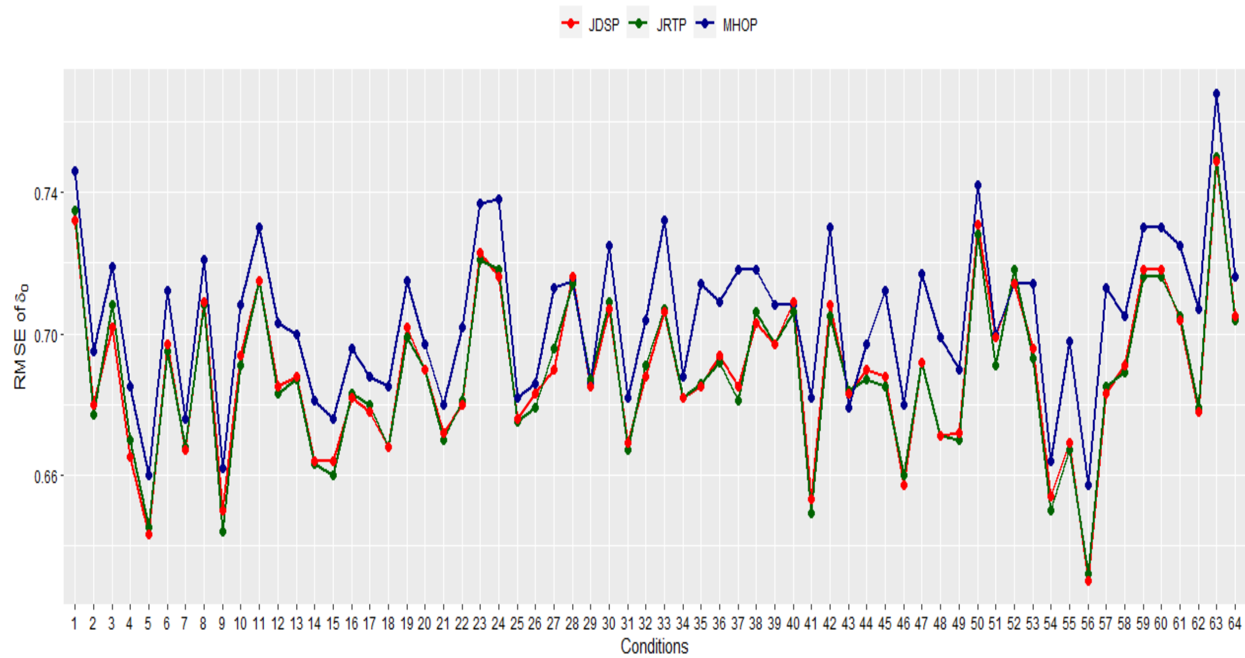


Figure 23 RMSE of δ_0 across simulation conditions – polytomous attribute configuration

Table 15

Exact and weighted classification accuracy rates with polytomous attributes

Model	Exact					Weighted				
	ACCR				PCCR	ACCR				PCCR
	α_1	α_2	α_3	α_4		α_1	α_2	α_3	α_4	
MHOP	0.165	0.181	0.213	0.201	0.012	0.494	0.479	0.483	0.441	0.095
JRTP	0.466	0.515	0.570	0.644	0.782	0.466	0.515	0.570	0.644	0.170
JDSP	0.464	0.514	0.569	0.644	0.169	0.848	0.862	0.869	0.887	0.580

The classification accuracies, in Table 15, improved considerably from the previous section, but the JDS-DINA shows considerably better values, as expected. The PCCR for the JDS model, though higher than others, is also low. This is unexpected since it is the true model. One possible reason could be the sample size used in the study. Given the complexity of the JDS model, larger sample size rapidly increases the computational burden. It is worth investigating further, to see if sample size alone explains the low PCCR that was observed in this study.

CHAPTER 5: DISCUSSION AND CONCLUSION

5.1 Summary of Findings

The aim of this study was multi-faceted. First, it proposed a new model that allows partial mastery and greater flexibility in incorporating response time in cognitive diagnostic models. Second, it assessed the performance of the new model under varying data conditions and compared it with existing models. Third, the study examined the effect of dichotomizing polytomous attributes using empirical and simulated data. From the simulated and real data analyses, several key findings were drawn.

5.1.1 Dichotomization

The dichotomization of polytomous attributes has implications for the accuracy of parameter estimates and skills diagnosis. The result showed very low classification accuracies when a binary classification model was imposed on polytomous attribute data. While binary classification models are relatively straightforward and easy to implement, this study, together with those from Karelitz (2004) and Zhan et al. (2019), has shown that dichotomizing attributes could lead to misleading results and wrong skills diagnosis. If attributes are meaningfully binary, artificially increasing their categorical levels to implement a polytomous attribute model would also be wrong. This study argues that if a set of attributes are meaningfully defined as polytomous, appropriate models that account for the ordinal category levels should be used.

5.1.2 Response time

Response time provides crucial supplementary information that can improve parameter estimation and classification accuracies. The real data analysis of the PISA computer-based data compared models with and without response time. The models with response time, though less parsimonious, showed better model fits and standard error of parameter estimates than the model that ignores response time. From the simulation study, results also showed that ignoring response time leads to poorer model performance. The typical testing situation imposes limited test time, even on supposedly power tests. This time limit introduces a new source of dependence in observed responses that are not accounted for in the traditional cognitive or item response models. This also means the all-important assumption of conditional independence is not satisfied with these data. The results of this study have shown that it is indeed important to account for time effect in modeling responses to tests for item calibration or for skills profile estimation.

5.1.3 Variable speed

The comparison between the JRT DINA with constant speed and the proposed JDS DINA with variable speed showed that both models performed equally well in recovering item and structural parameter estimates. In particular, the JRT-DINA recovered model parameter well, even when data was generated with a differential speed model. However, the effect of ignoring the differential speed is seen in the classification accuracies. This suggests that the flexibility provided by JDS DINA may be particularly important for correct classification, which is indeed the aim of cognitive diagnostic modeling. This result, however, is limited to the few data conditions that were explored. The influence on the item and structural parameters may be more pronounced with much higher or lower values for the variance of speed components.

5.1.4 Supplementary RT information

The results obtained from the 2012 computer-based Math test analysis showed that the data was more suited to a joint response time with constant speed (JRT DINA). As previously noted, this result is severely limited by the length of the test. The shortness of the test may not support a variable speed model, especially with a non-high-stakes test like PISA. It is very possible that students would keep a regulated speed of response through such a short test. That said, implementation of the variable speed model not only provides comparable parameter estimates with the constant speed model, but it also supplies additional information that could give insight to possible differences in test-taking behaviors and strategies. Figure 24 is a graphical depiction of what might be possible with the information obtained from a JDS model.

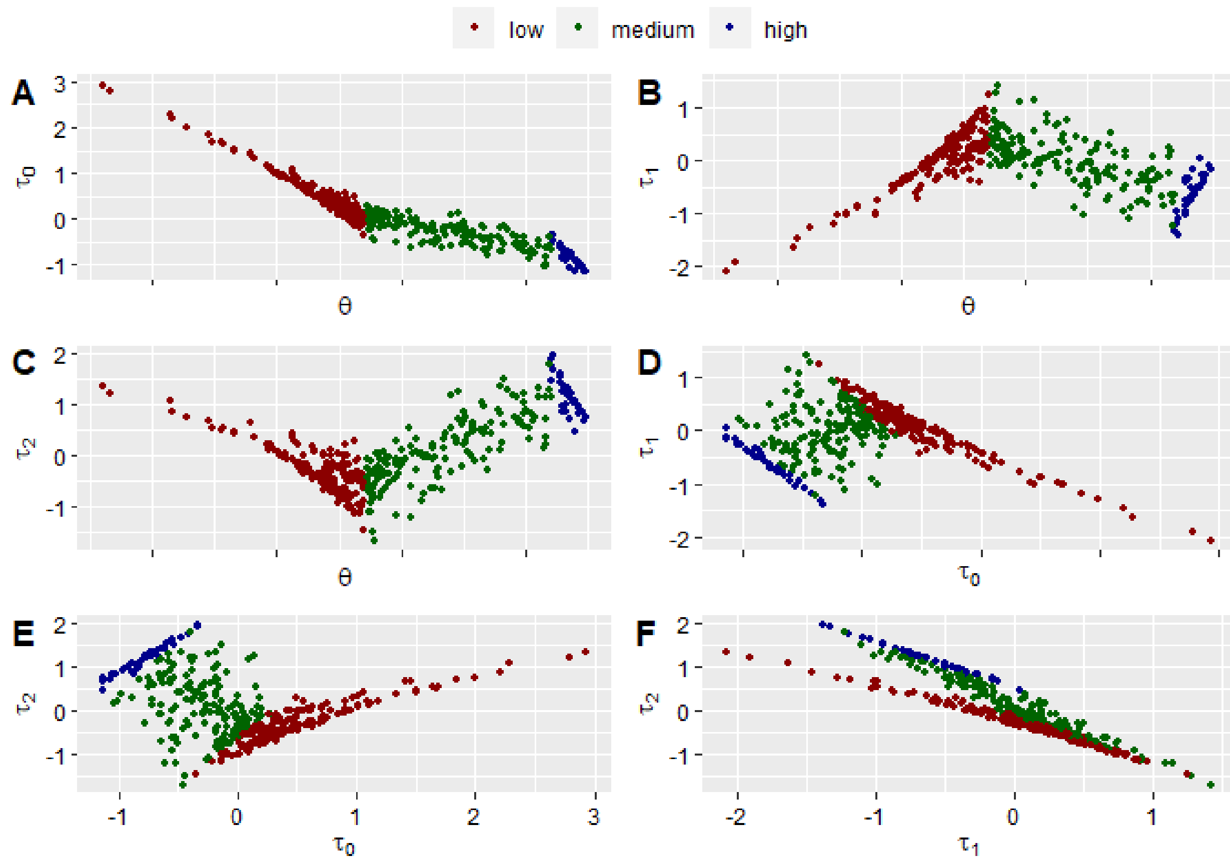


Figure 24 Relationships among person parameters from PISA computer-based Math test

In generating the figure above, only the data for the country of the USA was used. Observed patterns are similar across all the other countries in the data. The ability estimates were crudely split into three groups - low, medium, and high ability groups. The patterns show that the relationship among the person parameters differ across the three ability groups. This could indicate different test-taking strategies across these groups. For instance, panels A and B of

Figure 24 suggest that the high ability students are slow starters but increase speed very quickly, while the low to medium ability students start and proceed quickly through the test. Given PISA is not a high-stakes test, the observed pattern may be indicative of low motivation and (perhaps) guessing of answers among the low ability test-takers. Analysis of these patterns is beyond the scope of the current study. However, a detailed examination of this additional information could provide essential data to enhance skills diagnosis as well as item development and calibration for diagnostic purposes.

5.2 Limitations and future research

This section briefly discusses some limitations of the study while also offering directions for future research.

5.2.1 Test length

Research question 1 used the PISA data to examine the consequence of ignoring response time and the *speededness* effect. The result suggested that response time was important but not *speededness* effect. The item pool severely limited the results of this study. There were ten items used to assess seven attributes. By default, the associated Q-matrix is incomplete (Köhn & Chiu, 2018b), which may have affected the results. The nature of the test may have also played a role in the results. The PISA test a not a high-stakes test, and students are probably not highly motivated to finish the test. As such, the change in speed may not be very informative for skills diagnosis since the ability is further confounded with motivation. The findings, nonetheless, are interesting and promising. Future research should explore similar comparisons with a larger item pool to verify the findings of this study further.

5.2.2 Parameter values

The simulation study was used to address the last three questions of this study. Due to the limited number of studies on the relationship between *speededness* and cognitive ability, prior information on plausible values for the correlation between speed components and cognitive ability was not readily available. The empirical data available were not adequate to furnish reliable values either. The results of the current study may have been limited by the choice of the variance and correlation values used. A large item pool with response time information could be used to re-estimate these models and obtain more realistic true values for these parameters.

5.2.3 Simulation conditions

To reduce computational burden, a sample size of 200 was used in the simulation study. For a typical latent variable model, this is considered a small sample size. However, the study was kept at this sample size because of the computational burden of the JDS model. The number of iterations was also kept at 10,000 for the same reason. With 10,000 iterations, convergence for the person speed parameters in the variable speed model was relatively poor across the design conditions, with convergence rates between 73% and 84% for τ_0; 21% and 33% for τ_1, and between 26% and 43% for τ_2.

The rest of the parameters had a convergence rate of at least 94%, except for λ, which had between 69% and 100% convergence rates. For this reason, all model comparisons were restricted to item and structural parameters only. However, the poor convergence for these few parameters may have affected the classification accuracy rates obtained for the differential speed model. Given the low convergence for the person speed parameters, future studies would need to significantly increase the number of iterations to improve mixing for these parameters, taking note of the associated cost in computation time. Larger sample sizes should also be explored to understand the large sample behavior the variable speed model parameters.

5.2.4 Computational burden

Although the JDS-DINA offers more flexibility in the use of response time, the model, as defined in this study, is computationally intensive. In assessing these computation times, the JAGS estimation procedure was programmed to track all parameters and estimates in the model. Hence, the differences in estimation times are expected, especially because some of the additional estimates in the JDS model are incidental, increasing with sample size.

The current study was carried out using multiple computers with different specifications, and hence, estimation time across models could not be meaningfully compared. To show what is possible, the estimation time for one replication was obtained for each model using a computer with four cores, a base speed of 1.99GHz, and eight processors. Table 16 displays the computation times observed for one replication of the first design condition in the study. As expected, the polytomous configuration requires more time than binary attribute configuration, and computation time is very similar between MHO-DINA and JRT-DINA.

Table 16

Computation times (in minutes) for study models

	MHO-DINA	**JRT-DINA**	**JDS-DINA**
Binary	7.32	7.77	192.77
Polytomous	11.62	10.58	202.60

However, with JDS-DINA, there is a substantial increase in computation time. The computer-based PISA Mathematics data required 95.65 minutes to estimate the JDS-DINA model with 7 binary attributes, ten items, and 1,584 students. In the simulation study, sample size was reduced to 200, the number of attributes also dropped to 4, but the number of items increased from 10 to 30. These changes doubled the estimation time for the model with binary configuration. Hence, with a large item pool, the use of the variable speed model may be prohibitive.

Researchers should carefully consider the tradeoff between the flexibility and supplementary information offered by the differential speed model and the computational burden associated with its implementation, especially for large-scale assessment data. Alternative model specifications that offer the same amount of information and flexibility, but with lower time cost, could be explored. For instance, instead of a hierarchical model to relate response time to ability, one could use response time as a covariate for the ability level estimation via latent speed.

5.3 Summary

Literature is replete with studies that have proposed new models for analyzing cognitive diagnostic assessments. The call for transparency and accountability makes these research efforts expedient for enhancing the significance of educational assessments. However, most of the models in the literature have focused on the development of new diagnostic models to better reflect one or more specific test theories underlying a set of test responses. Only a few of these have investigated improving the outcome from these models by exploring information from response time.

This study explored a new model that expands existing models to incorporate response time and graded mastery levels in skills diagnosis. The examination of the model proved to be computationally demanding, but feasible. Comparison with existing models showed that incorporating response time with at least a constant speed is essential for item calibration. Extending response time to reflect variable speed may not significantly improve model parameter estimation, but it does improve attribute classification accuracy, which is the crux of diagnostic assessment.

The unavailability of cognitive diagnostic assessments for determining population parameter values qualified the outcome of this study. This is because most assessments are not designed for cognitive assessments, and current modeling attempts are restricted to retrofitting. More research efforts should be directed towards test construction and item calibration for diagnostic purposes to improve the modeling outcomes for these tests. Nonetheless, the results of the current study demonstrate great possibilities for using readily available response time to inform and enhance parameter estimation and classification accuracy in cognitive diagnostic modeling. Additional information supplied by this model can also provide insight into test behaviors that may compromise predicating test theory if ignored.

APPENDICES

APPENDIX A: SUPPLEMENTARY MATERIALS FOR RESEARCH QUESTION 4

Table A1

Bias and RMSE of λ by simulation design conditions – binary attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias			RMSE		
							MHOB	JRTB	JDSB	MHOB	JRTB	JDSB
1	0.1	0.1	0.1	0.3	-0.05	-0.1	--	-0.006	-0.005	--	0.055	0.055
2						-0.3	--	0.000	0.000	--	0.057	0.057
3					-0.10	-0.1	--	0.005	0.005	--	0.057	0.058
4						-0.3	--	0.000	0.000	--	0.056	0.056
5				0.7	-0.05	-0.1	--	0.003	0.002	--	0.056	0.056
6						-0.3	--	0.005	0.006	--	0.059	0.059
7					-0.10	-0.1	--	-0.003	-0.004	--	0.055	0.056
8						-0.3	--	0.000	0.001	--	0.057	0.058
9			0.5	0.3	-0.05	-0.1	--	0.001	0.002	--	0.060	0.061
10						-0.3	--	0.006	0.006	--	0.063	0.063
11					-0.10	-0.1	--	0.004	0.006	--	0.061	0.061
12						-0.3	--	-0.002	0.000	--	0.060	0.059
13				0.7	-0.05	-0.1	--	0.007	0.007	--	0.060	0.060
14						-0.3	--	-0.003	-0.003	--	0.060	0.060
15					-0.10	-0.1	--	0.003	0.003	--	0.063	0.063
16						-0.3	--	0.000	0.000	--	0.058	0.059
17		0.5	0.1	0.3	-0.05	-0.1	--	0.001	0.003	--	0.064	0.063
18						-0.3	--	-0.002	0.001	--	0.061	0.061
19					-0.10	-0.1	--	0.010	0.009	--	0.065	0.066
20						-0.3	--	-0.001	-0.002	--	0.060	0.060
21				0.7	-0.05	-0.1	--	-0.005	-0.004	--	0.062	0.064
22						-0.3	--	0.011	0.013	--	0.064	0.065
23					-0.10	-0.1	--	0.008	0.007	--	0.061	0.061
24						-0.3	--	-0.004	-0.003	--	0.062	0.063
25			0.5	0.3	-0.05	-0.1	--	0.009	0.009	--	0.068	0.068
26						-0.3	--	0.002	0.001	--	0.067	0.067
27					-0.10	-0.1	--	0.006	0.008	--	0.066	0.068
28						-0.3	--	0.003	0.004	--	0.067	0.067
29				0.7	-0.05	-0.1	--	0.002	0.002	--	0.065	0.065
30						-0.3	--	0.006	0.005	--	0.066	0.066
32					-0.10	-0.1	--	-0.004	-0.002	--	0.063	0.063
32						-0.3	--	-0.001	0.000	--	0.064	0.065

Table A1 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOB	Bias JRTB	Bias JDSB	RMSE MHOB	RMSE JRTP	RMSE JDSB
33	0.5	0.1	0.1	0.3	-0.05	-0.1	--	-0.010	-0.006	--	0.073	0.075
34						-0.3	--	0.006	0.005	--	0.073	0.076
35					-0.10	-0.1	--	-0.008	-0.007	--	0.074	0.077
36						-0.3	--	0.002	0.004	--	0.068	0.072
37				0.7	-0.05	-0.1	--	0.007	0.009	--	0.068	0.070
38						-0.3	--	0.018	0.017	--	0.072	0.072
39					-0.10	-0.1	--	0.000	0.000	--	0.070	0.075
40						-0.3	--	0.012	0.009	--	0.068	0.069
41			0.5	0.3	-0.05	-0.1	--	0.009	0.010	--	0.070	0.071
42						-0.3	--	0.001	-0.001	--	0.075	0.076
43					-0.10	-0.1	--	0.005	0.007	--	0.075	0.079
44						-0.3	--	0.011	0.011	--	0.069	0.071
45				0.7	-0.05	-0.1	--	0.009	0.008	--	0.078	0.079
46						-0.3	--	0.006	0.009	--	0.074	0.076
47					-0.10	-0.1	--	0.008	0.008	--	0.076	0.077
48						-0.3	--	-0.001	0.000	--	0.074	0.072
49		0.5	0.1	0.3	-0.05	-0.1	--	0.011	0.011	--	0.074	0.076
50						-0.3	--	0.007	0.007	--	0.070	0.075
51					-0.10	-0.1	--	0.006	0.002	--	0.083	0.081
52						-0.3	--	0.002	0.009	--	0.073	0.074
53				0.7	-0.05	-0.1	--	-0.007	-0.004	--	0.073	0.076
54						-0.3	--	0.003	0.004	--	0.080	0.079
55					-0.10	-0.1	--	0.002	0.003	--	0.081	0.079
56						-0.3	--	0.008	0.013	--	0.078	0.077
57			0.5	0.3	-0.05	-0.1	--	0.002	0.009	--	0.084	0.085
58						-0.3	--	0.006	0.006	--	0.072	0.073
59					-0.10	-0.1	--	0.004	0.009	--	0.083	0.086
60						-0.3	--	-0.002	-0.003	--	0.082	0.090
61				0.7	-0.05	-0.1	--	0.014	0.013	--	0.078	0.078
62						-0.3	--	-0.004	-0.005	--	0.078	0.079
63					-0.10	-0.1	--	0.000	0.004	--	0.076	0.078
64						-0.3	--	0.008	0.008	--	0.080	0.081
						Mean	--	0.003	0.004	--	0.068	0.069

Table A2

Bias and RMSE of δ_1 by simulation design conditions – binary attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOB	Bias JRTB	Bias JDSB	RMSE MHOB	RMSE JRTB	RMSE JDSB
1	0.1	0.1	0.1	0.3	-0.05	-0.1	-0.862	-0.875	-0.874	1.322	1.325	1.326
2						-0.3	-0.830	-0.838	-0.841	1.265	1.271	1.275
3					-0.10	-0.1	-0.849	-0.860	-0.856	1.275	1.284	1.282
4						-0.3	-0.849	-0.858	-0.855	1.281	1.281	1.279
5				0.7	-0.05	-0.1	-0.822	-0.832	-0.833	1.241	1.242	1.243
6						-0.3	-0.837	-0.851	-0.850	1.262	1.269	1.269
7					-0.10	-0.1	-0.882	-0.892	-0.891	1.288	1.293	1.292
8						-0.3	-0.881	-0.891	-0.887	1.331	1.331	1.330
9			0.5	0.3	-0.05	-0.1	-0.880	-0.889	-0.889	1.291	1.294	1.294
10						-0.3	-0.833	-0.846	-0.844	1.270	1.279	1.277
11					-0.10	-0.1	-0.832	-0.847	-0.846	1.258	1.265	1.265
12						-0.3	-0.885	-0.893	-0.890	1.303	1.306	1.306
13				0.7	-0.05	-0.1	-0.848	-0.864	-0.863	1.300	1.305	1.304
14						-0.3	-0.816	-0.831	-0.829	1.242	1.247	1.246
15					-0.10	-0.1	-0.872	-0.887	-0.887	1.285	1.292	1.293
16						-0.3	-0.862	-0.876	-0.874	1.310	1.315	1.314
17		0.5	0.1	0.3	-0.05	-0.1	-0.856	-0.868	-0.865	1.300	1.304	1.304
18						-0.3	-0.847	-0.854	-0.857	1.267	1.268	1.271
19					-0.10	-0.1	-0.863	-0.872	-0.872	1.314	1.316	1.317
20						-0.3	-0.840	-0.851	-0.850	1.279	1.284	1.285
21				0.7	-0.05	-0.1	-0.851	-0.867	-0.865	1.271	1.274	1.274
22						-0.3	-0.853	-0.868	-0.867	1.263	1.265	1.266
23					-0.10	-0.1	-0.833	-0.843	-0.843	1.282	1.283	1.282
24						-0.3	-0.896	-0.903	-0.904	1.334	1.337	1.338
25			0.5	0.3	-0.05	-0.1	-0.829	-0.842	-0.840	1.273	1.280	1.281
26						-0.3	-0.859	-0.870	-0.869	1.289	1.294	1.292
27					-0.10	-0.1	-0.857	-0.866	-0.865	1.302	1.307	1.306
28						-0.3	-0.846	-0.860	-0.860	1.280	1.287	1.288
29				0.7	-0.05	-0.1	-0.842	-0.855	-0.852	1.260	1.267	1.268
30						-0.3	-0.842	-0.853	-0.853	1.305	1.310	1.310
31					-0.10	-0.1	-0.854	-0.863	-0.862	1.273	1.279	1.277
32						-0.3	-0.853	-0.864	-0.865	1.292	1.295	1.297

Table A2 (cont'd).

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOB	Bias JRTB	Bias JDSB	RMSE MHOB	RMSE JRTB	RMSE JDSB
33	0.5	0.1	0.1	0.3	-0.05	-0.1	-0.828	-0.840	-0.840	1.295	1.299	1.298
34						-0.3	-0.829	-0.835	-0.832	1.283	1.284	1.283
35					-0.10	-0.1	-0.811	-0.824	-0.824	1.266	1.267	1.269
36						-0.3	-0.837	-0.853	-0.852	1.285	1.290	1.289
37				0.7	-0.05	-0.1	-0.865	-0.890	-0.892	1.295	1.302	1.303
38						-0.3	-0.853	-0.868	-0.869	1.297	1.304	1.305
39					-0.10	-0.1	-0.873	-0.881	-0.882	1.298	1.295	1.297
40						-0.3	-0.872	-0.889	-0.890	1.305	1.323	1.323
41			0.5	0.3	-0.05	-0.1	-0.836	-0.849	-0.845	1.271	1.273	1.273
42						-0.3	-0.798	-0.811	-0.814	1.273	1.281	1.282
43					-0.10	-0.1	-0.842	-0.849	-0.848	1.288	1.291	1.291
44						-0.3	-0.850	-0.863	-0.863	1.296	1.303	1.302
45				0.7	-0.05	-0.1	-0.834	-0.852	-0.852	1.278	1.280	1.280
46						-0.3	-0.867	-0.880	-0.878	1.294	1.298	1.297
47					-0.10	-0.1	-0.850	-0.859	-0.859	1.278	1.277	1.278
48						-0.3	-0.842	-0.849	-0.848	1.267	1.268	1.267
49		0.5	0.1	0.3	-0.05	-0.1	-0.844	-0.863	-0.861	1.284	1.286	1.284
50						-0.3	-0.843	-0.852	-0.852	1.293	1.295	1.296
51					-0.10	-0.1	-0.815	-0.828	-0.829	1.261	1.267	1.268
52						-0.3	-0.883	-0.890	-0.895	1.323	1.327	1.331
53				0.7	-0.05	-0.1	-0.857	-0.879	-0.878	1.303	1.312	1.311
54						-0.3	-0.835	-0.855	-0.853	1.276	1.284	1.283
55					-0.10	-0.1	-0.850	-0.867	-0.868	1.278	1.282	1.282
56						-0.3	-0.816	-0.827	-0.828	1.248	1.246	1.246
57			0.5	0.3	-0.05	-0.1	-0.808	-0.817	-0.815	1.253	1.251	1.251
58						-0.3	-0.869	-0.879	-0.879	1.298	1.301	1.302
59					-0.10	-0.1	-0.848	-0.861	-0.862	1.310	1.319	1.320
60						-0.3	-0.815	-0.823	-0.824	1.270	1.272	1.272
61				0.7	-0.05	-0.1	-0.845	-0.862	-0.861	1.271	1.277	1.279
62						-0.3	-0.860	-0.874	-0.875	1.305	1.305	1.308
63					-0.10	-0.1	-0.857	-0.878	-0.877	1.327	1.338	1.341
64						-0.3	-0.813	-0.823	-0.821	1.253	1.252	1.251
						Mean	-0.847	-0.859	-0.859	1.285	1.289	1.289

Table A3

Bias and RMSE of γ_{0k2} by simulation design conditions – binary attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOB	Bias JRTB	Bias JDSB	RMSE MHOB	RMSE JRTB	RMSE JDSB
1	0.1	0.1	0.1	0.3	-0.05	-0.1	-0.069	-0.066	-0.072	0.208	0.206	0.205
2						-0.3	-0.061	-0.063	-0.055	0.217	0.215	0.217
3					-0.10	-0.1	-0.106	-0.107	-0.114	0.221	0.221	0.221
4						-0.3	-0.088	-0.096	-0.095	0.213	0.213	0.212
5				0.7	-0.05	-0.1	-0.090	-0.092	-0.094	0.219	0.221	0.222
6						-0.3	-0.057	-0.052	-0.048	0.223	0.219	0.218
7					-0.10	-0.1	-0.147	-0.157	-0.147	0.235	0.232	0.235
8						-0.3	-0.081	-0.086	-0.083	0.209	0.209	0.208
9			0.5	0.3	-0.05	-0.1	-0.122	-0.130	-0.134	0.220	0.219	0.217
10						-0.3	-0.018	-0.018	-0.020	0.248	0.248	0.247
11					-0.10	-0.1	-0.085	-0.083	-0.080	0.231	0.228	0.228
12						-0.3	-0.067	-0.072	-0.078	0.224	0.225	0.223
13				0.7	-0.05	-0.1	-0.057	-0.048	-0.046	0.229	0.224	0.223
14						-0.3	-0.069	-0.067	-0.063	0.227	0.228	0.227
15					-0.10	-0.1	-0.057	-0.048	-0.041	0.217	0.217	0.220
16						-0.3	-0.122	-0.116	-0.107	0.211	0.212	0.215
17		0.5	0.1	0.3	-0.05	-0.1	-0.104	-0.110	-0.102	0.213	0.212	0.213
18						-0.3	-0.091	-0.095	-0.089	0.218	0.216	0.215
19					-0.10	-0.1	-0.152	-0.156	-0.150	0.222	0.221	0.221
20						-0.3	-0.085	-0.079	-0.084	0.209	0.211	0.211
21				0.7	-0.05	-0.1	-0.096	-0.101	-0.099	0.231	0.229	0.230
22						-0.3	-0.111	-0.116	-0.109	0.225	0.225	0.225
23					-0.10	-0.1	-0.058	-0.060	-0.058	0.238	0.237	0.237
24						-0.3	-0.096	-0.105	-0.108	0.227	0.225	0.225
25			0.5	0.3	-0.05	-0.1	-0.100	-0.098	-0.101	0.229	0.226	0.228
26						-0.3	-0.074	-0.077	-0.070	0.231	0.232	0.230
27					-0.10	-0.1	-0.041	-0.048	-0.040	0.233	0.233	0.231
28						-0.3	-0.064	-0.067	-0.062	0.226	0.226	0.225
29				0.7	-0.05	-0.1	-0.177	-0.175	-0.169	0.232	0.231	0.231
30						-0.3	-0.094	-0.101	-0.095	0.223	0.223	0.222
31					-0.10	-0.1	-0.018	-0.021	-0.016	0.226	0.226	0.226
32						-0.3	-0.098	-0.103	-0.102	0.213	0.212	0.213

Table A3 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOB	Bias JRTB	Bias JDSB	RMSE MHOB	RMSE JRTB	RMSE JDSB
33	0.5	0.1	0.1	0.3	-0.05	-0.1	-0.141	-0.144	-0.128	0.247	0.244	0.244
34						-0.3	-0.139	-0.150	-0.148	0.225	0.225	0.226
35					-0.10	-0.1	-0.093	-0.084	-0.082	0.236	0.234	0.235
36						-0.3	-0.134	-0.144	-0.131	0.215	0.215	0.213
37				0.7	-0.05	-0.1	-0.196	-0.182	-0.167	0.233	0.225	0.227
38						-0.3	-0.088	-0.089	-0.084	0.228	0.226	0.227
39					-0.10	-0.1	-0.049	-0.059	-0.049	0.202	0.205	0.206
40						-0.3	-0.128	-0.109	-0.126	0.208	0.208	0.208
41			0.5	0.3	-0.05	-0.1	-0.165	-0.162	-0.159	0.219	0.218	0.217
42						-0.3	-0.197	-0.190	-0.189	0.232	0.229	0.231
43					-0.10	-0.1	-0.089	-0.091	-0.086	0.216	0.215	0.216
44						-0.3	-0.001	-0.006	-0.005	0.222	0.222	0.224
45				0.7	-0.05	-0.1	-0.149	-0.147	-0.154	0.212	0.212	0.208
46						-0.3	-0.076	-0.077	-0.071	0.218	0.216	0.217
47					-0.10	-0.1	-0.145	-0.138	-0.141	0.223	0.228	0.228
48						-0.3	-0.071	-0.067	-0.067	0.230	0.229	0.226
49		0.5	0.1	0.3	-0.05	-0.1	-0.072	-0.071	-0.067	0.221	0.219	0.218
50						-0.3	-0.089	-0.088	-0.088	0.227	0.226	0.224
51					-0.10	-0.1	-0.064	-0.060	-0.048	0.223	0.225	0.224
52						-0.3	-0.086	-0.097	-0.078	0.234	0.231	0.231
53				0.7	-0.05	-0.1	-0.085	-0.083	-0.081	0.229	0.224	0.225
54						-0.3	-0.069	-0.069	-0.065	0.218	0.215	0.214
55					-0.10	-0.1	-0.112	-0.099	-0.091	0.221	0.223	0.223
56						-0.3	-0.121	-0.123	-0.104	0.219	0.219	0.219
57			0.5	0.3	-0.05	-0.1	-0.128	-0.137	-0.143	0.226	0.226	0.228
58						-0.3	-0.100	-0.097	-0.097	0.206	0.207	0.206
59					-0.10	-0.1	-0.093	-0.094	-0.088	0.210	0.211	0.211
60						-0.3	0.014	0.011	0.014	0.231	0.231	0.232
61				0.7	-0.05	-0.1	-0.137	-0.133	-0.126	0.224	0.221	0.218
62						-0.3	-0.215	-0.218	-0.209	0.227	0.227	0.227
63					-0.10	-0.1	-0.122	-0.121	-0.113	0.224	0.224	0.230
64						-0.3	-0.160	-0.164	-0.169	0.229	0.227	0.229
						Mean	-0.098	-0.099	-0.095	0.223	0.222	0.222

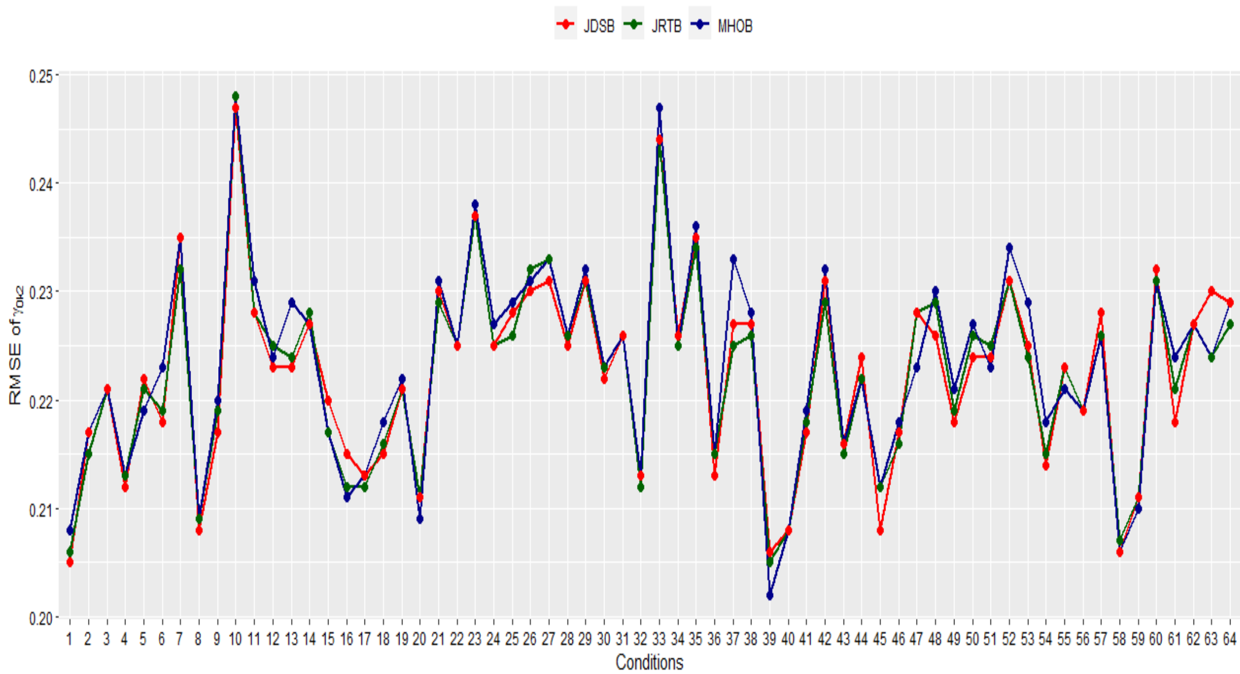


Figure A1 RMSE of γ_{0k2}across simulation conditions – polytomous attribute configuration

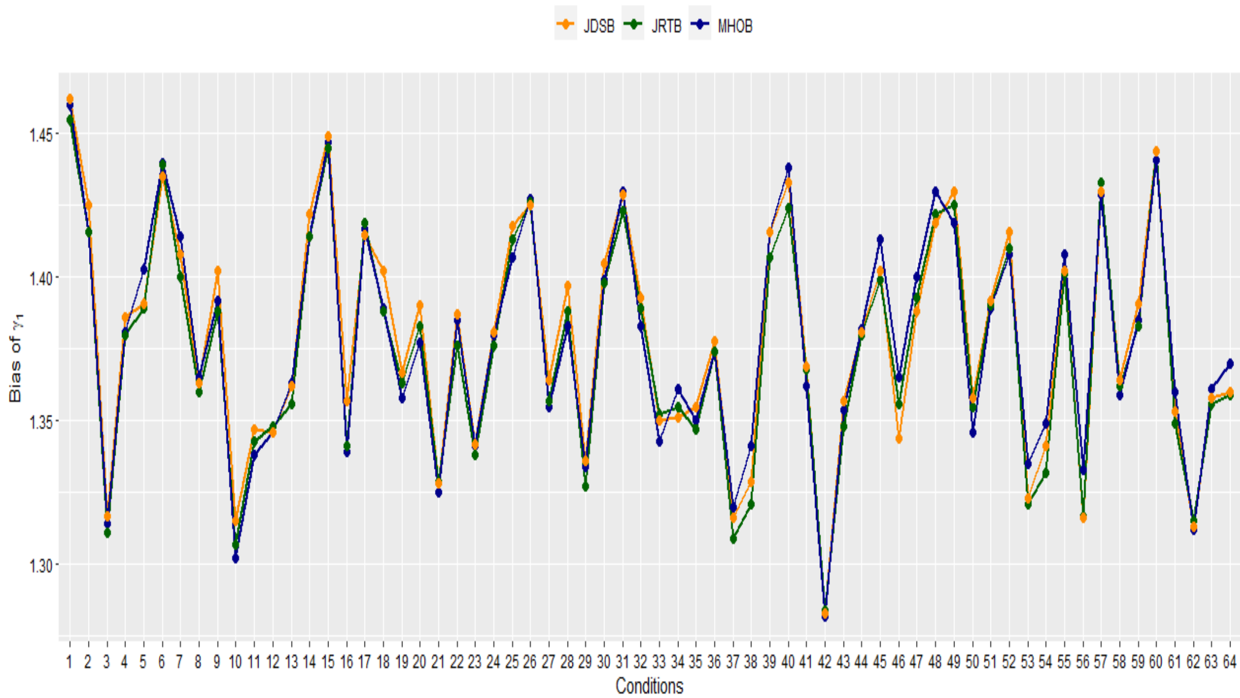


Figure A2 Bias of γ_1 across simulation conditions – polytomous attribute configuration

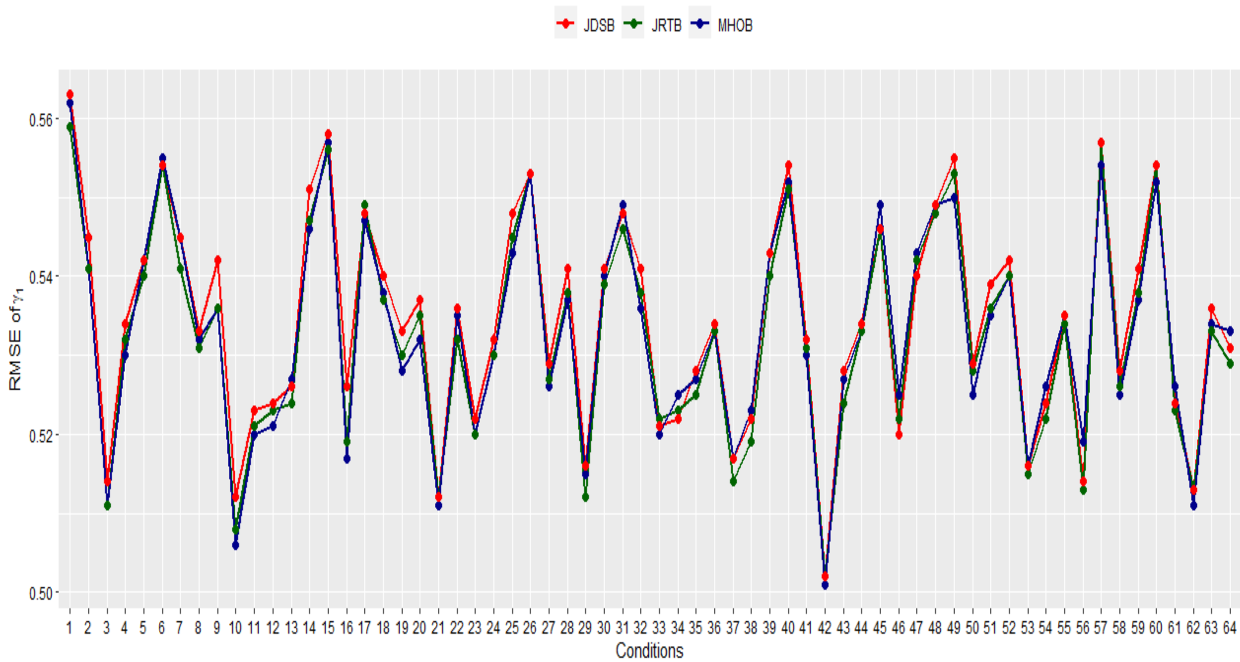


Figure A3 RMSE of γ_1 across simulation conditions – polytomous attribute configuration

APPENDIX B: SUPPLEMENTARY MATERIALS FOR RESEARCH QUESTION 5

Table B1

Bias and RMSE of λ by simulation design conditions – polytomous attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOP	Bias JRTP	Bias JDSP	RMSE MHOP	RMSE JRTP	RMSE JDSP
1	0.1	0.1	0.1	0.3	-0.05	-0.1	--	-0.007	-0.007	--	0.055	0.055
2						-0.3	--	0.000	0.001	--	0.057	0.058
3					-0.10	-0.1	--	0.004	0.003	--	0.057	0.057
4						-0.3	--	-0.001	-0.001	--	0.056	0.057
5				0.7	-0.05	-0.1	--	0.001	0.000	--	0.056	0.056
6						-0.3	--	0.003	0.003	--	0.058	0.059
7					-0.10	-0.1	--	-0.006	-0.006	--	0.055	0.056
8						-0.3	--	-0.002	-0.003	--	0.057	0.058
9			0.5	0.3	-0.05	-0.1	--	0.000	0.002	--	0.060	0.060
10						-0.3	--	0.004	0.005	--	0.063	0.063
11					-0.10	-0.1	--	0.004	0.004	--	0.061	0.061
12						-0.3	--	-0.002	-0.003	--	0.059	0.060
13				0.7	-0.05	-0.1	--	0.005	0.004	--	0.060	0.060
14						-0.3	--	-0.005	-0.005	--	0.060	0.060
15					-0.10	-0.1	--	-0.001	0.002	--	0.062	0.062
16						-0.3	--	-0.002	-0.001	--	0.059	0.058
17		0.5	0.1	0.3	-0.05	-0.1	--	0.001	0.003	--	0.063	0.064
18						-0.3	--	-0.004	-0.001	--	0.062	0.062
19					-0.10	-0.1	--	0.009	0.011	--	0.065	0.066
20						-0.3	--	-0.002	-0.001	--	0.060	0.060
21				0.7	-0.05	-0.1	--	-0.008	-0.007	--	0.063	0.063
22						-0.3	--	0.009	0.010	--	0.064	0.064
23					-0.10	-0.1	--	0.006	0.005	--	0.061	0.062
24						-0.3	--	-0.007	-0.007	--	0.063	0.063
25			0.5	0.3	-0.05	-0.1	--	0.008	0.007	--	0.068	0.068
26						-0.3	--	0.002	0.002	--	0.066	0.067
27					-0.10	-0.1	--	0.006	0.005	--	0.066	0.067
28						-0.3	--	0.004	0.003	--	0.067	0.068
29				0.7	-0.05	-0.1	--	0.000	0.002	--	0.064	0.066
30						-0.3	--	0.004	0.003	--	0.066	0.066
32					-0.10	-0.1	--	-0.006	-0.006	--	0.063	0.064
32						-0.3	--	-0.002	-0.001	--	0.064	0.064

Table B1 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias			RMSE		
							MHOP	JRTP	JDSP	MHOP	JRTP	JDSP
33	0.5	0.1	0.1	0.3	-0.05	-0.1	--	-0.011	-0.011	--	0.072	0.074
34						-0.3	--	0.003	0.005	--	0.074	0.074
35					-0.10	-0.1	--	-0.013	-0.011	--	0.076	0.078
36						-0.3	--	0.000	0.001	--	0.069	0.070
37				0.7	-0.05	-0.1	--	0.001	0.001	--	0.069	0.069
38						-0.3	--	0.006	0.006	--	0.071	0.074
39					-0.10	-0.1	--	-0.008	-0.004	--	0.070	0.073
40						-0.3	--	0.003	0.005	--	0.066	0.069
41			0.5	0.3	-0.05	-0.1	--	0.007	0.004	--	0.071	0.072
42						-0.3	--	-0.003	-0.001	--	0.075	0.075
43					-0.10	-0.1	--	0.005	0.003	--	0.076	0.077
44						-0.3	--	0.008	0.009	--	0.068	0.069
45				0.7	-0.05	-0.1	--	0.002	0.002	--	0.079	0.077
46						-0.3	--	0.001	0.002	--	0.072	0.073
47					-0.10	-0.1	--	0.001	-0.001	--	0.075	0.073
48						-0.3	--	-0.006	-0.006	--	0.072	0.073
49		0.5	0.1	0.3	-0.05	-0.1	--	0.008	0.011	--	0.071	0.074
50						-0.3	--	0.005	0.001	--	0.072	0.073
51					-0.10	-0.1	--	0.003	0.006	--	0.082	0.084
52						-0.3	--	0.003	0.003	--	0.073	0.073
53				0.7	-0.05	-0.1	--	-0.015	-0.013	--	0.076	0.077
54						-0.3	--	-0.005	-0.003	--	0.078	0.079
55					-0.10	-0.1	--	-0.007	-0.004	--	0.080	0.079
56						-0.3	--	0.000	-0.002	--	0.076	0.079
57			0.5	0.3	-0.05	-0.1	--	0.001	0.002	--	0.084	0.083
58						-0.3	--	0.004	0.003	--	0.071	0.075
59					-0.10	-0.1	--	0.004	0.005	--	0.084	0.086
60						-0.3	--	-0.004	-0.008	--	0.083	0.084
61				0.7	-0.05	-0.1	--	0.010	0.008	--	0.076	0.077
62						-0.3	--	-0.010	-0.011	--	0.080	0.082
63					-0.10	-0.1	--	-0.007	-0.004	--	0.078	0.076
64						-0.3	--	-0.001	0.000	--	0.081	0.083
						Mean	--	0.000	0.000	--	0.068	0.069

Table B2

Bias and RMSE of δ_1 by simulation design conditions – polytomous attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOP	Bias JRTP	Bias JDSP	RMSE MHOP	RMSE JRTP	RMSE JDSP
1	0.1	0.1	0.1	0.3	-0.05	-0.1	-0.006	-0.024	-0.025	0.842	0.834	0.831
2						-0.3	0.046	0.024	0.023	0.795	0.781	0.785
3					-0.10	-0.1	0.033	0.016	0.010	0.815	0.810	0.806
4						-0.3	0.046	0.030	0.023	0.797	0.786	0.782
5				0.7	-0.05	-0.1	0.024	0.000	-0.001	0.743	0.732	0.731
6						-0.3	0.015	-0.007	-0.001	0.799	0.786	0.788
7					-0.10	-0.1	-0.009	-0.019	-0.017	0.764	0.756	0.755
8						-0.3	0.006	-0.003	-0.005	0.805	0.795	0.795
9			0.5	0.3	-0.05	-0.1	-0.020	-0.039	-0.038	0.764	0.751	0.757
10						-0.3	-0.001	-0.026	-0.023	0.792	0.781	0.784
11					-0.10	-0.1	0.026	0.004	0.002	0.815	0.806	0.806
12						-0.3	0.026	0.010	0.012	0.791	0.778	0.779
13				0.7	-0.05	-0.1	0.016	-0.004	0.001	0.801	0.788	0.790
14						-0.3	0.048	0.026	0.030	0.773	0.758	0.759
15					-0.10	-0.1	0.029	0.010	0.009	0.767	0.761	0.763
16						-0.3	0.020	0.006	0.007	0.810	0.801	0.801
17		0.5	0.1	0.3	-0.05	-0.1	-0.001	-0.022	-0.022	0.799	0.793	0.792
18						-0.3	0.015	0.003	0.000	0.776	0.767	0.767
19					-0.10	-0.1	-0.001	-0.019	-0.019	0.815	0.805	0.808
20						-0.3	0.020	0.001	0.004	0.794	0.793	0.793
21				0.7	-0.05	-0.1	0.018	0.005	0.007	0.775	0.773	0.774
22						-0.3	0.044	0.020	0.021	0.785	0.769	0.769
23					-0.10	-0.1	0.027	0.005	0.012	0.821	0.809	0.810
24						-0.3	-0.027	-0.040	-0.043	0.828	0.815	0.814
25			0.5	0.3	-0.05	-0.1	0.000	-0.016	-0.016	0.775	0.771	0.771
26						-0.3	0.017	0.000	-0.001	0.778	0.773	0.779
27					-0.10	-0.1	0.020	0.001	0.001	0.816	0.807	0.801
28						-0.3	-0.002	-0.020	-0.015	0.814	0.818	0.820
29				0.7	-0.05	-0.1	0.050	0.035	0.038	0.780	0.781	0.780
30						-0.3	-0.002	-0.024	-0.021	0.816	0.811	0.809
31					-0.10	-0.1	0.021	0.007	0.008	0.782	0.773	0.776
32						-0.3	0.036	0.018	0.018	0.795	0.790	0.787

Table B2 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias			RMSE		
							MHOP	JRTP	JDSP	MHOP	JRTP	JDSP
33	0.5	0.1	0.1	0.3	-0.05	-0.1	0.025	0.001	-0.003	0.811	0.791	0.790
34						-0.3	-0.007	-0.032	-0.028	0.784	0.783	0.783
35					-0.10	-0.1	0.032	0.010	0.011	0.801	0.779	0.779
36						-0.3	0.032	0.004	0.011	0.803	0.793	0.795
37				0.7	-0.05	-0.1	-0.012	-0.039	-0.043	0.811	0.780	0.784
38						-0.3	-0.004	-0.031	-0.028	0.788	0.780	0.778
39					-0.10	-0.1	-0.016	-0.041	-0.036	0.795	0.784	0.783
40						-0.3	-0.004	-0.022	-0.019	0.801	0.801	0.804
41			0.5	0.3	-0.05	-0.1	0.039	0.025	0.027	0.784	0.759	0.762
42						-0.3	0.042	0.023	0.024	0.820	0.798	0.801
43					-0.10	-0.1	0.014	0.000	-0.004	0.781	0.787	0.787
44						-0.3	0.029	0.017	0.022	0.794	0.787	0.789
45				0.7	-0.05	-0.1	0.023	-0.002	-0.002	0.803	0.782	0.783
46						-0.3	-0.004	-0.022	-0.020	0.781	0.768	0.767
47					-0.10	-0.1	0.003	-0.018	-0.016	0.797	0.777	0.777
48						-0.3	0.027	0.015	0.016	0.792	0.769	0.768
49		0.5	0.1	0.3	-0.05	-0.1	0.055	0.032	0.030	0.788	0.774	0.775
50						-0.3	-0.017	-0.039	-0.041	0.818	0.811	0.813
51					-0.10	-0.1	0.019	-0.003	0.005	0.796	0.792	0.799
52						-0.3	-0.015	-0.033	-0.031	0.798	0.804	0.802
53				0.7	-0.05	-0.1	0.019	0.004	0.003	0.810	0.794	0.797
54						-0.3	0.008	-0.002	-0.001	0.762	0.756	0.760
55					-0.10	-0.1	0.033	0.006	0.010	0.786	0.763	0.764
56						-0.3	0.010	-0.005	-0.005	0.766	0.745	0.744
57			0.5	0.3	-0.05	-0.1	0.044	0.026	0.025	0.791	0.770	0.768
58						-0.3	0.001	-0.014	-0.016	0.786	0.776	0.778
59					-0.10	-0.1	0.021	0.002	-0.001	0.802	0.794	0.796
60						-0.3	0.003	-0.008	-0.009	0.817	0.806	0.809
61				0.7	-0.05	-0.1	0.030	0.015	0.013	0.806	0.795	0.793
62						-0.3	0.044	0.018	0.020	0.796	0.770	0.770
63					-0.10	-0.1	0.043	0.016	0.015	0.853	0.839	0.839
64						-0.3	0.052	0.033	0.032	0.817	0.810	0.812
						Mean	0.017	-0.002	-0.001	0.796	0.785	0.786

Table B3

Bias and RMSE of γ_{0k2} by simulation design conditions – polytomous attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOP	Bias JRTP	Bias JDSP	RMSE MHOP	RMSE JRTP	RMSE JDSP
1	0.1	0.1	0.1	0.3	-0.05	-0.1	0.075	0.077	0.072	0.160	0.164	0.163
2						-0.3	0.066	0.072	0.071	0.173	0.173	0.174
3					-0.10	-0.1	0.103	0.112	0.104	0.176	0.176	0.176
4						-0.3	0.012	0.017	0.014	0.162	0.161	0.163
5				0.7	-0.05	-0.1	0.087	0.085	0.088	0.185	0.185	0.185
6						-0.3	0.083	0.082	0.080	0.196	0.197	0.198
7					-0.10	-0.1	0.002	0.007	0.012	0.179	0.179	0.178
8						-0.3	0.077	0.089	0.082	0.181	0.181	0.182
9			0.5	0.3	-0.05	-0.1	0.073	0.074	0.080	0.180	0.179	0.176
10						-0.3	0.072	0.076	0.078	0.188	0.186	0.187
11					-0.10	-0.1	0.045	0.052	0.057	0.181	0.182	0.182
12						-0.3	-0.033	-0.033	-0.028	0.180	0.178	0.181
13				0.7	-0.05	-0.1	0.021	0.025	0.028	0.195	0.190	0.196
14						-0.3	0.026	0.029	0.035	0.187	0.182	0.184
15					-0.10	-0.1	0.055	0.053	0.057	0.175	0.172	0.171
16						-0.3	0.018	0.018	0.016	0.182	0.179	0.182
17		0.5	0.1	0.3	-0.05	-0.1	0.033	0.037	0.045	0.173	0.173	0.174
18						-0.3	0.057	0.064	0.068	0.172	0.167	0.168
19					-0.10	-0.1	0.041	0.048	0.045	0.174	0.172	0.172
20						-0.3	0.047	0.046	0.045	0.195	0.194	0.195
21				0.7	-0.05	-0.1	0.023	0.019	0.020	0.175	0.177	0.179
22						-0.3	0.087	0.092	0.089	0.196	0.193	0.193
23					-0.10	-0.1	0.075	0.079	0.083	0.181	0.180	0.182
24						-0.3	-0.014	-0.012	-0.018	0.168	0.166	0.169
25			0.5	0.3	-0.05	-0.1	0.032	0.043	0.049	0.177	0.177	0.176
26						-0.3	0.022	0.025	0.025	0.190	0.187	0.187
27					-0.10	-0.1	-0.012	-0.013	-0.016	0.193	0.193	0.195
28						-0.3	0.048	0.036	0.041	0.181	0.183	0.187
29				0.7	-0.05	-0.1	0.031	0.039	0.043	0.182	0.183	0.183
30						-0.3	0.059	0.062	0.059	0.174	0.171	0.175
31					-0.10	-0.1	0.071	0.073	0.072	0.191	0.187	0.185
32						-0.3	0.029	0.033	0.045	0.184	0.182	0.185

Table B3 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOP	Bias JRTP	Bias JDSP	RMSE MHOP	RMSE JRTP	RMSE JDSP
33	0.5	0.1	0.1	0.3	-0.05	-0.1	0.026	0.027	0.028	0.195	0.193	0.196
34						-0.3	0.068	0.070	0.071	0.182	0.182	0.186
35					-0.10	-0.1	-0.001	-0.007	0.004	0.170	0.167	0.168
36						-0.3	0.073	0.078	0.073	0.190	0.194	0.197
37				0.7	-0.05	-0.1	0.018	0.017	0.013	0.178	0.177	0.181
38						-0.3	0.055	0.069	0.071	0.179	0.179	0.180
39					-0.10	-0.1	0.017	0.013	0.019	0.189	0.185	0.188
40						-0.3	0.046	0.044	0.049	0.187	0.181	0.182
41			0.5	0.3	-0.05	-0.1	0.009	0.010	0.006	0.181	0.181	0.181
42						-0.3	0.043	0.044	0.042	0.191	0.187	0.192
43					-0.10	-0.1	0.067	0.073	0.081	0.172	0.170	0.171
44						-0.3	0.041	0.035	0.023	0.185	0.189	0.186
45				0.7	-0.05	-0.1	0.059	0.063	0.060	0.173	0.180	0.177
46						-0.3	0.069	0.067	0.069	0.176	0.176	0.174
47					-0.10	-0.1	0.053	0.069	0.060	0.169	0.172	0.169
48						-0.3	0.030	0.043	0.040	0.194	0.193	0.193
49		0.5	0.1	0.3	-0.05	-0.1	0.073	0.078	0.082	0.163	0.161	0.163
50						-0.3	0.043	0.058	0.059	0.196	0.190	0.190
51					-0.10	-0.1	0.060	0.068	0.067	0.175	0.174	0.177
52						-0.3	0.055	0.050	0.055	0.177	0.181	0.182
53				0.7	-0.05	-0.1	0.022	0.014	0.019	0.186	0.185	0.186
54						-0.3	0.062	0.047	0.047	0.176	0.176	0.177
55					-0.10	-0.1	0.023	0.022	0.027	0.180	0.175	0.176
56						-0.3	0.039	0.036	0.034	0.171	0.163	0.162
57			0.5	0.3	-0.05	-0.1	0.101	0.097	0.104	0.179	0.178	0.178
58						-0.3	0.052	0.051	0.054	0.182	0.183	0.185
59					-0.10	-0.1	0.056	0.046	0.050	0.190	0.189	0.188
60						-0.3	0.105	0.105	0.107	0.184	0.185	0.186
61				0.7	-0.05	-0.1	-0.005	-0.007	-0.006	0.191	0.189	0.187
62						-0.3	0.047	0.048	0.043	0.182	0.178	0.178
63					-0.10	-0.1	0.031	0.026	0.031	0.176	0.175	0.178
64						-0.3	0.076	0.075	0.077	0.178	0.174	0.173
						Mean	0.046	0.047	0.048	0.181	0.180	0.181

Table B4

Bias and RMSE of γ_{0k3} by simulation design conditions – polytomous attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOP	Bias JRTP	Bias JDSP	RMSE MHOP	RMSE JRTP	RMSE JDSP
1	0.1	0.1	0.1	0.3	-0.05	-0.1	-0.037	-0.039	-0.041	0.201	0.202	0.204
2						-0.3	-0.011	-0.013	-0.010	0.185	0.185	0.190
3					-0.10	-0.1	-0.023	-0.023	-0.027	0.190	0.189	0.190
4						-0.3	-0.009	-0.012	-0.010	0.195	0.198	0.197
5				0.7	-0.05	-0.1	0.030	0.034	0.032	0.205	0.206	0.209
6						-0.3	-0.070	-0.063	-0.063	0.197	0.192	0.192
7					-0.10	-0.1	-0.013	-0.017	-0.027	0.201	0.202	0.204
8						-0.3	-0.062	-0.070	-0.074	0.183	0.185	0.183
9			0.5	0.3	-0.05	-0.1	-0.029	-0.031	-0.028	0.207	0.204	0.204
10						-0.3	0.032	0.024	0.029	0.203	0.199	0.200
11					-0.10	-0.1	-0.013	-0.007	0.000	0.196	0.199	0.197
12						-0.3	0.019	0.022	0.015	0.189	0.189	0.191
13				0.7	-0.05	-0.1	0.047	0.058	0.055	0.205	0.207	0.208
14						-0.3	-0.046	-0.052	-0.048	0.189	0.191	0.189
15					-0.10	-0.1	0.043	0.050	0.047	0.204	0.202	0.202
16						-0.3	0.026	0.020	0.031	0.209	0.212	0.211
17		0.5	0.1	0.3	-0.05	-0.1	-0.003	-0.010	-0.004	0.184	0.182	0.185
18						-0.3	0.003	0.003	0.005	0.204	0.203	0.203
19					-0.10	-0.1	0.013	0.016	0.026	0.199	0.201	0.199
20						-0.3	-0.030	-0.022	-0.025	0.215	0.216	0.218
21				0.7	-0.05	-0.1	0.023	0.030	0.036	0.197	0.197	0.197
22						-0.3	-0.016	-0.016	-0.018	0.191	0.192	0.192
23					-0.10	-0.1	-0.006	0.000	-0.005	0.181	0.181	0.181
24						-0.3	0.020	0.026	0.018	0.190	0.191	0.192
25			0.5	0.3	-0.05	-0.1	-0.010	-0.011	-0.004	0.196	0.195	0.197
26						-0.3	-0.037	-0.034	-0.042	0.204	0.201	0.200
27					-0.10	-0.1	0.023	0.023	0.024	0.201	0.202	0.199
28						-0.3	-0.019	-0.013	-0.010	0.198	0.198	0.201
29				0.7	-0.05	-0.1	-0.027	-0.033	-0.026	0.184	0.183	0.186
30						-0.3	-0.039	-0.034	-0.037	0.198	0.201	0.201
31					-0.10	-0.1	-0.008	-0.008	-0.009	0.190	0.187	0.185
32						-0.3	-0.009	-0.009	-0.003	0.180	0.181	0.183

Table B4 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOP	Bias JRTP	Bias JDSP	RMSE MHOP	RMSE JRTP	RMSE JDSP
33	0.5	0.1	0.1	0.3	-0.05	-0.1	-0.055	-0.055	-0.052	0.212	0.212	0.211
34						-0.3	-0.014	-0.016	-0.017	0.185	0.188	0.187
35					-0.10	-0.1	0.004	0.003	0.000	0.191	0.187	0.187
36						-0.3	-0.054	-0.049	-0.050	0.193	0.192	0.193
37				0.7	-0.05	-0.1	-0.061	-0.049	-0.049	0.187	0.183	0.184
38						-0.3	0.035	0.024	0.023	0.190	0.188	0.186
39					-0.10	-0.1	-0.009	-0.013	-0.012	0.183	0.180	0.179
40						-0.3	-0.036	-0.030	-0.026	0.190	0.191	0.192
41			0.5	0.3	-0.05	-0.1	-0.040	-0.041	-0.051	0.181	0.185	0.191
42						-0.3	-0.027	-0.024	-0.024	0.193	0.192	0.191
43					-0.10	-0.1	0.022	0.020	0.033	0.204	0.199	0.201
44						-0.3	-0.006	-0.011	-0.017	0.207	0.209	0.209
45				0.7	-0.05	-0.1	-0.064	-0.062	-0.062	0.184	0.181	0.181
46						-0.3	-0.038	-0.030	-0.035	0.205	0.202	0.204
47					-0.10	-0.1	0.000	0.008	0.002	0.199	0.195	0.194
48						-0.3	0.006	0.010	0.004	0.198	0.199	0.200
49		0.5	0.1	0.3	-0.05	-0.1	-0.015	-0.014	-0.010	0.199	0.197	0.198
50						-0.3	-0.077	-0.079	-0.071	0.188	0.189	0.188
51					-0.10	-0.1	-0.057	-0.049	-0.047	0.179	0.178	0.179
52						-0.3	-0.013	-0.008	-0.009	0.202	0.201	0.201
53				0.7	-0.05	-0.1	-0.008	-0.019	-0.009	0.185	0.187	0.190
54						-0.3	-0.021	-0.011	-0.010	0.185	0.183	0.181
55					-0.10	-0.1	0.033	0.024	0.029	0.208	0.204	0.203
56						-0.3	-0.012	-0.019	-0.017	0.210	0.202	0.201
57			0.5	0.3	-0.05	-0.1	-0.075	-0.073	-0.072	0.205	0.203	0.206
58						-0.3	0.014	0.012	0.008	0.194	0.194	0.195
59					-0.10	-0.1	-0.089	-0.083	-0.078	0.205	0.202	0.205
60						-0.3	-0.012	-0.010	-0.005	0.199	0.199	0.199
61				0.7	-0.05	-0.1	0.032	0.035	0.027	0.209	0.207	0.207
62						-0.3	-0.080	-0.080	-0.077	0.187	0.186	0.187
63					-0.10	-0.1	-0.036	-0.025	-0.027	0.195	0.196	0.198
64						-0.3	-0.058	-0.057	-0.050	0.195	0.197	0.195
						Mean	-0.016	-0.015	-0.015	0.196	0.195	0.196

Table B5

Bias and RMSE of γ_{0k4} by simulation design conditions – polytomous attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias MHOP	Bias JRTP	Bias JDSP	RMSE MHOP	RMSE JRTP	RMSE JDSP
1	0.1	0.1	0.1	0.3	-0.05	-0.1	0.041	0.042	0.040	0.195	0.193	0.190
2						-0.3	-0.087	-0.085	-0.073	0.194	0.196	0.197
3					-0.10	-0.1	-0.087	-0.093	-0.092	0.187	0.185	0.186
4						-0.3	-0.092	-0.092	-0.092	0.201	0.202	0.201
5				0.7	-0.05	-0.1	-0.132	-0.134	-0.128	0.193	0.196	0.197
6						-0.3	-0.042	-0.044	-0.041	0.178	0.180	0.178
7					-0.10	-0.1	-0.057	-0.057	-0.051	0.191	0.194	0.194
8						-0.3	-0.075	-0.080	-0.081	0.175	0.177	0.178
9			0.5	0.3	-0.05	-0.1	-0.003	-0.007	-0.007	0.192	0.192	0.194
10						-0.3	-0.086	-0.079	-0.077	0.170	0.170	0.171
11					-0.10	-0.1	-0.092	-0.105	-0.089	0.192	0.193	0.192
12						-0.3	-0.036	-0.034	-0.034	0.184	0.188	0.190
13				0.7	-0.05	-0.1	-0.111	-0.130	-0.128	0.197	0.196	0.198
14						-0.3	-0.092	-0.099	-0.093	0.202	0.203	0.205
15					-0.10	-0.1	-0.020	-0.019	-0.006	0.194	0.190	0.191
16						-0.3	-0.094	-0.094	-0.096	0.207	0.205	0.206
17		0.5	0.1	0.3	-0.05	-0.1	-0.060	-0.053	-0.054	0.206	0.203	0.205
18						-0.3	-0.034	-0.046	-0.037	0.203	0.205	0.202
19					-0.10	-0.1	-0.149	-0.151	-0.148	0.185	0.185	0.186
20						-0.3	-0.028	-0.033	-0.038	0.186	0.192	0.189
21				0.7	-0.05	-0.1	-0.054	-0.063	-0.061	0.183	0.183	0.186
22						-0.3	-0.057	-0.055	-0.057	0.206	0.208	0.210
23					-0.10	-0.1	-0.051	-0.055	-0.056	0.185	0.188	0.187
24						-0.3	-0.073	-0.088	-0.084	0.200	0.201	0.203
25			0.5	0.3	-0.05	-0.1	-0.044	-0.051	-0.052	0.199	0.205	0.205
26						-0.3	-0.026	-0.026	-0.024	0.180	0.179	0.179
27					-0.10	-0.1	-0.066	-0.068	-0.064	0.177	0.178	0.175
28						-0.3	-0.025	-0.025	-0.017	0.185	0.186	0.186
29				0.7	-0.05	-0.1	-0.128	-0.139	-0.130	0.211	0.210	0.211
30						-0.3	-0.093	-0.104	-0.103	0.187	0.190	0.187
31					-0.10	-0.1	-0.021	-0.030	-0.024	0.197	0.193	0.193
32						-0.3	-0.090	-0.095	-0.085	0.169	0.171	0.168

Table B5 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias			RMSE		
							MHOP	**JRTP**	**JDSP**	**MHOP**	**JRTP**	**JDSP**
33	0.5	0.1	0.1	0.3	-0.05	-0.1	-0.101	-0.103	-0.094	0.180	0.179	0.487
34						-0.3	-0.145	-0.141	-0.143	0.183	0.184	0.506
35					-0.10	-0.1	-0.078	-0.084	-0.076	0.202	0.201	0.559
36						-0.3	-0.073	-0.070	-0.071	0.191	0.193	0.524
37				0.7	-0.05	-0.1	-0.072	-0.058	-0.054	0.177	0.178	0.490
38						-0.3	-0.077	-0.082	-0.079	0.195	0.200	0.556
39					-0.10	-0.1	-0.069	-0.080	-0.084	0.183	0.183	0.489
40						-0.3	-0.011	-0.015	-0.007	0.191	0.199	0.544
41			0.5	0.3	-0.05	-0.1	-0.050	-0.046	-0.048	0.202	0.207	0.561
42						-0.3	-0.140	-0.148	-0.147	0.187	0.190	0.515
43					-0.10	-0.1	-0.053	-0.052	-0.054	0.189	0.189	0.522
44						-0.3	-0.045	-0.045	-0.053	0.197	0.196	0.535
45				0.7	-0.05	-0.1	-0.055	-0.048	-0.047	0.187	0.194	0.537
46						-0.3	-0.058	-0.068	-0.065	0.181	0.179	0.503
47					-0.10	-0.1	-0.127	-0.123	-0.125	0.202	0.199	0.550
48						-0.3	-0.070	-0.074	-0.077	0.202	0.202	0.558
49		0.5	0.1	0.3	-0.05	-0.1	-0.023	-0.026	-0.025	0.194	0.195	0.532
50						-0.3	-0.105	-0.102	-0.104	0.199	0.192	0.539
51					-0.10	-0.1	-0.051	-0.050	-0.051	0.188	0.189	0.520
52						-0.3	-0.108	-0.112	-0.113	0.181	0.182	0.500
53				0.7	-0.05	-0.1	-0.041	-0.051	-0.050	0.210	0.210	0.571
54						-0.3	-0.091	-0.085	-0.090	0.201	0.202	0.555
55					-0.10	-0.1	-0.035	-0.029	-0.027	0.184	0.182	0.498
56						-0.3	-0.017	-0.011	-0.016	0.187	0.188	0.525
57			0.5	0.3	-0.05	-0.1	-0.081	-0.081	-0.079	0.200	0.198	0.557
58						-0.3	-0.095	-0.091	-0.096	0.193	0.193	0.539
59					-0.10	-0.1	-0.099	-0.093	-0.090	0.191	0.190	0.522
60						-0.3	-0.041	-0.043	-0.043	0.192	0.193	0.529
61				0.7	-0.05	-0.1	-0.091	-0.093	-0.095	0.200	0.201	0.551
62						-0.3	-0.145	-0.136	-0.132	0.201	0.200	0.551
63					-0.10	-0.1	-0.026	-0.045	-0.036	0.194	0.195	0.535
64						-0.3	-0.054	-0.058	-0.056	0.198	0.193	0.528
						Mean	-0.069	-0.071	-0.069	0.192	0.192	0.193

Table B6

Bias and RMSE of γ_1 by simulation design conditions – polytomous attributes

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	**Bias**			**RMSE**		
							MHOP	**JRTP**	**JDSP**	**MHOP**	**JRTP**	**JDSP**
1	0.1	0.1	0.1	0.3	-0.05	-0.1	0.188	0.195	0.200	0.162	0.163	0.166
2						-0.3	0.139	0.145	0.156	0.137	0.138	0.142
3					-0.10	-0.1	0.064	0.061	0.076	0.137	0.138	0.144
4						-0.3	0.103	0.107	0.118	0.145	0.145	0.148
5				0.7	-0.05	-0.1	0.119	0.119	0.126	0.141	0.142	0.143
6						-0.3	0.168	0.170	0.180	0.147	0.144	0.147
7					-0.10	-0.1	0.140	0.138	0.143	0.146	0.146	0.149
8						-0.3	0.076	0.072	0.081	0.138	0.141	0.142
9			0.5	0.3	-0.05	-0.1	0.140	0.141	0.146	0.152	0.151	0.156
10						-0.3	0.092	0.098	0.099	0.134	0.136	0.135
11					-0.10	-0.1	0.099	0.098	0.108	0.144	0.142	0.147
12						-0.3	0.099	0.099	0.108	0.142	0.143	0.144
13				0.7	-0.05	-0.1	0.146	0.146	0.149	0.154	0.151	0.152
14						-0.3	0.121	0.122	0.129	0.145	0.148	0.152
15					-0.10	-0.1	0.134	0.134	0.145	0.138	0.135	0.139
16						-0.3	0.088	0.087	0.093	0.142	0.139	0.142
17		0.5	0.1	0.3	-0.05	-0.1	0.107	0.115	0.120	0.137	0.139	0.143
18						-0.3	0.131	0.131	0.140	0.143	0.143	0.146
19					-0.10	-0.1	0.086	0.087	0.097	0.140	0.137	0.141
20						-0.3	0.112	0.119	0.118	0.138	0.140	0.141
21				0.7	-0.05	-0.1	0.078	0.075	0.079	0.128	0.127	0.131
22						-0.3	0.100	0.100	0.106	0.139	0.139	0.140
23					-0.10	-0.1	0.130	0.133	0.130	0.138	0.141	0.143
24						-0.3	0.126	0.123	0.131	0.140	0.141	0.143
25			0.5	0.3	-0.05	-0.1	0.137	0.139	0.141	0.139	0.139	0.141
26						-0.3	0.132	0.136	0.140	0.143	0.146	0.147
27					-0.10	-0.1	0.102	0.103	0.110	0.135	0.136	0.138
28						-0.3	0.143	0.147	0.152	0.145	0.145	0.150
29				0.7	-0.05	-0.1	0.065	0.064	0.077	0.130	0.127	0.132
30						-0.3	0.118	0.124	0.125	0.137	0.138	0.137
31					-0.10	-0.1	0.146	0.144	0.154	0.144	0.143	0.148
32						-0.3	0.095	0.101	0.104	0.131	0.131	0.133

Table B6 (cont'd)

Cond	$\sigma^2_{\tau_0}$	$\sigma^2_{\tau_1}$	$\sigma^2_{\tau_2}$	$\rho_{\theta\tau_0}$	$\rho_{\theta\tau_2}$	$\rho_{\theta\tau_2}$	Bias			RMSE		
							MHOP	JRTP	JDSP	MHOP	JRTP	JDSP
33	0.5	0.1	0.1	0.3	-0.05	-0.1	0.121	0.125	0.136	0.128	0.130	0.132
34						-0.3	0.095	0.104	0.108	0.135	0.135	0.140
35					-0.10	-0.1	0.136	0.141	0.146	0.162	0.163	0.163
36						-0.3	0.097	0.100	0.103	0.135	0.134	0.134
37				0.7	-0.05	-0.1	0.106	0.105	0.116	0.146	0.143	0.147
38						-0.3	0.108	0.106	0.107	0.143	0.140	0.140
39					-0.10	-0.1	0.156	0.153	0.155	0.151	0.148	0.149
40						-0.3	0.154	0.148	0.151	0.144	0.147	0.148
41			0.5	0.3	-0.05	-0.1	0.108	0.114	0.115	0.144	0.148	0.148
42						-0.3	0.073	0.075	0.080	0.139	0.141	0.141
43					-0.10	-0.1	0.125	0.127	0.133	0.142	0.145	0.146
44						-0.3	0.109	0.113	0.113	0.144	0.142	0.145
45				0.7	-0.05	-0.1	0.094	0.091	0.104	0.141	0.144	0.149
46						-0.3	0.131	0.125	0.129	0.141	0.139	0.141
47					-0.10	-0.1	0.134	0.136	0.139	0.144	0.143	0.142
48						-0.3	0.155	0.150	0.148	0.150	0.150	0.150
49		0.5	0.1	0.3	-0.05	-0.1	0.095	0.099	0.102	0.139	0.141	0.142
50						-0.3	0.125	0.132	0.133	0.142	0.141	0.142
51					-0.10	-0.1	0.123	0.131	0.132	0.147	0.149	0.151
52						-0.3	0.122	0.125	0.122	0.143	0.143	0.141
53				0.7	-0.05	-0.1	0.100	0.095	0.103	0.147	0.148	0.150
54						-0.3	0.105	0.105	0.111	0.138	0.136	0.141
55					-0.10	-0.1	0.125	0.127	0.126	0.137	0.138	0.137
56						-0.3	0.111	0.117	0.122	0.141	0.142	0.146
57			0.5	0.3	-0.05	-0.1	0.151	0.150	0.157	0.146	0.145	0.151
58						-0.3	0.093	0.099	0.101	0.136	0.134	0.140
59					-0.10	-0.1	0.112	0.115	0.125	0.144	0.140	0.143
60						-0.3	0.142	0.143	0.147	0.141	0.141	0.142
61				0.7	-0.05	-0.1	0.122	0.119	0.124	0.142	0.135	0.137
62						-0.3	0.068	0.077	0.087	0.146	0.149	0.152
63					-0.10	-0.1	0.094	0.091	0.098	0.146	0.146	0.149
64						-0.3	0.128	0.133	0.138	0.146	0.146	0.147
						Mean	0.116	0.118	0.123	0.142	0.142	0.144

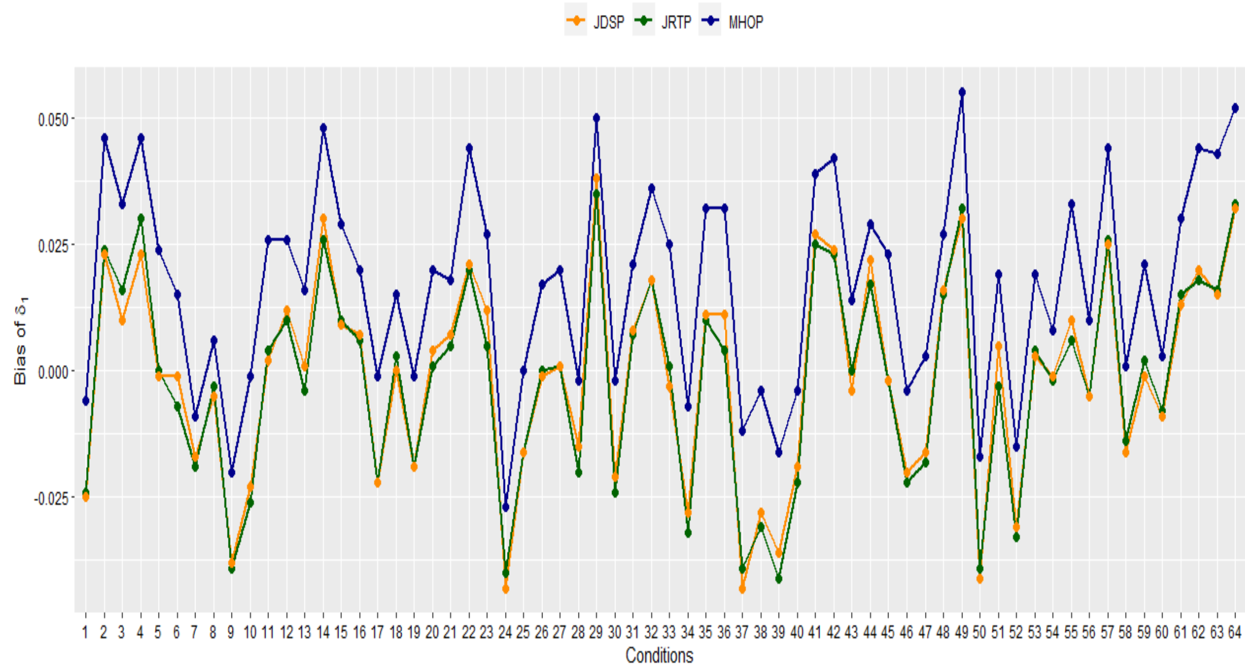


Figure B1 Bias of δ_1 across simulation conditions – polytomous attribute configuration

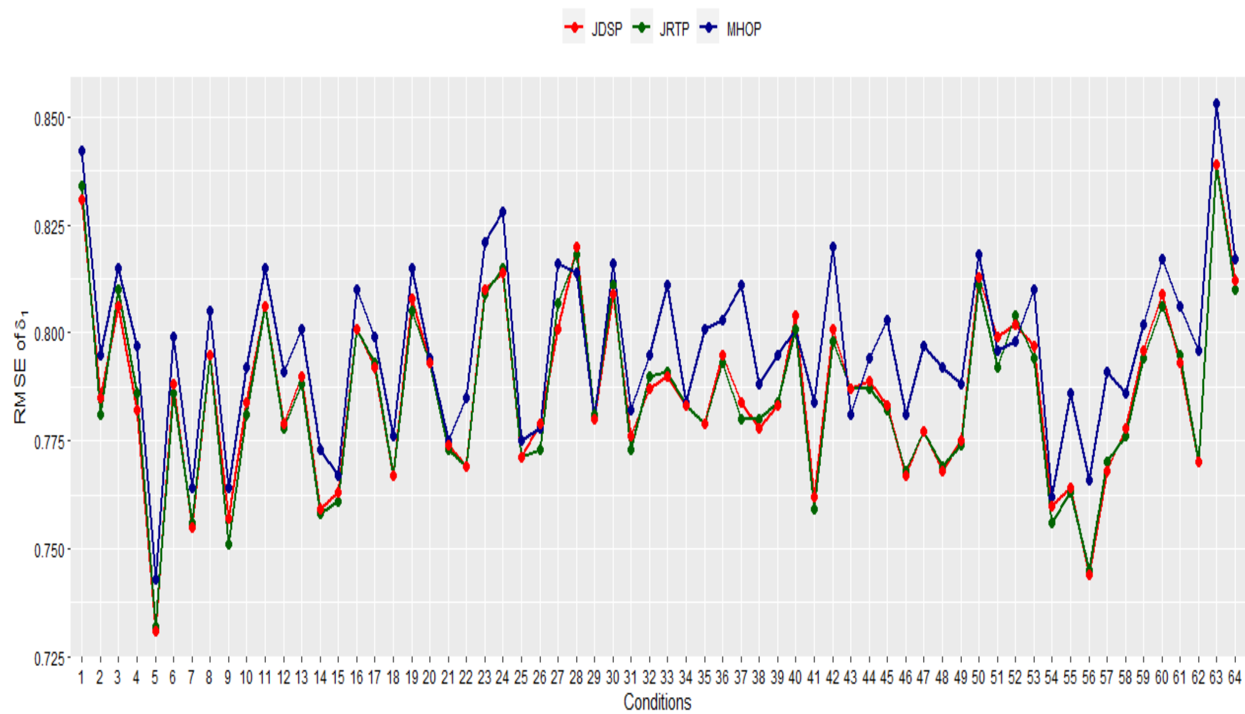


Figure B2 RMSE of δ_1 across simulation conditions – polytomous attribute configuration

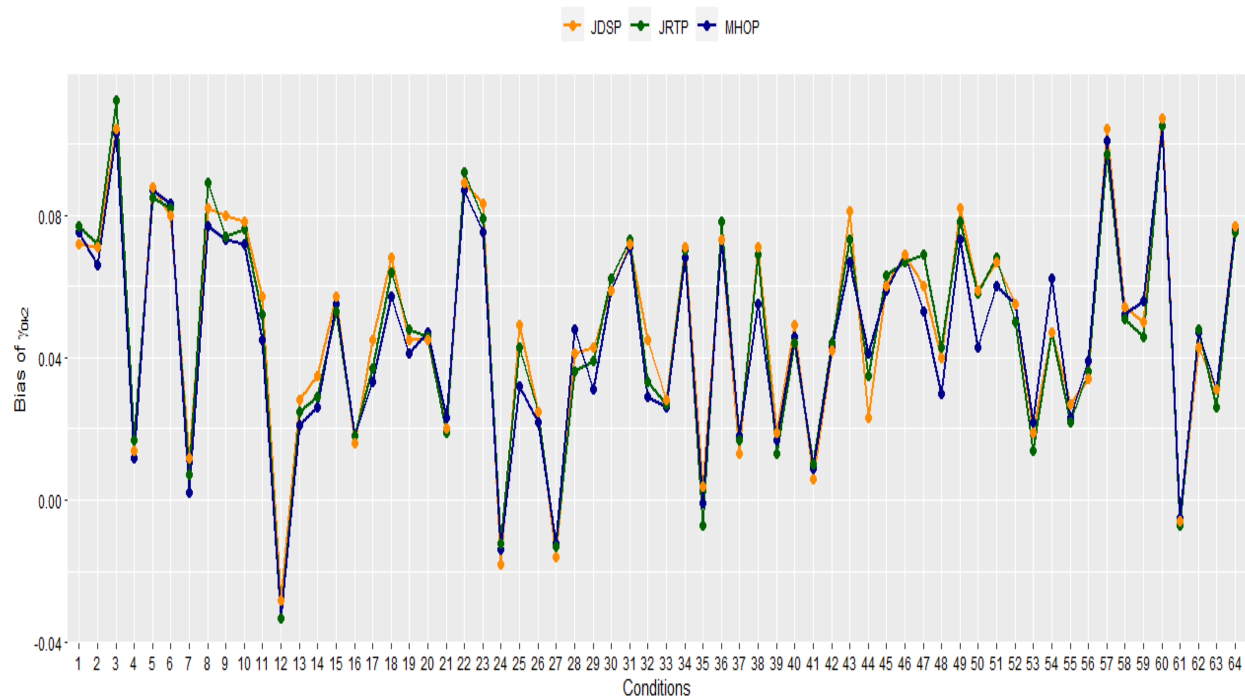


Figure B3 Bias of γ_{0k2} across simulation conditions – polytomous attribute configuration

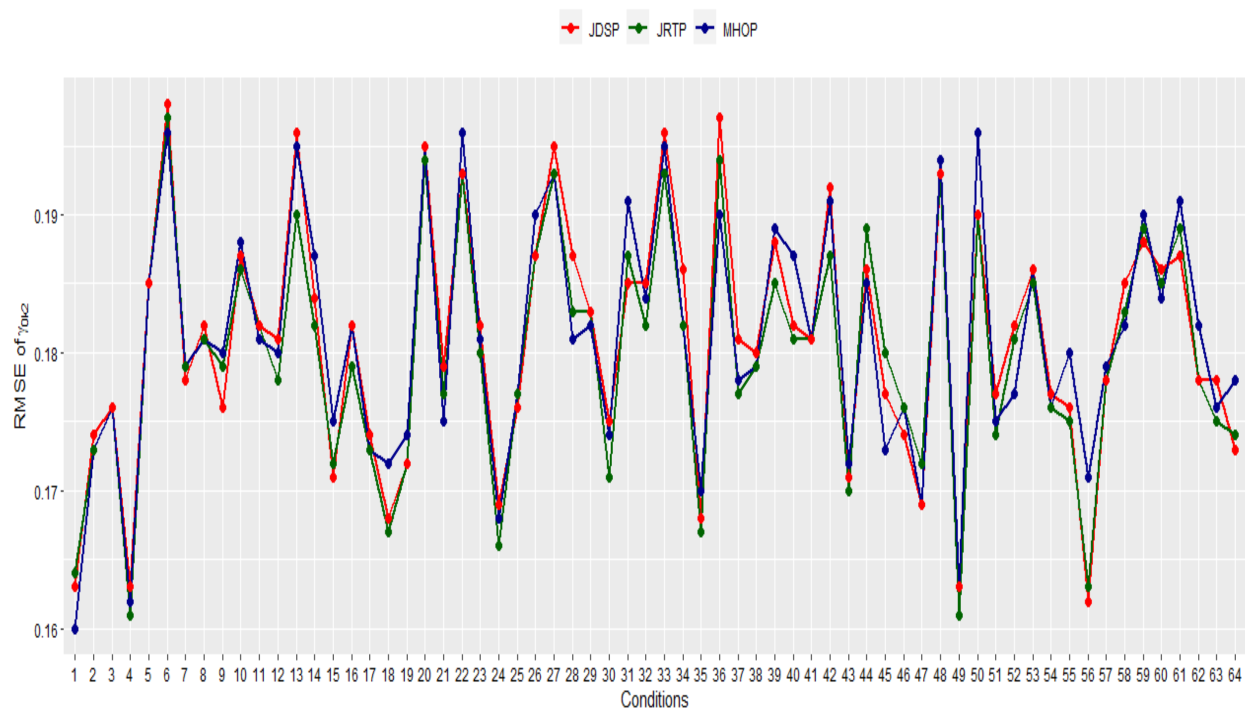


Figure B4 RMSE of γ_{0k2} across simulation conditions – polytomous attribute configuration

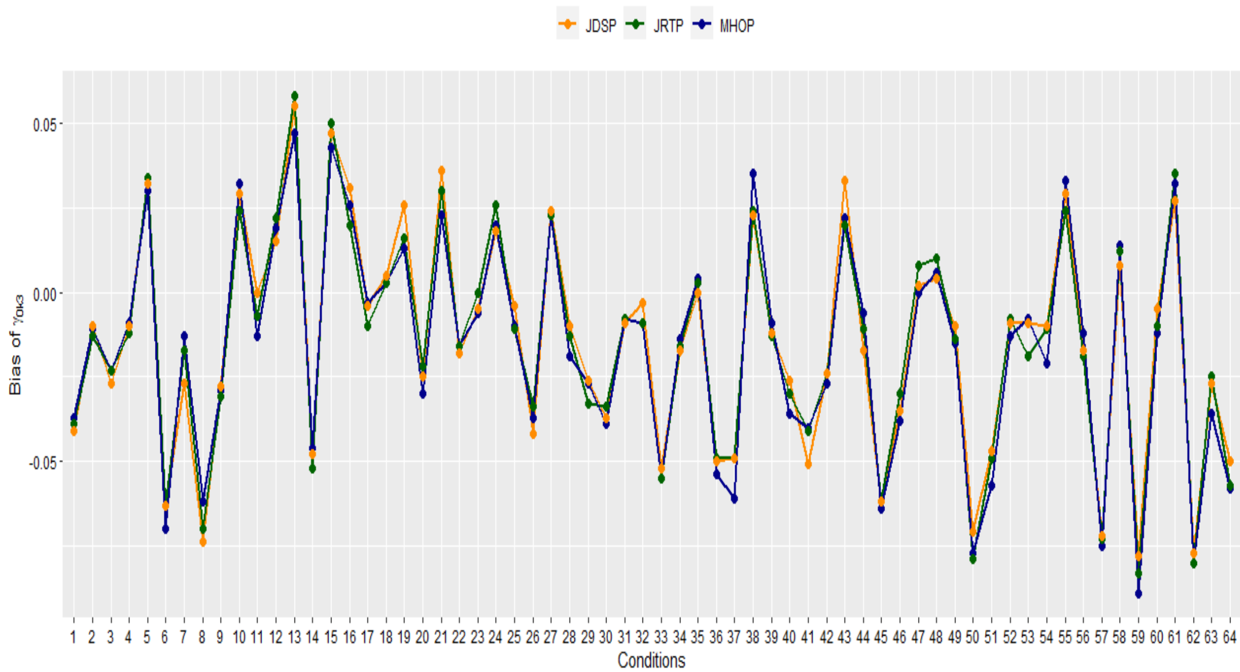


Figure B5 Bias of γ_{0k3} across simulation conditions – polytomous attribute configuration

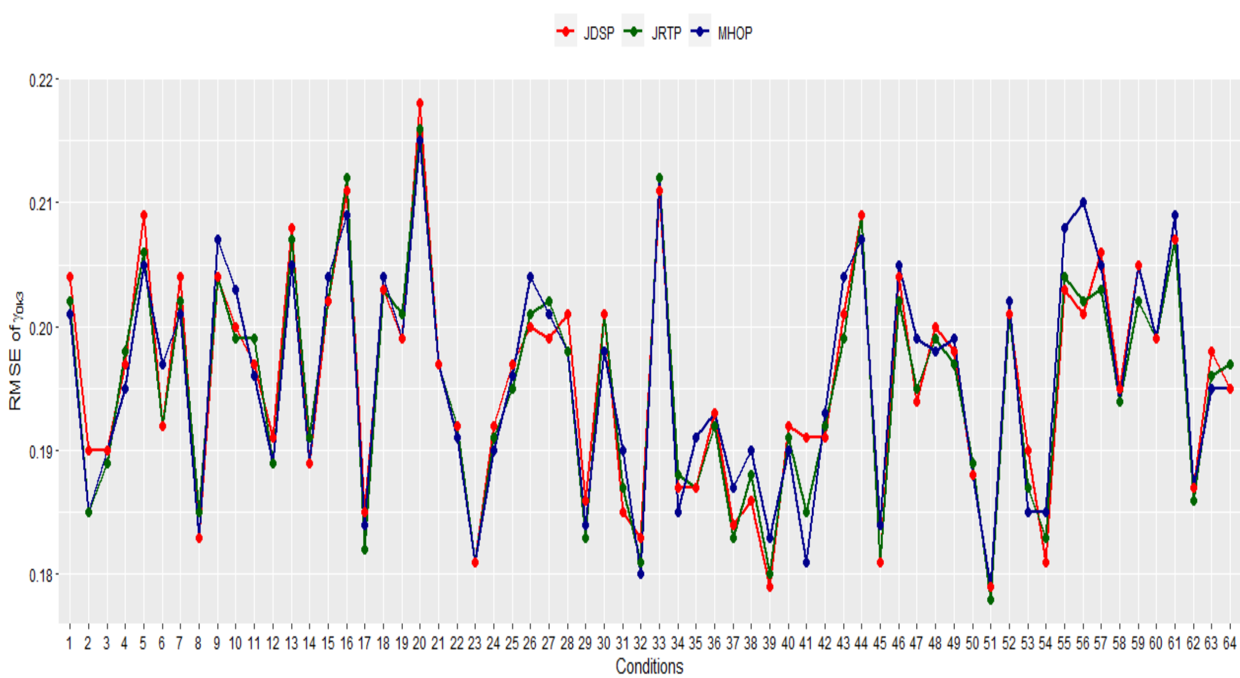


Figure B6 RMSE of γ_{0k3} across simulation conditions – polytomous attribute configuration

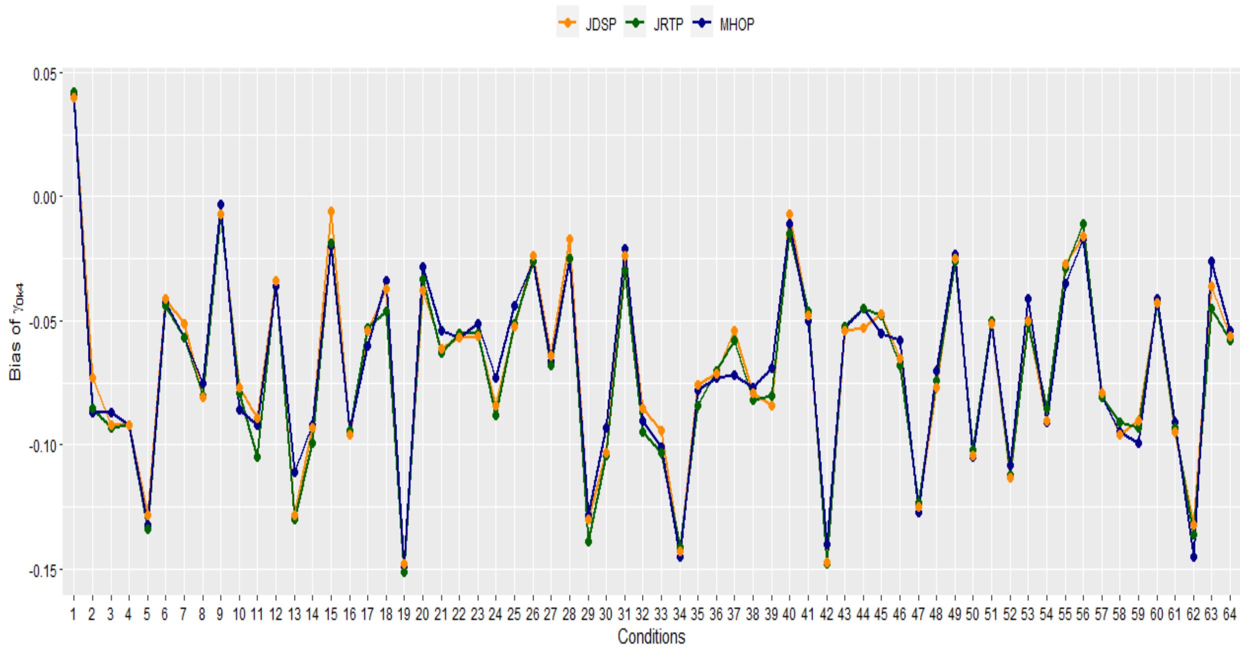


Figure B7 Bias of γ_{0k4} across simulation conditions – polytomous attribute configuration

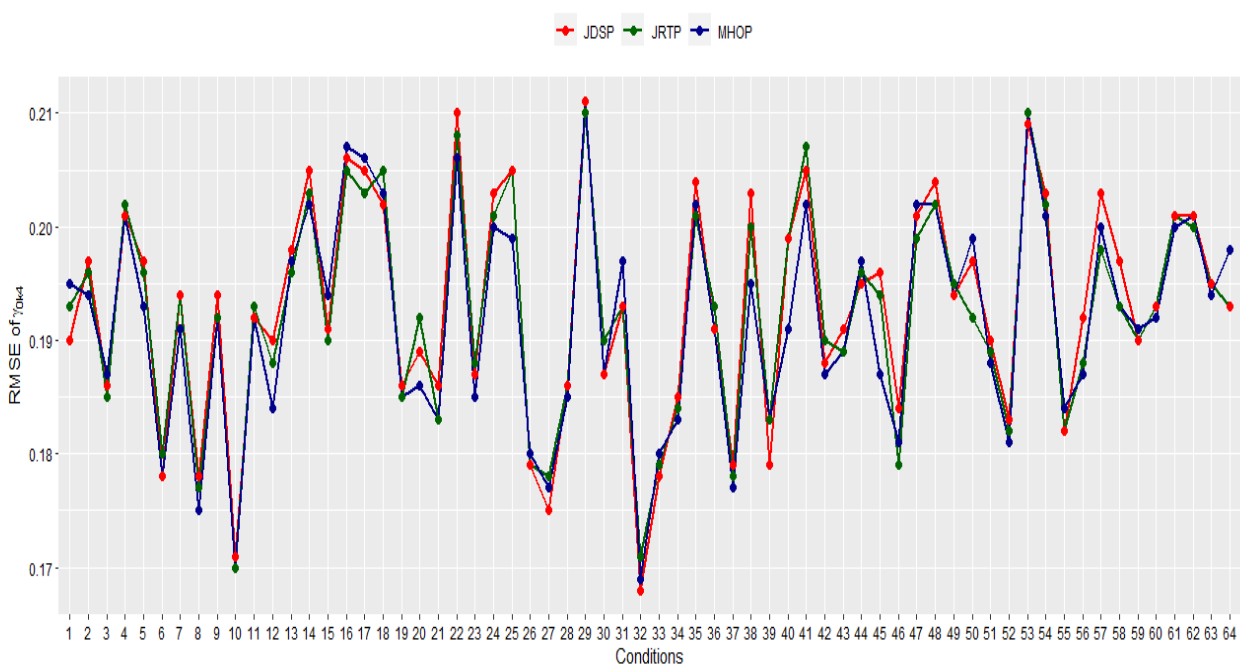


Figure B8 RMSE of γ_{0k4} across simulation conditions – polytomous attribute configuration

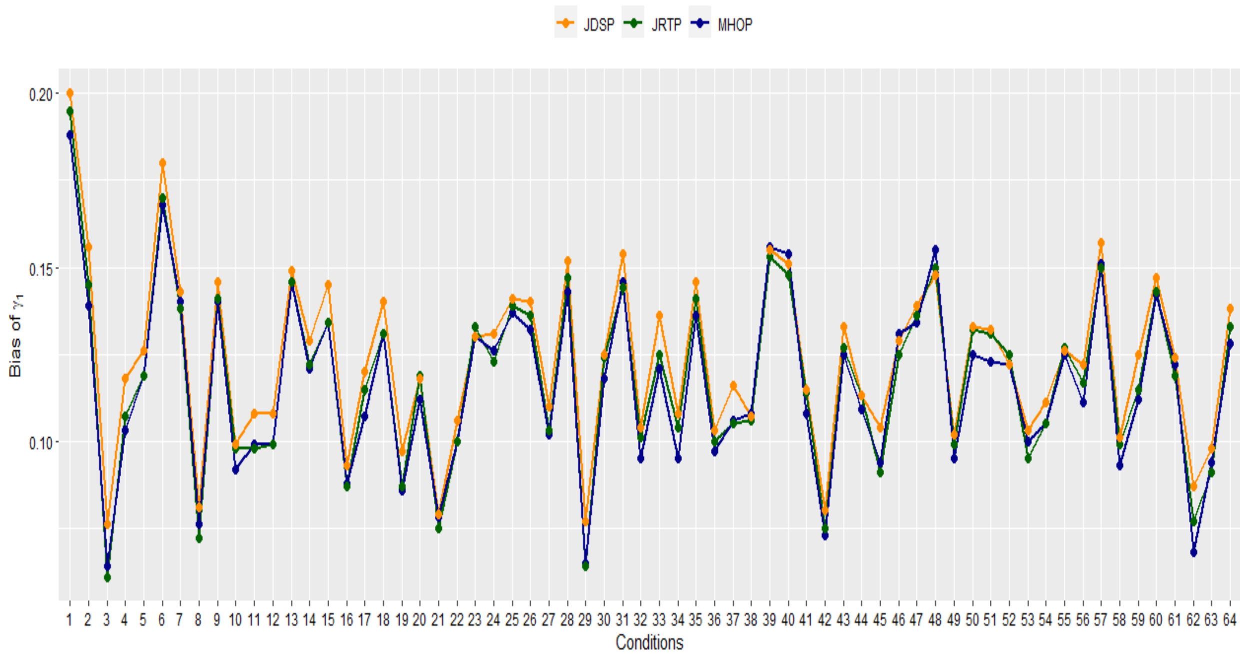


Figure B9 Bias of γ_1 across simulation conditions – polytomous attribute configuration

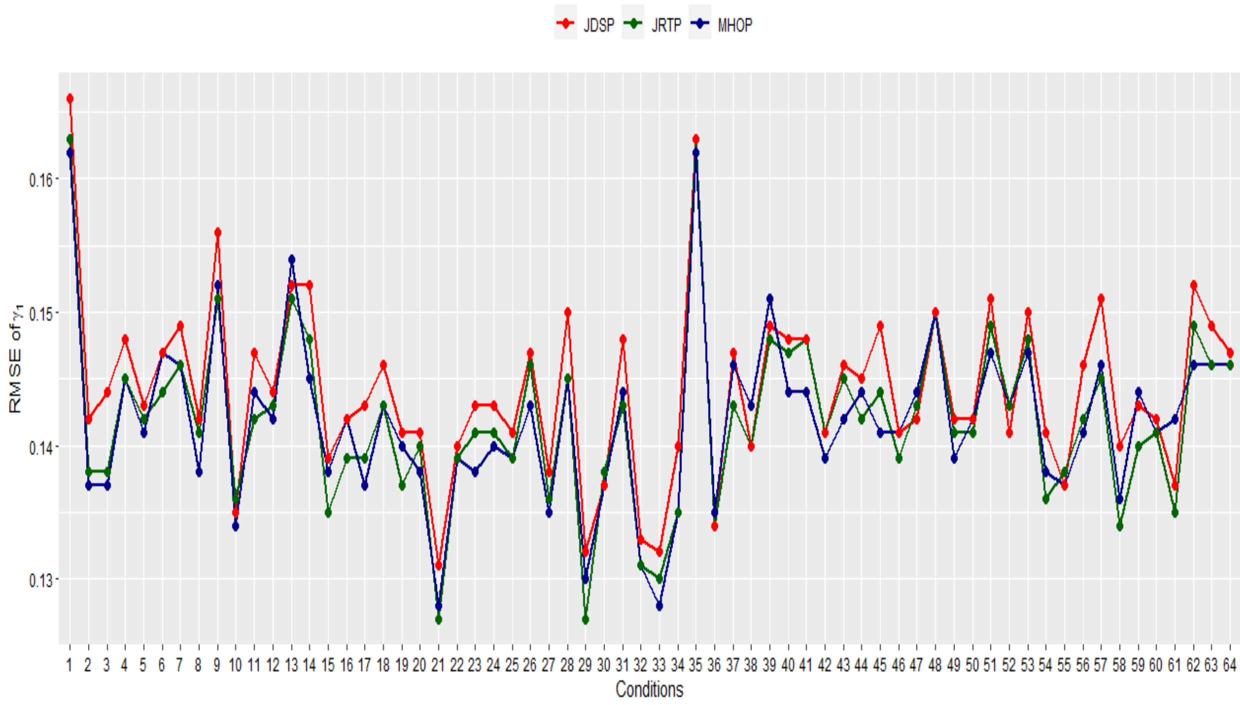


Figure B10 RMSE of γ_1 across simulation conditions – polytomous attribute configuration

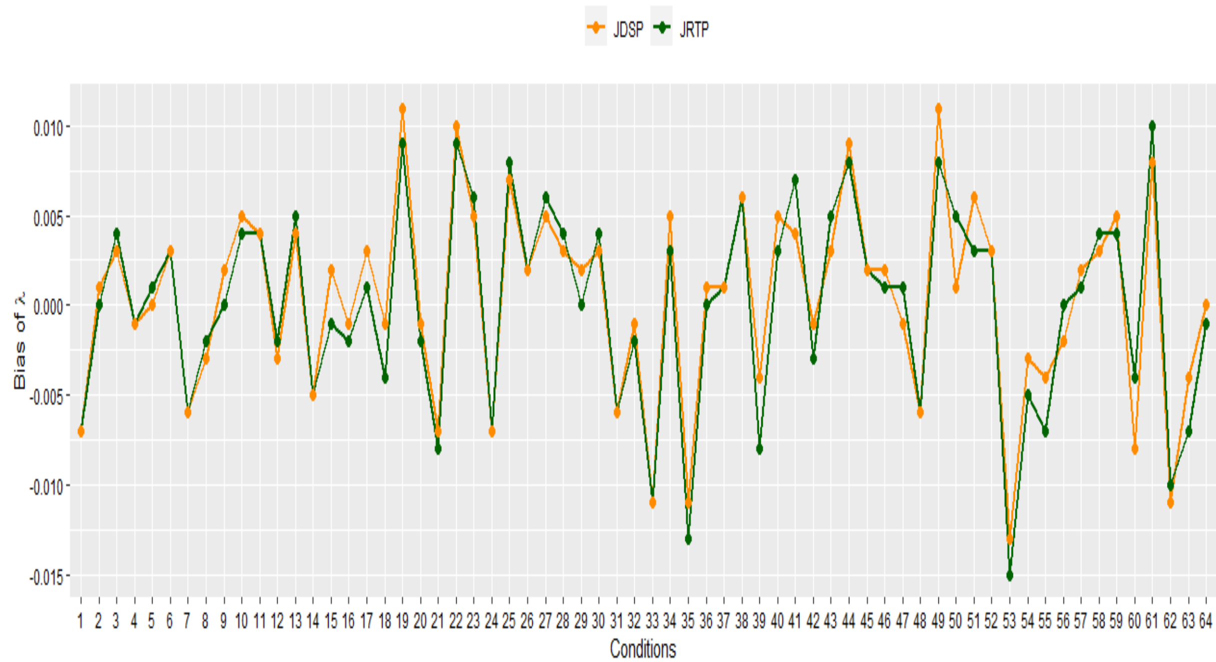


Figure B11 Bias of λ across simulation conditions – polytomous attribute configuration

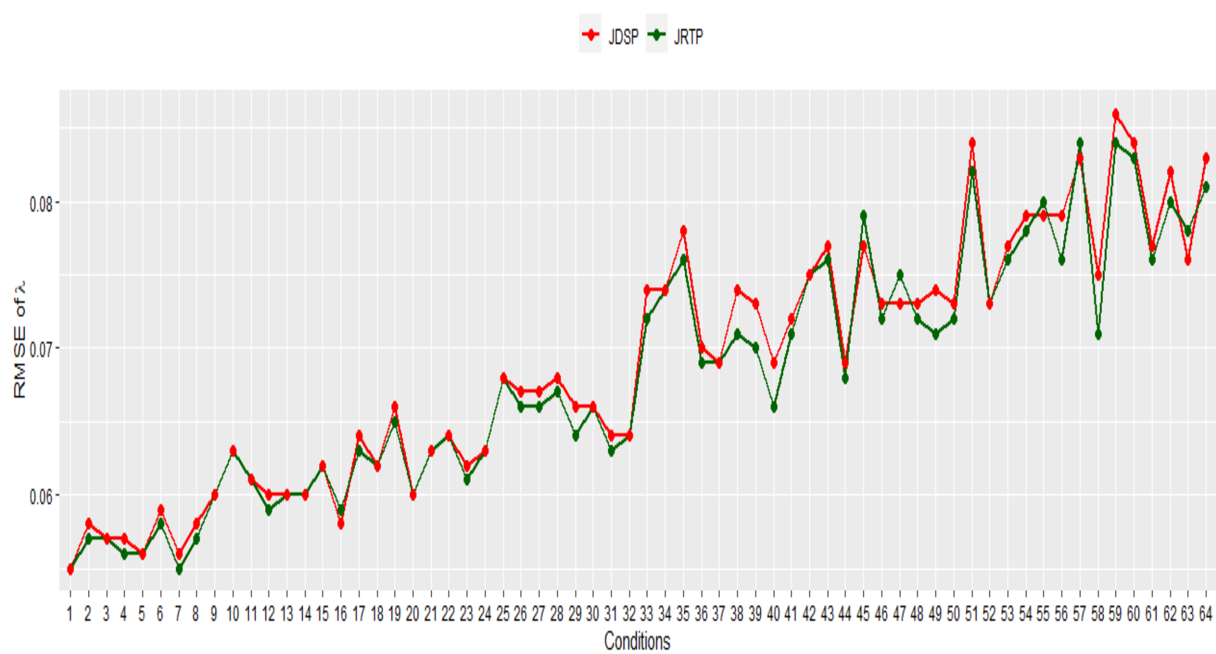


Figure B12 RMSE of λ across simulation conditions – polytomous attribute configuration

APPENDIX C: JAGS CODES FOR STUDY MODELS

C1. HO DINA model

```
HO.DINA <- function(){
   ##Partial mastery higher order model
   for (n in 1:N) {#examinee
     for (k in 1:K) {#attribute
       for (l in 1:L[k]){#level within attribute
         core[n, k,l] <- beta[k]*theta[n]-delta[k,l] #ltm for level l of attribute k
         sum.core[n,k,l]<-sum(core[n,k,1:l])#in parenthesis of numerator
         exp.sum.core[n,k,l]<-exp(sum.core[n,k,l])#numerator
         prob.a[n,k,l]<-exp.sum.core[n,k,l]/sum(exp.sum.core[n,k,1:L[k]])#probability of level l
     }#level within attribute
       alpha.star[n,k]~dcat(prob.a[n,k,1:L[k]])
       alpha[n,k] <- alpha.star[n,k]-1
   }#end of attribute
  }#end of examinee loop
##Measurement model
  for (n in 1:N){
   for (i in 1:I){
     for (k in 1:K){
       w[n,i,k]<-step(alpha[n,k]-Q[i,k])}
       eta[n,i]<-prod(w[n,i,])
       prob[n,i]<-g[i]+(1-s[i]-g[i])*eta[n,i]
       Score[n,i]~dbern(prob[n,i]) } }
       #Priors of the latent structural parameters
  for (k in 1:K){
   delta[k,1]<-0
   for (l in 2:L[k]){delta[k,l]~dnorm(0,0.5)}
   beta[k]~dnorm(0,0.5)%_%T(0,) }
     #prior of higher-order ability
  for (n in 1:N){
   theta[n]~dnorm(0,1)
  }
     #Priors of item parameters
  for (i in 1:I){
     s[i]~dbeta(1,1)
     g[i]~dbeta(1,1)%_%T(0,1-s[i])
   }}#End of model loop
```

C2. HO-RPa DINA model

```
RPA.DINA <- function(){
   ##Partial master higher order model
   for (n in 1:N) {#examinee
     for (k in 1:K) {#attribute
       for (l in 1:L[k]){#level within attribute
```

```
        core[n, k,l] <- beta[k]*theta[n]-delta[k,l] #ltm for level l of attribute k
        sum.core[n,k,l]<-sum(core[n,k,1:l])#in parenthesis of numerator
        exp.sum.core[n,k,l]<-exp(sum.core[n,k,l])#numerator
        prob.a[n,k,l]<-exp.sum.core[n,k,l]/sum(exp.sum.core[n,k,1:L[k]])#probability of level l
      }#level within attribute
       alpha.star[n,k]~dcat(prob.a[n,k,1:L[k]])
       alpha[n,k] <- alpha.star[n,k]-1
    }#end of attribute
  }#end of examinee loop
##Measurement model
  for (n in 1:N){
    for (i in 1:I){
      for (k in 1:K){
        w[n,i,k]<-step(alpha[n,k]-Q[i,k])}
        eta[n,i]<-prod(w[n,i,])
        prob[n,i]<-g[i]+(1-s[i]-g[i])*eta[n,i]
        Score[n,i]~dbern(prob[n,i]) } }
        #Priors of the latent structural parameters
  for (k in 1:K){
    delta[k,1]<-0
    for (l in 2:L[k]){delta[k,l]~dnorm(0,0.5) }
    beta[k]~dnorm(0,0.5)%_%T(0,)}
      #prior of higher-order ability
  for (n in 1:N){
    theta[n]~dnorm(0,1)
  }
      #Priors of item parameters
  for (i in 1:I){
      s[i]~dbeta(1,1)
      g[i]~dbeta(1,1)%_%T(0,1-s[i])
    }}#End of model loop
```

C3. MHO DINA for binary attribute configuration

```
MHOB <- function(){
  ##Partial mastery higher order model
  for (nn in 1:N) {#examinee
    for (k in 1:K) {#attribute
      for (l in 1:Lb[k]){#level within attribute
        core[nn, k,l] <- gam.1[k]*theta[nn]-gam.0[k,l] #ltm for level l of attribute k
        sum.core[nn,k,l]<-sum(core[nn,k,1:l])#in parenthesis of numerator
        exp.sum.core[nn,k,l]<-exp(sum.core[nn,k,l])#numerator
        prob.a[nn,k,l]<-exp.sum.core[nn,k,l]/sum(exp.sum.core[nn,k,1:Lb[k]])#probability of level l
      }#level within attribute
      att.star[nn,k]~dcat(prob.a[nn,k,1:Lb[k]])
      att[nn,k] <- att.star[nn,k]-1
    }#end of attribute
  }#end of examinee loop

  ##Measurement model
  for (nn in 1:N){
```

```
    for (i in 1:I){
      for (k in 1:K){
        w[nn,i,k]<-step(att[nn,k]-Q.bin[i,k])}
      eta[nn,i]<-prod(w[nn,i,])
      logit(prob[nn, i]) <- delta0[i] + delta1[i] * eta[nn,i]#DINA model
      Score[nn,i]~dbern(prob[nn,i]) } }
  #Priors of the latent structural parameters
  for (k in 1:K){
    gam.0[k,1]<-0
    for (l in 2:Lb[k]){ gam.0[k,l]~dnorm(0,0.5)}
    gam.1[k]~dnorm(0,0.5)%_%T(0,)  }
  ##Person parameters from joint distribution of response time and responses
  for (nn in 1:N) {
    person_parameter[nn]~ dnorm(person_mu, person_den)
    theta[nn] <- person_parameter[nn] }
  ##Item parameters from joint distribution of response time and responses
  for (i in 1:I) {
    item_parameter[i, 1:2]~ dmnorm(item_mu[1:2], item_den[1:2, 1:2])
    delta0[i]<-item_parameter[i, 1] #Item intercept from reparameterized DINA model
    delta1[i]<-item_parameter[i, 2] #Item interaction from reparameterized DINA model
    logit(g[i])<-delta0[i] #Item guessing parameter
    logit(ns[i])<-delta0[i] + delta1[i]#Solving for slipping parameter
    s[i] <- 1 - ns[i]  #Item slipping parameter
  }
  person_mu <- 0 #mean ability
  L_theta <- 1
  Sigma_theta <- L_theta
  person_den <- Sigma_theta

  #Hyper priors for miu of item parameters
  item_mu[1]~dnorm(-2.197, 0.5)#hyperprior of miu_delta0; item intercept for RDINA
  item_mu[2]~dnorm(4.394,0.5)%_%T(0,)#hyperprior of miu_delta is constrained to be +tive

  #Identity matrix for dsn of item covariance matrix
  R[1, 1] <- 1
  R[2, 2] <- 1
  R[1, 2] <- 0
  R[2, 1] <- 0
  item_den[1:2,1:2]~dwish(R[1:2,1:2],2) #hyper prior for Item covariance matrix
  Sigma_item[1:2,1:2]<-inverse(item_den[1:2,1:2])#Trasforming to inverse Wishart
}#End of model loop
```

C4. MHO DINA for polytomous attribute configuration

```
MHOP <- function(){
  ##Partial mastery higher order model
  for (nn in 1:N) {#examinee
    for (k in 1:K) {#attribute
      for (l in 1:L[k]){#level within attribute
        core[nn, k,l] <- gam.1[k]*theta[nn]-gam.0[k,l] #ltm for level l of attribute k
```

```
      sum.core[nn,k,l]<-sum(core[nn,k,1:l])#in parenthesis of numerator
      exp.sum.core[nn,k,l]<-exp(sum.core[nn,k,l])#numerator
      prob.a[nn,k,l]<-exp.sum.core[nn,k,l]/sum(exp.sum.core[nn,k,1:L[k]])#probability of level l
     }#level within attribute
     att.star[nn,k]~dcat(prob.a[nn,k,1:L[k]])
     att[nn,k] <- att.star[nn,k]-1
   }#end of attribute
 }#end of examinee loop

 ##Measurement model
 for (nn in 1:N){
   for (i in 1:I){
     for (k in 1:K){ w[nn,i,k]<-step(att[nn,k]-Q.poly[i,k])}
     eta[nn,i]<-prod(w[nn,i,])
     logit(prob[nn, i]) <- delta0[i] + delta1[i] * eta[nn,i]#DINA model
     Score[nn,i]~dbern(prob[nn,i]) } }
 #Priors of the latent structural parameters
 for (k in 1:K){
   gam.0[k,1]<-0
   for (l in 2:L[k]){ gam.0[k,l]~dnorm(0,0.5)   }
   gam.1[k]~dnorm(0,0.5)%_%T(0,)  }
 ##Person parameters from joint distribution of response time and responses
 for (nn in 1:N) {
   person_parameter[nn]~ dnorm(person_mu, person_den)
   theta[nn] <- person_parameter[nn] }
##Item parameters from joint distribution of response time and responses
 for (i in 1:I) {
   item_parameter[i, 1:2]~ dmnorm(item_mu[1:2], item_den[1:2, 1:2])
   delta0[i]<-item_parameter[i, 1] #Item intercept from reparameterized DINA model
   delta1[i]<-item_parameter[i, 2] #Item interaction from reparameterized DINA model
   logit(g[i])<-delta0[i] #Item guessing parameter
   logit(ns[i])<-delta0[i] + delta1[i]#Solving for slipping parameter
   s[i] <- 1 - ns[i]  #Item slipping parameter
 }
 person_mu <- 0 #mean ability
 L_theta <- 1
 Sigma_theta <- L_theta
 person_den <- Sigma_theta

 #Hyper priors for miu of item parameters
 item_mu[1]~dnorm(-2.197, 0.5)#hyperprior of miu_delta0; item intercept for RDINA
 item_mu[2]~dnorm(4.394,0.5)%_%T(0,)#hyperprior of miu_delta is constrained to be +tive

 #Identity matrix for dsn of item covariance matrix
 R[1, 1] <- 1
 R[2, 2] <- 1
 R[1, 2] <- 0
 R[2, 1] <- 0
 item_den[1:2,1:2]~dwish(R[1:2,1:2],2) #hyper prior for Item covariance matrix
 Sigma_item[1:2,1:2]<-inverse(item_den[1:2,1:2])#Trasforming to inverse Wishart
}#End of model loop
```

C5. JRT DINA for binary attribute configuration

```
JRTB <- function(){
 ##Partial mastery higher order model
 for (nn in 1:N) {#examinee
   for (k in 1:K) {#attribute
     for (l in 1:Lb[k]){#level within attribute
       core[nn, k,l] <- gam.1[k]*theta[nn]-gam.0[k,l] #ltm for level l of attribute k
       sum.core[nn,k,l]<-sum(core[nn,k,1:l])#in parenthesis of numerator
       exp.sum.core[nn,k,l]<-exp(sum.core[nn,k,l])#numerator
       prob.a[nn,k,l]<-exp.sum.core[nn,k,l]/sum(exp.sum.core[nn,k,1:Lb[k]])#probability of level l
     }#level within attribute
     att.star[nn,k]~dcat(prob.a[nn,k,1:Lb[k]])
     att[nn,k]<-att.star[nn,k]-1
   }#end of attribute
 }#end of examinee loop

 ##Measurement model
 for (nn in 1:N){
   for (i in 1:I){
     for (k in 1:K){
       w[nn,i,k]<-step(att[nn,k]-Q.bin[i,k])}
     eta[nn,i]<-prod(w[nn,i,])
     logit(prob[nn,i])<-delta0[i]+delta1[i]*eta[nn,i]#DINA model
     Score[nn,i]~dbern(prob[nn,i])
     logT[nn,i]~dnorm(lambda[i]-t0[nn],den_epsilon[i])#Draw resp time 4 item i & person nn
   } }
 #Priors of the latent structural parameters
 for (k in 1:K){
   gam.0[k,1]<-0
   for (l in 2:Lb[k]){gam.0[k,l]~dnorm(0,0.5)  }
   gam.1[k]~dnorm(0,0.5)%_%T(0,) }
 ##Person parameters from joint distribution of response time and responses
 for (nn in 1:N) {
   person_parameter[nn, 1:2]~ dmnorm(person_mu[1:2], person_den[1:2, 1:2])
   theta[nn] <- person_parameter[nn, 1]
   t0[nn] <- person_parameter[nn, 2] }
 ##Item parameters from joint distribution of response time and responses
 for (i in 1:I) {
   item_parameter[i, 1:3]~ dmnorm(item_mu[1:3], item_den[1:3, 1:3])
   lambda[i]<-item_parameter[i, 1] #Item time intensity from response time model
   delta0[i]<-item_parameter[i, 2] #Item intercept from reparameterized DINA model
   delta1[i]<-item_parameter[i, 3] #Item interaction from reparameterized DINA model
   logit(g[i])<-delta0[i] #Item guessing parameter
   logit(ns[i])<-delta0[i] + delta1[i]#Solving for slipping parameter
   s[i] <- 1 - ns[i]  #Item slipping parameter
   den_epsilon[i]~ dgamma(1, 1) #Error term from response time model
   Sigma_epsilon[i] <- 1/den_epsilon[i] #Item time discrimination parameter
 }
 person_mu[1] <- 0 #mean ability
```

```
  person_mu[2] <- 0 #mean initial speed
  L_theta[1, 1] <- 1
  L_theta[2, 2]~ dgamma(1, 1)
  L_theta[2, 1]~ dnorm(0,1)
  L_theta[1, 2] <- 0
  Sigma_theta <- L_theta %*% t(L_theta)
  person_den[1:2, 1:2] <- inverse(Sigma_theta[1:2, 1:2])

  #Hyper priors for miu of item parameters
  item_mu[1]~dnorm(3,0.5)# hyperprior of miu_lambda; item time discrimination
  item_mu[2]~dnorm(-2.197, 0.5)#hyperprior of miu_delta0; item intercept for RDINA
  item_mu[3]~dnorm(4.394,0.5)%_%T(0,)#hyperprior of miu_delta is constrained to be +tive

  #Identity matrix for dsn of item covariance matrix
  R[1, 1] <- 1
  R[2, 2] <- 1
  R[3, 3] <- 1
  R[1, 2] <- 0
  R[1, 3] <- 0
  R[2, 1] <- 0
  R[2, 3] <- 0
  R[3, 1] <- 0
  R[3, 2] <- 0
  item_den[1:3, 1:3]~ dwish(R[1:3, 1:3], 3) #hyper prior for Item covariance matrix
  Sigma_item[1:3, 1:3] <- inverse(item_den[1:3, 1:3])#Trasforming to inverse Wishart
  }#End of model loop
```

C6. JRT DINA for polytomous attribute configuration

```
JRTP<-function(){
  ##Partial mastery higher order model
  for (nn in 1:N) {#examinee
    for (k in 1:K) {#attribute
      for (l in 1:L[k]){#level within attribute
        core[nn, k,l] <- gam.1[k]*theta[nn]-gam.0[k,l] #ltm for level l of attribute k
        sum.core[nn,k,l]<-sum(core[nn,k,1:l])#in parenthesis of numerator
        exp.sum.core[nn,k,l]<-exp(sum.core[nn,k,l])#numerator
        prob.a[nn,k,l]<-exp.sum.core[nn,k,l]/sum(exp.sum.core[nn,k,1:L[k]])#probability of level l
      }#level within attribute
      att.star[nn,k]~dcat(prob.a[nn,k,1:L[k]])
      att[nn,k] <- att.star[nn,k]-1
    }#end of attribute
  }#end of examinee loop

  ##Measurement model
  for (nn in 1:N){
    for (i in 1:I){
      for (k in 1:K){
        w[nn,i,k]<-step(att[nn,k]-Q.poly[i,k])}
      eta[nn,i]<-prod(w[nn,i,])
```

```
    logit(prob[nn, i]) <- delta0[i] + delta1[i] * eta[nn,i]#DINA model
    Score[nn,i]~dbern(prob[nn,i])
    logT[nn,i]~dnorm(lambda[i]-t0[nn],den_epsilon[i])#Draw resp time 4 item i & person nn}}

#Priors of the latent structural parameters
for (k in 1:K){
  gam.0[k,1]<-0
  for (l in 2:L[k]){
    gam.0[k,l]~dnorm(0,0.5) }
  gam.1[k]~dnorm(0,0.5)%_%T(0,) }
  ##Person parameters from joint distribution of response time and responses
for (nn in 1:N) {
  person_parameter[nn, 1:2]~ dmnorm(person_mu[1:2], person_den[1:2, 1:2])
  theta[nn] <- person_parameter[nn, 1]
  t0[nn] <- person_parameter[nn, 2] }
  ##Item parameters from joint distribution of response time and responses
for (i in 1:I) {
  item_parameter[i, 1:3]~ dmnorm(item_mu[1:3], item_den[1:3, 1:3])
  lambda[i]<-item_parameter[i, 1] #Item time intensity from response time model
  delta0[i]<-item_parameter[i, 2] #Item intercept from reparameterized DINA model
  delta1[i]<-item_parameter[i, 3] #Item interaction from reparameterized DINA model
  logit(g[i])<-delta0[i] #Item guessing parameter
  logit(ns[i])<-delta0[i] + delta1[i]#Solving for slipping parameter
  s[i] <- 1 - ns[i]  #Item slipping parameter
  den_epsilon[i]~ dgamma(1, 1) #Error term from response time model
  Sigma_epsilon[i] <- 1/den_epsilon[i] #Item time discrimination parameter
}
person_mu[1] <- 0 #mean ability
person_mu[2] <- 0 #mean initial speed
L_theta[1, 1] <- 1
L_theta[2, 2]~ dgamma(1, 1)
L_theta[2, 1]~ dnorm(0,1)
L_theta[1, 2] <- 0
Sigma_theta <- L_theta %*% t(L_theta)
person_den[1:2, 1:2] <- inverse(Sigma_theta[1:2, 1:2])

#Hyper priors for miu of item parameters
item_mu[1]~dnorm(3,0.5)# hyperprior of miu_lambda; item time discrimination
item_mu[2]~dnorm(-2.197, 0.5)#hyperprior of miu_delta0; item intercept for RDINA
item_mu[3]~dnorm(4.394,0.5)%_%T(0,)#hyperprior of miu_delta is constrained to be +tive

#Identity matrix for dsn of item covariance matrix
R[1, 1] <- 1
R[2, 2] <- 1
R[3, 3] <- 1
R[1, 2] <- 0
R[1, 3] <- 0
R[2, 1] <- 0
R[2, 3] <- 0
R[3, 1] <- 0
R[3, 2] <- 0
```

```
  item_den[1:3, 1:3]~ dwish(R[1:3, 1:3], 3) #hyper prior for Item covariance matrix
  Sigma_item[1:3, 1:3] <- inverse(item_den[1:3, 1:3])#Trasforming to inverse Wishart
}#End of model loop
```

C7. JDS DINA for binary attribute configuration

```
JDSB <- function(){
  ##Partial mastery higher order model
  for (nn in 1:N) {#examinee
    for (k in 1:K) {#attribute
      for (l in 1:Lb[k]){#level within attribute
        core[nn, k,l] <- gam.1[k]*theta[nn]-gam.0[k,l] #ltm for level l of attribute k
        sum.core[nn,k,l]<-sum(core[nn,k,1:l])#in parenthesis of numerator
        exp.sum.core[nn,k,l]<-exp(sum.core[nn,k,l])#numerator
        prob.a[nn,k,l]<-exp.sum.core[nn,k,l]/sum(exp.sum.core[nn,k,1:Lb[k]])#probability of level l
      }#level within attribute
      att.star[nn,k]~dcat(prob.a[nn,k,1:Lb[k]])
      att[nn,k] <- att.star[nn,k]-1
    }#end of attribute
  }#end of examinee loop

  ##Measurement model
  for (nn in 1:N){
    for (i in 1:I){
      for (k in 1:K){w[nn,i,k]<-step(att[nn,k]-Q.bin[i,k])}
      eta[nn,i]<-prod(w[nn,i,])
      logit(prob[nn, i]) <- delta0[i] + delta1[i] * eta[nn,i]#DINA model
      Score[nn,i]~dbern(prob[nn,i])
      logT[nn,i]~dnorm(lambda[i]-speed[nn,1]-speed[nn,2]-speed[nn,3],den_epsilon[i])#Draw resp time 4
item i & person nn
    }}
  #Priors of the latent structural parameters
  for (k in 1:K){
    gam.0[k,1]<-0
    for (l in 2:Lb[k]){
      gam.0[k,l]~dnorm(0,0.5)
    }
    gam.1[k]~dnorm(0,0.5)%_%T(0,)
  }
  ##Person parameters from joint distribution of response time and responses
  for (nn in 1:N) {
    person_parameter[nn, 1:4]~ dmnorm(person_mu[1:4], person_den[1:4, 1:4])
    theta[nn] <- person_parameter[nn, 1]
    t0[nn] <- person_parameter[nn, 2]
    t1[nn] <- person_parameter[nn, 3]
    t2[nn] <- person_parameter[nn, 4]  }
  speed<-person_parameter[,2:4]%*%t(Xn)

  ##Item parameters from joint distribution of response time and responses
  for (i in 1:I) {
```

```
  item_parameter[i, 1:3]~ dmnorm(item_mu[1:3], item_den[1:3, 1:3])
  lambda[i]<-item_parameter[i, 1] #Item time intensity from response time model
  delta0[i]<-item_parameter[i, 2] #Item intercept from reparameterized DINA model
  delta1[i]<-item_parameter[i, 3] #Item interaction from reparameterized DINA model
  logit(g[i])<-delta0[i] #Item guessing parameter
  logit(ns[i])<-delta0[i] + delta1[i]#Solving for slipping parameter
  s[i] <- 1 - ns[i]  #Item slipping parameter
  den_epsilon[i]~ dgamma(1, 1) #Error term from response time model
  Sigma_epsilon[i] <- 1/den_epsilon[i] #Item time discrimination parameter
 }
 person_mu[1] <- 0 #mean ability
 person_mu[2] <- 0 #mean initial speed
 person_mu[3] <- 0 #mean slope
 person_mu[4] <- 0 #mean quadratic term
 L_theta[1, 1] <- 1
 L_theta[2, 2]~ dgamma(1, 1)
 L_theta[3, 3]~ dgamma(1, 1)
 L_theta[4, 4]~ dgamma(1, 1)
 L_theta[2, 1]~ dnorm(0,1)
 L_theta[3, 1]~ dnorm(0,1)
 L_theta[4, 1]~ dnorm(0,1)
 L_theta[3, 2]~ dnorm(0,1)
 L_theta[4, 2]~ dnorm(0,1)
 L_theta[4, 3]~ dnorm(0,1)
 L_theta[1, 2] <- 0
 L_theta[1, 3] <- 0
 L_theta[1, 4] <- 0
 L_theta[2, 3] <- 0
 L_theta[2, 4] <- 0
 L_theta[3, 4] <- 0
 Sigma_theta <- L_theta %*% t(L_theta)
 person_den[1:4, 1:4] <- inverse(Sigma_theta[1:4, 1:4])

 #Hyper priors for miu of item parameters
 item_mu[1]~dnorm(3,0.5)# hyperprior of miu_lambda; item time discrimination
 item_mu[2]~dnorm(-2.197, 0.5)#hyperprior of miu_delta0; item intercept for RDINA
 item_mu[3]~dnorm(4.394,0.5)%_%T(0,)#hyperprior of miu_delta1 is constrained to be +tive

 #Identity matrix for dsn of item covariance matrix
 R[1, 1] <- 1
 R[2, 2] <- 1
 R[3, 3] <- 1
 R[1, 2] <- 0
 R[1, 3] <- 0
 R[2, 1] <- 0
 R[2, 3] <- 0
 R[3, 1] <- 0
 R[3, 2] <- 0
 item_den[1:3, 1:3]~ dwish(R[1:3, 1:3], 3) #hyper prior for Item covariance matrix
 Sigma_item[1:3, 1:3] <- inverse(item_den[1:3, 1:3])#Trasforming to inverse Wishart
}#End of model loop
```

C8. JDS DINA for polytomous attributes

```
JDSP <- function(){
  ##Partial mastery higher order model
  for (nn in 1:N) {#examinee
    for (k in 1:K) {#attribute
      for (l in 1:L[k]){#level within attribute
        core[nn, k,l] <- gam.1[k]*theta[nn]-gam.0[k,l] #ltm for level l of attribute k
        sum.core[nn,k,l]<-sum(core[nn,k,1:l])#in parenthesis of numerator
        exp.sum.core[nn,k,l]<-exp(sum.core[nn,k,l])#numerator
        prob.a[nn,k,l]<-exp.sum.core[nn,k,l]/sum(exp.sum.core[nn,k,1:L[k]])#probability of level l
      }#level within attribute
      att.star[nn,k]~dcat(prob.a[nn,k,1:L[k]])
      att[nn,k] <- att.star[nn,k]-1
    }#end of attribute
  }#end of examinee loop

  ##Measurement model
  for (nn in 1:N){
    for (i in 1:I){
      for (k in 1:K){w[nn,i,k]<-step(att[nn,k]-Q.poly[i,k])}
      eta[nn,i]<-prod(w[nn,i,])
      logit(prob[nn, i]) <- delta0[i] + delta1[i] * eta[nn,i]#DINA model
      Score[nn,i]~dbern(prob[nn,i])
      logT[nn,i]~dnorm(lambda[i]-speed[nn,1]-speed[nn,2]-speed[nn,3],den_epsilon[i])#Draw resp time 4
item i & person nn
    } }
  #Priors of the latent structural parameters
  for (k in 1:K){
    gam.0[k,1]<-0
    for (l in 2:L[k]){
      gam.0[k,l]~dnorm(0,0.5)}
    gam.1[k]~dnorm(0,0.5)%_%T(0,)}

  ##Person parameters from joint distribution of response time and responses
  for (nn in 1:N) {
    person_parameter[nn, 1:4]~ dmnorm(person_mu[1:4], person_den[1:4, 1:4])
    theta[nn] <- person_parameter[nn, 1]
    t0[nn] <- person_parameter[nn, 2]
    t1[nn] <- person_parameter[nn, 3]
    t2[nn] <- person_parameter[nn, 4] }
  speed<-person_parameter[,2:4]%*%t(Xn)

  ##Item parameters from joint distribution of response time and responses
  for (i in 1:I) {
    item_parameter[i, 1:3]~ dmnorm(item_mu[1:3], item_den[1:3, 1:3])
    lambda[i]<-item_parameter[i, 1] #Item time intensity from response time model
    delta0[i]<-item_parameter[i, 2] #Item intercept from reparameterized DINA model
    delta1[i]<-item_parameter[i, 3] #Item interaction from reparameterized DINA model
    logit(g[i])<-delta0[i] #Item guessing parameter
    logit(ns[i])<-delta0[i] + delta1[i]#Solving for slipping parameter
```

```
    s[i] <- 1 - ns[i]  #Item slipping parameter
    den_epsilon[i]~ dgamma(1, 1) #Error term from response time model
    Sigma_epsilon[i] <- 1/den_epsilon[i] #Item time discrimination parameter
  }
  person_mu[1] <- 0 #mean ability
  person_mu[2] <- 0 #mean initial speed
  person_mu[3] <- 0 #mean slope
  person_mu[4] <- 0 #mean quadratic term
  L_theta[1, 1] <- 1
  L_theta[2, 2]~ dgamma(1, 1)
  L_theta[3, 3]~ dgamma(1, 1)
  L_theta[4, 4]~ dgamma(1, 1)
  L_theta[2, 1]~ dnorm(0,1)
  L_theta[3, 1]~ dnorm(0,1)
  L_theta[4, 1]~ dnorm(0,1)
  L_theta[3, 2]~ dnorm(0,1)
  L_theta[4, 2]~ dnorm(0,1)
  L_theta[4, 3]~ dnorm(0,1)
  L_theta[1, 2] <- 0
  L_theta[1, 3] <- 0
  L_theta[1, 4] <- 0
  L_theta[2, 3] <- 0
  L_theta[2, 4] <- 0
  L_theta[3, 4] <- 0
  Sigma_theta <- L_theta %*% t(L_theta)
  person_den[1:4, 1:4] <- inverse(Sigma_theta[1:4, 1:4])

  #Hyper priors for miu of item parameters
  item_mu[1]~dnorm(3,0.5)# hyperprior of miu_lambda; item time discrimination
  item_mu[2]~dnorm(-2.197, 0.5)#hyperprior of miu_delta0; item intercept for RDINA
  item_mu[3]~dnorm(4.394,0.5)%_%T(0,)#hyperprior of miu_delta1 is constrained to be +tive

  #Identity matrix for dsn of item covariance matrix
  R[1, 1] <- 1
  R[2, 2] <- 1
  R[3, 3] <- 1
  R[1, 2] <- 0
  R[1, 3] <- 0
  R[2, 1] <- 0
  R[2, 3] <- 0
  R[3, 1] <- 0
  R[3, 2] <- 0
  item_den[1:3, 1:3]~ dwish(R[1:3, 1:3], 3) #hyper prior for Item covariance matrix
  Sigma_item[1:3, 1:3] <- inverse(item_den[1:3, 1:3])#Trasforming to inverse Wishart
}#End of model loop.
```

REFERENCES

REFERENCES

Akaike, H., Parzen, E., Tanabe, K., & Kitagawa, G. (1998). *Selected papers of Hirotugu Akaike*. New York: Springer.

Anders, R., Alario, F., & Van Maanen, L. (2016). The shifted Wald distribution for response time data analysis. *Psychological methods*, *21*(3), 309.

Almond, R. G., Mislevy, R. J., Steinberg, L. S., Yan, D., & Williamson, D. M. (2015). Bayesian networks in educational assessment. New York, NY: Springer.

Bradshaw, L., Izsák, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understandings of rational numbers: Building a multidimensional test within the diagnostic classification framework. Educational measurement*: Issues and practice*, *33*(1), 2-14.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, *7*(4), 434-455.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician*, *46*(3), 167-174.

Chen, J., & de la Torre, J. (2014). A Procedure for diagnostically modeling extant large-scale assessment data: The case of the programme for international student assessment in reading. Psychology, 5(18), 1967.

Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. Applied Psychological Measurement, 37, 419–437. https://doi.org/10.1177/0146621613479818.

Chiu, C. Y. (2013). Statistical Refinement of the Q-matrix in Cognitive Diagnosis. Applied Psychological Measurement, 37(8), 598-618.

Chiu, C. Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, *74*(4), 633.

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. Journal of educational measurement, 45(4), 343-362.

de la Torre, J. (2009). DINA model and parameter estimation: A didactic. Journal of educational and behavioral statistics, 34(1), 115-130.

de la Torre, J., & Chiu, C. Y. (2016). A general method of empirical Q-matrix validation. Psychometrika, 81(2), 253-273.

de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. Psychometrika, 69(3), 333-353.

De Boeck, P., & Jeon, M. (2019). An Overview of Models for Response Times and Processes in Cognitive Tests. Frontiers in psychology, 10, 102.

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. Applied Psychological Measurement, 35(1), 8-26.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. *Cognitively diagnostic assessment*, *361389*.

Dittmar, D. (2013). Slice sampling. *TU Darmstadt. URL https://www.ias.informatik.tu-darmstadt.de/uploads/Teaching/RobotLearningSeminar/Dittmar_RLS_2013.pdf*.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, N.J: L. Erlbaum Associates.

Fox, J. P. (2010). *Bayesian item response modeling: Theory and applications*. Springer Science & Business Media.

Fox, J. P., & Marianti, S. (2016). Joint modeling of ability and differential speed using responses and response times. *Multivariate behavioral research*, *51*(4), 540-553.

Fox, J. P., Klein Entink, R. H., & van der Linden, W. J. (2007). Modeling of responses and response times with the package cirt. *Journal of Statistical Software*, *20*(7), 1-14.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, *85*(410), 398-409.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457-472.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. Boca Raton: CRC Press.

Geman, S., & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 721-741.

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In Bayesian Statistics.

Gierl, M. J., Alves, C., & Majeau, R. T. (2010). Using the attribute hierarchy method to make diagnostic inferences about examinees' knowledge and skills in mathematics: An operational implementation of cognitive diagnostic assessment. *International Journal of Testing*, *10*(4), 318-341.

Gitomer, D. H., & Yamamoto, K. (1991). Performance modeling that integrates latent trait and class theory. *Journal of Educational Measurement*, *28*(2), 173-189.

Gulliksen, H. (1950). *Theory of mental tests.* New York: Wiley.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. Journal of Educational Measurement, 26(4), 301-321.

Hartz, S. M. (2002). *A Bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality* (Order No. 3044108). Available from Dissertations & Theses @ CIC Institutions; ProQuest Dissertations & Theses Global. (305590285).

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.

Henson, R. A. (2009). Diagnostic classification models: Thoughts and future directions. Measurement: Interdisciplinary Research and Perspectives, 7, 34– 36

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. Psychometrika, 74(2), 191.

Hong, H., Wang, C., Lim, Y. S., & Douglas, J. (2015). Efficient models for cognitive diagnosis with continuous and mixed-type latent variables. Applied psychological measurement, 39(1), 31-43.

Houts, C. R., & Cai, L. (2013). flexMIRTR: Flexible Multilevel Multidimensional Item Analysis and Test Scoring User's Manual Version 2.0.

Hu, J., Miller, M. D., Huggins-Manley, A. C., & Chen, Y. H. (2016). Evaluation of model fit in cognitive diagnosis models. International Journal of Testing, 16(2), 119-141.

Huang, H. Y. (2019). Utilizing response times in cognitive diagnostic computerized adaptive testing under the higher-order deterministic input, noisy 'and'gate model. British Journal of Mathematical and Statistical Psychology.

Huebner, A. (2010). An overview of recent development in cognitive diagnostic computer adaptive assessments. Practical Assessment, Research & Evaluation, 15, 1-7. http://pareonline.net/getvn.asp?v=15&n=3

Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement, 71*(2), 407-419.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258-272. http://dx.doi.org/10.1177/01466210122032064

Jurich, D. P., & Bradshaw, L. P. (2014). An illustration of diagnostic classification modeling in student learning outcomes assessment. International Journal of Testing, 14(1), 49-72.

Kahraman, N., Cuddy, M. M., & Clauser, B. E. (2013). Modeling pacing behavior and test speededness using latent growth curve models. *Applied Psychological Measurement*, *37*(5), 343-360.

Karelitz, T. M. (2004). Ordered category attribute coding framework for cognitive assessments (unpublished doctoral dissertation). University of Illinois at Urbana–Champaign.

Karelitz, T. M. (2008). How binary skills obscure the transition from non-mastery to mastery. *Measurement: Interdisciplinary Research & Perspective*, 6: 268–272.

Kim, J. S., & Bolt, D. M. (2007). Estimating item response theory models using Markov chain Monte Carlo methods. *Educational Measurement: Issues and Practice*, *26*(4), 38-51.

Kim, S. H. (2001). Towards a statistical foundation in combining structures of decomposable graphical models. Tech. Report 1-2, Division of Applied Mathematics, KAIST, Daejeon, 305-701, South Korea.

Klein Entink, R. H., Fox, J. P., & van der Linden, W. J. (2009a). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, *74*(1), 21.

Klein Entink, R. H., van der Linden, W. J., & Fox, J. P. (2009b). A Box–Cox normal model for response times. British Journal of Mathematical and Statistical Psychology, 62(3), 621-640.

Köhn, HF. & Chiu, CY. J (2018a) Attribute Hierarchy Models in Cognitive Diagnosis: Identifiability of the Latent Attribute Space and Conditions for Completeness of the Q-Matrix. Journal of Classification, 1-25.

Köhn, H. F., & Chiu, C. Y. (2018b). How to build a complete Q-matrix for a cognitively diagnostic test. *Journal of Classification*, *35*(2), 273-299.

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models. Journal of Educational Measurement, 49(1), 59-81.

Lee, Y. H., & Chen, H. (2011). A review of recent response-time analyses in educational testing. *Psychological Test and Assessment Modeling*, *53*(3), 359.

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. Journal of educational measurement, 41(3), 205-237.

Leighton, J., & Gierl, M. (Eds.). (2007). Cognitive diagnostic assessment for education: Theory and applications. Cambridge University Press.

Liu, J., Xu, G., & Ying, Z. (2012). Data-driven learning of Q-matrix. Applied psychological measurement, 36(7), 548-564.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171.

Loeys, T., Rosseel, Y., & Baten, K. (2011). A joint modeling approach for reaction time and accuracy in psycholinguistic experiments. *Psychometrika*, *76*(3), 487-503.

Lu, Y., & Sireci, S. G. (2007). Validity issues in test speededness. *Educational Measurement: Issues and Practice*, *26*(4), 29-37.

Maris, E. (1993). Additive and multiplicative models for gamma distributed random variables, and their application as psychometric models for response times. *Psychometrika*, *58*(3), 445-469.

Maris, E. (1995). Psychometric latent response models. Psychometrika, 60(4), 523-547.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*(2), 187-212.

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, *49*(4), 305-328.

Molenaar, D., Oberski, D., Vermunt, J., & De Boeck, P. (2016). Hidden Markov item response theory models for responses and response times. Multivariate behavioral research, 51(5), 606-626.

Molenaar, D., Tuerlinckx, F., & van der Maas, H. L. (2015). A bivariate generalized linear item response theory modeling framework to the analysis of responses and response times. *Multivariate Behavioral Research*, *50*(1), 56-74.

Muthén, L. K., & Muthén, B. (2010). Growth modeling with latent variables using Mplus: Introductory and intermediate growth models. *Mplus Short Course Topic 3*.

Muthen, L. K., & Muthen, B. O. (2018). Mplus Version 8. *Los Angeles, CA: Muthen & Muthen*.

Naumann, J., & Goldhammer, F. (2017). Time-on-task effects in digital reading are non-linear and moderated by persons' skills and tasks' demands. *Learning and Individual Differences, 53*, 1-16.

Neal, R. M. (2003). Slice sampling. *The annals of statistics*, *31*(3), 705-767.

Neal, R. M. (1997). Markov chain Monte Carlo methods based on slicing the density function. *Preprint*.

OECD (2014). PISA 2012 technical report. Paris, France: Author.

Partchev, I., & De Boeck, P. (2012). Can fast and slow intelligence be differentiated? *Intelligence*, *40*(1), 23-32.

Plummer, M. (2012). JAGS Version 3.3. 0 user manual. *International Agency for Research on Cancer, Lyon, France*.

Raftery, A. E., & Lewis, S.M. (1992a). One long run with diagnostics: implementation strategies for Markov chain Monte Carlo. Statistical Science,7, 493–497

Raftery, A. E., & Lewis, S.M. (1992b). How many iterations in the Gibbs sampler? In J.M. Bernardo, J.O. Berger, A.P. Dawid, & A.F.M. Smith (Eds.), Bayesian statistics (Vol. 4, pp. 765–776). Oxford, UK: Oxford University Press.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873-922.

Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. Journal of Psychoeducational Assessment, 34(8), 782-799.

Robert, C. P., & Casella, G. (1999). Monte Carlo Statistical Methods. New York, NY: Springer New York.

Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. *Markov chain Monte Carlo in practice*, *57*.

Roussos, L. A., Templin, J. L., & Henson, R. A. (2007). Skills diagnosis using IRT-based latent class models. Journal of Educational Measurement, 44(4), 293-311.

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, *6*(4), 219-262.

Rupp, A. A., & Mislevy, R. J. (2007). Cognitive foundations of structured item response models.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). Diagnostic measurement: Theory, methods, and applications. Guilford Press.

Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, *34*(3), 213-232.

Schnipke, D. L., & Scrams, D. J. (2002). Exploring issues of examinee behavior: Insights gained from response-time analyses. *Computer-based testing: Building the foundation for future assessments*, 237-266.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Sen, R. (2012). *Structural Equation Model Approach to the use of Response times for Improving Estimation in Item Response Models*. University of Connecticut.

Sessoms, J., & Henson, R. A. (2018). Applications of Diagnostic Classification Models: A Literature Review and Critical Commentary. Measurement: Interdisciplinary Research and Perspectives, 16(1), 1-17.

Sheehan, K., & Mislevy, R. J. (1990). Integrating cognitive and psychometric models to measure document literacy. Journal of Educational Measurement, 27(3), 255-272.

Simonetto, A. (2011). Using structural equation and item response models to assess relationship between latent traits. *Journal of Applied Quantitative Methods*, *6*(4).

Sinharay, S., & Haberman, S. J. (2009). How much can we reliably know about what examinees know? Measurement: Interdisciplinary Research and Perspectives, 7, 46– 49.

Su, Y. L. (2013). Cognitive diagnostic analysis using hierarchically structured skills.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. Diagnostic monitoring of skill and knowledge acquisition, 453-488.

Tatsuoka, K. K. (2009). Cognitive assessment: An introduction to the rule space method. Routledge.

Tatsuoka, K., & Tatsuoka, M. (1979). *A Model for Incorporating Response-Time Data in Scoring Achievement Tests (No.* CERL-E-7). Illinois Univ at Urbana-Champaign Computer-Based Education Research Lab.

Templin, J. L. (2004). Generalized linear mixed proficiency models. *Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign*.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287-305. doi: http://dx.doi.org.proxy1.cl.msu.edu/10.1037/1082-989X.11.3.287

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. Psychometrika, 79(2), 317-339.

Terzi, R., & de la Torre, J. (2018). An iterative method for empirically based Q-matrix validation. International Journal of Assessment Tools in Education, 5(2), 248-262.

Terzi, R., & Sen, S. (2019). A Nondiagnostic Assessment for Diagnostic Purposes: Q-Matrix Validation and Item-Based Model Fit Evaluation for the TIMSS 2011 Assessment. SAGE Open, 9(1), 2158244019832684.

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*(2), 181-204.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. Psychometrika, 72(3), 287.

van der Linden, W. J. (2016). Lognormal response-time model. In *Handbook of Item Response Theory, Volume One* (pp. 289-310). Chapman and Hall/CRC.

van der Linden, W. J., & Fox, J. P. (2015). Joint hierarchical modeling of responses and response times. In W. J. van der Linden (Ed.), Handbook of Item Response Theory: Vol 1. Models. Boca Raton: FL: Chapman & Hall/CRC.

van der Linden, W. J., Breithaupt, K., Chuah, S. C., & Zhang, Y. (2007). Detecting differential speededness in multistage testing. *Journal of Educational Measurement*, *44*(2), 117-130.

van der Linden, W. J., Scrams, D. J., & Schnipke, D. L. (1999). Using response-time constraints to control for differential speededness in computerized adaptive testing. *Applied Psychological Measurement*, *23*(3), 195-210.

Verhelst, N. D., Verstralen, H. H., & Jansen, M. G. H. (1997). A logistic model for time-limit tests. In *Handbook of modern item response theory* (pp. 169-185). Springer, New York, NY.

Von Davier, M. (2005). A general diagnostic model applied to language testing data. ETS Research Report Series, 2005(2), i-35.

Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. British Journal of Mathematical and Statistical Psychology, 68(3), 456-477.

Wang, T., & Hanson, B. A. (2005). Development and calibration of an item response model that incorporates response time. *Applied Psychological Measurement*, *29*(5), 323-339.

Wollack, J. A., Cohen, A. S., & Wells, C. S. (2003). A method for maintaining scale stability in the presence of test speededness. *Journal of Educational Measurement*, *40*(4), 307-330.

Xu, G. (2017). Identifiability of restricted latent class models with binary responses. The Annals of Statistics, 45(2), 675-707.

Yan, D., Mislevy, R. J., & Almond, R. G. (2003). Design and analysis in a cognitive assessment (ETS Research Report Series, RR-03-32). Princeton, NJ: ETS.

Zhan, P., Jiao, H., & Liao, D. (2018a). Cognitive diagnosis modelling incorporating item response times. British Journal of Mathematical and Statistical Psychology, 71(2), 262-286.

Zhan, P., Liao, M., & Bian, Y. (2018b). Joint testlet cognitive diagnosis modeling for paired local item dependence in response times and response accuracy. Frontiers in psychology, 9, 607.

Zhan, P., Ma, W., Jiao, H., & Ding, S. (2019a). A Sequential Higher Order Latent Structural Model for Hierarchical Attributes in Cognitive Diagnostic Assessments. Applied Psychological Measurement, 0146621619832935

Zhan, P., Wang, W. C., & Li, X. (2019b). A partial mastery, higher-order latent structural model for polytomous attributes in cognitive diagnostic assessments. Journal of Classification, 1-24.

Zhang, J. (2015). Analyzing Hierarchical Data with the DINA-HC Approach (Doctoral dissertation, Teachers College).