THE ROLE OF COVARIATES ON INFERRING THE $\boldsymbol{Q}$-MATRIX AND LEARNING TRAJECTORY

BY

HULYA DUYGU YIGIT

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Educational Psychology
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Professor Steven Culpepper, Chair and Director of Research
Assistant Professor Justin Kern, Co-Chair
Professor Jeffrey Douglas
Professor Jimmy de la Torre

Abstract

Chapter 2: In learning environments, understanding the longitudinal path of learning is one of the main goals. Cognitive diagnostic models (CDMs) for measurement combined with a transition model for mastery may be beneficial for providing fine-grained information about students' knowledge profiles over time. An efficient algorithm to estimate model parameters would augment the practicality of this combination. In this study, the Expectation-Maximization (EM) algorithm is presented for the estimation of student learning trajectories with the GDINA (generalized deterministic inputs, noisy, "and" gate) and some of its submodels for the measurement component, and a first-order Markov model for learning transitions are implemented. A simulation study is conducted to investigate the efficiency of the algorithm in estimation accuracy of student and model parameters under several factors—sample size, number of attributes, number of time points in a test, and complexity of the measurement model. Attribute- and vector-level agreement rates as well as the root mean square error rates of the model parameters are investigated. In addition, the computer run times for converging are recorded. The result shows that for a majority of the conditions, the accuracy rates of the parameters are quite promising in conjunction with relatively short computation times. Only for the conditions with relatively low sample sizes and high numbers of attributes, the computation time increases with a reduction parameter recovery rate. An application using spatial reasoning data is given. Based on the Bayesian information criterion (BIC), the model fit analysis shows that the DINA (deterministic inputs, noisy, "and" gate) model is preferable to the GDINA with these data.

Chapter 3: The rise of online learning platforms requires new approaches for developing

formative assessments that provide accurate, fine-grained information on student learning profiles. Restricted latent classification models (RLCMs) serve a central role in the development and implementation of formative assessments. The latent structure for RLCMs is defined by the $\boldsymbol{Q}$ matrix, which is a binary matrix that specifies the relationship between underlying attributes and observed responses. Recent research developed fully exploratory approaches for inferring the RLCM $\boldsymbol{Q}$ matrix. Although exploratory methods exist for uncovering the latent structure educational researchers are also interested in understanding the role of intervention effects and student covariates on item performance and skill mastery. Consequently, the purpose of our project is to extend the exploratory RLCM framework to jointly uncover the latent structure and assess the role of student covariates on item performance and attribute mastery. We consider a general modeling framework for including covariates and consider two special cases which correspond to different research settings. Our models provide researchers with tools for evaluating intervention effects aimed at enhancing learning outcomes and documenting the extent to which the relationship between the latent structure and responses is invariant to student background characteristics. We develop a new Bayesian formulation to estimate model parameters and report Monte Carlo evidence pertaining to accurate recovery of $\boldsymbol{Q}$ and other model parameters. We apply the methods to a dataset including 516 students' performance on a spatial rotation test (Culpepper & Balamuta, 2017). In addition, including covariates also benefits us by providing insights about the relationships between the covariates and the item success and attribute mastery probabilities.

Chapter 4: In educational environments and online learning platforms, formative assessments can yield valuable information about students' knowledge profiles. Knowing which attribute a student has been mastered versus has not been yet will help educators provide well-targeted instructions. In this respect, exploratory restricted latent class models have significantly been used to estimate students learning profiles from their response patterns. Although students' response patterns are the primary source for estimating students' item

performance and skill mastery profiles, students' covariates may also provide beneficial information in the process. However, one main challenge is to decide which covariates to include in the model when many covariates are available. Thus, this chapter applies a "spike-slab" variable selection algorithm on covariates in an exploratory RCLM, which simultaneously estimates a mapping between items and the attributes. We develop a Bayesian formulation to estimate model parameters while imposing a variable selection algorithm on covariates. We report Monte Carlo evidence pertaining to accurate recovery of $\boldsymbol{Q}$ and other model parameters while correctly identifying the active covariates from inactive ones.

To my family and friends, for their love and support.

Acknowledgments

Love of learning and pure curiosity encouraged me to start my PhD journey. When compared to other journeys undertaken in life, pursuing a PhD was one of the hardest, but also the most rewarding. Through my PhD, I was able to view the world from new lenses and gained new abilities to positively impact the world. Along the way, I've met and worked with a series of outstanding collaborators and fellow students that quickly became life-long friends.

I would like to express my deepest appreciation to my advisers, Dr. Steven Culpepper and Dr. Jeff Douglas. They have always supported and nurtured my ideas over the years. When times were tough, neither advisor wavered in their support. Instead, each advisor provided me with new ideas, encouragement, and patience throughout my PhD journey. In the later years, Dr. Culpepper further challenged me and helped me to grow to the point that this thesis exhausts all the letters in the Greek alphabet. The existence of this thesis is a direct result of their combined kindness and beliefs in me.

I also would like to express my deepest appreciation to my committee members Dr. Steven Culpepper, Dr. Justin Kern, Dr. Jeff Douglas, and Dr. Jimmy de la Torre for their time and their generosity in sharing research ideas and for their tremendous amount of guidance on my research. I would like to especially thank Dr. Jimmy de la Torre for his guidance, mentorship, and warm acceptance to his research group when I pursued my Master's in Education Measurement at Rutgers University.

I must also thank the Turkish Government for providing me this unparalleled opportunity to come and study abroad. Without this scholarship, I could not make my dream come true. I truly appreciate this opportunity.

My growth is highly indebted to my professors in the Department of Statistics. Every class I took moved me one more step closer to my goals and helped prepare a strong foundation for my future life. Through the knowledge gained, I feel that I can reshape my world and the world of others around me. A special thanks goes out to Dr. Alexey Stepanov who was one of my most adamant teachers. I feel extremely lucky to have the privilege of learning from him. Not only did I learn a great deal of knowledge from him but also, he provided me with valuable advice and insightful suggestions for surviving through my journey.

I appreciated the many lively discussions with my peers. With their great statistical knowledge and great explanation capacity, Anamitra Chaudhuri and Neha Agarwal broadened my understanding of the concepts deeply. Moreover, James Balamuta opened an entirely new world of efficient and reproducible computing as well as providing constructive advice on programming skills, which pushed me to new heights both professionally and personally.

This journey has also brought me several life-long friends. While being an entire ocean away from my family, my friends become the extended family that I always wanted. Together, we went through hard times and celebrated each accomplishment. I cannot begin to express my thanks to Apurv Garg, who never wavered in his support of me during this journey and always managed to bring a smile to my face. Together, on-campus and later off, we had many happy adventures and memories, from our on-campus shenanigans that originated from my tech plaza office next to the garbage chute to sunnier beach and farmers market runs in California. Moreover, Na'ama Av-Shalom was a highly influential and amazing person. She provided me with calm and reassuring guidance when I first came to the United States. Moreover, she was always available on the other side of a call whenever I needed some courage, or I needed another perspective. Lastly, I need to thank Sahil Kumar, who become like a brother to me. Through the existence of my friends, I never felt truly alone in this journey on the other side of the world.

My path also provided opportunities to form many new connections with great people.

I'd also like to extend my gratitude to Charles Iaconangelo for showing me the ways of applying my research to another domain. Charles provided me with encouragement, patience throughout the duration of our project, and a profound belief in my abilities. Further, I appreciated the guidance, patience, and communications from my co-author Miguel A. Sorrel in my first ever paper. Moreover, thanks should also go to Mehmet Kaplan, Tuba Arabacı Atlamaz and Umit Atlamaz for all their help, endless support, and advice on how I should approach graduate school life on another continent. My life would have been a lot harder had our paths not crossed.

Finally, I want to thank my family for always being with me throughout this process. Even though this journey was not at all a familiar experience for them, they always supported me and tried to understand what I was going through. I know they will always believe profoundly in my abilities and work. I'm incredibly fortunate to have such a family: from my father Müfit, who encouraged me to explore and learn both at home and an ocean away, to my sister Derya, who has an impeccable fashion taste, to my brother Niyazi, who become a medical doctor before I became a doctor of philosophy, to my sister in-spirit Aydan, who helped me start this adventure and, lastly, to my mother Meliha, who was with me for both the moments of joy and sadness throughout my PhD.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

RLCM	Restricted Latent Class Models
CDM	Cognitive Diagnostic Models
DINA	Deterministic Input, Noisy "AND" Gate Model
DINO	Deterministic Input, Noisy "OR" Gate Model
GDINA	Generalized Deterministic Inputs, Noisy "AND" Gate Model
DIF	Differential Item Functioning
AERA	American Educational Research Association
APA	American Psychological Association
NCME	National Council on Measurement in Education
MCMC	Markov chain Monte Carlo
MH	Metropolis–Hastings Algorithm
SSVS	Stochastic Search Variable Selection
RT	Response Times
MAE	Mean Absolute Errors
AAR	Attribute Agreement Rate
PAR	Attribute Pattern Agreement Rate
RMSE	Root Mean Squared Error
EM	Expectation-Maximization
BIC	Bayesian Information Criterion
LOO	Leave-one-out Cross-validation
GD	Gradient Descent
PSVT-R	Purdue Spatial Visualization Test-Revised

Chapter 1

Introduction

With the increasing emphasis on formative assessment in education, the value of fine-grained feedback on student learning and understanding the longitudinal path of these learning are becoming crucial for traditional classroom assignments. In fact, learning technology algorithms leverage the formative assessments by tailoring learning interventions to adapt to individual students' capabilities and needs (e.g., Chen, Li, Liu, & Ying, 2018; Han, Chen, & Tan, 2020; Huang et al., 2019; X. Li, Xu, Zhang, & Chang, 2021; Tan, Han, Ye, & Chen, 2020) and to track skill development (e.g., F. Li, Cohen, Bottge, & Templin, 2015; Madison & Bradshaw, 2018; Studer, 2012; S. Wang, Yang, Culpepper, & A., 2018; S. Wang, Zhang, Douglas, & Culpepper, 2018; Ye, Fellouris, Culpepper, & Douglas, 2016a; Ye et al., 2016a; S. Zhang & Chang, 2020).

In this respect, diagnostic models, which are also known as restricted latent classification models (RLCMs), cognitive diagnostic models, or diagnostic classification models, provide a statistical framework for designing formative assessments by classifying student knowledge profiles according to a collection of fine-grained attributes (de la Torre & Douglas, 2004; Rupp, Templin, & Henson, 2010; Templin, 2020; von Davier, 2008). One of the main component of RCLMs is a domain-specific item-skill map (i.e., the $\mathbf{Q}$ matrix). Based on the construction of the $\mathbf{Q}$ matrix, RCLMs can be built in two different ways: confirmatory and exploratory. In this regard, there have been many studies conducted on confirmatory RLCMs which need practitioners to establish the $\mathbf{Q}$ matrix and assume that it is a known structure prior to the estimation of the models (e.g. de la Torre, 2011; Haertel, 1989; Henson, Templin, & Willse, 2009; Junker & Sijtsma, 2001; Tatsuoka, 1985). However, in the case

of the misspecification of this structure, model-fit issues may arise alongside potential misclassifications of examinees into latent classes Chen, Liu, Xu, and Ying (2015a). Recently, a significant amount of research has been developed on exploratory RLCMs which infers the underlying structure of $\mathbf{Q}$ matrix with the rest of the model parameters (e.g. Chen, Culpepper, Chen, & Douglas, 2018; Chen, Culpepper, & Liang, 2020; Chen, Liu, Xu, & Ying, 2015b; Culpepper, 2019b; Culpepper & Chen, 2019; G. Xu & Shang, 2018).

As previously indicated, RLCMs are beneficial to estimate a student learning profile, and can only provide static information at a given time. However, by combining with a learning model, they can also provide a longitudinal perspective for student's learning profiles and information for attributes' transition probabilities from non-mastery to mastery states. In this respect, there have been conducted several studies to investigate the RCLMs combined with a learning model (e.g. F. Li et al., 2015; Madison & Bradshaw, 2018; Studer, 2012; S. Wang, Yang, et al., 2018; S. Wang, Zhang, et al., 2018; Ye et al., 2016a; S. Zhang & Chang, 2020). In these studies, Markov chain Monte Carlo (MCMC) algorithm has been commonly used for estimation of the model parameters. The common reasons for using MCMC are a difficulty in formulating a tractable form of the marginal likelihood function for model parameters, and the burden of high dimensionality. In high dimensional parameter spaces, MCMC adopts a sampling process to explore the posterior distribution of parameters after many iterations. Although MCMC is a feasible estimation method in high dimensional scenarios, determining convergence is never certain with finite sample size, and it can require carefully chosen starting values and often a tremendous number of iterations. Thus, a more computationally efficient algorithm may make this process more accessible in classroom settings. As an alternative to MCMC, marginal maximum likelihood estimation with the Expectation-Maximization (EM) can be applied since in many cases it requires shorter computation time and also can better guarantee convergence.

Besides the eminence in the longitudinal perspectives of learning, student covariate is another important component in educational settings. Covariates can be incorporated into

RCLMs from two different angles: as a part of the measurement model possibly affecting students' item performance probabilities, and/or as a component affecting the mastery of the attributes. There is a long-standing tradition and interest in incorporating covariates into psychometric measurement models (Meredith & Millsap, 1992). In fact, including covariates in the measurement models is closely connected with evaluating differential item functioning (DIF), which is important for ensuring the validity of test scores and subsequent diagnostic classifications. Concretely, the presence of DIF means that when two students have the same knowledge and skills, their test results may differ because of irrelevant variables such as race, gender, national origin, or test-taking ability. The implications of DIF are that teachers may incorrectly infer students skills. For example, in artificial intelligence based online learning system, unfamiliarity with the technological interfaces or tools (e.g., calculators, dictionaries, spell-check, etc.) may create differential performance that could be interpreted as target content skill deficits rather than differences in construct irrelevant variables. Furthermore, if online learning systems, that rely upon RCLMs, assume items are DIF-free then the assigned learning modules may target attributes that have already been mastered thereby making the intervention ineffective. We can evaluate the extent to which there is DIF by incorporating covariates into the measurement model. This enables researchers to disentangle the effects of construct-irrelevant covariates from the true underlying latent structure.

Moreover, incorporating covariates effects on the attribute mastery can identify the relevance of the covariates and reveal the association of covariates (if any) with the attribute mastery probabilities. Using possible relevance of the covariates with the attribute mastery may yield more accurate class membership probability estimates. In classroom settings, recognizing the possible associations between the covariates and attributes mastery probabilities may help identify at-risk students and benefit educators to design student-tailored interventions that accelerate skill development. Beyond the educational settings, these insights about the association between covariates and attribute possession probabilities may have a valuable application in clinical studies. Covariates may carry important information

about the subjects' susceptibility to disease and provide as much information as the patient's answers to diagnostic tests.

When including covariates into a model, one challenge is to distinguish active covariates, which relate to the outcome variable, from inactive ones. To address this issue, one can apply a regularization technique to the loss function or incorporate a variable selection mechanism into the priors. In this regard, Iaconangelo (2017) used a regularization technique (i.e., the $\mathcal{L}1$ penalty) to select the active covariates in the structural part of the three-step diagnostic regression approach. In this thesis, we use a "spike-slab" variable selection procedure on both item- and attribute-level covariates. The spike-slab variable selection procedure assumes that the priors of each covariate coefficient follows a mixture of two normal distributions. The variance of the prior distribution is governed by the activeness of the coefficients.

Given the light of the information above, the purpose of this thesis is to improve RCLMs from two different points. In particular, Chapter 2 applies an EM algorithm to increase the efficiency in estimating model parameters of RCLMs combined with a learning model. Chapter 3 develops an exploratory RCLMs model in which the effect of students covariates on item performance and attribute mastery are investigated while simultaneously uncovering the latent structure. Moreover, instead of assuming the independence assumption among attributes, which is difficult to satisfy when fine-grained attributes are measured, the Chapter 3 proposed a structural parameter (i.e., $\boldsymbol{\psi}$) to govern the relationships among the attributes. Chapter 4 extends the methodology in Chapter 3 by incorporating a "spike-slab" variable selection procedure on item- and attribute-level covariates to distinguish active covariates from inactivates one, and yield more parsimony models in the existence of inactive covariates.

Chapter 2

First-Order Learning Models With the GDINA: Estimation With the EM Algorithm and Applications

2.1 Introduction

With the increasing emphasis on formative assessment in education, the value of fine-grained feedback on student learning is becoming crucial for classroom assignments. Moreover, with the great growth in online learning systems, embedded assignments - which can lead students to individualized learning practice - are in great need. To estimate student learning skills, cognitive diagnosis models can be effective. Several studies on CDMs in educational settings, such as English language proficiency (Chiu & Köhn, 2015; Templin & Hoffman, 2013), proportional reasoning (Tjoe & de la Torre, 2014), fraction subtraction (de la Torre & Douglas, 2004), and a few in psychology, such as pathological gambling (Templin & Henson, 2006), social anxiety disorders (Chen, Liu, Xu, & Ying, 2015c), mental disorders (de la Torre, van der Ark, & Rossi, 2018) have been conducted. Although all these studies gave the fine-grained information about the examinees' current status, they can only provide a static picture at a given time-point. However, especially in education, progress in learning material is the entire goal, and assuming that it is a static quantity ignores the purpose of education. Although applying traditional CDMs sequentially at several time-points appears to be a possible solution, this cannot capture the longitudinal perspective of learning directly. Thus, the traditional CDMs should be combined with some methods that can consider the change of skills over time. In this respect, several studies (e.g., Chen, Culpepper, Wang, & Douglas, 2018; Kaya & Leite, 2017; F. Li et al., 2015; Studer, 2012; S. Wang, Yang, et al., 2018; Ye, Fellouris, Culpepper, & Douglas, 2016b) have considered CDMs from a dynamic

perspective. Studer (2012) proposed two models as dynamic CDMs by incorporating CDMs into the knowledge tracing framework (Colbett, Anderson, & OBrien, 1995) and the parameter driven process for change methods. The former one traces the changes at the skill-level and is contrasted with knowledge tracing by considering more than one skill for each item at once and gives more flexibility in choosing a CDM. The latter one traces the change on the level of latent classes rather than the skill levels which is more parsimonious than the previous method. Moreover, while Kaya and Leite (2017) and F. Li et al. (2015) performed the latent transition analysis with CDMs as a measurement model, S. Wang, Yang, et al. (2018) modeled students' learning trajectories with a higher-order Markov model by incorporating several latent and observed covariates, and Chen, Culpepper, Wang, and Douglas (2018) gave a first-order hidden Markov model and applied it to a spatial reasoning training module. Taking another approach, Corbett and Anderson (1994) performed the method of Knowledge Tracing to track students' skill transitions dynamically in intelligent tutoring systems. By integrating assessment items into the learning materials, they tried to monitor and enhance students' skill transitions from unlearned states to the learned states using a probabilistic framework. Although initially Corbett and Anderson (1994) set up the measurement design in a way that every item can only check one skill at a time and were restricted to having identical item parameters, later several studies were designed to remove these limitations (Gonzalez-Brenes, Huang, & Brusilovsky, 2014; Gonzalez-Brenes & Mostow, 2013; Pardos & Heffernan, 2010; Y. Xu & Mostow, 2012). In spite of the different methodologies having been used to model learning trajectories, Markov chain Monte Carlo (MCMC) has been commonly used for estimation of the model parameters. Common reasons are a difficulty in formulating a tractable form of the marginal likelihood function for model parameters, and the burden of high dimensionality which complicates this. In high dimensional parameter spaces, MCMC adopts a sampling process to explore the posterior distribution of parameters after many iterations. Although MCMC is a feasible estimation method in high dimensional scenarios, determining convergence is never certain, and it can

require carefully chosen starting values and often a tremendous number of iterations. An alternative is marginal maximum likelihood estimation with the Expectation-Maximization (EM) algorithm. By treating the latent parameters as "missing values", the objective is to maximize the marginal likelihood of the complete data which includes the observed data and the current estimates of the latent variables based on the observed data and some provisional model parameters. The purpose of this article is to study the EM algorithm in a longitudinal framework with saturated and restricted CDMs as measurement models, and a first-order Markov model for learning transitions. The contribution of this study is to offer a different estimation procedure for the model parameters which can better guarantee convergence and in many cases require shorter computation time. The effectiveness of estimation with the EM-algorithm is investigated by considering different scenarios such as model complexity, number of attributes, and sample size. Moreover, a data example of a module designed to train students in spatial rotation skills is conducted to see the practical outcomes. The rest of paper is organized into five sections. First a brief review of CDM and learning models is given, then a section is devoted to the details of parameter estimation. Later a section on the effectiveness of the procedure with simulated data is given, followed by the results of a data analysis. Finally a discussion section is provided to summarize our findings and discuss possible future directions.

2.2 Model Description

The proposed model can be considered the combination of two submodels: the measurement model and the learning model. The cognitive diagnosis measurement model is used to estimate the examinees' latent skill classes by using observed responses, which is the static part of the model, and the learning model is used for modeling skill acquisitions over time. Following is a brief review of CDMs with learning.

2.2.1 Cognitive Diagnosis Models

Many CDMs with different assumptions and constraints have been proposed to investigate the mastery status of student knowledge on a fine-grained basis. These models can be categorized based on their constraints in several ways. One way is to categorize a CDM as compensatory or noncompensatory. Noncompensatory models require an examinee to master all the attributes required by an item in order to correctly answer it, whereas compensatory models allow some compensation for missing attributes by use of the others. Moreover, based on how heavily the model is parameterized, they can be anywhere from very restrictive to saturated. Saturated models naturally give better fit by requiring more parameters to estimate than the restricted models, which are usually easier to interpret and often more practical.

Let a binary response vector denote an examinee's answer to items, with 1 indicating a correct answer and 0 otherwise. In CDMs with binary attributes there are 2^K possible latent classes which can be denoted by $\boldsymbol{\alpha}_i = \{\alpha_{ik}\}$ for $k = 1, 2, \ldots, K$, with K being the number of attributes assessed. The entries in the latent classes specify the mastery status for an examinee on the K skills, with 1 indicating mastery and 0 nonmastery. Let $\boldsymbol{Q}$ denote a matrix with J x K dimensions in order to map the items with the attributes. Each row represents an item and includes 1s and 0s depending on whether the item requires the attribute or not, respectively. A number of CDMs with different assumptions and constraints on the data such as the deterministic inputs, noisy "and" gate model (DINA; Junker & Sijtsma, 2001), the deterministic inputs, noisy "or" gate model (DINO; Templin & Henson, 2006), and the generalized DINA model (G-DINA; de la Torre, 2011) have been proposed. In the present study the saturated GDINA and a restricted version of it, the DINA model, are used as the measurement models.

The DINA model is one of the most commonly used CDMs because of its easy interpretation and practicality. In this model, each examinee has an ideal response vector, $\boldsymbol{\eta}_i = \{\eta_{ij}\}$ for $j = 1, \ldots, J$, where each entry is either $\eta_{ij} = 1$ or $\eta_{ij} = 0$ depending on whether the ex-

aminee has all the attributes required by the item or lacks at least one of them, respectively. The entries of the ideal response vector can be calculated for item j and examinee i:

$$\eta_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}}.$$

Moreover, each item has two item parameters referring to either guessing or slipping. The guessing parameter is the probability of giving a correct answer to the item when the ideal response of an examinee is $\eta_{ij} = 0$, $g_j = P(X_{ij} = 1|\eta_{ij} = 0)$, and the slipping parameter is the probability of giving an incorrect answer to the item when the examinee's ideal response is $\eta_{ij} = 1$, $s_j = P(X_{ij} = 0|\eta_{ij} = 1)$. The item response function for examinee i and item j is

$$P(X_{ij} = 1|\boldsymbol{\alpha}_i, s_j, g_j, \boldsymbol{q}_j) = (1 - s_j)^{\eta_{ij}} g_j^{(1-\eta_{ij})}. \tag{2.1}$$

The DINA model cannot distinguish among examinees who have the ideal response $\eta_{ij} = 0$ because it assigns the same probability regardless of how many attributes the examinee has not mastered (Rupp & Templin, 2008). A generalized version of the DINA overcomes this limitation by including many more item parameters. In the GDINA, examinees are partitioned into $2^{K_j^*}$ reduced latent groups by an item j where K_j^* is the number of attributes that the item j requires. The response function of GDINA is,

$$P(X_{ij} = 1|\boldsymbol{\alpha}_{ij}^*, \boldsymbol{\delta}) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{lk}\alpha_{lk'} + ... + \delta_{j12...K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}. \tag{2.2}$$

where δ_{j0} is the baseline effect which can be interpreted as the probability of correctly answering an item for those who have mastered none of the attributes required by the item. Parameter δ_{jk} is the main effect for the kth attribute on item j, which is the change of the probability of giving a correct answer to an item once the examinee has achieved mastery on the single attribute. $\delta_{jkk'}$ has the same interpretation as δ_{jk} except that it is a change due to having mastery on both attributes, and delta $\delta_{j12...K_j^*}$ is the interaction effect due to

having multiple attributes at once.

2.2.2 Learning models

Cognitive diagnosis models provide a clear picture about the examinees' learning profiles on a fine-grained level, but they can only provide this information at a specific time-point. However, during practice or a learning intervention, it is likely that some changes in students' attribute profiles take place at each point in time from t to $t+1$ where $t = 1, 2, 3, ..., T$. Learning models may be used to capture changes between any two discrete recorded time-points during a longitudinal learning assessment. Under this framework, examinee i, $i = 1, 2, ..., I$ has a latent class vector (i.e., will be referred to as learning trajectory after this point) which includes the latent classes that the examinee is classified for each time-point $\boldsymbol{\alpha}_i = (\boldsymbol{\alpha}_{i1}, \boldsymbol{\alpha}_{i2}, \boldsymbol{\alpha}_{i3}, ..., \boldsymbol{\alpha}_{iT})$ where $\boldsymbol{\alpha}_{it} = (\alpha_{it1}, \alpha_{it2}, \alpha_{it3}, ..., \alpha_{itK})^\top$, and K is the number of attributes. Moreover, each examinee also has an observed response vector which includes the response vectors at each time-point $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \boldsymbol{X}_{i3}, ..., \boldsymbol{X}_{iT})$ with the tth entry $\boldsymbol{X}_{it} = (X_{it1}, X_{it2}, X_{it3}, ..., X_{itJ(t))})^\top$. At each time- point, there might be multiple items administered to the examinees. Thus, $J(t)$ represents the number of items in each time-point, and J denotes the total number of items in the entire assessment, $J = \sum_{t=1}^{T} J(t)$. The number of items to be administered might be different from time-point to time-point but in the present study, it will be the same at all time-points. Moreover, under learning models, $\boldsymbol{Q}_t$ denotes a $\boldsymbol{Q}$ matrix with the $J(t)$ x K dimension at each time-point, which possesses the same interpretation under measurement models. Under the current study, $\boldsymbol{\beta}_t$ stands for the item parameter vector for the items administered at time-point t, and they are assumed to be time invariant. Specifically, it is assumed that the item parameters do not vary depending on which time-points they are administered.

To model the attribute transitions between the learning states (i.e., mastery or non-mastery), a Markov model with a monotonicity constraint is used. The monotonic trend under this frameworks assumes that the probability of transition from a non-mastery to

mastery state of a skill does not depend on which states any other skill has been. Moreover, another constraint has been imposed: Once a transition occurs on a skill from a non-mastery state to a mastery state, the probability of transitioning back to non-mastery state is zero for an examinee i and an attribute k $\exists t, t'$, and $t, t' \geq 1, \forall k$,

$$P(\alpha_{it'k} = 1 | \alpha_{itk} = 1) = 1 \ \forall t^{'} > t, \text{ and } P(\alpha_{it'k} = 0 | \alpha_{itk} = 1) = 0 \ \forall t^{'} > t.$$

In addition, the transition probabilities are time invariant, $\forall t, t'$, and $t, t' \geq 1$,

$$\tau_k = P(\alpha_{i(t+1)k} = 1 | \alpha_{itk} = 0) = P(\alpha_{i(t'+1)k} = 1 | \alpha_{it'k} = 0).$$

Based on the assumptions above, the probability that examinee i has $\boldsymbol{\alpha}_{l(t+1)}$ at time $t+1$ given examinee i has $\boldsymbol{\alpha}_{l'(t)}$ at time t, and the skill transition probabilities $\boldsymbol{\tau}$ where $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_K)$ is

$$P(\boldsymbol{\alpha}_{i(t+1)} = \boldsymbol{\alpha}_{l(t+1)} | \boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_{l't}, \boldsymbol{\tau}) = \prod_{k=1}^{K} P(\alpha_{i(t+1)k} = \alpha_{l(t+1)k} | \alpha_{itk} = \alpha_{l'tk}, \tau_k). \tag{2.3}$$

2.3 Parameter Estimation

To estimate the model and examinees' parameters the expectation-maximization algorithm is used. The initial estimates of the unknown population probabilities of the possible learning trajectories are obtained by using random starting values for transition probabilities and attribute prevalence. Once the probabilities of learning trajectories are determined, the log marginalized likelihood is calculated across response data and initial item parameters. Later, the item parameters and the skill transition probabilities are updated based on maximization of the log-likelihood and updated probabilities of the learning trajectories respectively. After several iterations on the skill transition probabilities and model parameters, the algorithm stops if the difference of the marginalized log-likelihood between two consecutive iterations

is under a certain cut point. The E- and M- steps of the algorithm will be discuss in detail as follows.

E-Step: The likelihood of examinee i's responses to items $J(t)$ at time t given the $\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_{lt}$ and item parameters $\boldsymbol{\beta}_t$ is,

$$P(\boldsymbol{X}_{it}|\boldsymbol{\alpha}_{lt}, \boldsymbol{\beta}_t) = \prod_{j=1}^{J(t)} P(X_{ijt} = 1|\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_{lt}, \boldsymbol{\beta}_t)^{X_{ijt}}(1 - P(X_{ijt} = 1|\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_{lt}, \boldsymbol{\beta}_t))^{1-X_{ijt}}.$$

where $P(X_{ijt} = 1|\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_{lt}, \boldsymbol{\beta}_t)$ is calculated depending on which measurement model is used. Assuming item responses are independent over time and independent given the learning trajectory, the likelihood of the response data of examinee i with the attribute trajectory $\boldsymbol{\alpha}_l$ is,

$$P(\boldsymbol{X}_i|\boldsymbol{\alpha}_l, \boldsymbol{\beta}) = \prod_{t=1}^{T}\prod_{j=1}^{J(t)} P(X_{ijt} = 1|\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_{lt}, \boldsymbol{\beta}_t)^{X_{ijt}}(1 - P(X_{ijt} = 1|\boldsymbol{\alpha}_{it} = \boldsymbol{\alpha}_{lt}, \boldsymbol{\beta}_t))^{1-X_{ijt}}.$$

The likelihood of the response data of all examinees is,

$$L(\boldsymbol{\alpha}, \boldsymbol{\beta}; \boldsymbol{X}) = \prod_{i=1}^{I} P(\boldsymbol{X}_i|\boldsymbol{\alpha}_l, \boldsymbol{\beta})$$

The marginalized likelihood of the response data is,

$$L(\boldsymbol{\beta}; \boldsymbol{X}) = \prod_{i=1}^{I}\sum_{l=1}^{2^{KT}} P(\boldsymbol{X}_i|\boldsymbol{\alpha}_l, \boldsymbol{\beta})\Big[\prod_{k=1}^{K} P(\alpha_{1k})\prod_{t=2}^{T} P(\alpha_{ltk}|\alpha_{l(t-1)k})\Big],$$

and the log-marginalized likelihood of the response data is,

$$l(\boldsymbol{\beta}; \boldsymbol{X}) = \sum_{i=1}^{I} log\Big[\sum_{l=1}^{2^{KT}} P(\boldsymbol{X}_i|\boldsymbol{\alpha}_l, \boldsymbol{\beta})\big[\prod_{k=1}^{K} P(\alpha_{1k})\prod_{t=2}^{T} P(\alpha_{ltk}|\alpha_{l(t-1)k})\big]\Big].$$

M-Step: The solution to partial derivatives of the log-marginalized likelihood with respect to item parameters produces the intermediate item parameters' estimates. Currently, the mod-

els under consideration have closed-form solutions to the derivatives of the log-marginalized likelihood. In the case of adopting a CDM model which does not have closed-form solution (e.g., additive model), any standard optimization algorithm may be used. For example, Gradient Descent (GD; Cauchy, 1847) can be performed to find the item parameter estimates that can maximize the the log-marginalized likelihood. With a standard optimization algorithm, the extension is straightforward because the item parameters are assumed not to vary depending on which time-points they are administered. In the present study, the item parameters are updated by using the closed-form solutions with the aid of intermediate estimates for examinees' probability distributions over all the learning trajectories from the E-step. First, the estimate of having a correct answer to item j given the reduced attribute group $\boldsymbol{\alpha}_{ij}^{*}$ is $\hat{P}(\boldsymbol{\alpha}_{ij}^{*}) = \frac{\boldsymbol{R}_{\boldsymbol{\alpha}_{lj}^{*}}}{\boldsymbol{I}_{\boldsymbol{\alpha}_{lj}}}$, where $\boldsymbol{I}_{\boldsymbol{\alpha}_{lj}}$ is the expected number of examinees to be in the reduced latent group $\boldsymbol{\alpha}_{lj}$,and $\boldsymbol{R}_{\boldsymbol{\alpha}_{lj}}$ is the expected number of examinees to answer item j correctly in the reduced latent group $\boldsymbol{\alpha}_{lj}$. By denoting the set of examinees who are administered item j at timepoint t as $\{\bar{N}_{ijt}\}$, $\boldsymbol{I}_{\boldsymbol{\alpha}_{lj}}$ and $\boldsymbol{R}_{\boldsymbol{\alpha}_{lj}}$ can be given by,

$$\boldsymbol{I}_{\boldsymbol{\alpha}_{lj}} = \sum_{t=1}^{T} \left[\sum\nolimits_{i' \in \{\bar{N}_{ijt}\}} P(\boldsymbol{\alpha}_{ltj}^{*} | \boldsymbol{X}_{i't}) \right],$$

$$\boldsymbol{R}_{\boldsymbol{\alpha}_{lj}} = \sum_{t=1}^{T} \left[\sum\nolimits_{i' \in \{\bar{N}_{ijt}\}} \left[P(\boldsymbol{\alpha}_{ltj}^{*} | \boldsymbol{X}_{i't}) \; X_{i'tj} \right] \right].$$

After updating $\hat{P}(\boldsymbol{\alpha}_{ij}^{*})$, one can find the $\boldsymbol{\delta}_j$ coefficients of the GDINA model by using an appropriate design matrix explained in de la Torre (2011). Moreover, by applying the appropriate design matrices and after some algebraic manipulations, the item parameters of the DINA model, can be derived from GDINA parameters.

The next step is to recalculate the prevalence of the learning trajectories. First, the

expected number, n_l, of students in learning trajectory $\boldsymbol{\alpha_l}$ is calculated as,

$$n_l = \sum_{i=1}^{I} P(\boldsymbol{\alpha}_l|\boldsymbol{X}_i;\boldsymbol{\beta}).$$

Then the learning trajectory distribution is updated by,

$$P(\boldsymbol{\alpha}_l|\boldsymbol{\beta}) = \frac{n_l}{I}, l = 1, 2, ..., L.$$

With the updated learning trajectory distribution, the attributes' prevalence at the initial time-point and the transition probabilities for each attribute can be estimated respectively as it follows.

The probabilities of having and not having attribute k at initial time-point $t = 1$ are updated by,

$$\Pi_{k1} = \sum_{l=1}^{2^{KT}} \mathbb{1}[\alpha_{l1k} = 1]\ P(\boldsymbol{\alpha_l}|\boldsymbol{\beta}),$$

$$\Pi_{k0} = \sum_{l=1}^{2^{KT}} \mathbb{1}[\alpha_{l1k} = 0]\ P(\boldsymbol{\alpha_l}|\boldsymbol{\beta}).$$

The probability of transitioning from non-mastery to non-mastery on attribute k at any timepoint is,

$$\tau_{k00} = \frac{\sum_{t=1}^{T-1}\left[\sum_{l=1}^{2^{KT}} \mathbb{1}[\alpha_{l(t+1)k} = 0|\alpha_{ltk} = 0]\ P(\boldsymbol{\alpha_l}|\boldsymbol{\beta})\right]}{\sum_{r=0}^{1}\sum_{t=1}^{T-1}\left[\sum_{l=1}^{2^{KT}} \mathbb{1}[\alpha_{l(t+1)k} = r|\alpha_{ltk} = 0]\ P(\boldsymbol{\alpha_l}|\boldsymbol{\beta})\right]}.$$

The probability of transitioning from non-mastery to mastery on attribute k at any time-

point is,

$$\tau_{k01} = 1 - \tau_{k01} = \frac{\sum_{t=1}^{T-1}\left[\sum_{l=1}^{2^{KT}} \mathbb{1}[\alpha_{l(t+1)k} = 1|\alpha_{ltk} = 0]\ P(\boldsymbol{\alpha_l}|\boldsymbol{\beta})\right]}{\sum_{r=0}^{1}\sum_{t=1}^{T-1}\left[\sum_{l=1}^{2^{KT}} \mathbb{1}[\alpha_{l(t+1)k} = r|\alpha_{ltk} = 0]\ P(\boldsymbol{\alpha_l}|\boldsymbol{\beta})\right]}.$$

2.4 Simulation Study

A Monte Carlo simulation study with 50 independent replications is performed to investigate the efficiency of the algorithm on the recovery rate of student and model parameters. Overall, the sample size, number of attributes, length of test blocks, and the true model under which the data are generated are treated as factors. In detail, to investigate the effect of sample size on the accuracy of parameter estimation, three different sample sizes — 250, 500 and 1000 - are considered. Due to the negative effect of increasing the number of attributes, the conditions are explored with four and six attributes to study the effects on computation time and estimation accuracy. Two different time-points $T = 3$ and $T = 5$ which result in different total numbers of items and item parameters are considered. Thus, the study has four different $\boldsymbol{Q}$ matrices as they are presented in Table 2.1.

The fist 30 items (i.e., 10 items per time point) are the common items in the $\boldsymbol{Q}$ matrices. Thus, the $\boldsymbol{Q}$ matrix for $T = 3$ is nested in the $\boldsymbol{Q}$ matrix for $T = 5$. Two different CDMs are fitted, DINA and GDINA, and their response functions can be found in Equation 1 and Equation 2, respectively. Under the DINA model, guessing and slipping parameters are drawn from $U(0.05, 0.15)$. For the GDINA model, we categorize the item parameters under three groups: *guessing*, *slipping*, and *others*. The *guessing* item parameters are sampled from $U(0.05, 0.15)$ for the latent group lacking all required attributes, the *slipping* item parameters are drawn from $U(0.05, 0.15)$ to cover the latent group with all required attributes, and, lastly, the *others* group that includes the remaining latent groups have their item parameters picked from $U(0, 1)$ to represent the range of the values in real applications.

Table 2.1: Q matrix

Item	$K = 4$				$K = 6$					
	K1	K2	K3	K4	K1	K2	K3	K4	K5	K6
Item1	1	0	0	0	1	0	0	0	0	0
Item2	0	1	0	0	0	1	0	0	0	0
Item3	0	0	1	0	0	0	1	0	0	0
Item4	0	0	0	1	0	0	0	1	0	0
Item5	1	0	0	0	0	0	0	0	1	0
Item6	0	1	0	0	0	0	0	0	0	1
Item7	0	0	1	0	1	1	0	0	0	0
Item8	0	0	0	1	1	0	1	0	0	0
Item9	1	1	0	0	1	0	0	1	0	0
Item10	1	0	1	0	1	0	0	0	1	0
Item11	1	0	0	1	1	0	0	0	0	1
Item12	0	1	1	0	0	1	1	0	0	0
Item13	0	1	0	1	0	1	0	1	0	0
Item14	0	0	1	1	0	1	0	0	1	0
Item15	1	1	0	0	0	1	0	0	0	1
Item16	1	0	1	0	0	0	1	1	0	0
Item17	1	0	0	1	0	0	1	0	1	0
Item18	0	1	1	0	0	0	1	0	0	1
Item19	0	1	0	1	0	0	0	1	1	0
Item20	0	0	1	1	0	0	0	1	0	1
Item21	1	1	1	0	0	0	0	0	1	1
Item22	1	1	0	1	1	1	1	0	0	0
Item23	1	0	1	1	1	1	0	1	0	0
Item24	0	1	1	1	1	0	1	1	0	0
Item25	1	1	1	0	1	0	0	1	0	1
Item26	1	1	0	1	0	1	1	1	0	0
Item27	1	0	1	1	0	1	1	0	1	0
Item28	0	1	1	1	0	1	0	0	1	1
Item29	1	0	0	0	0	0	1	1	1	0
Item30	0	1	0	0	0	0	0	1	1	1
Item31	0	1	0	1	1	1	0	0	0	1
Item32	0	0	1	1	1	0	1	1	0	0
Item33	1	1	0	0	1	0	1	0	1	0
Item34	1	0	1	0	1	0	1	0	0	1
Item35	1	0	0	1	1	0	0	1	1	0
Item36	0	1	1	0	1	0	0	1	0	1
Item37	0	1	0	1	1	0	0	0	1	1
Item38	0	0	1	1	0	1	1	1	0	0
Item39	1	1	1	0	0	1	1	0	1	0
Item40	1	1	0	1	0	1	1	0	0	1
Item41	1	0	1	1	0	1	0	1	1	0
Item42	0	1	1	1	0	1	0	1	0	1
Item43	1	1	1	0	0	1	0	0	1	1
Item44	1	1	0	1	0	0	1	1	1	0
Item45	1	0	1	1	0	0	1	1	0	1
Item46	0	1	1	1	0	0	1	0	1	1
Item47	1	1	1	0	0	0	0	1	1	1
Item48	1	1	0	1	0	0	1	1	0	0
Item49	1	0	1	1	1	1	0	0	0	0
Item50	0	1	1	1	0	0	0	1	1	1

A first-order hidden Markov model is assumed for learning transitions with a constraint that the transition probability from mastery to non-mastery at each stage is 0. The prevalence of each attribute is set to 0.5 at the initial time point. In addition, the attribute transition probability from non-mastery to mastery for all attributes is 0.25. For each examinee, there is a learning transition profile generated under the following procedure. First, the probabilities for every possible trajectory on the attribute level are calculated based on the transition and the initial prevalence probabilities. With these attribute level trajectory probabilities, K attribute level trajectories are randomly sampled for each examinee. To obtain each examinee's learning trajectory, the K attribute level trajectories are rearranged. To evaluate the efficiency of the model, attribute agreement rate (AAR), $\sum_{i=1}^{N}\sum_{k=1}^{K}\frac{\hat{\alpha}_{ikt}=\alpha_{ikt}}{NK}$, and attribute pattern agreement rate (PAR), $\sum_{i=1}^{N}\frac{I|\hat{\boldsymbol{\alpha}}_{it}=\boldsymbol{\alpha}_{\boldsymbol{it}}|}{N}$, are chosen as criteria to investigate person parameters in addition to the root mean squared error (RMSE) for the model parameters. The average time that the algorithm takes to converge on each Monte Carlo sample is reported for different conditions on the minute time unit. Due to space constraints, we present GDINA item parameters along with their RMSE across three different categories previously defined as: *guessing*, *slipping*, and *others*. Within the *others* category, RMSE is taken to be to the averaged RMSE values for the rest of the item parameters associated with the remaining latent groups.

Table 2.2: Attribute pattern agreement rates

Model	N	K	$T=3$			$T=5$				
			1	2	3	1	2	3	4	5
GDINA	250	4	0.823 (0.004)	0.877 (0.003)	0.871 (0.003)	0.833 (0.004)	0.893 (0.003)	0.924 (0.002)	0.939 (0.003)	0.931 (0.002)
		6	0.570 (0.007)	0.668 (0.006)	0.667 (0.005)	0.584 (0.006)	0.710 (0.004)	0.782 (0.004)	0.822 (0.004)	0.812 (0.005)
	500	4	0.832 (0.003)	0.890 (0.003)	0.877 (0.003)	0.840 (0.003)	0.902 (0.002)	0.933 (0.002)	0.946 (0.002)	0.937 (0.002)
		6	0.589 (0.005)	0.681 (0.004)	0.677 (0.005)	0.601 (0.005)	0.723 (0.003)	0.792 (0.003)	0.832 (0.003)	0.819 (0.003)
	1000	4	0.835 (0.003)	0.893 (0.002)	0.881 (0.002)	0.843 (0.002)	0.906 (0.001)	0.933 (0.001)	0.945 (0.001)	0.937 (0.001)
		6	0.592 (0.004)	0.687 (0.003)	0.684 (0.003)	0.616 (0.003)	0.737 (0.003)	0.799 (0.002)	0.836 (0.002)	0.824 (0.002)
DINA	250	4	0.794 (0.005)	0.897 (0.002)	0.909 (0.003)	0.781 (0.004)	0.893 (0.003)	0.944 (0.002)	0.966 (0.001)	0.966 (0.002)
		6	0.609 (0.005)	0.759 (0.004)	0.803 (0.003)	0.602 (0.004)	0.771 (0.004)	0.854 (0.003)	0.906 (0.003)	0.914 (0.002)
	500	4	0.794 (0.003)	0.896 (0.002)	0.911 (0.002)	0.786 (0.002)	0.893 (0.002)	0.941 (0.001)	0.964 (0.001)	0.967 (0.001)
		6	0.619 (0.003)	0.764 (0.002)	0.806 (0.002)	0.609 (0.003)	0.776 (0.003)	0.866 (0.002)	0.915 (0.002)	0.919 (0.002)
	1000	4	0.794 (0.002)	0.893 (0.002)	0.912 (0.002)	0.784 (0.002)	0.894 (0.001)	0.945 (0.001)	0.966 (0.001)	0.968 (0.001)
		6	0.627 (0.003)	0.766 (0.002)	0.806 (0.002)	0.605 (0.003)	0.772 (0.002)	0.869 (0.002)	0.916 (0.001)	0.922 (0.001)

As shown in Table 2.2, the EM algorithm can provide accurate attribute pattern agreement rates. As the model complexity increases, the classification rates decrease slightly. Also the correct classification rates decline as the number of attributes increases from 4 attributes to 6 attributes. The increase in the number of the time points, which involves an increase in the total number of items, results in a rise in accuracy, especially in the 6-attribute conditions.

Table 2.3: Averaged attribute agreement rates

Model	N	K	$T = 3$			$T = 5$				
			1	2	3	1	2	3	4	5
GDINA	250	4	0.947 (0.001)	0.966 (0.001)	0.963 (0.001)	0.950 (0.001)	0.970 (0.001)	0.979 (0.001)	0.984 (0.001)	0.981 (0.001)
		6	0.899 (0.002)	0.928 (0.001)	0.928 (0.001)	0.899 (0.001)	0.937 (0.001)	0.956 (0.000)	0.965 (0.000)	0.963 (0.000)
	500	4	0.951 (0.001)	0.969 (0.001)	0.965 (0.001)	0.953 (0.006)	0.973 (0.004)	0.982 (0.003)	0.986 (0.003)	0.983 (0.003)
		6	0.905 (0.001)	0.932 (0.001)	0.931 (0.001)	0.906 (0.001)	0.941 (0.001)	0.959 (0.001)	0.967 (0.001)	0.965 (0.001)
	1000	4	0.951 (0.001)	0.970 (0.001)	0.967 (0.001)	0.953 (0.001)	0.974 (0.000)	0.982 (0.000)	0.985 (0.000)	0.983 (0.000)
		6	0.906 (0.001)	0.933 (0.001)	0.932 (0.001)	0.910 (0.001)	0.944 (0.001)	0.960 (0.000)	0.968 (0.000)	0.966 (0.001)
DINA	250	4	0.930 (0.002)	0.969 (0.001)	0.973 (0.001)	0.919 (0.002)	0.964 (0.001)	0.983 (0.001)	0.990 (0.000)	0.991 (0.000)
		6	0.898 (0.001)	0.943 (0.001)	0.956 (0.001)	0.892 (0.001)	0.944 (0.001)	0.968 (0.001)	0.981 (0.001)	0.983 (0.001)
	500	4	0.931 (0.001)	0.968 (0.001)	0.974 (0.001)	0.922 (0.001)	0.964 (0.001)	0.982 (0.000)	0.990 (0.000)	0.991 (0.000)
		6	0.901 (0.001)	0.944 (0.001)	0.956 (0.001)	0.893 (0.001)	0.945 (0.001)	0.971 (0.001)	0.983 (0.000)	0.984 (0.000)
	1000	4	0.931 (0.001)	0.967 (0.001)	0.974 (0.001)	0.921 (0.001)	0.964 (0.000)	0.983 (0.000)	0.990 (0.000)	0.991 (0.000)
		6	0.903 (0.001)	0.945 (0.001)	0.957 (0.001)	0.892 (0.001)	0.945 (0.001)	0.971 (0.000)	0.983 (0.000)	0.985 (0.000)

Table 2.3 presents the attribute agreement rates which are averaged over the attributes and the Monte Carlo standard deviations of the replications. The patterns are quite similar to vector-wise classification rates. As the model complexity increases the classification rates slightly decrease in the 6-attribute conditions, but in the 4-attribute conditions the classification rates are mostly the same.

Table 2.4: Root mean square error (RMSE) transition probabilities from 0 to 1

Model	N	K	$T = 3$	$T = 5$
GDINA	250	4	0.033 (0.002)	0.023 (0.001)
		6	0.042 (0.002)	0.028 (0.001)
	500	4	0.022 (0.001)	0.016 (0.001)
		6	0.031 (0.001)	0.019 (0.001)
	1000	4	0.017 (0.001)	0.012 (0.001)
		6	0.020 (0.001)	0.012 (0.001)
DINA	250	4	0.031 (0.002)	0.024 (0.001)
		6	0.037 (0.001)	0.027 (0.001)
	500	4	0.021 (0.001)	0.018 (0.001)
		6	0.026 (0.001)	0.017 (0.001)
	1000	4	0.016 (0.001)	0.013 (0.001)
		6	0.017 (0.001)	0.013 (0.001)

In Table 2.4, the RMSE of probabilities from the non-mastery state to mastery state are presented. As the number of time points increases, the RMSEs decrease slightly. The change in the sample size does have a decreasing effect on the RMSEs. Moreover, the increase in the number of attributes does not cause any change in the RMSEs. Model complexity does not affect the RMSEs of transition probabilities. The computer run times per iteration for converging to the maximum likelihood estimate under different conditions are recorded and presented in Table 2.5.

Table 2.5: Average time (minutes) to until the chain converges

Model	N	K	$T = 3$	$T = 5$
GDINA	250	4	0.060	0.232
			(0.002)	(0.009)
		6	1.110	15.231
			(0.074)	(0.654)
	500	4	0.093	0.439
			(0.004)	(0.017)
		6	1.823	28.121
			(0.052)	(0.906)
	1000	4	0.153	0.776
			(0.008)	(0.032)
		6	3.519	48.804
			(0.098)	(1.330)
DINA	250	4	0.017	0.062
			(0.001)	(0.004)
		6	0.294	3.863
			(0.018)	(0.216)
	500	4	0.030	0.118
			(0.000)	(0.002)
		6	0.428	5.408
			(0.013)	(0.116)
	1000	4	0.058	0.239
			(0.001)	(0.003)
		6	0.866	11.172
			(0.011)	(0.143)

Overall, the algorithm seems quite efficient with respect to the computational run time. The longest time the algorithm takes is on the 6-attribute conditions with 1000 sample size (48.80 minutes). This is because more iterations are needed to converge in this condition. Table 2.6 presents the RMSE of the *guessing* and *slipping* parameters for both DINA and GDINA, and averaged RMSE values for the rest of the item parameters associated with the remaining latent groups in GDINA, *other*. As expected, the RMSEs decrease as the sample size increases, and this decreasing pattern is clearer in the conditions with GDINA. The increase in the number of attributes usually causes a decrease on the accuracy of the parameter estimates, especially with the GDINA with the small sample size. The reason that RMSEs are relatively high in the GDINA model might be due to the number of parameters

to be estimated in GDINA are higher than the DINA model.

Table 2.6: Root mean square error (RMSE) of guessing and slipping

Model	N	K	$T = 3$			$T = 5$		
			Guessing	Slipping	Other	Guessing	Slipping	Other
GDINA	250	4	0.061 (0.002)	0.036 (0.001)	0.085 (0.001)	0.075 (0.002)	0.028 (0.000)	0.094 (0.001)
		6	0.089 (0.003)	0.043 (0.001)	0.104 (0.002)	0.111 (0.003)	0.032 (0.000)	0.113 (0.002)
	500	4	0.044 (0.001)	0.025 (0.000)	0.057 (0.001)	0.052 (0.001)	0.020 (0.000)	0.062 (0.001)
		6	0.057 (0.002)	0.029 (0.001)	0.070 (0.001)	0.073 (0.001)	0.023 (0.000)	0.074 (0.001)
	1000	4	0.031 (0.001)	0.017 (0.000)	0.040 (0.001)	0.037 (0.001)	0.014 (0.000)	0.044 (0.001)
		6	0.042 (0.001)	0.021 (0.000)	0.049 (0.001)	0.053 (0.001)	0.016 (0.000)	0.052 (0.001)
DINA	250	4	0.029 (0.001)	0.034 (0.001)	N/A	0.031 (0.000)	0.028 (0.000)	N/A
		6	0.030 (0.001)	0.038 (0.001)	N/A	0.032 (0.001)	0.030 (0.001)	N/A
	500	4	0.020 (0.000)	0.023 (0.000)	N/A	0.022 (0.000)	0.020 (0.000)	N/A
		6	0.020 (0.000)	0.270 (0.000)	N/A	0.022 (0.000)	0.021 (0.000)	N/A
	1000	4	0.014 (0.000)	0.017 (0.000)	N/A	0.015 (0.000)	0.014 (0.000)	N/A
		6	0.014 (0.000)	0.018 (0.000)	N/A	0.015 (0.000)	0.015 (0.000)	N/A

2.5 Analysis of Spatial Rotation Data

This section is devoted to a real data application of the proposed model on the data collected by S. Wang, Yang, et al. (2018). By extending and embedding a revised version of the Purdue Spatial Visualization Test (PSVT-R; Yoon, 2011) into an online learning platform, S. Wang, Yang, et al. (2018) created an interactive learning tool for the students' rotation skills. The dataset with 50 items was collected from the 351 subjects studied in the Department

of Psychology at the University of Illinois at Urbana-Champaign. The assessment with intervention is to administer to a block of cognitively-based questions to students and later provide relevant interactive learning materials in order to boost learning. A total of 5 different item blocks each containing 10 spatial reasoning items were administered at 5 different time-points. In order to eliminate the effect of time on the item parameter estimates (i.e., effect of mastering the attributes over time), each test block was administered to examinees randomly at different time-points so that they were distributed evenly over item positions. The questions were constructed to check the presence of four attributes related to the rotation skills: (a) 90 ° x axis, (b) 90° y axis, (c) 180 ° x axis, and (d) 180 ° y axis. The complexity of the questions was varied depending on how many skills were being checked simultaneously. In present data set, the maximum number of skills an item can possibly include is up to 3. Each item is comprised of an object with two different rotations, and students are expected to figure out which rotation is applied. The latent classes of students at each time-point are estimated under the GDINA and DINA models. With the same monotonicity constraint constraints on the attribute transitions, the transition parameters of 4 attributes are estimated with a first-order hidden Markov model. Figure 2.1 presents the estimated guessing and $1-$slipping parameters under the DINA and GDINA models. Although the item parameters are not identical in the two models, they are considerably close to each other. The slipping parameter estimates appear to be reasonable whereas the guessing parameters are relatively high with some of them exceeding 0.5. This problem likely resulted from the multiple choice format in which some of the wrong asnwers could easily be eliminated from consideration. Table 2.7 shows the initial state probabilities for the mastery state, it can be seen that over half of the students have mastered all four attributes at time-point one. Furthermore, Table 2.8 reflects that the transition probabilities from the non-mastery state to mastery state are quite low. Particularly, the possibility that learning attribute three during the intervention is zero.

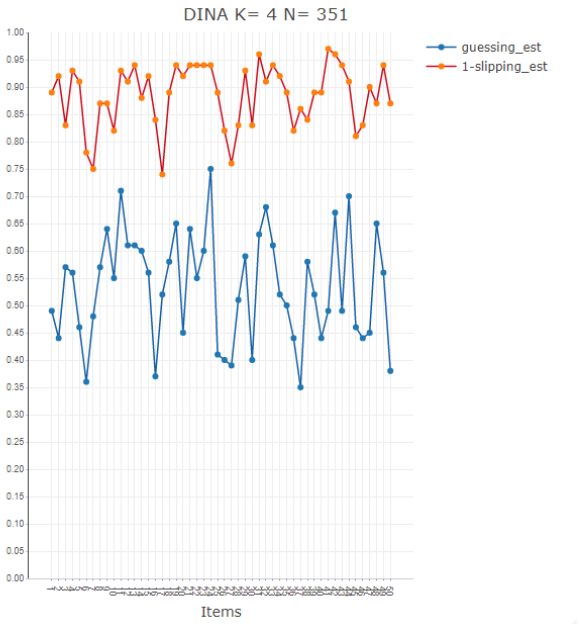

Figure 2.1: Real data estimated slipping and guessing parameters under the DINA and GDINA model

Table 2.7: Estimated initial probabilities for the mastery state

Model	Attribute			
	1	2	3	4
GDINA	0.588	0.628	0.564	0.522
DINA	0.698	0.634	0.630	0.569

Table 2.8: Estimated transition probabilities from non-mastery to mastery state

Model	Attribute			
	1	2	3	4
GDINA	0.0002	0.0413	0.0000	0.1478
DINA	0.0002	0.0711	0.0000	0.1805

For a further analysis on real data, we define a couple discrepancy measures to investigate the model fit versus actual observations that span various important aspects of the model, one corresponding to the measurement model and the other corresponding to learning transitions. The behavior of the model for item parameters is investigated by the discrepancy between the proportion of people who answered items correctly under the real data and the proportion expected under the fitted model, and do so over time for the learning aspect. Later, we use the parametric bootstrap method to approximate an observed significance level using these discrepancy measures (for an overview of parametric bootstrap, see; Cheng, Holland, & Hughes, 1996). After various analyses, the results show that under both the DINA and GDINA models the discrepancy between the proportion of people who answered items correctly under the real data and the proportion expected under the fitted model are 66 percent of the time less than the discrepancy between the proportion of people who answered items correctly under the bootstrap samples and the proportion expected under the fitted model on the bootstrap samples, which indicates 33 out of 50 items gives an acceptable model fit results under 10 percent significance level. Furthermore, the misfitting items were generally the same ones under both models. The results of the discrepancy measure analysis for the transition probabilities showed that the GDINA model yields

a discrepancy on the real data 88 percent of the times less than the parametric bootstrap simulations, which means the parametric bootstrap p-value is 0.88, indicating no reason to believe misfit. By comparison, the parametric bootstrap p-value for transition probabilities when using the DINA measurement model was a smaller but still insignificant 0.33. Thus, it can be concluded that both methods explain the Spatial Rotation data well in terms of describing the transition probabilities. Alongside the approximate p values for discrepancy measures, we also investigate the Bayesian information criterion (BIC) for model selection between the DINA and GDINA models. The DINA model results lower BIC value (i.e., 14392.41) then the GDINA model (i.e., 14480.88). Thus, the DINA model is preferable to GDINA on Spatial Rotation data.

2.6 Discussion

In education, the importance of formative assessment has been emphasized more and more. Tracking what students have learned in detail during learning interventions could be very helpful for designing individualized learning environments and accelerating productivity. CDMs can be seen as the models answering this need efficiently. However, traditional CDMs cannot provide any information on the transitions of attributes or on factors that have clear effects on learning practices. Under this instance, modeling the transitions and possible co-variates can be beneficial. However, the parameter estimation process can be quite challenging because of the increase on the number of parameters to be estimated.

In the current literature of modeling transition probabilities (e.g., Chen, Culpepper, Wang, & Douglas, 2018; Kaya & Leite, 2017; S. Wang, Yang, et al., 2018), MCMC is widely used because of the complexity of the marginal likelihood, and the relative simplicity of the complete likelihood. However, MCMC can require long run times, and there is never complete certainty that it has reached equilibrium so that parameter estimates converge to the true posterior mean. The Expectation-Maximization algorithm can be seen as an

alternative to MCMC for parameter estimation, that more directly targets a point estimate.

In the present study, we examine the EM algorithm by varying several factors which are complexity of measurement model, number of attributes, sample size, and number of time points. The computation run times are studied. Results show that estimation accuracy for the patternwise and attribute wise classification rates are promising for a majority of the conditions. For the condition with the relatively low sample size and high number of attributes, the correct classification rate drops significantly, and the RMSE of parameter estimates increases, which is expected. In conjunction with the accurate results, the computation run time is also quite short. Except for only three conditions, all others are under 5 minutes, even some taking as little as a few seconds. One reason for the relatively high computation time of the three conditions is the greater number of iterations required to converge to the maximum likelihood estimate.

Chapter 3

Extending Exploratory Diagnostic Classification Models: Inferring the Effect of Covariates

3.1 Introduction

Formative assessments have been considered as a foremost component of the traditional classroom setting. Diagnostic models provide a statistical framework for designing formative assessments by classifying student knowledge profiles according to a collection of fine-grained attributes (de la Torre & Douglas, 2004; Rupp et al., 2010; Templin, 2020; von Davier, 2008). In fact, learning technology algorithms leverage the diagnostic modeling framework by tailoring learning interventions to adapt to individual students' capabilities and needs (e.g., Chen, Li, et al., 2018; Han et al., 2020; Huang et al., 2019; X. Li et al., 2021; Tan et al., 2020) and to track skill development (e.g. F. Li et al., 2015; Madison & Bradshaw, 2018; Studer, 2012; S. Wang, Yang, et al., 2018; S. Wang, Zhang, et al., 2018; Ye et al., 2016a; S. Zhang & Chang, 2020).

The context by which students learn may be important to understand when making diagnostic decisions. For instance, students may receive distinct interventions that target specific learning objections or students may be multiskilled with prior knowledge and experiences that inform mastery on the target attributes. It is therefore important to understand how a broader collection of student characteristics, which we refer to as covariates, shape performance and attribute mastery. The existing research demonstrates the value of using covariates to shape formative assessments in education. The advantages of including the students' covariates into diagnostic models can be clustered around three primary purposes. Covariates can detect the possible differential item functioning (DIF) to provide fair test

designs across diverse populations (e.g. F. Li, 2008; X. Li & Wang, 2015; Z. Wang, Guo, & Bian, 2014; W. Zhang, 2006). Covariates are also used to explain individual differences or serve as indicators to evaluate educational intervention effects (e.g. Ayers, Rabe-Hesketh, & Nugent, 2013; Iaconangelo, 2017; Minchen, de la Torre, & Liu, 2017; S. Wang, Yang, et al., 2018; Zhan, Jiao, & Liao, 2018). Thus, exploring covariates might help to identify at-risk students and benefit educators to design student-tailored interventions that accelerate skill development. Moreover,using covariates, like response times (RTs) or other student experiences associated with the target attributes, might improve the accuracy of the model parameter estimates and attribute classification rates (e.g. S. Wang, Zhang, et al., 2018; S. Wang, Zhang, & Shen, 2020).

All the existing studies above analyze covariates using the restricted latent class models (i.e., RLCMs) in a confirmatory fashion. The application of confirmatory methods is appropriate in cases where the underlying structure in terms of how attributes relate to observed performance is precisely known (i.e., the $\boldsymbol{Q}$ matrix). However, using an incorrect latent structure can adversely impact classification decisions (Henson & Templin, 2007), so exploratory methods that infer the latent structure from student responses may be preferred. Recently, a significant amount of research has also developed exploratory RLCMs which infer the underlying structure of the Q-matrix (e.g. Chen, Culpepper, Chen, & Douglas, 2018; Chen et al., 2020, 2015b; Culpepper, 2019b; Culpepper & Chen, 2019; G. Xu & Shang, 2018). However, one short-coming of existing exploratory RLCMs is that current methods are not designed to incorporate covariates, which may limit the types of inferences and insights that are available to applied researchers and educators.

In this paper, we offer new methods for including covariates in exploratory RLCMs. Specifically, we develop three general strategies for including covariates in exploratory RLCMs. For instance, we consider models that include covariates at the item level (i.e., the $\boldsymbol{\Gamma}$ model), the attribute level (i.e., the $\boldsymbol{\zeta}$ model), or both item and attribute levels (i.e., the saturated model) simultaneously. For the $\boldsymbol{\Gamma}$ model, we investigate the effect of the covari-

ates on the probability of correctly answering an item. Similarly, the $\boldsymbol{\zeta}$ model considers the relationship between the covariates and the probability that a student possesses a particular attribute. We present a Markov chain Monte Carlo (MCMC) algorithm using a Metropolis-within-Gibbs algorithm for approximating the model parameter posterior distribution.

The layout of the rest of the paper is as follows. First, we introduce our novel framework for incorporating covariates into exploratory RLCMs. Second, we present a Bayesian model specification and discuss details related to posterior inference. Third, we report a Monte Carlo simulation study design and discuss its results concerning the accuracy of model parameter recovery. Fourth, we apply our method to a spatial rotation test and demonstrate the types of inferences that are available with our modeling framework. Finally, we provide a discussion section with the final remarks and future research directions.

3.2 Model Specification

We denote an individual with i; $i = 1, \ldots, N$, an item with j; $j = 1, \ldots, J$, and an attribute with k; $k = 1, \ldots, K$. We let $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)^\top$ denote the $N \times J$ response matrix of random binary responses with the ith row defined as $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ})^\top$ and we let the observed value be $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)^\top$ where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})^\top$ and $y_{ij} \in \{0, 1\}$. We consider the case where observed responses are described by an underlying set of attributes. That is, each individual respondent has a latent binary attribute pattern $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK})^\top$ in which each entry will be 1 if the individual has mastered that attribute and 0 otherwise.

The purpose of this section is to introduce our general modeling framework for incorporating covariates into the exploratory RLCM framework. We consider two types of covariates, which are covariates that relate to: 1) item responses; and 2) attributes. Moreover, $\boldsymbol{X}$ is an item-related matrix of covariates with the dimension $N \times V$ where V is the number of covariates, and the ith row for an individual i is $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iV})^\top$. Similarly, $\boldsymbol{Z}$ is an attribute-related covariate matrix with the dimension $N \times L$ where L is the

number of covariates included in $\boldsymbol{Z}$, and the ith row is denoted by $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iL})^\top$. It is important to note that $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are free to include the same covariates.

3.2.1 Model for Item-Related Covariates

To simplify the formulation, we introduce a design matrix such that $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N)^\top$ is a matrix with the dimension $N \times 2^K$, where $\boldsymbol{a}_i$ is a design vector for individual i based on the $\boldsymbol{\alpha}_i$. As an example, let's assume $K = 3$. The attribute pattern for an individual i will be in the format $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3})^\top$. Thus, the design matrix will be $\boldsymbol{a}_i = (1, \alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i1}\alpha_{i2}, \alpha_{i1}\alpha_{i3}, \alpha_{i2}\alpha_{i3}, \alpha_{i1}\alpha_{i2}\alpha_{i3})^\top$ where the first element (i.e., "1") stands for the intercept and the remaining elements correspond with main-effects and interaction terms. We consider the following item response for the probability that individual i correctly respond to item j,

$$\theta_{ij} = P(Y_{ij} = 1 | \boldsymbol{\alpha}_i, \boldsymbol{\gamma}_j, \boldsymbol{\beta}_j) = \Phi\big(\boldsymbol{a}_i^\top \boldsymbol{\beta}_j + \boldsymbol{x}_i^\top \boldsymbol{\gamma}_j\big). \tag{3.1}$$

where Φ is the cumulative distribution function of the standard normal distribution. In Equation 3.1, the relationship between attributes and item responses is denoted by the vector of regression coefficients, $\boldsymbol{\beta}_j$. $\boldsymbol{\beta}_j$ is the 2^K vector of regression coefficients indexed by $\{\beta_{jp}\}_{p=0}^{2^K-1}$. The relationship between the item-level covariates for individual i, $\boldsymbol{x}_i$, and response probabilities is $\boldsymbol{\gamma}_j$, which is a V vector of coefficients for item j.

3.2.2 Model for Attribute-Related Covariates

Previous research modeled dependence among the attributes using a variety of strategies. The most common approach is to use an unstructured $\boldsymbol{\pi}$ vector, and other options include using a more parsimonious structure (e.g. Chen & Culpepper, 2020; de la Torre & Douglas, 2004; Henson et al., 2009; Templin, Henson, Templin, & Roussos, 2008). We do not assume the attributes are independent when constructing the latent class probabilities. Thus, the

joint distribution for the attribute profile for individual i can be factored as

$$p(\boldsymbol{\alpha}_i) = p(\alpha_{i1})p(\alpha_{i2}|\alpha_{i1}) \dots p(\alpha_{iK}|\alpha_{i1}, \dots, \alpha_{i,K-1}). \tag{3.2}$$

We use the factored representation of the joint distribution for attributes to model the effect student covariates have on the probability of latent class membership. We define a parameter $\boldsymbol{\psi} = (\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_K)^\top$ as a $K \times 2^{K-1}$ matrix that describes the structure of the relationship among the attributes. In addition, we incorporate the covariates at the attribute level. Thus, the probability examinee i belongs to an attribute class $\boldsymbol{\alpha}_i$ is influenced by both examinee's covariates as well as the relations among attributes. Specifically, we use the following probit link to model the joint distribution for the probability of class membership for individual i,

$$p(\boldsymbol{\alpha}_i|\boldsymbol{\psi}, \boldsymbol{\zeta}) = \prod_{k=1}^{K} \Phi\left(\boldsymbol{a}_{i,K-1}^\top \boldsymbol{\psi}_k + \boldsymbol{z}_i^\top \boldsymbol{\zeta}_k\right)^{\alpha_{ik}} \left(1 - \Phi\left(\boldsymbol{a}_{i,K-1}^\top \boldsymbol{\psi}_k + \boldsymbol{z}_i^\top \boldsymbol{\zeta}_k\right)\right)^{1-\alpha_{ik}} \tag{3.3}$$

where $\boldsymbol{a}_{i,K-1}$ is a design vector defined by a subset of $\boldsymbol{\alpha}_i$ pattern (i.e., $(\alpha_{i1}, \dots, \alpha_{i,K-1})$). For example, for $K = 3$, the design matrix will be $\boldsymbol{a}_{i,K-1} = (1, \alpha_{i1}, \alpha_{i2}, \alpha_{i1}\alpha_{i2})^\top$ where the first element (i.e., 1) stands for the intercept. The rows of $\boldsymbol{\psi}$ include a fixed pattern of zeros based upon the order the attribute distribution is factored (e.g., see Equation 3.2). Thus, the number of non-zero elements in $\boldsymbol{\psi}$ equals to $2^K - 1$. For example, for $K = 3$, $\boldsymbol{\psi}_1 = (\psi_{11}, 0, 0, 0)^\top$, $\boldsymbol{\psi}_2 = (\psi_{21}, \psi_{22}, 0, 0)^\top$, and $\boldsymbol{\psi}_3 = (\psi_{31}, \psi_{32}, \psi_{33}, \psi_{34})^\top$. The relationship between the covariates and attributes are denoted by the $K \times L$ matrix of coefficients $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_K)^\top$ where for individual i, $\boldsymbol{\zeta}_k$ is the relationship between $\boldsymbol{z}_i$ and the conditional attribute mastery probabilities.

3.2.3 Likelihood for the Saturated Model

We next describe the likelihood for the saturated model. Note we discuss several important special cases of our model in the next subsection. The conditional likelihood of observing $\boldsymbol{y}_i$

under the local independence given the item-level and attribute-level covariates is

$$p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i, \boldsymbol{\Gamma}, \boldsymbol{B}) = \prod_{j=1}^{J} \theta_{ij}^{y_{ij}} (1-\theta_{ij})^{1-y_{ij}} \tag{3.4}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_J)^\top$ is the $J \times V$ item-level regression coefficients matrix and $\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)^\top$ denotes the $J \times 2^K$ matrix of the regression coefficients.

Thus, the likelihood for individual i is

$$p(\boldsymbol{y}_i|\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\psi}, \boldsymbol{\zeta}) = \sum_{c=0}^{2^K-1} \pi_{ic} p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{B}, \boldsymbol{\Gamma}) \tag{3.5}$$

where we define $\pi_{ic} = P(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c|\boldsymbol{\psi}, \boldsymbol{\zeta})$ to be the conditional probability that individual i belongs to class c and we note the use of the vector $\boldsymbol{v} = (2^{\{K-1\}}, 2^{\{K-2\}}, \ldots, 1)^\top$ to create a bijection between the binary attribute pattern and a integer c, such that $c = 0, \ldots, 2^K - 1$. The likelihood function for a sample of N independent observations can be formulated as the product of N respondents' likelihoods,

$$p(\boldsymbol{y}|\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\psi}, \boldsymbol{\zeta}) = \prod_{i=1}^{N} \sum_{c=0}^{2^K-1} \pi_{ic} p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{B}, \boldsymbol{\Gamma}) \tag{3.6}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)^\top$ is a $N \times J$ matrix of responses.

3.2.4 Special Cases

From the saturated model, we derived two different parsimonious models. One model (i.e., the $\boldsymbol{\Gamma}$ model) associates covariates with the probability of correctly answering an item, but not the attributes, and the other model (i.e., the $\boldsymbol{\zeta}$ model) associates covariates with the probability of mastering an attribute, but not the item responses. Moreover, in order to investigate the benefit of the covariates on uncovering the latent structure, we also consider the base model, which excludes attributes. The connection between the saturated model

and the special cases are presented below:

- Saturated model where both $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ are estimated.
- Base model where $\boldsymbol{\Gamma} = \mathbf{0}$ and $\boldsymbol{\zeta} = \mathbf{0}$.
- $\boldsymbol{\Gamma}$ model where $\boldsymbol{\Gamma}$ is estimated and $\boldsymbol{\zeta} = \mathbf{0}$.
- $\boldsymbol{\zeta}$ model where $\boldsymbol{\Gamma} = \mathbf{0}$ and $\boldsymbol{\zeta}$ is estimated.

The various special cases may be relevant for different researchers. For instance, researchers interested assessing DIF may be more interested in the $\boldsymbol{\Gamma}$ model to assess whether covariates relate to item responses. In contrast, intervention studies may focus on the $\boldsymbol{\zeta}$ model given the goal is to test hypotheses about how experimental conditions or student characteristics relate to attribute mastery. Additionally, researchers would deploy the base model in the absence of covariates.

3.3 Bayesian Inference of the Saturated Model

We next discuss a Bayesian model for inferring the model parameters. We first discuss the prior specifications. Then introduce the full conditional distributions and describe a Metropolis-within-Gibbs sampling algorithm for approximating the posterior distribution.

3.3.1 Bayesian Formulation

The following subsection introduces our novel Bayesian formulation. The following five subsections discuss: 1) data augmentation for item responses; 2) priors for the effects of covariates for item responses; 3) data augmentation for attributes; 4) priors for the effects of covariates for attribute mastery probabilities; and 5) priors for inferring the latent structure related parameters.

Item Augmented Data

We use the classic probit data augmentation approach (Albert & Chib, 1993). That is, we define a deterministic relationship between the observed binary responses and a continuous augmented random variable as $Y_{ij} = \mathbb{1}(Y_{ij}^* > 0)$. Next, in order to augment our model we specify the following normal distribution for the augmented data,

$$Y_{ij}^* | \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j, \boldsymbol{x}_i, \boldsymbol{\gamma}_j \sim \mathcal{N}(\boldsymbol{a}_i^\top \boldsymbol{\beta}_j + \boldsymbol{x}_i^\top \boldsymbol{\gamma}_j, 1). \tag{3.7}$$

Item-Level Covariate Coefficients

We adopt a Zellner (1986) g-prior for the covariate regression coefficients, $\boldsymbol{\gamma}_j$. That is, the prior for $\boldsymbol{\gamma}_j$ is the following multivariate normal distribution,

$$\boldsymbol{\gamma}_j | v_\gamma \sim \mathcal{N}_V(\mathbf{0}, (v_\gamma^{-1})(\boldsymbol{X}^\top \boldsymbol{X})^{-1}) \tag{3.8}$$

where the mean is a vector of zeros and the prior variance-covariance matrix is a function of the predictor cross-products (i.e., $\boldsymbol{X}^\top \boldsymbol{X}$) and a scalar precision, v_γ. Further note we choose a gamma prior for the hyper-parameter v_γ, $v_\gamma \sim \text{gamma}(a_{\gamma 0}, b_{\gamma 0})$.

Attributes and Augmented Data

We use the probit data augmentation for attributes, as well. Specifically, we use the following formation for α_{ik},

$$\begin{aligned} \alpha_{ik} &= \mathbb{1}(\alpha_{ik}^* > 0) \\ \alpha_{ik}^* | \alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{i,k-1}, \boldsymbol{\zeta}_k, \boldsymbol{\psi}_k &\sim \mathcal{N}(\boldsymbol{a}_{i,K-1}^\top \boldsymbol{\psi}_k + \boldsymbol{z}_i^\top \boldsymbol{\zeta}_k, 1) \end{aligned} \tag{3.9}$$

where α_{ik}^* is a continuous, augmented version of α_{ik} that has a normal distribution conditioned on the first $k-1$ attributes, $(\alpha_{i1}, \ldots, \alpha_{i,k-1})$, and the coefficients $\boldsymbol{\zeta}_k$ and $\boldsymbol{\psi}_k$.

Attribute-Level Covariate Coefficients

The $\boldsymbol{\zeta}$ parameter reveals the relationship between the attribute-level covariates and attribute mastery probabilities. We specify a g-prior for $\boldsymbol{\zeta}_k$ in the following multivariate normal distribution with a vector of zeros for the mean and a variance-covariance matrix, which is a function of the covariate cross-products and a scalar precision v_ζ,

$$\boldsymbol{\zeta}_k | v_\zeta \sim \mathcal{N}_L(\mathbf{0}, (v_\zeta^{-1})(\boldsymbol{Z}^\top \boldsymbol{Z})^{-\mathbf{1}}) \tag{3.10}$$

where v_ζ is sampled from the gamma distribution, $v_\zeta \sim \text{gamma}(a_{\zeta 0}, b_{\zeta 0})$, and $\boldsymbol{Z}$ is an attribute-related covariate matrix as described in the model selection section.
We specified a multivariate normal prior for the $2^K - 1$ non-zero elements of $\boldsymbol{\psi}_k$'s to model the relationships among the attributes,

$$\boldsymbol{\psi}_k | v_\psi \sim \mathcal{N}_{2^{k-1}}(\mathbf{0}_{2^{k-1}}, (v_\psi^{-1})\boldsymbol{I}_{2^{k-1}}) \tag{3.11}$$

where the variance-covariance matrix assumes attributes are *a priori* independent with v_ψ as a hyper parameter sampled from the gamma distribution $v_\psi \sim \text{gamma}(a_{\psi 0}, b_{\psi 0})$.

Latent structure related parameters

In exploratory RLCMs, researchers need to infer an underlying structure between attributes and the binary, domain-specific item-skill map— i.e., the $\boldsymbol{Q}$ matrix. Assuming that the elements of $\boldsymbol{Q}$ are conditionally independent and follow Bernoulli distributions—as $q_{jk}|\nu \sim \text{Bernoulli}(\nu)$ where ν has a beta prior $\nu \sim \text{beta}(a_\nu, b_\nu)$, the joint prior distribution for $\boldsymbol{Q}$ is,

$$p(\boldsymbol{Q}|\nu) \propto \Big(\prod_{j=1}^{J} \prod_{k=1}^{K} \nu^{q_{jk}} (1-\nu)^{1-q_{jk}} \Big) \mathbb{1}(\boldsymbol{Q} \in \mathbb{Q}) \tag{3.12}$$

where $\mathbb{Q}$ denotes the identifiable space of the $\boldsymbol{Q}$ matrix. Culpepper (2019a) introduced a fully Bayesian model for inferring $\boldsymbol{Q}$ while applying a "spike-slab" prior for the $\boldsymbol{\beta}_j$, but one limita-

tion of the approach is that the formulation imposed a more restrictive monotonicity condition. They introduced a structure, $\tilde{q}_j = (1, q_{j1}, \ldots, q_{jK}, q_{j1}q_{j2}, \ldots, q_{j(K-1)}q_{jK}, \ldots, \prod_{k=1}^{K} q_{jk})'$, to define the activeness of each β_{jp}. Later, Chen et al. (2020) introduced a sparse latent class model with the latent structure defined by a $\boldsymbol{\Delta} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_J)$ matrix rather than a $\boldsymbol{Q}$ matrix. In this setting, the elements of $\boldsymbol{\delta}_j$ denote the activeness of each β_{jp} parameter, $\boldsymbol{\delta}_j = (1, \delta_{j1}, \ldots, \delta_{jK}, \delta_{j12}, \ldots, \delta_{j(K-1)K}, \ldots, \delta_{j1\ldots K})' \in \{0, 1\}^{2^K}$. Thus, $\boldsymbol{\delta}_j$ and $\boldsymbol{\beta}_j$ are connected in a way that a δ_{jp} is 1 if the corresponding parameter β_{jp} is active, and zero otherwise. Chen et al. (2020) relaxed the restrictive monotonicity condition of Culpepper (2019a), but the approach does not provide a mechanism for specifying expert knowledge about the $\boldsymbol{Q}$ matrix. In addition, a method proposed by Balamuta and Culpepper (2021) allows the inclusion of expert knowledge about $\boldsymbol{Q}$ links $\boldsymbol{\beta}_j$ to $\boldsymbol{q}_j$ through the structure of $\boldsymbol{\delta}_j$ as $p(\boldsymbol{\beta}|\boldsymbol{Q}) = \sum_{\text{all}\boldsymbol{\Delta}} p(\boldsymbol{\beta}|\boldsymbol{\Delta}) \times p(\boldsymbol{\Delta}|\boldsymbol{Q})$. To establish a stochastic relationship between $\boldsymbol{\delta}_j$ and $\boldsymbol{q}_j$, Balamuta and Culpepper (2021) used a confirmatory DINA model with common guessing, g, and slipping, s, parameters across all $\boldsymbol{q}_j$.

Thus, the prior distribution of δ_{jp} is parameterized as follow

$$p(\delta_{jp}|\boldsymbol{q}_j, g, s) \sim g^{(1-\tilde{q}_{pj})}(1-s)^{\tilde{q}_{pj}} \tag{3.13}$$

where $\tilde{q}_{jp}$ is the pth entry of the $\tilde{\boldsymbol{q}}_{\boldsymbol{j}}$ in Culpepper (2019a).

Additionally, we deploy the following truncated, conditionally normal prior for $\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp}$ as follows,

$$p(\beta_{jp}|\delta_{jp}, \boldsymbol{\beta}_{j(p)}) \sim v_p^{-1/2} \exp\Big(-\frac{1}{2}\beta_{jp}^2/v_p\Big)\mathbb{1}(\beta_{jp} > L_{jp}) \tag{3.14}$$

where $\boldsymbol{\beta}_{j(p)}$ excluded β_{jp} from $\boldsymbol{\beta}_j$, L_{jp} is the lower bound for the β_{jp}, and δ_{jp} is the activeness parameter associated with p. Note this is a stochastic search variable selection (SSVS) prior (Culpepper, 2019a; George & McCulloch, 1993) such that $v_{jp} = \delta_{jp}/w_1 + (1-\delta_{jp})/w_0$ where the precisions for the spike and slab are w_0 and w_1, respectively.

3.3.2 Posterior Inference

Next, we specify the full conditional distributions for the parameters under consideration for our Metropolis-within-Gibbs sampling algorithm. Similar to the previous section, we present the posterior specifications under five subsections:1) data augmentation for item responses; 2) posterior for the effects of covariates for item responses; 3) data augmentation for attributes; 4) posterior for the effects of covariates for attribute mastery probabilities; 5) posterior for inferring the latent structure related parameters.

Item Augmented Data

The full conditional distribution of the augmented response data follows a truncated normal distribution,

$$Y_{ij}^*|Y_{ij} = y_{ij}, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j, \boldsymbol{\gamma}_j \sim \mathcal{N}(\boldsymbol{a}_i^\top \boldsymbol{\beta}_j + \boldsymbol{x}_i^\top \boldsymbol{\gamma}_j, 1)\mathbb{1}(Y_{ij}^* > 0)^{y_{ij}}\mathbb{1}(Y_{ij}^* \leq 0)^{1-y_{ij}}. \tag{3.15}$$

where $\boldsymbol{\beta}_j$ denotes the relationship between attributes and item responses and $\boldsymbol{a}_i$ is a design vector for individual i based on the $\boldsymbol{\alpha}_i$. $\boldsymbol{x}_i$ is the ith student's item-related covariate vector, and $\boldsymbol{\gamma}_j$ is the item-related covariate coefficient vector associated item j.

Item-Level Covariate Coefficients

The full conditional distribution of $\boldsymbol{\gamma}_j$ is a multivariate normal distribution with a V-dimensional covariate space,

$$\boldsymbol{\gamma}_j|\boldsymbol{y}_{1:N,j}^*, \boldsymbol{\alpha}_{1:N}, \boldsymbol{\beta}_j \sim \mathcal{N}_V(\boldsymbol{\mu}_{\gamma j}, \boldsymbol{\Sigma}_\gamma) \tag{3.16}$$

where $\boldsymbol{y}_{1:N,j}^* = (y_{1j}^*, \ldots, y_{Nj}^*)'$, $\boldsymbol{\alpha}_{1:N} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N)$, $\boldsymbol{\Sigma}_\gamma = ((1+v_\gamma)\boldsymbol{X}^\top\boldsymbol{X})^{-1}$ and $\boldsymbol{\mu}_{\gamma j} = \boldsymbol{\Sigma}_\gamma \boldsymbol{X}^\top \left(\boldsymbol{y}_{1:N,j}^* - \boldsymbol{A}\boldsymbol{\beta}_j\right)$. In addition, the full conditional distribution of the hyper-parameter v_γ is $v_\gamma|\boldsymbol{\Gamma} \sim \text{gamma}(a_{\gamma 1}, b_{\gamma 1})$ where $a_{\gamma 1} = \frac{JV}{2} + a_{\gamma 0}$ and $b_{\gamma 1} = \frac{1}{2}\text{tr}\left(\boldsymbol{\Gamma}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\Gamma}^\top\right) + b_{\gamma 0}$.

Attributes and Augmented Data

The full conditional distribution for an individual's alpha pattern $\boldsymbol{\alpha}_i$ is a categorical distribution with the conditional posterior probability of membership in class c being proportional to,

$$p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c|\boldsymbol{y}_i, \boldsymbol{\Gamma}, \boldsymbol{B}, \boldsymbol{\zeta}, \boldsymbol{\psi}) \propto \pi_{ic} \prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{\gamma}_j, \boldsymbol{\beta}_j). \tag{3.17}$$

The full conditional distribution for the augmented attribute data is a truncated normal distribution,

$$\alpha_{ik}^*|\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{i,k-1}, \boldsymbol{\zeta}_k, \boldsymbol{\psi}_k \sim \mathcal{N}(\boldsymbol{a}'_{i,K-1}\boldsymbol{\psi}_k + \boldsymbol{z}_i^\top \boldsymbol{\zeta}_k, 1)\mathbb{1}(\alpha_{ik}^* > 0)^{\alpha_{ik}}\mathbb{1}(\alpha_{ik}^* \leq 0)^{1-\alpha_{ik}}. \tag{3.18}$$

where $\boldsymbol{a}_{i,K-1}$ is a design vector defined by a subset of $\boldsymbol{\alpha}_i$ pattern (i.e., $(\alpha_{i1}, \ldots, \alpha_{i,K-1})$), the non-zero elements of the $\boldsymbol{\psi}_k$ characterizes the structure of the relationship among the attributes, and $\boldsymbol{\zeta}_k$ is covariate coefficients between $\boldsymbol{z}_i$ and the attribute mastery probabilities.

Attribute-Level Covariate Coefficients

We specify the full conditional distribution of the attribute-related parameters $\boldsymbol{\zeta}$ as

$$\boldsymbol{\zeta}_k|\ \boldsymbol{\alpha}_{1:N,k}^*, \boldsymbol{\alpha}_{1:N,1}, \ldots, \boldsymbol{\alpha}_{1:N,k-1}, \boldsymbol{\psi}_k \sim \mathcal{N}_L(\boldsymbol{\mu}_{\zeta k}, \boldsymbol{\Sigma}_\zeta) \tag{3.19}$$

where $\boldsymbol{\alpha}_{1:N,k}^* = (\alpha_{1k}^*, \ldots, \alpha_{Nk}^*)^\top$, $\boldsymbol{\Sigma}_\zeta = ((1+v_\zeta)\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}$, and $\boldsymbol{\mu}_{\zeta k} = \boldsymbol{\Sigma}_\zeta \boldsymbol{Z}^\top\big(\boldsymbol{\alpha}_{1:N,k}^* - \boldsymbol{A}_{K-1}\boldsymbol{\psi}_{\boldsymbol{k}}\big)$, $\boldsymbol{A}_{K-1} = (\boldsymbol{a}_{1,K-1} \ldots \boldsymbol{a}_{N,K-1})^\top$ is a design matrix with the dimension $N \times 2^{K-1}$. Moreover, the full conditional distribution of the hyper-parameter v_ζ is $v_\zeta|\boldsymbol{\zeta} \sim \text{gamma}(a_{\zeta 1}, b_{\zeta 1})$ where $a_{\zeta 1} = \frac{NL}{2} + a_{\zeta 0}$ and $b_{\zeta 1} = \frac{1}{2}\text{tr}\left(\boldsymbol{\zeta}\boldsymbol{Z}^\top\boldsymbol{Z}\boldsymbol{\zeta}^\top\right) + b_{\zeta 0}$. The posterior distribution of ψ_{ks} for an attribute k and a parameter s is specified to be normally distributed. As previously stated, we use the parameter, ψ_{ks}, to model the structural relationships among the attributes, which is used to model the attribute structural probabilities. This yields the dimension of the $\boldsymbol{\psi}$ as $K \times 2^{K-1}$ with fixed pattern of zeros based upon the order the attribute distribution is factored (e.g.,

see subsection "Model for Attribute-Related Covariates"). Thus, the conditional posterior distribution of each element is

$$\psi_{ks}|\boldsymbol{\alpha}^*_{1:N,k}, \boldsymbol{\zeta}_k, \boldsymbol{\psi}_{k(s)} \sim \mathcal{N}(\mu_{ks}, \sigma^2_{ks}) \tag{3.20}$$

where $\boldsymbol{\alpha}^*_{1:N,k} = (\alpha^*_{1k}, \ldots, \alpha^*_{Nk})^\top$, $\boldsymbol{\psi}_{k(s)}$ is $\boldsymbol{\psi}_k$ without element s, $\sigma^2_{ks} = (\boldsymbol{A}^\top_{K-1,s}\boldsymbol{A}_{K-1,s} + v_\psi)^{-1}$, and $\mu_{ks} = \sigma^2_{ks}\ \boldsymbol{A}^\top_{K-1,s}\big(\boldsymbol{\alpha}^*_{1:N,k} - \boldsymbol{A}_{K-1,(s)}\boldsymbol{\psi}_{k(s)} - \boldsymbol{Z}\boldsymbol{\zeta}_k\big)$. Also, $\boldsymbol{A}_{K-1} = (\boldsymbol{a}_{1,K-1} \ldots \boldsymbol{a}_{N,K-1})^\top$ is a design matrix with the dimension $N \times 2^{K-1}$, $\boldsymbol{A}_{K-1,s}$ is column s of $\boldsymbol{A}_{K-1}$, and $\boldsymbol{A}_{K-1,(s)}$ excludes column s of $\boldsymbol{A}_{K-1}$. In addition, the full conditional distribution of the hyper-parameter v_ψ is $v_\psi|\boldsymbol{\psi} \sim \text{gamma}(a_{\psi 1}, b_{\psi 1})$ where $a_{\psi 1} = \frac{2^K-1}{2} + a_{\psi 0}$ and $b_{\psi 1} = \frac{1}{2}\text{tr}\left(\boldsymbol{\psi}\boldsymbol{\psi}^\top\right) + b_{\psi 0}$.

Latent structure related parameters

The full conditional distribution for β_{jp} is

$$\beta_{jp}|\ \boldsymbol{Y}^*_j, \boldsymbol{\alpha}, \boldsymbol{\beta}_{j(p)}, \delta_{jp}, \boldsymbol{\gamma}_j, \boldsymbol{q}_j \sim \begin{cases} \mathcal{N}(\mu_{jp}, \sigma^2_{jp})\mathbb{1}(\beta_{jp} > 0) & 1 \leq p \leq K \\ \mathcal{N}(\mu_{jp}, \sigma^2_{jp}) & \text{otherwise} \end{cases} \tag{3.21}$$

where $\sigma^2_{jp} = (\boldsymbol{A}^\top_p \boldsymbol{A}_p + v^{-1}_{jp})^{-1}$ and $\mu_{jp} = \sigma^2_{jp}\ \boldsymbol{A}^\top_p\big(\boldsymbol{Y}^*_j - \boldsymbol{A}_{(p)}\boldsymbol{\beta}_{j(p)} - \boldsymbol{X}\boldsymbol{\gamma}_j\big)$. $\boldsymbol{A}_p$ is the pth column of the design matrix $\boldsymbol{A}$, $\boldsymbol{A}_{(p)}$ is a matrix with all of $\boldsymbol{A}$, but the pth column, and $\boldsymbol{\beta}_{j(p)}$ is the $2^K - 1$ vector that omits $\boldsymbol{\beta}_{jp}$.

As in Balamuta and Culpepper (2021), the entries of the $\boldsymbol{Q}$ matrix are updated with the Metropolis-Hastings (MH) algorithm. While updating an entry, we always propose a change to its current state. Thus, if $q^{(t)}_{jk}$ is the current value of the entry of $\boldsymbol{Q}$, then the proposed candidate value is $q' = 1 - q^{(t)}_{jk}$. Let $T(q', q^{(t)}_{jk})$ be

$$T\left(q', q^{(t)}_{jk}\right) = \left(\frac{p(\boldsymbol{\delta}_j|\boldsymbol{q}_{j(k)}, q', s, g)p(q'|\boldsymbol{Q}_{(j,k)}, \nu)}{p(\boldsymbol{\delta}_j|\boldsymbol{q}_{j(k)}, q^{(t)}_{jk}, s, g)p(q^{(t)}_{jk}|\boldsymbol{Q}_{(j,k)}, \nu)}\right) \tag{3.22}$$

where $\boldsymbol{Q}_{(j,k)}$ indicates all other elements of $\boldsymbol{Q}$. The decision rule for the MH algorithm

is

$$q_{jk}^{(t+1)} = \begin{cases} q' & \min\left(1, T\left(q', q_{jk}^{(t)}\right)\right) > U \\ q_{jk}^{(t)} & \text{otherwise} \end{cases} \tag{3.23}$$

where U is drawn from uniform$(0, 1)$. The full conditional distribution of ν is

$$\nu|\boldsymbol{Q} \sim Beta\left(\sum_{j=1}^{J}\sum_{k=1}^{K} q_{jk} + a_\nu, JK - \sum_{j=1}^{J}\sum_{k=1}^{K} q_{jk} + b_\nu\right). \tag{3.24}$$

Another model parameter that is updated with the MH algorithm is the $\boldsymbol{\delta}_j$ parameters. Similar to the procedure for updating entries of $\boldsymbol{q}_j$, we use a MH sampler that always proposes a change in the value at the current state of δ_{jp}. Let $\delta_{jp}^{(t)}$ be the current value of the $\boldsymbol{\delta}_j$ p'th entry, and δ' be the proposed value such that $\delta' = 1 - \delta_{jp}^{(t)}$. $T(\delta', \delta_{jp}^{(t)})$ is defined as

$$T\left(\delta^\top, \delta_{jp}^{(t)}\right) = \left(\frac{p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta')p(\delta'|\boldsymbol{q}_j, s, g)}{p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp}^{(t)})p(\delta_{jp}^{(t)}|\boldsymbol{q}_j, s, g)}\right). \tag{3.25}$$

The decision rule for the MH algorithm is

$$\delta_{jp}^{(t+1)} = \begin{cases} \delta' & \min\left(1, T\left(\delta', \delta_{jp}^{(t)}\right)\right) > U \\ \delta_{jp}^{(t)} & \text{otherwise} \end{cases}. \tag{3.26}$$

3.4 Monte Carlo Simulation Study

3.4.1 Overview

We conducted several Monte Carlo simulations to examine parameter recovery for our new MCMC procedures for models with item- and attribute-level covariates. Three general data-generating simulation conditions corresponding to the three ways to include covariates are explored: 1) saturated model with item- and attribute-level covariates; 2) $\boldsymbol{\zeta}$ model that only includes attribute-level covariates; and 3) $\boldsymbol{\Gamma}$ model that only includes item-level covariates.

For the simulation studies, we generate data from each of the three models and compare parameter recovery with a base model that excludes the covariates in order to demonstrate the implications of neglecting to incorporate relevant covariates in exploratory RLCMs.

In the simulation study design, the sample size is varied (i.e., $N = 500, 1000, 2000$ and 4000). In addition, to ensure our simulation is representative of real data, we generated covariates effects (i.e., $\mathbf{\Gamma}$ and $\boldsymbol{\zeta}$) and attribute structural parameters (i.e., $\boldsymbol{\psi}$) that are consistent with our application (e.g., see the next section). Specifically, we sampled the elements of $\mathbf{\Gamma}$ and $\boldsymbol{\zeta}$ from uniform$(-1, 1)$, and we set the hyper-parameters of the prior precision to 1 (i,e. $a_{\gamma 0} = b_{\gamma 0} = 1$ for the $\mathbf{\Gamma}$ parameter and $a_{\zeta 0} = b_{\zeta 0} = 1$ for the $\boldsymbol{\zeta}$ parameter). Moreover, the non-zero elements of the attribute structural parameters (i.e., $\boldsymbol{\psi}$) are sampled from uniform$(-0.4, 0.4)$, and the hyper-parameters of the prior precision is set to 1 (i.e. $a_{\psi 0} = b_{\psi 0} = 1$). A single binary covariate for each respondent is sampled from a Bernoulli(0.5) distribution. Moreover, we used the strictly identified $\boldsymbol{Q}$ matrix with $J = 20$ for $K = 3$ and with $J = 30$ for $K = 5$ as presented below:

$$\boldsymbol{Q}_{K=3} = \begin{pmatrix} 0&0&1\\ 0&0&1\\ 0&0&1\\ 0&1&0\\ 0&1&0\\ 0&1&0\\ 1&0&0\\ 1&0&0\\ 1&0&0\\ 0&1&1\\ 0&1&1\\ 1&0&1\\ 1&0&1\\ 1&0&1\\ 1&1&0\\ 1&1&0\\ 1&1&0\\ 1&1&1\\ 1&1&1\\ 1&1&1 \end{pmatrix}. \qquad \boldsymbol{Q}_{K=5} = \begin{pmatrix} 0&0&0&0&1\\ 0&0&0&0&1\\ 0&0&0&1&0\\ 0&0&0&1&0\\ 0&0&1&0&0\\ 0&0&1&0&0\\ 0&1&0&0&0\\ 0&1&0&0&0\\ 1&0&0&0&0\\ 1&0&0&0&0\\ 0&0&0&1&1\\ 0&0&1&0&1\\ 0&0&1&1&0\\ 0&1&0&0&1\\ 0&1&0&1&0\\ 0&1&1&0&0\\ 1&0&0&0&1\\ 1&0&0&1&0\\ 1&0&1&0&0\\ 1&1&0&0&0\\ 0&0&1&1&1\\ 0&1&0&1&1\\ 0&1&1&0&1\\ 0&1&1&1&0\\ 1&0&0&1&1\\ 1&0&1&0&1\\ 1&0&1&1&0\\ 1&1&0&0&1\\ 1&1&0&1&0\\ 1&1&1&0&0 \end{pmatrix}.$$

We generate $\boldsymbol{B}$ using the approach of G. Xu and Shang (2018) where the success probabilities associated with the latent groups $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\alpha} = \mathbf{1}$ are set to 0.2 and 0.8, respectively. The remaining success probabilities are varied between 0.2 and 0.8 with values defined by $0.2+(0.8-0.2)\times K_j'/K_j$ where K_j' denotes the number of the required attributes in

$\boldsymbol{q}_j = (q_{j1}, \ldots q_{jK})$. Moreover, both prior hyper-parameters of ν are set to 1 (i.e. $a_\nu = b_\nu = 1$)

In each simulation, 100 Monte Carlo replications are generated. We varied the number of attributes, $K = 3, 5$, and set the "spike" precision parameter to 100 and "slab" scale parameter to 1. In each chain, we ran $50,000$ iterations, of which we discarded the first $20,000$ as the burn-in period. To assess the recovery rates of the model parameters, we calculated the average mean absolute errors (MAE) of the $\boldsymbol{B}$ parameters using

$$MAE(\boldsymbol{B}) = \frac{1}{RJP} \sum_{r=1}^{R} \sum_{j=1}^{J} \sum_{p=1}^{P} |\hat{\beta}_{jp}^{(r)} - \beta_{jp}| \tag{3.27}$$

where $P = 2^K$ denotes the number of regression coefficients and $\hat{\beta}_{jp}^{(r)}$ is the posterior mean from replication r. Similarly, the MAE for the $\boldsymbol{\Gamma}$ parameters is, $\frac{1}{RJV} \sum_{r=1}^{R} \sum_{j=1}^{J} \sum_{v=1}^{V} |\hat{\gamma}_{jv}^{(r)} - \gamma_{jv}|$ where V is the number of item-related covariate coefficients and $\hat{\gamma}_{jv}^{(r)}$ is the posterior mean from replication r.

For the $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}$ parameters, we reported the average mean absolute errors of the attribute class distribution, $\boldsymbol{\pi}_i = (\pi_{i0}, \ldots, \pi_{i,2^K-1})'$, of each examinee i. In Equation 3, π_{ic} is denoted as the probability of an examinee i belonging to an latent class c. Thus, the average mean absolute value for $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_N$ can be calculated as $\frac{1}{RCN} \sum_{r=1}^{R} \sum_{c=0}^{C} \sum_{i=1}^{N} |\hat{\pi}_{ic}^{(r)} - \pi_{ic}^{(r)}|$ where $\hat{\pi}_{ic}^{(r)}$ and $\pi_{ic}^{(r)}$ are the posterior mean and data generating valuing for replication r. In addition, the average mean absolute errors of the probability of correctly answering an item, which is a function of covariates for the saturated and $\boldsymbol{\Gamma}$ models, are reported as follows. Recall that $\theta_{ij} = P(Y_{ij} = 1 | \boldsymbol{\alpha}_i, \boldsymbol{\gamma}_j, \boldsymbol{\beta}_j)$ is the item j mastery probability of examinee i as in Equation 1. Then, the average mean absolute errors of θ_{ij} can be calculated as $\frac{1}{RJN} \sum_{r=1}^{R} \sum_{j=1}^{J} \sum_{i=1}^{N} |\hat{\theta}_{ij}^{(r)} - \theta_{ij}^{(r)}|$.

We evaluated the model performance on estimating the latent structure by reporting the element-wise agreement rates for $\boldsymbol{Q}$, which is defined as $\frac{1}{R} \sum_{r=1}^{R} \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} \mathbb{1}(\hat{q}_{jk}^{(r)} = q_{jk})$, and matrix-wise agreement rates, $\frac{1}{R} \sum_{r=1}^{R} \mathbb{1}\left(\hat{\boldsymbol{Q}}^{(r)} = \boldsymbol{Q}\right)$. In the rth replication, $\hat{\boldsymbol{Q}}^{(r)}$ is obtained after discretizing the element-wise mean of the sampled binary $\boldsymbol{Q}$ matrices as 1 if the

corresponding entry is greater than 0.5, and 0 otherwise, and $\hat{q}_{jk}^{(r)}$ represents the element in the jth row and kth column of the $\hat{\boldsymbol{Q}}^{(r)}$. In addition, we evaluated the attribute- and pattern-wise agreement rates of the examinee latent classes with $\frac{1}{R}\sum_{r=1}^{R}\frac{1}{NK}\sum_{N=1}^{N}\sum_{k=1}^{K}\mathbb{1}(\hat{\alpha}_{ik}^{(r)}=\alpha_{ik})$, and $\frac{1}{R}\sum_{r=1}^{R}\frac{1}{N}\sum_{N=1}^{N}\mathbb{1}(\hat{\boldsymbol{\alpha}}_{i}^{(r)}=\boldsymbol{\alpha}_{ik})$, respectively.

3.4.2 Results

For the case with a saturated data-generating model, Table 3.1 presents the average mean absolute errors of the parameters $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, θ_{ij}, and $\boldsymbol{\pi}_i$ which denotes the probability of latent class membership for individual i, the matrix- and element-wise recovery rates of the $\boldsymbol{Q}$ matrix and attribute- and pattern-wise recovery rates of examinees' latent classes. Table 3.1 shows that the parameters are recovered with low error rates in the saturated model, and these errors become smaller as the sample size increases. Including the covariates in the attribute level in addition to the item level provides leverage by yielding lower error rates in the $\boldsymbol{\pi}_i$ parameter and yields higher latent class recovery rates than in the base model (i.e., the misspecified model that incorrectly excludes covariates). For example, in recovering the pattern- and attribute-wise attribute classes, the saturated model yields over 75% and 90% accuracy rates, respectively. In addition, the $\boldsymbol{Q}$ matrix is recovered better when we fit the data with the saturated model than the base model. For instance, the saturated model recoveries the $\boldsymbol{Q}$ matrix exactly in over 85% of the replications compared to 18% for the base model when the sample size was $N = 2000$ and $K = 3$. The relative performance of the saturated and base model are consistent in the case with $K = 5$ attributes. However, in both models, as the number of attributes increases, the majority of the parameter recovery rates decrease slightly. In addition, the matrix-wise recovery rates of $\boldsymbol{Q}$ equaled 0 for almost all the conditions when the number of attributes increases, which may be expected given it will be difficult to recover the entire matrix out of the 2^{KJ} possible $\boldsymbol{Q}$ matrices.

Table 3.1: Summary of the Monte Carlo replications for saturated model

Fitted Model	K	Sample size	Average MAE				$\boldsymbol{Q}$ Accuracy		Class Acccuracy	
			$\boldsymbol{B}$	$\boldsymbol{\Gamma}$	θ_{ij}	$\boldsymbol{\pi}_i$	Element	Matrix	Element	Pattern
Saturated Model	3	500	0.1395	0.1143	0.0499	0.0216	0.9590	0.08	0.9066	0.7602
		1000	0.1012	0.0770	0.0362	0.0164	0.9865	0.46	0.9123	0.7752
		2000	0.0743	0.0569	0.0270	0.0125	0.9977	0.86	0.9140	0.7797
		4000	0.0551	0.0386	0.0202	0.0097	1.0000	1.00	0.9154	0.7823
Base Model		500	0.1657	NA	0.0947	0.0826	0.9308	0.01	0.8864	0.7118
		1000	0.1290	NA	0.0885	0.0783	0.9612	0.08	0.8926	0.7273
		2000	0.1039	NA	0.0850	0.0817	0.9738	0.18	0.8935	0.7305
		4000	0.0867	NA	0.0829	0.0816	0.9882	0.49	0.8957	0.7350
Saturated Model	5	500	0.0834	0.1107	0.0862	0.0143	0.8921	0.00	0.8642	0.4824
		1000	0.0708	0.0809	0.0704	0.0128	0.9282	0.00	0.8717	0.5067
		2000	0.0600	0.0561	0.0580	0.0113	0.9638	0.00	0.8791	0.5319
		4000	0.0511	0.0388	0.0493	0.0100	0.9909	0.21	0.8831	0.5471
Base Model		500	0.1095	NA	0.1427	0.0284	0.8027	0	0.7742	0.2990
		1000	0.1077	NA	0.1430	0.0295	0.8136	0	0.7626	0.2752
		2000	0.1070	NA	0.1446	0.0306	0.8295	0	0.7585	0.2639
		4000	0.1101	NA	0.1480	0.0308	0.8216	0	0.7485	0.2491

Average MAE = Average mean absolute errors for parameters $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, mastery probability of an item θ_{ij} and latent class membership probability $\boldsymbol{\pi}_i$; $\boldsymbol{Q}$ Accuracy Matrix = matrix-wise accuracy rate for $\boldsymbol{Q}$; $\boldsymbol{Q}$ Accuracy Element = element-wise accuracy for $\boldsymbol{Q}$; Class Accuracy Element = element-wise accuracy for examinee' s latent class profile; Class Accuracy Pattern = pattern-wise accuracy for examinee's latent class profile; $\boldsymbol{\Gamma}$ parameter is not applicable in base model; Results are based upon 100 replications

Table 3.2 reports the parameter recovery rates when the data-generating model includes the covariates only on the attribute level (i.e., $\boldsymbol{\zeta}$ model). Similar to the saturated model, the model performance becomes better as the sample size increases. The $\boldsymbol{\zeta}$ model performs well on recovering the model and examinee parameters as well as the underlying latent structure. For example, the $\boldsymbol{Q}$ matrix element-wise recovery rate is quite high even with the small sample size (i.e., $N = 500$). And when $N = 2000$, the $\boldsymbol{Q}$ matrix is recovered exactly in 83% of the 100 replications in the $K = 3$. Compared to the base model, which ignores covariates in the setting with a data generating $\boldsymbol{\zeta}$ model, the $\boldsymbol{\zeta}$ model produces much lower mean absolute errors for the estimates of the probability of latent class membership (i.e., $\boldsymbol{\pi}_i$). In

addition, the base model yields almost equivalent pattern- and attribute-wise latent class accuracy rates to the $\boldsymbol{\zeta}$ model. Overall, increasing the number of attributes yields slightly lower accuracy rates in many parameters; however, as in the saturated model case, the matrix-wise $\boldsymbol{Q}$ matrix accuracy rates significantly drop in both the $\boldsymbol{\zeta}$ and base models. Note that the pattern-and element-wise recovery rates of the $\boldsymbol{Q}$ matrix is comparable between the $\boldsymbol{\zeta}$ and base models, which provides evidence that in this simulation setting that excluding covariates when the truth is a $\boldsymbol{\zeta}$ model does not affect the recovery of the latent structure.

Table 3.2: Summary of the Monte Carlo replications for $\boldsymbol{\zeta}$ model

Fitted Model	K	Sample size	Average MAE				$\boldsymbol{Q}$ Accuracy		Class Acccuracy	
			$\boldsymbol{B}$	$\boldsymbol{\Gamma}$	θ_{ij}	$\boldsymbol{\pi}_i$	Element	Matrix	Element	Pattern
Zeta Model	3	500	0.1282	NA	0.0473	0.0207	0.9583	0.10	0.9183	0.7880
		1000	0.0958	NA	0.0344	0.0151	0.9862	0.40	0.9218	0.7968
		2000	0.0705	NA	0.0257	0.0119	0.9968	0.83	0.9238	0.8018
		4000	0.0519	NA	0.0193	0.0088	1.0000	1.00	0.9249	0.8042
Base Model		500	0.1305	NA	0.0481	0.0806	0.9605	0.11	0.9142	0.7793
		1000	0.0973	NA	0.0348	0.0748	0.9855	0.40	0.9185	0.7900
		2000	0.0713	NA	0.0261	0.0801	0.9965	0.83	0.9209	0.7961
		4000	0.0526	NA	0.0196	0.0785	1.0000	1.00	0.9218	0.7979
Zeta Model	5	500	0.0820	NA	0.0885	0.0144	0.8962	0.00	0.8710	0.5015
		1000	0.0681	NA	0.0707	0.0122	0.9322	0.00	0.8822	0.5373
		2000	0.0580	NA	0.0588	0.0110	0.9677	0.01	0.8874	0.5566
		4000	0.0500	NA	0.0508	0.0099	0.9916	0.31	0.8902	0.5673
Base Model		500	0.0788	NA	0.0820	0.0287	0.8905	0.00	0.8685	0.4999
		1000	0.0677	NA	0.0688	0.0294	0.9299	0.00	0.8782	0.5279
		2000	0.0584	NA	0.0587	0.0288	0.9661	0.02	0.8832	0.5449
		4000	0.0518	NA	0.0530	0.0292	0.9887	0.30	0.8837	0.5494

Average MAE = Average mean absolute errors for parameters $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, mastery probability of an item θ_{ij} and latent class membership probability $\boldsymbol{\pi}_i$; $\boldsymbol{Q}$ Accuracy Matrix = matrix-wise accuracy rate for $\boldsymbol{Q}$; $\boldsymbol{Q}$ Accuracy Element = element-wise accuracy for $\boldsymbol{Q}$; Class Accuracy Element = element-wise accuracy for examinee' s latent class profile; Class Accuracy Pattern = pattern-wise accuracy for examinee's latent class profile; $\boldsymbol{\Gamma}$ parameter is not applicable in $\boldsymbol{\zeta}$ and base models; Results are based upon 100 replications

For the $\boldsymbol{\Gamma}$ model, Table 3.3 shows the mean absolute errors and the accuracy rates of

the same parameters we consider in saturated and $\boldsymbol{\zeta}$ models. The results suggest that larger sample sizes improve the accuracy rates for all of the model parameters. The recovery rates of the parameter estimates resulting from the $\boldsymbol{\Gamma}$ model are always better than those in the base model. Especially, the error rates in $\boldsymbol{B}$ and θ_{ij} are significantly lower when we fit the data with the $\boldsymbol{\Gamma}$ model versus the misspecified base model. Moreover, the $\boldsymbol{\Gamma}$ model is superior to the base model on recovering the matrix- and element-wise $\boldsymbol{Q}$ matrix. For example, while the base model recovers the $\boldsymbol{Q}$ matrix exactly in only 28% the replications, the $\boldsymbol{\Gamma}$ model can recover the $\boldsymbol{Q}$ matrix exactly in around 98% times when $N = 2000$. The pattern- and attribute-wise recovery rates of the latent classes are higher in the $\boldsymbol{\Gamma}$ model than in the base model, and the difference becomes more evident when the number of attributes increases. In addition, in majority conditions, increasing the number of attributes yields lower accuracy rates in parameter estimation in both $\boldsymbol{\Gamma}$ and base models. Once again, the matrix-wise $\boldsymbol{Q}$ matrix accuracy rates significantly drop with the higher number of attributes.

Table 3.3: Summary of the Monte Carlo replications for $\boldsymbol{\Gamma}$ model

Fitted Model	K	Sample size	Average MAE				$\boldsymbol{Q}$ Accuracy		Class Acccuracy	
			$\boldsymbol{B}$	$\boldsymbol{\Gamma}$	θ_{ij}	$\boldsymbol{\pi}_i$	Element	Matrix	Element	Pattern
Gamma Model	3	500	0.1315	0.1148	0.0488	0.0197	0.9630	0.12	0.8989	0.7426
		1000	0.0950	0.0785	0.0355	0.0147	0.9887	0.57	0.9041	0.7569
		2000	0.0719	0.0562	0.0265	0.0126	0.9987	0.92	0.9070	0.7631
		4000	0.0540	0.0387	0.0200	0.0101	0.9997	0.98	0.9078	0.7659
Base Model		500	0.1569	NA	0.0929	0.0392	0.9325	0.01	0.8812	0.6997
		1000	0.1227	NA	0.0877	0.0354	0.9612	0.08	0.8866	0.7137
		2000	0.1010	NA	0.0846	0.0330	0.9785	0.28	0.8902	0.7216
		4000	0.0853	NA	0.0827	0.0341	0.9865	0.44	0.8912	0.7242
Gamma Model	5	500	0.0864	0.1112	0.0907	0.0149	0.8970	0.00	0.8507	0.4439
		1000	0.0708	0.0811	0.0721	0.0130	0.9347	0.00	0.8625	0.4804
		2000	0.0589	0.0550	0.0597	0.0117	0.9750	0.04	0.8691	0.4999
		4000	0.0499	0.0396	0.0496	0.0103	0.9938	0.45	0.8739	0.5197
Base Model		500	0.1066	NA	0.1381	0.0224	0.8204	0	0.7850	0.3110
		1000	0.1046	NA	0.1374	0.0234	0.8379	0	0.7831	0.3059
		2000	0.1018	NA	0.1397	0.0239	0.8461	0	0.7692	0.2813
		4000	0.1071	NA	0.1456	0.0245	0.8262	0	0.7586	0.2587

Average MAE = Average mean absolute errors for parameters $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, mastery probability of an item θ_{ij} and latent class membership probability $\boldsymbol{\pi}_i$; $\boldsymbol{Q}$ Accuracy Matrix = matrix-wise accuracy rate for $\boldsymbol{Q}$; $\boldsymbol{Q}$ Accuracy Element = element-wise accuracy for $\boldsymbol{Q}$; Class Accuracy Element = element-wise accuracy for examinee' s latent class profile; Class Accuracy Pattern = pattern-wise accuracy for examinee's latent class profile; $\boldsymbol{\Gamma}$ parameter is not applicable in base model; Results are based upon 100 replications.

3.5 Application to Spatial Rotation Dataset

This section reports results from an application of our developed methods to a real data set collected by Culpepper and Balamuta (2017). In this data set, 516 student-participants were administered 30 spatial rotation items (i.e., the Purdue Spatial Visualization Test: Rotations (PSVT-R; Yoon, 2011)) and a survey to collect the participants' covariates such as demographic information and background information that may be associated with spatial reasoning skills. Two categorical items from the Survey of Spatial Representation and Activities (Terlecki & Newcombe, 2005) were collected and included as covariates of the spa-

tial rotation latent structure: 1) "How proficient or skilled do you believe you are at using maps?" (i.e., "not skilled", "not very skilled", "moderately skilled", "very skilled"); and 2) "How often do you use maps?" (i.e., "once every few years to not much at all", "one to two times a year", "one to two times in six months", "one to two times a month", "weekly", "daily"). Table 3.4 reports the frequency distribution for the covariates. Note for the map skill variable that we combined the "not skilled" and "not very skilled" levels into a "not skilled/not very skilled" category for our analysis as only twelve respondents reported "not skilled". The covariates were collected as categorical variables and yield seven distinct main-effect variables in a dummy-coded design matrix (note the references levels are "not skill/not very skilled" and "once every few years to not much at all"). In order to demonstrate our models we set the number of attributes to $K = 3$. Similar to simulation study design, we set the hyper-parameters of the prior precision to 1. (i,e. $a_{\zeta_0} = b_{\zeta 0} = 1$ for the $\boldsymbol{\zeta}$ parameters and $a_{\psi_0} = b_{\psi 0} = 1$ for the $\boldsymbol{\psi}$ parameters). Moreover, we set the prior hyper-parameters of ν as 1 (i.e. $a_\nu = b_\nu = 1$).

Table 3.4: Descriptive statistics of covariates for spatial rotation dataset.

Covariate	Category	Frequency
Map skill	not skilled	13
	not very skilled	98
	moderately skilled	297
	very skilled	108
Map usage	Once every few years to not much at all	66
	1 to 2 times a year	49
	1 to 2 times in six months	62
	1 to 2 times in a month	152
	weekly	159
	daily	28

Note we consider the same set of covariates at both the item level and attribute level. We compared the four models (i.e., the saturated, $\boldsymbol{\Gamma}$, $\boldsymbol{\zeta}$ and base models) using the relative model fit index by calculating leave-one-out cross-validation (LOO; Geisser & Eddy, 1979;

Vehtari, Gelman, & Gabry, 2017). The estimated LOO values were 16776.56, 16763.78, 16810.5, and 16815.69 for the saturated, $\boldsymbol{\zeta}$, $\boldsymbol{\Gamma}$ and base models, respectively. The LOO index favors the $\boldsymbol{\zeta}$ model (i.e., the model embeds the covariates only at the attribute level).

In Table 3.5, we present element-wise means for the inferred $\boldsymbol{Q}$ matrix as well as latent class response probabilities. Table 3.5 provides evidence that the $\boldsymbol{\zeta}$ model yields a dense $\boldsymbol{Q}$ matrix estimate as most of the element-wise probabilities exceed 0.50. Chen et al. (2020) noted that dense $\boldsymbol{Q}$ matrices, which are generically identified, are indicative of a general unstructured mixture model. Therefore, the estimated $\boldsymbol{Q}$ matrix suggests that the data support an unstructured mixture where most items require all eight classes to describe response patterns. In addition, Table 3.5 shows that the item response probabilities associated with the latent groups $\boldsymbol{\alpha} = \mathbf{0}$ equal values between 0.076 and 0.687 with most values falling below 0.4. The majority of the $\theta_{j,111}$ parameters associated with the latent groups $\boldsymbol{\alpha} = \mathbf{1}$ are around 0.95.

Table 3.5: Estimated posterior means of the $\boldsymbol{Q}$ matrix and item parameters for spatial rotation data.

Items	$\overline{\boldsymbol{Q}}$			$\bar{\theta}_j$							
	Q1	Q2	Q3	"000"	"001"	"010"	"011"	"100"	"101"	"110"	"111"
1	0.687	0.651	0.612	0.641	0.746	0.805	0.935	0.846	0.949	0.959	0.997
2	0.680	0.624	0.470	0.687	0.734	0.833	0.882	0.892	0.932	0.987	0.997
3	0.687	0.694	0.637	0.613	0.802	0.886	0.958	0.902	0.967	0.989	0.999
4	0.574	0.550	0.465	0.646	0.733	0.789	0.860	0.819	0.879	0.917	0.953
5	0.695	0.698	0.700	0.443	0.801	0.766	0.950	0.762	0.957	0.949	0.998
6	0.682	0.644	0.696	0.455	0.798	0.683	0.934	0.739	0.950	0.898	0.997
7	0.667	0.663	0.667	0.374	0.706	0.707	0.903	0.738	0.920	0.940	0.991
8	0.526	0.543	0.614	0.540	0.640	0.621	0.809	0.601	0.834	0.742	0.981
9	0.672	0.620	0.681	0.252	0.556	0.525	0.798	0.673	0.879	0.900	0.995
10	0.687	0.648	0.706	0.454	0.706	0.630	0.892	0.662	0.918	0.859	0.995
11	0.656	0.582	0.665	0.414	0.695	0.607	0.860	0.705	0.904	0.888	0.990
12	0.717	0.712	0.728	0.233	0.770	0.502	0.958	0.498	0.960	0.866	0.999
13	0.686	0.686	0.692	0.374	0.591	0.635	0.903	0.650	0.893	0.853	0.992
14	0.707	0.692	0.673	0.215	0.391	0.555	0.835	0.663	0.892	0.920	0.998
15	0.563	0.603	0.687	0.290	0.630	0.443	0.792	0.401	0.787	0.694	0.979
16	0.704	0.697	0.705	0.249	0.654	0.522	0.912	0.537	0.913	0.865	0.996
17	0.670	0.670	0.661	0.335	0.527	0.566	0.861	0.616	0.823	0.846	0.980
18	0.598	0.527	0.530	0.316	0.438	0.481	0.637	0.585	0.721	0.773	0.896
19	0.714	0.710	0.739	0.255	0.724	0.555	0.943	0.566	0.936	0.801	0.997
20	0.657	0.672	0.719	0.343	0.693	0.533	0.914	0.532	0.883	0.795	0.995
21	0.695	0.700	0.724	0.215	0.581	0.496	0.892	0.473	0.872	0.756	0.993
22	0.685	0.650	0.691	0.311	0.509	0.446	0.778	0.480	0.791	0.753	0.980
23	0.568	0.633	0.713	0.361	0.611	0.532	0.871	0.482	0.804	0.763	0.977
24	0.556	0.580	0.625	0.358	0.604	0.536	0.781	0.515	0.758	0.694	0.879
25	0.607	0.627	0.710	0.267	0.584	0.420	0.805	0.399	0.787	0.602	0.979
26	0.561	0.633	0.728	0.262	0.463	0.394	0.814	0.355	0.717	0.566	0.984
27	0.598	0.645	0.669	0.415	0.530	0.526	0.846	0.501	0.784	0.756	0.992
28	0.671	0.670	0.662	0.204	0.319	0.360	0.701	0.334	0.745	0.573	0.988
29	0.577	0.574	0.633	0.190	0.301	0.277	0.556	0.278	0.530	0.417	0.787
30	0.581	0.629	0.645	0.076	0.165	0.175	0.421	0.140	0.381	0.332	0.691

Table 3.6 includes attribute-level covariate coefficient estimates associated with the 90% credible intervals. The credible intervals of the attribute-level coefficients suggest that there is a high probability that two of the "map skill" coefficients are positively associated with underlying latent class membership. More specifically, students that report themselves as "very skilled" in using the maps are more likely to master the second and third attributes than the students who reports themselves in the "not skilled/not very skilled" category.

Table 3.6: $\boldsymbol{\zeta}$ coefficient estimates and credible intervals for spatial rotation data application for the $\boldsymbol{\zeta}$ model.

Coefficients	Estimates			Credible Interval			
	α_1	α_2	α_3	α_1	α_2	α_3	
Moderately skilled	0.248	0.187	0.358	-0.073	-0.170	-0.019	5%
				0.560	0.541	0.729	95%
Very skilled	0.425	0.572	0.721	-0.088	0.008	0.229	5%
				1.044	1.148	1.218	95%
1 to 2 times a year map usage	-0.024	0.054	-0.089	-0.463	-0.427	-0.570	5%
				0.447	0.585	0.412	95%
1 to 2 times in six months map usages	-0.173	-0.200	-0.426	-0.628	-0.675	-0.913	5%
				0.257	0.259	0.073	95%
1 to 2 times in a month map usage	-0.221	-0.212	-0.193	-0.554	-0.561	-0.534	5%
				0.109	0.141	0.151	95%
Weekly map usage	-0.096	-0.097	-0.329	-0.454	-0.476	-0.721	5%
				0.259	0.290	0.073	95%
Daily map usage	-0.045	0.109	-0.057	-0.655	-0.571	-0.680	5%
				0.635	0.841	0.593	95%

Table 3.7 reports the averaged probability of belonging to each latent class $\boldsymbol{\alpha}$ for all students, in addition to disaggregated averages for students in the "very skilled" and "not skilled/not very skilled" levels of the map skill covariate. In doing so, we first calculated the probability of belonging to the latent class c for each examinee i as in Equation 3, and then we took the average over the students in the corresponding group. The results indicate that the students in the reference group "not skilled/not very skilled" are more likely to belong to the class $\boldsymbol{\alpha} = \mathbf{0}$. Meanwhile, the students who report themselves as "very skilled" are more likely to be classified into a skill profile with two or more attributes.

Table 3.7: Estimated latent class membership distribution for spatial rotation data application for the $\boldsymbol{\zeta}$ model.

	$\boldsymbol{\pi}_c$							
	"000"	"001"	"010"	"011"	"100"	"101"	"110"	"111"
Total sample	0.170	0.105	0.114	0.102	0.174	0.099	0.154	0.081
"Very skilled" group	0.080	0.092	0.096	0.147	0.119	0.117	0.191	0.158
Reference group	0.256	0.105	0.135	0.075	0.196	0.070	0.125	0.037

$\boldsymbol{\pi}_c$ denotes the average latent class membership probability of students in the corresponding groups; "Reference group" is formed by students who report themselves as "not skilled/not very skilled".

3.6 Discussion

Formative assessments are important for providing fine-grained information for students' learning and progress. Exploratory RLCMs remain important tools for broadening the applicability of diagnostic models for developing formative assessments. We considered new methods for including covariates into exploratory RLCMs. Collateral information—here, covariates—about students might benefit educators in several ways. First, covariates may be a source of undesirable interactions between students and items such as DIF; thus, investigating the covariates can uncover and help prevent these undesirable interactions and construct irrelevant variance. Second, covariates might affect the difficulty level of accruing attributes for students. Knowing these effects can help educators to identify and support at-risk students. Covariates can also be used to evaluate educational intervention effects. Third, significant covariates can reduce the uncertainty in the model parameters by providing additional information about students.

Fine-grained information about a student's knowledge profile is another important component in educational environments. Recently, a significant amount of research has been conducted on exploratory RLCMs that can infer the underlying structure between attributes and a binary domain-specific item-skill map. Despite the common interest in exploratory RLCMs and covariates, no study has been conducted on incorporating the covariates in RLCMs. Thus, we proposed three models to include covariates in the exploratory RLCMs—a

saturated model and two of its constrained versions. In the saturated model, we explore the covariates' effects on the probability of correctly answering an item and the probability of mastering an attribute. The $\boldsymbol{\zeta}$ model, one of the constrained models, links covariates with the probability of mastering an attribute, while the $\boldsymbol{\Gamma}$ model links the covariates with the probability of correctly answering an item.

The findings suggest that all three models were able to recover the model parameters well. In addition, the latent structure recovery rates in all three models were quite promising—all proposed models recover the entire $\boldsymbol{Q}$ matrix accurately when $N = 4000$. We also investigate the condition when covariates are ignored in the model, the base model, even though the true data generating model included a covariate. Ignoring the covariates increases the mean absolute errors in almost all conditions. Moreover, it yields significantly lower latent structure recovery rates in the base model than in the saturated and gamma models. In the $\boldsymbol{\zeta}$ model, the findings suggest that including the attribute-level covariates in the model improves the recovery of $\boldsymbol{\pi}$, which may be used in practice to classify students. However, we did not find evidence in the current simulation design that omitting attribute-level covariates impacts inference about the latent structure (e.g., $\boldsymbol{Q}$ and $\boldsymbol{B}$). In general, including covariates benefits us not only by providing insights about the relationships between the covariates and the item success and attribute mastery probabilities, but also significantly increases the parameter recovery rates. In the real data application, the $\boldsymbol{\zeta}$ model yields the best model fit on the spatial rotation data. The findings suggest that two of the attribute-level covariates are significant.

In addition to incorporating covariates into RLCMs, we also introduce a novel way to describe the structure of the relationship among the attributes (see subsection "Model for Attribute-Related covariates"). This approach provides great flexibility in structuring the relationships among attributes. In the present study, we assume an unstructured relationship among attributes and estimate the intercept, main-effects, and interaction terms. In the future, researchers may consider employing a regularization technique to infer structure re-

lated to relationships among attributes to possibly reveal underlying higher-order structures. Moreover, with the innovation in learning systems, an increasing number of covariates, such as process information from log files or response times, can be recorded during assessments. However, without a regularization technique, deciding which covariates will reveal significant information about students can present challenges. For example, Iaconangelo (2017) investigated the relationship between covariates and attribute mastery under the confirmatory RLCM setting, and applied the $L1$ penalty to the attribute-level regression log-likelihood to shrink the non-informative covariate coefficients to zero. Thus, another future direction might include a variable selection algorithm in the proposed models under the exploratory RLCM setting.

In conclusion, exploratory RLCMs are important tools for developing and refining formative assessments to accelerate student learning. We addressed the problem of incorporating information about students' context in the form of covariates to strengthen the applicability of these methods for applied research.

Chapter 4

Variable Selection for Exploratory Restricted Latent Class Models with Covariates

4.1 Introduction

Formative assessments have been considered as a foremost component of the traditional classroom setting. Diagnostic models provide a statistical framework for designing formative assessments by classifying student knowledge profiles according to a collection of fine-grained attributes (de la Torre & Douglas, 2004; Rupp et al., 2010; von Davier, 2008). In fact, learning technology algorithms leverage the diagnostic modeling framework by tailoring learning interventions to adapt to individual students' capabilities and needs (e.g. Chen, Li, et al., 2018; Han et al., 2020; Huang et al., 2019; X. Li et al., 2021; Tan et al., 2020; S. Zhang & Chang, 2016) and to track skill development (e.g. F. Li et al., 2015; Madison & Bradshaw, 2018; Studer, 2012; S. Wang, Yang, et al., 2018; S. Wang, Zhang, et al., 2018; Ye et al., 2016a; Yigit & Douglas, 2021; S. Zhang & Chang, 2020).

The restricted latent class models can be formulated into two main frameworks— confirmatory and exploratory RLCMs. The application of confirmatory methods is appropriate in cases where the underlying structure in terms of how attributes relate to observed performance is precisely known (i.e., the $\boldsymbol{Q}$ matrix). However, using an incorrect latent structure can adversely impact classification decisions (Henson & Templin, 2007), so exploratory methods that infer the latent structure from student responses may be preferred. Recently, a significant amount of research has also developed exploratory RLCMs which infer the underlying structure of the $\boldsymbol{Q}$ matrix (e.g. Chen, Culpepper, Chen, & Douglas, 2018; Chen et al., 2020, 2015b; Culpepper, 2019b; Culpepper & Chen, 2019; G. Xu & Shang, 2018).

The context by which students learn may be important to understand when making diagnostic decisions. For instance, students may receive distinct interventions that target specific learning objections or students may be multiskilled with prior knowledge and experiences that inform mastery on the target attributes. It is therefore important to understand how a broader collection of student characteristics, which we refer to as covariates, shape performance and attribute mastery. The existing research demonstrates the value of using covariates to shape formative assessments in education. The advantages of including the students' covariates into diagnostic models can be clustered around three primary purposes. First, covariates can detect the possibility of differential item functioning (DIF) to provide fair test designs across diverse populations (e.g. F. Li, 2008; X. Li & Wang, 2015; Z. Wang et al., 2014; W. Zhang, 2006). Second, covariates are also used to explain individual differences or serve as indicators to evaluate educational intervention effects (e.g. Ayers et al., 2013; Iaconangelo, 2017; Minchen et al., 2017; S. Wang, Yang, et al., 2018; Zhan et al., 2018). Thus, exploring covariates might help to identify at-risk students and benefit educators to design student-tailored interventions that accelerate skill development. Third, using covariates, like response times (RTs) or other student experiences associated with the target attributes, might improve the accuracy of the model parameter estimates and attribute classification rates (e.g. S. Wang, Zhang, et al., 2018; S. Wang et al., 2020). When including covariates into a model, one challenge is to distinguish active covariates, which relate to the outcome variable, from inactive ones. To address this issue, one can apply a regularization technique to the loss function or incorporate a variable selection mechanism into the priors. In this regard, Iaconangelo (2017) used a regularization technique (i.e., the $\mathcal{L}1$ penalty) to select the active covariates in the structural part of the three-step diagnostic regression approach.

Despite the common interest in exploratory RLCMs and covariates, no study has been conducted on incorporating covariates and a variable selection mechanism in RLCMs, which may limit the types of inferences and insights available to applied researchers and educators. Thus, in this paper, we offer new methods for including covariates in exploratory RLCMs.

Specifically, three general strategies are developed for including covariates—at the item level, attribute level, or both item and attribute levels. In the rest of the paper, these models are referred to as $\mathbf{\Gamma}$, $\boldsymbol{\zeta}$, and saturated models, respectively. For the $\mathbf{\Gamma}$ model, we investigate the effect of the covariates on the probability of correctly answering an item. Similarly, the $\boldsymbol{\zeta}$ model considers the relationship between the covariates and the probability that a student possesses a particular attribute. In order to approximate the model parameter posterior distribution, we present a Markov chain Monte Carlo (MCMC) algorithm using a Metropolis-within-Gibbs algorithm. Moreover, we apply a variable selection mechanism on both item- and attribute-level covariates to infer the active covariates in the model.

Note that including covariates at the item level explicitly assumes that both attributes and covariates may impact observed responses. One consequence of active item-level covariates is that the the assumption of conditional independence of responses given attributes no longer holds. There is a long-standing tradition and interest in incorporating covariates into psychometric measurement models (Meredith & Millsap, 1992). In fact, including covariates in the measurement models is closely connected with evaluating differential item functioning (DIF), which is important for ensuring the validity of test scores and subsequent diagnostic classifications. The Equal Educational Opportunities Act of 1974 requires equal access to education regardless of students' race, gender, ethnicity, national origin, or other protected identity. Moreover, AERA/APA/NCME Standard 3.2 (AERA, APA, & NCME, 2014) emphasizes that "Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics." Thus, eliminating construct-irrelevant variance that contributes to DIF is an essential task to guarantee a bias-free assessment. Whereas covariates may be unnecessary in some applications, there are diagnostic testing circumstances where including item-level covariates is advantageous. For instance, in an artificial intelligent based online learning system, unfamiliarity with the technological interfaces can create differential performance

that could be interpreted as target content skill deficits rather than differences in construct irrelevant variables. Thus, considering covariates in the measurement model can be a solution for disentangling the effects of construct-irrelevant covariates from the true underlying latent structure.

The layout of the rest of the paper is as follows. First, we introduce our novel framework for incorporating covariates into exploratory RLCMs. Second, a Bayesian model specification is presented along with details about posterior inference. Third, we report results from two Monte Carlo simulation studies concerning the accuracy of model parameter recovery. The first simulation examines recovery in the case where there is a single active covariate. One concern is that researchers may have many irrelevant covariates, and it would be necessary to identify predictor variables that actively relate to both item responses and attributes. In order to address this issue, we extend our framework and incorporate variable selection procedures for covariates at the item and attribute levels and assess recovery in a second simulation study. Fourth, we apply our method to a spatial rotation test and demonstrate the types of inferences that are available with our modeling framework. Finally, we provide a discussion section with the final remarks and future research directions.

4.2 Model Specification

We denote an individual with i; $i = 1, \ldots, N$, an item with j; $j = 1, \ldots, J$, and an attribute with k; $k = 1, \ldots, K$. Let $\boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)^\top$ denote the $N \times J$ response matrix of random binary responses with the ith row defined as $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \ldots, Y_{iJ})^\top$ and let the observed value be $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)^\top$ where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{iJ})$ and $y_{ij} \in \{0, 1\}$. We consider the case where observed responses are described by an underlying set of attributes. That is, each individual respondent has a latent binary attribute pattern $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{iK})^\top$ in which each entry will be 1 if the individual has mastered that attribute and 0 otherwise.

The purpose of this section is to introduce our general modeling framework for in-

corporating covariates into the exploratory RLCM framework. We consider two types of covariates, which are covariates that relate to: 1) item responses; and 2) attributes. Moreover, $\boldsymbol{X}$ is an item-related matrix of covariates with the dimension $N \times V$ where V is the number of covariates, and the ith row for an individual i is $\boldsymbol{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iV})^\top$. Similarly, $\boldsymbol{Z}$ is an attribute-related covariate matrix with the dimension $N \times L$ where L is the number of covariates included in $\boldsymbol{Z}$, and the ith row is denoted by $\boldsymbol{z}_i = (z_{i1}, z_{i2}, \ldots, z_{iL})^\top$. It is important to note that $\boldsymbol{x}_i$ and $\boldsymbol{z}_i$ are free to include the same covariates.

4.2.1 Model for Item-Related Covariates

To simplify the formulation, we introduce a design matrix such that $\boldsymbol{A} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N)^\top$ is a matrix with the dimension $N \times 2^K$, where $\boldsymbol{a}_i$ is a design vector for individual i based on the $\boldsymbol{\alpha}_i$. As an example, if $K = 3$, the attribute pattern for an individual i will be in the format $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \alpha_{i3})^\top$ and the design vector will be $\boldsymbol{a}_i = (1, \alpha_{i1}, \alpha_{i2}, \alpha_{i3}, \alpha_{i1}\alpha_{i2}, \alpha_{i1}\alpha_{i3}, \alpha_{i2}\alpha_{i3}, \alpha_{i1}\alpha_{i2}\alpha_{i3})^\top$ where the first element (i.e., "1") stands for the intercept and the remaining elements correspond with main-effects and interaction terms. We consider the following item response for the probability that individual i correctly respond to item j,

$$\theta_{ij} = P(Y_{ij} = 1 | \boldsymbol{\alpha}_i, \boldsymbol{\gamma}_j, \boldsymbol{\beta}_j) = \Phi\big(\boldsymbol{a}_i^\top \boldsymbol{\beta}_j + \boldsymbol{x}_i^\top \boldsymbol{\gamma}_j\big) \tag{4.1}$$

where Φ is the cumulative distribution function of the standard normal distribution. In Equation 4.1, the relationship between attributes and item responses is denoted by the vector of regression coefficients, $\boldsymbol{\beta}_j$. $\boldsymbol{\beta}_j$ is the 2^K vector of regression coefficients indexed by $\{\beta_{jp}\}_{p=0}^{2^K-1}$. In addition, the relationship between the item-level covariates for individual i, $\boldsymbol{x}_i$, and response probabilities is $\boldsymbol{\gamma}_j$, which is a V-dimensional vector of coefficients for item j.

4.2.2 Model for Attribute-Related Covariates

Previous research modeled dependence among the attributes using a variety of strategies. The most common approach is to use an unstructured $\boldsymbol{\pi}$ vector, and other options include using a more parsimonious structure (e.g. Chen & Culpepper, 2020; de la Torre & Douglas, 2004; Henson et al., 2009; Templin et al., 2008). In the present study, we do not assume the attributes are independent when constructing the latent class probabilities. Thus, the joint distribution for the attribute profile for individual i can be factored as

$$p(\boldsymbol{\alpha}_i) = p(\alpha_{i1})p(\alpha_{i2}|\alpha_{i1})\dots p(\alpha_{iK}|\alpha_{i1},\dots,\alpha_{i,K-1}). \tag{4.2}$$

We use this factored representation of the joint distribution for attributes to model the effect student covariates have on the probability of latent class membership. A parameter $\boldsymbol{\psi} = (\boldsymbol{\psi}_1,\dots,\boldsymbol{\psi}_K)^\top$ as a $K \times 2^{K-1}$ matrix is defined to describe the structure of the relationship among the attributes. In addition, the covariates are incorporated at the attribute level. Thus, the probability examinee i belongs to an attribute class $\boldsymbol{\alpha}_i$ is influenced by both examinee's covariates as well as the relations among attributes. Specifically, we use the following probit link to model the joint distribution for the probability of class membership for individual i,

$$p(\boldsymbol{\alpha}_i|\boldsymbol{\psi},\boldsymbol{\zeta}) = \prod_{k=1}^{K} \Phi\left(\boldsymbol{a}_{i,K-1}^\top\boldsymbol{\psi}_k + \boldsymbol{z}_i^\top\boldsymbol{\zeta}_k\right)^{\alpha_{ik}} \left(1 - \Phi\left(\boldsymbol{a}_{i,K-1}^\top\boldsymbol{\psi}_k + \boldsymbol{z}_i^\top\boldsymbol{\zeta}_k\right)\right)^{1-\alpha_{ik}} \tag{4.3}$$

where $\boldsymbol{a}_{i,K-1}$ is a design vector defined by a subset of $\boldsymbol{\alpha}_i$ pattern (i.e., $(\alpha_{i1},\dots,\alpha_{i,K-1})$). For example, for $K = 3$, the design vector will be $\boldsymbol{a}_{i,K-1} = (1, \alpha_{i1}, \alpha_{i2}, \alpha_{i1}\alpha_{i2})^\top$ where the first element (i.e., 1) stands for the intercept. The rows of $\boldsymbol{\psi}$ include a fixed pattern of zeros based upon the order the attribute distribution is factored (e.g., see Equation 4.2). Thus, the number of non-zero elements in $\boldsymbol{\psi}$ equals to $2^K - 1$. For example, for $K = 3$, $\boldsymbol{\psi}_1 = (\psi_{11}, 0, 0, 0)^\top$, $\boldsymbol{\psi}_2 = (\psi_{21}, \psi_{22}, 0, 0)^\top$, and $\boldsymbol{\psi}_3 = (\psi_{31}, \psi_{32}, \psi_{33}, \psi_{34})^\top$. The relationship

between the covariates and attributes are denoted by the $K \times L$ matrix of coefficients $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_K)^\top$ where for individual i, $\boldsymbol{\zeta}_k$ is the relationship between $\boldsymbol{z}_i$ and the conditional attribute mastery probabilities.

4.2.3 Likelihood for the Saturated Model

We next describe the likelihood for the saturated model and discuss several important special cases of our model in the next subsection. The conditional likelihood of observing $\boldsymbol{y}_i$ given the attributes and item-level covariates is

$$p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i, \boldsymbol{\Gamma}, \boldsymbol{B}) = \prod_{j=1}^{J} \theta_{ij}^{y_{ij}} (1 - \theta_{ij})^{1-y_{ij}} \tag{4.4}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_J)^\top$ is the $J \times V$ item-level regression coefficients matrix, and $\boldsymbol{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J)^\top$ denotes the $J \times 2^K$ matrix of the regression coefficients.

Thus, the likelihood for individual i is

$$p(\boldsymbol{y}_i|\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\psi}, \boldsymbol{\zeta}) = \sum_{c=0}^{2^K-1} \pi_{ic} p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{B}, \boldsymbol{\Gamma}) \tag{4.5}$$

where $\pi_{ic} = P(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c|\boldsymbol{\psi}, \boldsymbol{\zeta})$ is defined to be the conditional probability that individual i belongs to class c. We note the use of the vector $\boldsymbol{v} = (2^{\{K-1\}}, 2^{\{K-2\}}, \ldots, 1)^\top$ to create a bijection between the binary attribute pattern and a integer c, such that $c = 0, \ldots, 2^K - 1$. The likelihood function for a sample of N independent observations can be formulated as the product of N respondents' likelihoods,

$$p(\boldsymbol{y}|\boldsymbol{B}, \boldsymbol{\Gamma}, \boldsymbol{\psi}, \boldsymbol{\zeta}) = \prod_{i=1}^{N} \sum_{c=0}^{2^K-1} \pi_{ic} p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{B}, \boldsymbol{\Gamma}) \tag{4.6}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N)^\top$ is a $N \times J$ matrix of responses.

4.2.4 Special Cases

From the saturated model, we derived two different parsimonious models. One model (i.e., the $\boldsymbol{\Gamma}$ model) associates covariates with the probability of correctly answering an item, but not the attributes, and the other model (i.e., the $\boldsymbol{\zeta}$ model) associates covariates with the probability of mastering an attribute, but not the item responses. Moreover, in order to investigate the benefit of the covariates on uncovering the latent structure, we also consider the base model, which excludes attributes. The connection between the saturated model and the special cases are presented below:

- Saturated model where both $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ are estimated.
- Base model where $\boldsymbol{\Gamma} = \mathbf{0}$ and $\boldsymbol{\zeta} = \mathbf{0}$.
- $\boldsymbol{\Gamma}$ model where $\boldsymbol{\Gamma}$ is estimated and $\boldsymbol{\zeta} = \mathbf{0}$.
- $\boldsymbol{\zeta}$ model where $\boldsymbol{\Gamma} = \mathbf{0}$ and $\boldsymbol{\zeta}$ is estimated.

The various special cases may be relevant for different researchers. For instance, researchers interested assessing DIF may be more interested in the $\boldsymbol{\Gamma}$ model to assess whether covariates relate to item responses. In contrast, intervention studies may focus on the $\boldsymbol{\zeta}$ model given the goal is to test hypotheses about how experimental conditions or student characteristics relate to attribute mastery. Additionally, researchers would deploy the base model in the absence of covariates.

4.3 Bayesian Inference of the Saturated Model

This section discusses A Bayesian model for inferring the model parameters. First, we discuss the prior specifications, and then introduce the full conditional distributions. Later we describe a Metropolis-within-Gibbs sampling algorithm for approximating the posterior distribution.

4.3.1 Bayesian Formulation

The following subsection introduces our novel Bayesian formulation. The following five subsections discuss: 1) data augmentation for item responses; 2) priors for the effects of covariates for item responses; 3) data augmentation for attributes; 4) priors for the effects of covariates for attribute mastery probabilities; and 5) priors for inferring the latent structure related parameters.

Item Augmented Data

We use the classic probit data augmentation approach (Albert & Chib, 1993). That is, we define a deterministic relationship between the observed binary responses and a continuous augmented random variable as $Y_{ij} = \mathbb{1}(Y_{ij}^* > 0)$. Next, in order to augment our model we specify the following normal distribution for the augmented data,

$$Y_{ij}^* | \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j, \boldsymbol{\gamma}_j \sim \mathcal{N}(\boldsymbol{a}_i^\top \boldsymbol{\beta}_j + \boldsymbol{x}_i^\top \boldsymbol{\gamma}_j, 1). \tag{4.7}$$

Item-Level Covariate Coefficients

To apply a variable selection algorithm on item-level covariates, we use the "spike-slab" hierarchical formulation (George & McCulloch, 1993) on the prior of the $\boldsymbol{\gamma}_j$ parameters. George and McCulloch (1993) noted that under the hierarchical modeling framework, the prior distribution becomes equivalent to a generalized version of a Zellner (1986) g-prior. In detail, each parameter in $\boldsymbol{\gamma}_j$ is assumed to follow a "spike-slab" mixture of two normal distributions, which $\gamma_{jv} \sim (1 - \xi_{\gamma jv})\ \mathrm{N}\left(0, s_{\gamma jv}^{-2} \tau_{\gamma_j}^{-1} (\boldsymbol{x_v}^\top \boldsymbol{x_v})^{-1}\right) + \xi_{\gamma jv}\ \mathrm{N}\left(0, \tau_{\gamma_j}^{-1} (\boldsymbol{x_v}^\top \boldsymbol{x_v})^{-1}\right)$, and $P(\xi_{\gamma jv} = 1) = 1 - P(\xi_{\gamma jv} = 0) = p_{\gamma j}$. Here, $\xi_{\gamma jv}$ is a binary latent variable defining the activeness of the parameter. That is, the prior for γ_{jv} for $\xi_{\gamma jv} = 0$ is a normal distribution concentrated near zero (i.e., X_{iv} is unrelated to Y_{ij}) and the prior for $\xi_{\gamma jv} = 1$ corresponds with a normal with a larger variance (i.e., X_{iv} relates to Y_{ij}). Thus, the conditional prior for

$\boldsymbol{\gamma}_j$ is the following multivariate normal distribution,

$$\boldsymbol{\gamma}_j|\boldsymbol{\xi}_{\gamma j} \sim \mathcal{N}_V(\mathbf{0}, \tau_{\gamma j}^{-1}(\boldsymbol{K}_{\boldsymbol{\gamma j}}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{K}_{\boldsymbol{\gamma j}})^{-1}) \tag{4.8}$$

where we set $\boldsymbol{\xi}_{\gamma j} = (\xi_{\gamma j1} \dots \xi_{\gamma jV})^\top$ as the binary latent vector, and $\boldsymbol{K}_{\gamma j} = \text{diag}[s_1, \dots, s_V]$ where $s_v = m$,$m > 1$ if $\xi_{\gamma_{jv}} = 0$, and $s_v = 1$ if $\xi_{\gamma_{jv}} = 1$. Note that m is chosen as a large constant that shrinks the variance for inactive predictors [1] The mean is a vector of zeros and the prior variance-covariance matrix is a function of the predictor cross-products (i.e., $\boldsymbol{X}^\top\boldsymbol{X}$) and a $\boldsymbol{K}_{\gamma j}$. Moreover, $\xi_{\gamma jv}$ follows a Bernoulli distribution with a probability $p_{\gamma j}$ for item j, $\xi_{\gamma jv}|p_{\gamma j} \sim \text{Bernoulli}(p_{\gamma j})$, and $p_{\gamma j}$ follows a $\text{Beta}(a_\gamma, b_\gamma)$ distribution. We choose a gamma prior for the precision hyper-parameter $\tau_{\gamma j}$, $\tau_{\gamma j} \sim \text{gamma}(a_{\gamma 0}, b_{\gamma 0})$. Further note setting $\boldsymbol{K}_{\gamma j}$ to an identity matrix (i.e., setting all the latent binary variables to one), one can make the prior distribution to be equivalent to a Zellner (1986) g-prior without a variable selection mechanism.

Attributes and Augmented Data

We use the probit data augmentation for attributes, as well. Specifically, we use the following formulation for α_{ik},

$$\begin{aligned} \alpha_{ik} &= \mathbb{1}(\alpha_{ik}^* > 0) \\ \alpha_{ik}^*|\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{i,k-1}, \boldsymbol{\zeta}_k, \boldsymbol{\psi}_k &\sim \mathcal{N}(\boldsymbol{a}_{i,K-1}^\top\boldsymbol{\psi}_k + \boldsymbol{z}_i^\top\boldsymbol{\zeta}_k, 1) \end{aligned} \tag{4.9}$$

where α_{ik}^* is a continuous, augmented version of α_{ik} that has a normal distribution conditioned on the first $k-1$ attributes, $(\alpha_{i1}, \dots, \alpha_{i,k-1})$, and the coefficients $\boldsymbol{\zeta}_k$ and $\boldsymbol{\psi}_k$.

[1] We follow the recommendation of O'Hara and Sillanpää (2009) and set $m = \sqrt{1000}$, which indicates that the prior variance for an inactive predictor v is 1000 times smaller than the prior variance for the active predictor.

Attribute-Level Covariate Coefficients

The $\boldsymbol{\zeta}$ parameter reveals the relationship between the attribute-level covariates and attribute mastery probabilities. Similar to the item-level covariates, we specify a generalized g-prior for $\boldsymbol{\zeta}_k$ such that we can apply a variable selection algorithm on the attribute-level covariate coefficients. With a binary activeness latent variable, we assume that each parameter in $\boldsymbol{\zeta}_k$ follows a mixture of two normal distributions, where $\zeta_{kl} \sim (1-\xi_{\zeta kl})\ \mathrm{N}\left(0, s_{\zeta kl}^{-2}\tau_{\zeta k}^{-1}(\boldsymbol{z_l}^\top \boldsymbol{z_l})^{-\mathbf{1}}\right) + \xi_{\zeta kl}\ \mathrm{N}\left(0, \tau_{\zeta k}^{-1}(\boldsymbol{z_l}^\top \boldsymbol{z_l})^{-\mathbf{1}}\right)$, and $P(\xi_{\zeta kl}=1) = 1 - P(\xi_{\zeta kl}=0) = p_{\zeta k}$. Moreover, we set the prior of the hyper-parameter, $\xi_{\zeta kl}$, to follow a Bernoulli distribution with a probability $p_{\zeta k}$, $\xi_{\zeta kv}|p_{\zeta k} \sim \text{Bernoulli}(p_{\zeta k})$, and $p_{\zeta k}$ follows a Beta(a_ζ, b_ζ) distribution. Thus, the prior of $\boldsymbol{\zeta}_k$ follows a multivariate normal distribution as,

$$\boldsymbol{\zeta}_k|\boldsymbol{v}_{\zeta k} \sim \mathcal{N}_V(\mathbf{0}, \tau_{\zeta k}^{-1}(\boldsymbol{K_{\zeta k}}\boldsymbol{Z}^\top \boldsymbol{Z}\boldsymbol{K_{\zeta k}})^{-1}) \tag{4.10}$$

where $\boldsymbol{\xi}_{\zeta k} = (\xi_{\zeta k1} \dots \xi_{\zeta kL})^\top$ is the binary latent vector with an entry being 1 if the corresponding parameter is active, and 0 otherwise. In addition, a hyper-parameter, τ_{ζ_k}, denotes the prior' precision of the parameter ζ_{kl}. $\boldsymbol{K}_{\zeta k} = \text{diag}[s_1, \dots, s_L]$ where we set $s_l = n, n > 1$ if $\xi_{\zeta kl} = 0$, and $s_v = 1$ if $\xi_{\zeta kl} = 1$. Moreover, the prior for $\tau_{\zeta k}$ is a gamma distribution for each k, $\tau_{\zeta k} \sim \text{Gamma}(a_{\zeta_0}, b_{\zeta_0})$. Note that setting $\boldsymbol{K}_{\zeta k}$ to an identity matrix (i.e., setting all the latent binary variables to one), one can exclude the variable selection algorithm from the procedure.

Latent structure related parameters

In exploratory RLCMs, researchers need to infer an underlying structure between attributes and the binary, domain-specific item-skill map— i.e., the $\boldsymbol{Q}$ matrix. Assuming that the elements of $\boldsymbol{Q}$ are conditionally independent and follow Bernoulli distributions—as $q_{jk}|\nu \sim$

Bernoulli(ν) where ν has a beta prior $\nu \sim \text{beta}(a_\nu, b_\nu)$, the joint prior distribution for $\boldsymbol{Q}$ is,

$$p(\boldsymbol{Q}|\nu) \propto \Big(\prod_{j=1}^{J} \prod_{k=1}^{K} \nu^{q_{jk}} (1-\nu)^{1-q_{jk}} \Big) \mathbb{1}(\boldsymbol{Q} \in \mathbb{Q}) \tag{4.11}$$

where $\mathbb{Q}$ denotes the identifiable space of the $\boldsymbol{Q}$ matrix. Culpepper (2019a) introduced a fully Bayesian model for inferring $\boldsymbol{Q}$ while applying a "spike-slab" prior for the $\boldsymbol{\beta}_j$, but one limitation of the approach is that the formulation imposed a more restrictive monotonicity condition. They introduced a structure, $\tilde{q}_j = (1, q_{j1}, \ldots, q_{jK}, q_{j1}q_{j2}, \ldots, q_{j(K-1)}q_{jK}, \ldots, \prod_{k=1}^{K} q_{jk})^\top$, to define the activeness of each β_{jp}. Later, Chen et al. (2020) introduced a sparse latent class model with the latent structure defined by a $\boldsymbol{\Delta} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_J)$ matrix rather than a $\boldsymbol{Q}$ matrix. In this setting, the elements of $\boldsymbol{\delta}_j$ denote the activeness of each β_{jp} parameter, $\boldsymbol{\delta}_j = (1, \delta_{j1}, \ldots, \delta_{jK}, \delta_{j12}, \ldots, \delta_{j(K-1)K}, \ldots, \delta_{j1\ldots K})^\top \in \{0,1\}^{2^K}$. Thus, $\boldsymbol{\delta}_j$ and $\boldsymbol{\beta}_j$ are connected in a way that a δ_{jp} is 1 if the corresponding parameter β_{jp} is active, and zero otherwise. Chen et al. (2020) relaxed the restrictive monotonicity condition of Culpepper (2019a), but the approach does not provide a mechanism for specifying expert knowledge about the $\boldsymbol{Q}$ matrix. In addition, a method proposed by Balamuta and Culpepper (2021) allows the inclusion of expert knowledge about $\boldsymbol{Q}$ links $\boldsymbol{\beta}_j$ to $\boldsymbol{q}_j$ through the structure of $\boldsymbol{\delta}_j$ as $p(\boldsymbol{\beta}|\boldsymbol{Q}) = \sum_{\text{all}\boldsymbol{\Delta}} p(\boldsymbol{\beta}|\boldsymbol{\Delta}) \times p(\boldsymbol{\Delta}|\boldsymbol{Q})$. To establish a stochastic relationship between $\boldsymbol{\delta}_j$ and $\boldsymbol{q}_j$, Balamuta and Culpepper (2021) used a confirmatory DINA model with common guessing, g, and slipping, s, parameters across all $\boldsymbol{q}_j$.

Thus, the prior distribution of δ_{jp} is parameterized as follow

$$p(\delta_{jp}|\boldsymbol{q}_j, g, s) \propto g^{(1-\tilde{q}_{pj})}(1-s)^{\tilde{q}_{pj}} \tag{4.12}$$

where $\tilde{q}_{jp}$ is the pth entry of the $\tilde{\boldsymbol{q}}_{\boldsymbol{j}}$ in Culpepper (2019a). Additionally, we deploy the

following truncated, conditionally normal prior for $\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp}$ as follows,

$$p(\beta_{jp}|\delta_{jp}, \boldsymbol{\beta}_{j(p)}) \propto v_{jp}^{-1/2} \exp\Big(-\frac{1}{2}\beta_{jp}^2/v_{jp}\Big) \mathbb{1}(\beta_{jp} > L_{jp}) \tag{4.13}$$

where $\boldsymbol{\beta}_{j(p)}$ excludes β_{jp} from $\boldsymbol{\beta}_j$, L_{jp} is the lower bound for the β_{jp}, and δ_{jp} is the activeness parameter associated with p. Note this is a stochastic search variable selection prior (Culpepper, 2019a; George & McCulloch, 1993) such that $v_{jp} = \delta_{jp}/w_1 + (1-\delta_{jp})/w_0$ where the precisions for the spike and slab are w_0 and w_1, respectively.

4.3.2 Posterior Inference

Next, we specify the full conditional distributions for the parameters under consideration for our Metropolis-within-Gibbs sampling algorithm. Similar to the previous section, we present the posterior specifications under five subsections:1) data augmentation for item responses; 2) posterior for the effects of covariates for item responses; 3) data augmentation for attributes; 4) posterior for the effects of covariates for attribute mastery probabilities; 5) posterior for inferring the latent structure related parameters.

Item Augmented Data

The full conditional distribution of the augmented response data follows a truncated normal distribution,

$$Y_{ij}^*|Y_{ij} = y_{ij}, \boldsymbol{\alpha}_i, \boldsymbol{\beta}_j, \boldsymbol{\gamma}_j \sim \mathcal{N}(\boldsymbol{a}_i^\top \boldsymbol{\beta}_j + \boldsymbol{x}_i^\top \boldsymbol{\gamma}_j, 1)\mathbb{1}(Y_{ij}^* > 0)^{y_{ij}} \mathbb{1}(Y_{ij}^* \leq 0)^{1-y_{ij}}. \tag{4.14}$$

Item-Level Covariate Coefficients

The full conditional distribution of $\boldsymbol{\gamma}_j$ is a multivariate normal distribution with a V-dimensional covariate space,

$$\boldsymbol{\gamma}_j | \boldsymbol{y}^*_{1:N,j}, \boldsymbol{\alpha}_{1:N}, \boldsymbol{\beta}_j, \boldsymbol{v}_{\gamma j} \sim \mathcal{N}_V(\boldsymbol{\mu}_{\gamma j}, \boldsymbol{\Sigma}_{\gamma j}) \tag{4.15}$$

where $\boldsymbol{y}^*_{1:N,j} = (y^*_{1j}, \ldots, y^*_{Nj})^\top$, $\boldsymbol{\alpha}_{1:N} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N)$, $\boldsymbol{\Sigma}_{\gamma j} = (\boldsymbol{X}^\top \boldsymbol{X} + \tau_{\gamma j} \boldsymbol{K}_{\gamma j} (\boldsymbol{X}^\top \boldsymbol{X}) \boldsymbol{K}_{\gamma j})^{-1}$ and $\boldsymbol{\mu}_{\gamma j} = \boldsymbol{\Sigma}_{\gamma j} \boldsymbol{X}^\top \left(\boldsymbol{y}^*_{1:N,j} - \boldsymbol{A}\boldsymbol{\beta}_j \right)$ where $K_{\gamma j} = \text{diag}[s_1, \ldots, s_V]$ where $s_v = m$, $m > 1$ if $\xi_{\gamma_{jv}} = 0$, and $s_v = 1$ if $\xi_{\gamma_{jv}} = 1$. Moreover, the full conditional distribution of the precision parameter, $\tau_{\gamma j}$, follows a gamma distribution such that $\tau_{\gamma j} \sim \text{gamma}(a_{\gamma 1}, b_{\gamma 1})$ where $a_{\gamma 1} = \frac{V}{2} + a_{\gamma 0}$, and $b_{\gamma 1} = \frac{1}{2}(\boldsymbol{\gamma}_{\boldsymbol{j}}^\top \boldsymbol{K}_{\boldsymbol{\gamma j}} \boldsymbol{X}^\top \boldsymbol{X} \boldsymbol{K}_{\boldsymbol{\gamma j}} \boldsymbol{\gamma}_{\boldsymbol{j}}) + b_{\gamma 0}$.

Similar to the procedure for updating entries of $\boldsymbol{q}_j$ and $\boldsymbol{\delta}_j$, we use a MH sampler always proposing a change in the value at the current state of $\xi_{\gamma jv}$. Let $\xi^{(t)}_{\gamma jv}$ be the current value of the $\xi_{\gamma jv}$ v'th entry, and $\xi'_{\gamma jv}$ be the proposed value such that $\xi'_{\gamma jv} = 1 - \xi^{(t)}_{\gamma jv}$. $T(\xi'_{\gamma jv}, \xi^{(t)}_{\gamma jv})$ is defined as

$$T\left(\xi'_{\gamma jv}, \xi^{(t)}_{\gamma jv}\right) = \left(\frac{p(\boldsymbol{\gamma}_j | \boldsymbol{\xi}'_{\gamma j}) p(\boldsymbol{\xi}'_{\gamma j} | p_{\gamma j})}{p(\boldsymbol{\gamma}_j | \boldsymbol{\xi}^{(t)}_{\gamma j}) p(\boldsymbol{\xi}^{(t)}_{\gamma j} | p_{\gamma j})} \right). \tag{4.16}$$

where $\boldsymbol{\xi}^{(t)}_{\gamma j}$ and $\boldsymbol{\xi}^{'}_{\gamma j}$ are the same on all entries but the vth entry, and the decision rule for the MH algorithm is

$$\xi^{(t+1)}_{\gamma jp} = \begin{cases} \xi'_{\gamma jv} & \min\left(1, T\left(\xi'_{\gamma jv}, \xi^{(t)}_{\gamma jv}\right)\right) > U \\ \xi^{(t)}_{\gamma jv} & \text{otherwise} \end{cases}. \tag{4.17}$$

Moreover, the full conditional posterior distribution of $p_{\gamma j}$ can be denoted as $p_{\gamma j} | \boldsymbol{\xi}_{\gamma_j} \sim \text{Beta}(\sum_{v=1}^V \xi_{\gamma_{jv}} + a_\gamma, V - \sum_{v=1}^V \xi_{\gamma_{jv}} + b_\gamma)$.

Attributes and Augmented Data

The full conditional distribution for an individual's alpha pattern $\boldsymbol{\alpha}_i$ is a categorical distribution with the conditional posterior probability of membership in class c being proportional

to,

$$p(\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c|\boldsymbol{y}_i, \boldsymbol{\Gamma}, \boldsymbol{B}, \boldsymbol{\zeta}, \boldsymbol{\psi}) \propto \pi_{ic} \prod_{j=1}^{J} p(y_{ij}|\boldsymbol{\alpha}_i^\top \boldsymbol{v} = c, \boldsymbol{\gamma}_j, \boldsymbol{\beta}_j). \tag{4.18}$$

The full conditional distribution for the augmented attribute data is a truncated normal distribution,

$$\alpha_{ik}^*|\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{i,k-1}, \boldsymbol{\zeta}_k, \boldsymbol{\psi}_k \sim \mathcal{N}(\boldsymbol{a}_{i,K-1}^\top \boldsymbol{\psi}_k + \boldsymbol{z}_i^\top \boldsymbol{\zeta}_k, 1)\mathbb{1}(\alpha_{ik}^* > 0)^{\alpha_{ik}}\mathbb{1}(\alpha_{ik}^* \leq 0)^{1-\alpha_{ik}}. \tag{4.19}$$

where $\boldsymbol{a}_{i,K-1}$ is a design vector defined by a subset of $\boldsymbol{\alpha}_i$ pattern (i.e., $(\alpha_{i1}, \ldots, \alpha_{i,K-1})$), the non-zero elements of the $\boldsymbol{\psi}_k$ characterizes the structure of the relationship among the attributes, and $\boldsymbol{\zeta}_k$ is covariate coefficients between $\boldsymbol{z}_i$ and the attribute mastery probabilities.

Attribute-Level Covariate Coefficients

The full conditional distribution for the attribute-related parameters, $\boldsymbol{\zeta}_k$, is

$$\boldsymbol{\zeta}_k|\ \boldsymbol{\alpha}_{1:N,k}^*, \boldsymbol{\alpha}_{1:N,1}, \ldots, \boldsymbol{\alpha}_{1:N,k-1}, \boldsymbol{\psi}_k, \boldsymbol{v}_{\zeta k} \sim \mathcal{N}_L(\boldsymbol{\mu}_{\zeta k}, \boldsymbol{\Sigma}_{\zeta k}) \tag{4.20}$$

where $\boldsymbol{\alpha}_{1:N,k}^* = (\alpha_{1k}^*, \ldots, \alpha_{Nk}^*)^\top$, $\boldsymbol{\Sigma}_{\zeta k} = (\boldsymbol{Z}^\top \boldsymbol{Z} + \tau_{\zeta k} \boldsymbol{K}_{\zeta k}(\boldsymbol{Z}^\top \boldsymbol{Z})\boldsymbol{K}_{\zeta k})^{-1}$ and $\boldsymbol{\mu}_{\zeta k} = \boldsymbol{\Sigma}_{\zeta k}\boldsymbol{Z}^\top\big(\boldsymbol{\alpha}_{1:N,k}^* - \boldsymbol{A}_{K-1}\boldsymbol{\psi}_{\boldsymbol{k}}\big)$, $\boldsymbol{A}_{K-1} = (\boldsymbol{a}_{1,K-1} \ldots \boldsymbol{a}_{N,K-1})^\top$ is a design matrix with the dimension $N \times 2^{K-1}$, and $K_{\zeta k} = \text{diag}[s_1, \ldots, s_L]$ where $s_l = m$, $m > 1$ if $\xi_{\zeta kl} = 0$, and $s_l = 1$ if $\xi_{\zeta kl} = 1$.

In addition, the hyper-parameter $\xi_{\zeta kl}$ is updated with the MH algorithm. We always propose a change in the value at the current state of $\xi_{\zeta kl}$. Define $\xi_{\zeta kl}^{(t)}$ as the current value of the lth entry of the latent binary variable, and let the proposed value be $\xi'_{\zeta kl} = 1 - \xi_{\zeta kl}^{(t)}$. We can define a function $T\left(\xi'_{\zeta kl}, \xi_{\zeta kl}^{(t)}\right)$ defined as

$$T\left(\xi'_{\zeta kl}, \xi_{\zeta kl}^{(t)}\right) = \left(\frac{p(\boldsymbol{\zeta}_k|\boldsymbol{\xi}'_{\zeta k})p(\boldsymbol{\xi}'_{\zeta k}|p_{\zeta k})}{p(\boldsymbol{\zeta}_k|\boldsymbol{\xi}_{\zeta k}^{(t)})p(\boldsymbol{\xi}_{\zeta k}^{(t)}|p_{\zeta k})}\right). \tag{4.21}$$

where $\boldsymbol{\xi}_{\zeta k}^{(t)}$ and $\boldsymbol{\xi}'_{\zeta k}$ are the same on all entries but the lth entry. The decision rule for the MH algorithm is

$$\xi_{\zeta kp}^{(t+1)} = \begin{cases} \xi'_{\zeta kl} & \min\left(1, T\left(\xi'_{\zeta kl}, \xi_{\zeta kl}^{(t)}\right)\right) > U \\ \xi_{\zeta kl}^{(t)} & \text{otherwise} \end{cases}. \tag{4.22}$$

The full conditional distribution of the precision parameter $\tau_{\zeta k}$ follows a gamma distribution, $\tau_{\zeta k} \sim \text{Gamma}(a_{\zeta 1}, b_{\zeta 1})$ where $a_{\zeta 1} = \frac{L}{2} + a_{\zeta 0}$ and $b_{\zeta 1} = \frac{1}{2}(\boldsymbol{\zeta_k}^\top \boldsymbol{K_{\zeta k}} \boldsymbol{Z}^\top \boldsymbol{Z} \boldsymbol{K_{\zeta k}} \boldsymbol{\zeta_k}) + b_{\zeta 0}$, $K_{\zeta k} = \text{diag}[s_1, \ldots, s_L]$ where $s_l = m$, $m > 1$ if $\xi_{\zeta kl} = 0$. Moreover, $p_{\zeta k}$'s full conditional posterior distribution can be denoted as $p_{\zeta k}|\boldsymbol{\xi}_{\zeta_k} \sim \text{Beta}(\sum_{l=1}^{L} \xi_{\zeta_{kl}} + a, L - \sum_{l=1}^{L} \xi_{\zeta_{kl}} + b)$.

The posterior distribution of ψ_{ks} for an attribute k and a parameter s is specified to be normally distributed. As previously stated, we use the parameter, ψ_{ks}, to model the structural relationships among the attributes, which is used to model the attribute structural probabilities. This yields the dimension of the $\boldsymbol{\psi}$ as $K \times 2^{K-1}$ with fixed pattern of zeros based upon the order the attribute distribution is factored (e.g., see subsection "Model for Attribute-Related Covariates"). Thus, the conditional posterior distribution of each element is

$$\psi_{ks}|\boldsymbol{\alpha}_{1:N,k}^*, \boldsymbol{\zeta}_k, \boldsymbol{\psi}_{k(s)} \sim \mathcal{N}(\mu_{ks}, \sigma_{ks}^2) \tag{4.23}$$

where $\boldsymbol{\alpha}_{1:N,k}^* = (\alpha_{1k}^*, \ldots, \alpha_{Nk}^*)^\top$, $\boldsymbol{\psi}_{k(s)}$ is $\boldsymbol{\psi}_k$ without element s, $\sigma_{ks}^2 = (\boldsymbol{A}_{K-1,s}^\top \boldsymbol{A}_{K-1,s} + v_\psi)^{-1}$, and $\mu_{ks} = \sigma_{ks}^2 \ \boldsymbol{A}_{K-1,s}^\top \big(\boldsymbol{\alpha}_{1:N,k}^* - \boldsymbol{A}_{K-1,(s)} \boldsymbol{\psi}_{k(s)} - \boldsymbol{Z}\boldsymbol{\zeta}_k\big)$. Also, $\boldsymbol{A}_{K-1} = (\boldsymbol{a}_{1,K-1} \ldots \boldsymbol{a}_{N,K-1})^\top$ is a design matrix with the dimension $N \times 2^{K-1}$, $\boldsymbol{A}_{K-1,s}$ is column s of $\boldsymbol{A}_{K-1}$, and $\boldsymbol{A}_{K-1,(s)}$ excludes column s of $\boldsymbol{A}_{K-1}$. In addition, the full conditional distribution of the hyper-parameter v_ψ is $v_\psi|\boldsymbol{\psi} \sim \text{gamma}(a_{\psi 1}, b_{\psi 1})$ where $a_{\psi 1} = \frac{2^K - 1}{2} + a_{\psi 0}$ and $b_{\psi 1} = \frac{1}{2}\text{tr}\left(\boldsymbol{\psi}\boldsymbol{\psi}^\top\right) + b_{\psi 0}$.

Latent structure related parameters

The full conditional distribution for β_{jp} is

$$\beta_{jp}|\ \boldsymbol{Y}_j^*, \boldsymbol{\alpha}, \boldsymbol{\beta}_{j(p)}, \delta_{jp}, \boldsymbol{\gamma}_j, \boldsymbol{q}_j \sim \begin{cases} \mathcal{N}(\mu_{jp}, \sigma_{jp}^2)\mathbb{1}(\beta_{jp} > 0) & 1 \leq p \leq K \\ \mathcal{N}(\mu_{jp}, \sigma_{jp}^2) & \text{otherwise} \end{cases} \tag{4.24}$$

where $\sigma_{jp}^2 = (\boldsymbol{A}_p^\top \boldsymbol{A}_p + v_{jp}^{-1})^{-1}$ and $\mu_{jp} = \sigma_{jp}^2\ \boldsymbol{A}_p^\top \big(\boldsymbol{Y}_j^* - \boldsymbol{A}_{(p)}\boldsymbol{\beta}_{j(p)} - \boldsymbol{X}\boldsymbol{\gamma}_j\big)$. $\boldsymbol{A}_p$ is the pth column of the design matrix $\boldsymbol{A}$, $\boldsymbol{A}_{(p)}$ is a matrix with all of $\boldsymbol{A}$, but the pth column, and $\boldsymbol{\beta}_{j(p)}$ is the $2^K - 1$ vector that omits $\boldsymbol{\beta}_{jp}$.

As in Balamuta and Culpepper (2021), the entries of the $\boldsymbol{Q}$ matrix are updated with the Metropolis-Hastings (MH) algorithm. While updating an entry, we always propose a change to its current state. Thus, if $q_{jk}^{(t)}$ is the current value of the entry of $\boldsymbol{Q}$, then the proposed candidate value is $q' = 1 - q_{jk}^{(t)}$. Let $T(q', q_{jk}^{(t)})$ be

$$T\left(q', q_{jk}^{(t)}\right) = \left(\frac{p(\boldsymbol{\delta}_j|\boldsymbol{q}_{j(k)}, q', s, g)p(q'|\boldsymbol{Q}_{(j,k)}, \nu)}{p(\boldsymbol{\delta}_j|\boldsymbol{q}_{j(k)}, q_{jk}^{(t)}, s, g)p(q_{jk}^{(t)}|\boldsymbol{Q}_{(j,k)}, \nu)}\right) \tag{4.25}$$

where $\boldsymbol{Q}_{(j,k)}$ indicates all other elements of $\boldsymbol{Q}$. The decision rule for the MH algorithm is

$$q_{jk}^{(t+1)} = \begin{cases} q' & \min\left(1, T\left(q', q_{jk}^{(t)}\right)\right) > U \\ q_{jk}^{(t)} & \text{otherwise} \end{cases} \tag{4.26}$$

where U is drawn from uniform$(0, 1)$. The full conditional distribution of ν is

$$\nu|\boldsymbol{Q} \sim Beta\left(\sum_{j=1}^{J}\sum_{k=1}^{K} q_{jk} + a_\nu, JK - \sum_{j=1}^{J}\sum_{k=1}^{K} q_{jk} + b_\nu\right). \tag{4.27}$$

Another model parameter that is updated with the MH algorithm is the $\boldsymbol{\delta}_j$ parameters. Similar to the procedure for updating entries of $\boldsymbol{q}_j$, we use a MH sampler that always proposes a change in the value at the current state of δ_{jp}. Let $\delta_{jp}^{(t)}$ be the current value of the $\boldsymbol{\delta}_j$ p'th

entry, and δ' be the proposed value such that $\delta' = 1 - \delta_{jp}^{(t)}$. $T(\delta', \delta_{jp}^{(t)})$ is defined as

$$T\left(\delta', \delta_{jp}^{(t)}\right) = \left(\frac{p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta')p(\delta'|\boldsymbol{q}_j, s, g)}{p(\beta_{jp}|\boldsymbol{\beta}_{j(p)}, \delta_{jp}^{(t)})p(\delta_{jp}^{(t)}|\boldsymbol{q}_j, s, g)}\right). \tag{4.28}$$

The decision rule for the MH algorithm is

$$\delta_{jp}^{(t+1)} = \begin{cases} \delta' & \min\left(1, T\left(\delta', \delta_{jp}^{(t)}\right)\right) > U \\ \delta_{jp}^{(t)} & \text{otherwise} \end{cases}. \tag{4.29}$$

4.4 Monte Carlo Simulation Study

4.4.1 Overview

A hundred Monte Carlo simulations are conducted to assess the performance of the parameter selection algorithms on the $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ parameters and examine its effects on the parameter recovery rates. Three general data-generating simulation conditions corresponding to the three ways to include covariates are explored: 1) saturated model with item- and attribute-level covariates; 2) $\boldsymbol{\zeta}$ model that only includes attribute-level covariates; and 3) $\boldsymbol{\Gamma}$ model that only includes item-level covariates. For the simulation studies, we generate data by incorporating the covariate effect on the measurement and/or structural part of the model. We estimate three models with each generated dataset. The first model is the base model that excludes the covariates to demonstrate the implications of neglecting to incorporate relevant covariates in exploratory RLCMs. The second and third models are designed to quantify the role covariates (e.g., the $\boldsymbol{\zeta}$, $\boldsymbol{\Gamma}$, or saturated models) with or with the variable selection framework. Note we estimate either the $\boldsymbol{\zeta}$, $\boldsymbol{\Gamma}$, or saturated model based on which part of the model we incorporate the covariates in the data-generation process.

In the simulation study design, the sample size is varied (i.e., $N = 1000$ and 4000), and the number of attributes is set to $K = 3$. In addition, to ensure our simulation is representative of real data, we generated covariates effects (i.e., $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$) and attribute

structural parameters (i.e., $\boldsymbol{\psi}$) that are consistent with our application (e.g., see the next section). Specifically, we sampled the elements of $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ from uniform$(-1, 1)$, and we set the hyper-parameters of the prior precision to 1 (i,e. $a_{\gamma 0} = b_{\gamma 0} = 1$ for the $\boldsymbol{\Gamma}$ parameter and $a_{\zeta 0} = b_{\zeta 0} = 1$ for the $\boldsymbol{\zeta}$ parameter). Moreover, the non-zero elements of the attribute structural parameters (i.e., $\boldsymbol{\psi}$) are sampled from uniform$(-0.4, 0.4)$, and the hyper-parameters of the prior precision is set to 1 (i.e. $a_{\psi 0} = b_{\psi 0} = 1$). We compared the performance of two different "spike" parameters (i.e., $m = 10$ and $\sqrt{1000}$) for the prior variance of irrelevant coefficients. However, the results only differ in the third decimal; thus, we only presented the results from the $m = \sqrt{1000}$ condition. We generated 10 binary item- and attribute-level covariates for each respondent from a Bernoulli(0.5) distribution. To evaluate the performance of the variable selection algorithm, we set 20% of the true $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ coefficients to a non-zero value (i.e., 0.75 or -0.75)—representing the relevant covariates—and the remaining 80% to 0—representing the irrelevant covariates. Moreover, we used the strictly identified $\boldsymbol{Q}$ matrix with $J = 20$ for $K = 3$ as presented below:

$$\boldsymbol{Q}_{K=3} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

$\boldsymbol{B}$ are generated using the approach of G. Xu and Shang (2018) where the success probabilities associated with the latent groups $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\alpha} = \mathbf{1}$ are set to 0.2 and 0.8, respectively. The remaining success probabilities are varied between 0.2 and 0.8 with values defined by $0.2 + (0.8 - 0.2) \times K_j^{\top} / K_j$ where $K_j^{\top}$ denotes the number of the required attributes in $\boldsymbol{q}_j = (q_{j1}, \ldots q_{jK})$. Moreover, both prior hyper-parameters of ν are set to 1 (i.e. $a_\nu = b_\nu = 1$)

In each simulation, 100 Monte Carlo replications are generated. In each chain, we ran $50,000$ iterations, of which we discarded the first $20,000$ as the burn-in period. To assess the recovery rates of the model parameters, we calculated the average mean absolute errors (MAE) of the $\boldsymbol{B}$ parameters using

$$MAE(\boldsymbol{B}) = \frac{1}{RJP}\sum_{r=1}^{R}\sum_{j=1}^{J}\sum_{p=1}^{P}|\hat{\beta}_{jp}^{(r)} - \beta_{jp}| \tag{4.30}$$

where $P = 2^K$ denotes the number of regression coefficients and $\hat{\beta}_{jp}^{(r)}$ is the posterior mean from replication r. Similarly, the MAE for the $\boldsymbol{\Gamma}$ parameters is, $\frac{1}{RJV}\sum_{r=1}^{R}\sum_{j=1}^{J}\sum_{v=1}^{V}|\hat{\gamma}_{jv}^{(r)} - \gamma_{jv}|$ where V is the number of item-related covariate coefficients and $\hat{\gamma}_{jv}^{(r)}$ is the posterior mean from replication r. Moreover, we reported the element-wise agreement rate of the binary activeness parameters of $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ coefficients as $\frac{1}{R}\sum_{r=1}^{R}\frac{1}{JV}\sum_{j=1}^{J}\sum_{v=1}^{V}\mathbb{1}(\hat{\xi}_{\gamma jv}^{(r)} = \xi_{\gamma jv})$ and $\frac{1}{R}\sum_{r=1}^{R}\frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}\mathbb{1}(\hat{\xi}_{\zeta kl}^{(r)} = \xi_{\zeta kl})$.

For the $\boldsymbol{\zeta}$ and $\boldsymbol{\psi}$ parameters, we reported the average mean absolute errors of the attribute class distribution, $\boldsymbol{\pi}_i = (\pi_{i0}, \ldots, \pi_{i,2^K-1})^\top$, of each examinee i. In Equation 3, π_{ic} is denoted as the probability of an examinee i belonging to an latent class c. Thus, the average mean absolute value for $\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_N$ can be calculated as $\frac{1}{RCN}\sum_{r=1}^{R}\sum_{c=0}^{C}\sum_{i=1}^{N}|\hat{\pi}_{ic}^{(r)} - \pi_{ic}^{(r)}|$ where $\hat{\pi}_{ic}^{(r)}$ and $\pi_{ic}^{(r)}$ are the posterior mean and data generating valuing for replication r. In addition, the average mean absolute errors of the probability of correctly answering an item, which is a function of covariates for the saturated and $\boldsymbol{\Gamma}$ models, are reported as follows. Recall that $\theta_{ij} = P(Y_{ij} = 1|\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_j, \boldsymbol{\beta}_j)$ is the item j mastery probability of examinee i as in Equation 1. Then, the average mean absolute errors of θ_{ij} can be calculated as $\frac{1}{RJN}\sum_{r=1}^{R}\sum_{j=1}^{J}\sum_{i=1}^{N}|\hat{\theta}_{ij}^{(r)} - \theta_{ij}^{(r)}|$.

We evaluated the model performance on estimating the latent structure by reporting the element-wise agreement rates for $\boldsymbol{Q}$, which is defined as $\frac{1}{R}\sum_{r=1}^{R}\frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbb{1}(\hat{q}_{jk}^{(r)} = q_{jk})$, and matrix-wise agreement rates, $\frac{1}{R}\sum_{r=1}^{R}\mathbb{1}\left(\hat{\boldsymbol{Q}}^{(r)} = \boldsymbol{Q}\right)$. In the rth replication, $\hat{\boldsymbol{Q}}^{(r)}$ is obtained after discretizing the element-wise mean of the sampled binary $\boldsymbol{Q}$ matrices as 1 if the

corresponding entry is greater than 0.5, and 0 otherwise, and $\hat{q}_{jk}^{(r)}$ represents the element in the jth row and kth column of the $\hat{\boldsymbol{Q}}^{(r)}$. In addition, we evaluated the attribute- and pattern-wise agreement rates of the examinee latent classes with $\frac{1}{R}\sum_{r=1}^{R}\frac{1}{NK}\sum_{N=1}^{N}\sum_{k=1}^{K}\mathbb{1}(\hat{\alpha}_{ik}^{(r)}=\alpha_{ik})$, and $\frac{1}{R}\sum_{r=1}^{R}\frac{1}{N}\sum_{N=1}^{N}\mathbb{1}(\hat{\boldsymbol{\alpha}}_{i}^{(r)}=\boldsymbol{\alpha}_{ik})$, respectively.

4.4.2 Results

In this section, we presented the results to report the performance of the variable selection mechanism on the item- and attribute-level covariates. For the case with a saturated data-generating model, Table 4.1 presents the results when we fit the data to the models with and without a variable selection algorithm on the $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ parameters as well as the base model (i.e., the misspecified model that incorrectly excludes covariates). The results show that when many irrelevant covariates are introduced to the model (i.e., 80% of our covariates are irrelevant), for the majority of the parameters, the saturated model with the variable selection mechanisms provides slightly better accuracy rates than the one without the variable selection mechanism. In addition, the $\boldsymbol{\Gamma}$ parameter recovery rates are significantly better in the saturated model with the variable selection mechanism than the saturated model without one. Moreover, the model with the parameter selection algorithm correctly recovers the $\boldsymbol{\Gamma}$ parameters' activeness indicator over 95% of the time and the $\boldsymbol{\zeta}$ parameters' activeness indicator over 90% of the time. Compared to the base model, the saturated model, regardless of including the parameter selection mechanism, is superior to the base model on recovering all the parameters in interest. In detail, the base model performs very poorly on recovering the underlying latent structure when the true generating data includes covariates on both item and attribute levels.

Table 4.1: Summary of the Monte Carlo replications for saturated model

Fitted Model	Sample size	Average MAE				$\boldsymbol{Q}$ Accuracy		Class Accuracy		$\boldsymbol{\xi}$ Accuracy	
		$\boldsymbol{B}$	$\boldsymbol{\Gamma}$	θ_{ij}	$\boldsymbol{\pi}_i$	Element	Matrix	Element	Pattern	ξ_γ	ξ_ζ
W/O variable selection	1000	0.1431	0.0878	0.0597	0.0261	0.9807	0.32	0.9138	0.7713	NA	NA
	4000	0.0713	0.0421	0.0301	0.0140	0.9995	0.97	0.9205	0.7884	NA	NA
W/ variable	1000	0.1236	0.0385	0.0494	0.0201	0.9790	0.24	0.9162	0.7777	0.9611	0.9033
	4000	0.0652	0.0170	0.0254	0.0111	0.9993	0.96	0.9210	0.7897	0.9934	0.9873
Base model	1000	0.2514	NA	0.1544	0.1043	0.7773	0	0.7830	0.52	NA	NA
	4000	0.2646	NA	0.1508	0.1046	0.7830	0	0.7811	0.52	NA	NA

Average MAE = Average mean absolute errors for parameters $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, mastery probability of an item θ_{ij} and latent class membership probability $\boldsymbol{\pi}_i$; $\boldsymbol{Q}$ Accuracy Matrix = matrix-wise accuracy rate for $\boldsymbol{Q}$; $\boldsymbol{Q}$ Accuracy Element = element-wise accuracy for $\boldsymbol{Q}$; ξ_γ, ξ_ζ Accuracy = average accuracy rate of latent binary activeness of the $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ parameters,respectively; $\boldsymbol{\Gamma}$ parameter is not applicable in base model; ξ_γ and ξ_ζ are not applicable in saturated model without a variable selection mechanism and base model. Results are based upon 100 replications.

To investigate the variable selection mechanism only on the $\boldsymbol{\Gamma}$ parameters, we first generate the data by introducing covariates only on the item level. Later, we fit the data to the models with and without a variable selection algorithm on the $\boldsymbol{\Gamma}$ parameters and the base model. Finally, we presented the results in Table 4.2. The model with the variable selection mechanism performs slightly better than the one without the mechanism for all the parameters but $\boldsymbol{\Gamma}$. For the $\boldsymbol{\Gamma}$, the model with the variable selection mechanism performs significantly better than the one without the mechanism. In addition, the $\boldsymbol{\Gamma}$ parameters' activeness indicators are restored almost perfectly (i.e., over 97 percent). On the other hand, the base model shows significantly low recovery rates of underlying latent structure and only mild recovery rates of the examinee-related parameters compared to the $\boldsymbol{\Gamma}$ models regardless of the selection mechanism.

Table 4.2: Summary of the Monte Carlo replications for $\boldsymbol{\Gamma}$ model

Fitted Model	Sample size	Average MAE				$\boldsymbol{Q}$ Accuracy		Class Accuracy		$\boldsymbol{\xi}$ Accuracy	
		$\boldsymbol{B}$	$\boldsymbol{\Gamma}$	θ_{ij}	$\boldsymbol{\pi}_i$	Element	Matrix	Element	Pattern	ξ_γ	ξ_ζ
W/O variable selection	1000	0.1189	0.0865	0.0500	0.0187	0.9852	0.40	0.8827	0.7030	NA	NA
	4000	0.0625	0.0418	0.0255	0.0119	0.9997	0.98	0.8905	0.7214	NA	NA
W/ variable selection	1000	0.1021	0.0369	0.0392	0.0185	0.9857	0.44	0.8853	0.7095	0.9676	NA
	4000	0.0570	0.0167	0.0205	0.0119	0.9997	0.98	0.8911	0.7228	0.9942	NA
Base model	1000	0.2000	NA	0.1386	0.0416	0.8797	0	0.8215	0.5866	NA	NA
	4000	0.1786	NA	0.1335	0.0380	0.8892	0	0.8266	0.6006	NA	NA

Average MAE = Average mean absolute errors for parameters $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, mastery probability of an item θ_{ij} and latent class membership probability $\boldsymbol{\pi}_i$; $\boldsymbol{Q}$ Accuracy Matrix = matrix-wise accuracy rate for $\boldsymbol{Q}$; $\boldsymbol{Q}$ Accuracy Element = element-wise accuracy for $\boldsymbol{Q}$; ξ_γ, ξ_ζ Accuracy = average accuracy rate of latent binary activeness of the $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ parameters,respectively; $\boldsymbol{\Gamma}$ parameter is not applicable in base model; ξ_γ is not applicable in saturated model without a variable selection mechanism and base model, and ξ_ζ is not applicable for none of the models. Results are based upon 100 replications.

Finally, we apply the variable selection mechanism on only the $\boldsymbol{\zeta}$ parameter. We first generate the data by including the covariate's effects on the attribute level. Similar to the first two models, only 20% of the variables have non-zero coefficients. Table 4.3 shows that the $\boldsymbol{\zeta}$ model, regardless of including variable selection mechanism, performs similarly to the base model on recovering the underlying latent structure. With respect to recovering the latent class membership probabilities, similar to the first study, the $\boldsymbol{\zeta}$ model shows promising results over the base model. Moreover, the $\boldsymbol{\zeta}$ model with a variable selection mechanism results in the highest recovery rates for the examinees' latent class profiles with around 2% more in pattern-wise recovery rates than the base model. In distinguishing the relevant covariates from the irrelevant ones on the attribute level, the variable selection mechanism can correctly recover the parameters' activeness indicators more than 95% of the time. Thus, we can conclude that in the case of incorporating any irrelevant covariates on the attribute-level covariates, the selection mechanism on the $\boldsymbol{\zeta}$ parameter can correctly identify those covariates.

Table 4.3: Summary of the Monte Carlo replications for $\boldsymbol{\zeta}$ model

Fitted Model	Sample size	Average MAE				$\boldsymbol{Q}$ Accuracy		Class Accuracy		$\boldsymbol{\xi}$ Accuracy	
		$\boldsymbol{B}$	$\boldsymbol{\Gamma}$	θ_{ij}	$\boldsymbol{\pi}_i$	Element	Matrix	Element	Pattern	ξ_γ	ξ_ζ
W/O variable selection	1000	0.1067	NA	0.0432	0.0238	0.9858	0.40	0.9367	0.8294	NA	NA
	4000	0.0581	NA	0.0240	0.0125	0.9995	0.97	0.9405	0.8383	NA	NA
W/ variable selection	1000	0.1058	NA	0.0429	0.0178	0.9858	0.41	0.9372	0.8308	NA	0.9533
	4000	0.0580	NA	0.0239	0.0095	0.9995	0.97	0.9406	0.8388	NA	0.9907
Base model	1000	0.1111	NA	0.0446	0.1012	0.9868	0.43	0.9302	0.8119	NA	NA
	4000	0.0610	NA	0.0251	0.1046	0.9995	0.97	0.9340	0.8211	NA	NA

Average MAE = Average mean absolute errors for parameters $\boldsymbol{B}$, $\boldsymbol{\Gamma}$, mastery probability of an item θ_{ij} and latent class membership probability $\boldsymbol{\pi}_i$; $\boldsymbol{Q}$ Accuracy Matrix = matrix-wise accuracy rate for $\boldsymbol{Q}$; $\boldsymbol{Q}$ Accuracy Element = element-wise accuracy for $\boldsymbol{Q}$; ξ_γ, ξ_ζ Accuracy = average accuracy rate of latent binary activeness of the $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ parameters,respectively; $\boldsymbol{\Gamma}$ and ξ_γ parameter are not applicable in any models; ξ_ζ is not applicable for $\boldsymbol{\zeta}$ model without a variable selection mechanism and base model. Results are based upon 100 replications.

4.5 Application to Spatial Rotation Dataset

This section reports results from an application of our developed methods to a real data set collected by Culpepper and Balamuta (2017). In this data set, 516 student-participants were administered 30 spatial rotation items (i.e., the Purdue Spatial Visualization Test: Rotations (PSVT-R; Yoon, 2011)) and a survey to collect the participants' covariates such as demographic information and background information that may be associated with spatial reasoning skills. Two categorical items from the Survey of Spatial Representation and Activities (Terlecki & Newcombe, 2005) were collected and included as covariates of the spatial rotation latent structure: 1) "How proficient or skilled do you believe you are at using maps?" (i.e., "not skilled", "not very skilled", "moderately skilled", "very skilled"); and 2) "How often do you use maps?" (i.e., "once every few years to not much at all", "one to two times a year", "one to two times in six months", "one to two times a month", "weekly", "daily"). In addition, we include gender as a covariate in our application. Table 4.4 reports the frequency distribution for the covariates. Note for the map skill variable that we combined the "not skilled" and "not very skilled" levels into a "not skilled/not very skilled" category for our

analysis as only twelve respondents reported "not skilled". The covariates were collected as categorical variables and yield eight distinct main-effect variables in a dummy-coded design matrix (note the references levels are "not skill/not very skilled", "once every few years to not much at all" and "Female"). Similar to simulation study design, we set the hyper-parameters of the prior precision to 1. (i,e. $a_{\zeta_0} = b_{\zeta 0} = 1$ for the $\boldsymbol{\zeta}$ parameters and $a_{\psi_0} = b_{\psi 0} = 1$ for the $\boldsymbol{\psi}$ parameters). Moreover, we set the prior hyper-parameters of ν as 1 (i.e. $a_\nu = b_\nu = 1$).

Table 4.4: Descriptive statistics of covariates for spatial rotation dataset.

Covariate	Category	Frequency
Map skill	not skilled	13
	not very skilled	98
	moderately skilled	297
	very skilled	108
Map usage	Once every few years to not much at all	66
	1 to 2 times a year	49
	1 to 2 times in six months	62
	1 to 2 times in a month	152
	weekly	159
	daily	28
Gender	Male	179
	Female	337

Note we consider the same set of covariates at both the item level and attribute level. To decide the number of attributes in the model, we fit the data to the saturated model with a variable selection mechanism on item- and attribute-level covariates for $K = 3$ and $K = 5$. Later, we use the relative model fit index by calculating 13-fold cross-validation (LOO; Geisser & Eddy, 1979; Vehtari et al., 2017). The LOO index favors the saturated model with the $K = 3$ condition (LOO for $K = 3$ 1474.754 and for $K = 5$ 1486.894). Thus, we presented the results for only the $K = 3$ condition in this section.

Table 4.5: Estimated posterior means of the $\boldsymbol{Q}$ matrix and item regression coefficients for spatial rotation data.

Items	$\overline{\boldsymbol{Q}}$			$\bar{\boldsymbol{\beta}}_j$							
	Q1	Q2	Q3	β_0	β_3	β_2	β_{23}	β_1	β_{13}	β_{12}	β_{123}
1	0.6203	0.7130	0.6423	0.3555	0.3736	0.7320	0.5099	0.4090	0.4213	0.2604	0.3586
2	0.5328	0.7111	0.5120	0.4552	0.2264	0.9487	0.5336	0.3789	0.1647	0.4556	0.2826
3	0.6516	0.7005	0.6472	0.2570	0.6619	1.1591	0.1868	0.8113	0.1047	0.2248	0.4025
4	0.4693	0.6132	0.4876	0.3488	0.3120	0.6797	0.0020	0.3165	0.0592	0.0213	-0.0229
5	0.6819	0.7184	0.7143	-0.1717	1.0502	0.9630	0.3846	0.8027	0.0390	0.1408	0.3602
6	0.6072	0.7142	0.6965	-0.1704	0.9948	0.9355	0.3788	0.4994	0.1864	0.1756	0.4326
7	0.6555	0.6776	0.6659	-0.3411	0.8096	1.0511	0.0585	0.7862	-0.0074	0.1489	0.1149
8	0.5694	0.5229	0.5996	0.1021	0.2186	0.1468	0.6548	0.2135	0.3471	0.2536	0.4629
9	0.5284	0.7159	0.6868	-0.7306	0.8288	1.4152	0.4588	0.4498	0.1330	0.2300	0.3669
10	0.6465	0.7195	0.7077	-0.1137	0.6256	0.5884	0.6052	0.4435	0.3468	0.2955	0.4153
11	0.5410	0.6999	0.6697	-0.2441	0.7185	0.9354	0.2942	0.3728	0.1750	0.1887	0.2966
12	0.7109	0.7336	0.7324	-0.7315	1.5039	0.8530	0.6085	0.8114	0.4161	0.4717	0.2334
13	0.6852	0.7002	0.6730	-0.3501	0.5064	0.7750	0.3401	0.6139	0.6304	0.0947	0.3160
14	0.6363	0.7386	0.7259	-0.8391	0.6468	1.3244	0.8224	0.5734	0.5489	0.2505	0.2590
15	0.6270	0.5163	0.6655	-0.5450	0.8086	0.2859	0.1966	0.4941	0.1049	0.3816	0.5022
16	0.7014	0.7156	0.7102	-0.6930	1.0048	0.8320	0.2965	0.7821	0.3114	0.3475	0.3483
17	0.6846	0.6765	0.5949	-0.4364	0.4127	0.7808	0.0606	0.7013	0.6462	0.1901	0.0795
18	0.4832	0.6527	0.5365	-0.5084	0.3373	0.8143	0.0736	0.2999	0.1135	0.0903	0.0360
19	0.7126	0.7150	0.7329	-0.6869	1.2263	0.9125	0.3618	0.7962	0.5280	0.0467	0.3964
20	0.7038	0.6577	0.7107	-0.4181	0.8209	0.5067	0.1814	0.6140	0.6731	0.3375	0.3904
21	0.7000	0.6884	0.7210	-0.8051	0.9789	0.7430	0.1752	0.7551	0.4663	0.0949	0.4989
22	0.6678	0.6960	0.6835	-0.5669	0.4516	0.4374	0.3784	0.3722	0.4845	0.4770	0.3657
23	0.6742	0.5218	0.6580	-0.3179	0.4972	0.2284	0.3324	0.7159	0.4999	0.3435	0.0018
24	0.5811	0.5516	0.6104	-0.3445	0.5914	0.3822	0.0335	0.4655	0.0852	0.0692	-0.0846
25	0.6691	0.5985	0.7082	-0.6264	0.7533	0.3863	0.2680	0.4718	0.4990	0.1765	0.5875
26	0.6965	0.5228	0.7095	-0.6263	0.4498	0.2017	0.3448	0.4683	0.9758	0.2886	0.4517
27	0.6922	0.5647	0.6478	-0.2403	0.2621	0.1720	0.4641	0.4136	0.8561	0.4176	0.4082
28	0.6166	0.7021	0.6868	-0.8569	0.3675	0.4201	1.0416	0.3420	0.6083	0.2876	0.5200
29	0.6046	0.5490	0.6112	-0.8638	0.2963	0.2732	0.2409	0.3024	0.4650	0.1238	-0.0153
30	0.6477	0.5481	0.6407	-1.4380	0.4371	0.2565	0.4336	0.5160	0.2543	0.1650	-0.1524

In Table 4.5, we present element-wise means for the inferred $\boldsymbol{Q}$ matrix as well as regression coefficients, $\boldsymbol{\beta}_j$,—denoting the relationship between attributes and item responses. Table 4.5 provides evidence that the saturated model yields a dense $\boldsymbol{Q}$ matrix estimate as most of the element-wise probabilities exceed 0.50. Chen et al. (2020) noted that dense $\boldsymbol{Q}$ matrices, which are generically identified, are indicative of a general unstructured mixture model. Therefore, the estimated $\boldsymbol{Q}$ matrix suggests that the data support an unstructured mixture where most items require all eight classes to describe response patterns. In addition, Table 4.5 shows the estimated regression coefficients for the relationship between attributes and item responses. The majority of $\boldsymbol{\beta}$ parameters have large estimated coefficients, which suggests an unstructured latent class model, and only several coefficients have been estimated closer to 0.

Table 4.6 includes the item-level coefficient estimates, and Table 4.7 shows the activeness probabilities associated with each of the coefficients. In Table 4.6, we present the estimated coefficients in bold, which are estimated as active based on a cutoff value 0.5 on the probability of being active. The majority of the $\boldsymbol{\Gamma}$ coefficients are estimated close to zero and inactive, and most of the active coefficients have around 0.5 probability of being active. Under the current setting, an active coefficient on the item-level covariates can be interpreted as the item shows a possible DIF. The majority of our real data application items have one or two active coefficients, with the probability of being active is close to our 0.5 threshold. Thus, we can conclude that there is a sign of having a DIF for many items, although the evidence is not overwhelming. However, we suggest practitioners be more cautious about using item 12 since all the coefficients associated with it are estimated as active with relatively higher probabilities.

Table 4.6: $\boldsymbol{\Gamma}$ coefficient estimates for spatial rotation data application for the saturated model.

Item	Map Usage					Skill Level		Gender
	1 to 2 times in a month	1 to 2 times a year	1 to 2 times in six months	Daily	Weekly	Moderately	Very	Male
1	**-0.0579**	-0.0294	0.0065	0.0048	0.0107	0.0128	0.0393	0.0055
2	-0.0032	0.0138	-0.0454	-0.0856	**0.0801**	0.0153	0.0173	0.0005
3	-0.0014	0.0501	0.0048	-0.0037	-0.0410	0.0379	0.0265	-0.0175
4	-0.0209	**0.0726**	**0.0657**	-0.0662	0.0133	-0.0343	**-0.0757**	0.0300
5	-0.0025	-0.0211	-0.0122	-0.0203	0.0046	0.0033	-0.0036	0.0185
6	0.0058	-0.0154	0.0235	-0.0001	-0.0087	0.0227	0.0187	0.0206
7	-0.0395	**0.0742**	0.0002	0.0391	-0.0213	0.0237	0.0332	-0.0175
8	0.0331	**-0.0939**	-0.0218	-0.0280	-0.0190	0.0134	0.0122	-0.0065
9	0.0022	-0.0052	-0.0404	0.0704	0.0163	-0.0133	0.0065	-0.0124
10	-0.0328	-0.0334	-0.0061	-0.0293	-0.0103	0.0267	0.0436	-0.0015
11	-0.0073	-0.0163	-0.0066	0.0025	-0.0017	0.0114	0.0001	-0.0067
12	**-0.0150**	**-0.0952**	**-0.1374**	**-0.0073**	**-0.0863**	**0.0406**	**-0.0039**	**0.0589**
13	**0.0759**	0.0038	-0.0295	0.0698	-0.0054	-0.0128	0.0010	-0.0363
14	-0.0141	0.0022	0.0547	0.0424	0.0291	0.0049	-0.0137	0.0188
15	-0.0212	-0.0384	-0.0282	0.0183	0.0098	0.0096	**0.0688**	0.0262
16	-0.0210	-0.0015	-0.0121	-0.0577	-0.0073	-0.0046	0.0416	0.0141
17	0.0138	0.0281	0.0444	**-0.0963**	0.0375	**-0.0721**	0.0003	**0.0362**
18	-0.0108	**-0.0956**	-0.0287	0.0165	-0.0298	0.0308	0.0085	-0.0060
19	0.0114	0.0378	0.0262	-0.0440	**-0.1051**	0.0222	0.0310	0.0154
20	0.0221	**-0.1326**	0.0320	0.0028	0.0005	0.0201	-0.0010	0.0134
21	-0.0096	0.0189	0.0091	0.0426	0.0235	-0.0068	-0.0041	-0.0086
22	0.0567	0.0018	0.0078	-0.0910	0.0301	**0.1198**	**-0.0875**	-0.0104
23	**-0.0591**	**0.0857**	-0.0325	0.0230	-0.0026	0.0040	**0.0496**	-0.0240
24	-0.0062	-0.0185	**-0.1096**	0.0296	0.0206	-0.0069	-0.0033	0.0113
25	-0.0126	**0.0875**	-0.0295	-0.0086	-0.0254	0.0092	0.0012	0.0002
26	-0.0002	0.0177	-0.0057	0.0482	0.0070	-0.0132	0.0281	-0.0267
27	-0.0048	-0.0324	-0.0195	0.0665	0.0276	**0.0500**	0.0118	-0.0251
28	-0.0057	0.0075	0.0244	0.0313	-0.0202	-0.0150	-0.0035	0.0286
29	-0.0120	0.0526	-0.0209	-0.0493	-0.0065	0.0055	-0.0290	-0.0188
30	0.0004	0.0124	0.0099	0.0521	0.0269	-0.0184	0.0475	-0.0248

The coefficients in bold are estimated as active based on a 0.5 cutoff value on the probability of being an active parameter.

Table 4.7: $\boldsymbol{\Gamma}$ binary activeness indicators estimates for spatial rotation data application for the saturated model.

Item	Map Usage					Skill Level		Gender
	1 to 2 times in a month	1 to 2 times a year	1 to 2 times in six months	Daily	Weekly	Moderately	Very	Male
1	0.5221	0.4388	0.4332	0.4376	0.4402	0.4270	0.4741	0.4300
2	0.4473	0.4497	0.4743	0.4871	0.5558	0.4509	0.4529	0.4392
3	0.4234	0.4654	0.4282	0.4294	0.4727	0.4692	0.4411	0.4397
4	0.4814	0.5312	0.5249	0.4999	0.4742	0.4952	0.5488	0.4948
5	0.4056	0.4105	0.3959	0.4010	0.4027	0.3955	0.4008	0.4204
6	0.4002	0.4017	0.4116	0.3963	0.3971	0.4335	0.4195	0.4304
7	0.4797	0.5180	0.4358	0.4637	0.4473	0.4596	0.4652	0.4461
8	0.4788	0.5479	0.4309	0.4293	0.4401	0.4250	0.4218	0.4239
9	0.4162	0.4153	0.4532	0.4549	0.4289	0.4110	0.4217	0.4196
10	0.4357	0.4242	0.4011	0.4152	0.3975	0.4299	0.4517	0.4042
11	0.3692	0.3686	0.3625	0.3674	0.3639	0.3774	0.3736	0.3784
12	0.5062	0.5650	0.6230	0.5034	0.5798	0.5265	0.5017	0.5758
13	0.5762	0.4289	0.4616	0.4792	0.4389	0.4470	0.4415	0.4983
14	0.4235	0.4140	0.4728	0.4338	0.4447	0.4103	0.4169	0.4362
15	0.4259	0.4362	0.4261	0.4130	0.4126	0.4171	0.5324	0.4596
16	0.4088	0.3941	0.3884	0.4353	0.3934	0.3966	0.4447	0.4082
17	0.4644	0.4694	0.4918	0.5291	0.4921	0.5818	0.4745	0.5133
18	0.4297	0.5451	0.4391	0.4333	0.4481	0.4578	0.4297	0.4338
19	0.4570	0.4812	0.4702	0.4752	0.6425	0.4704	0.4730	0.4746
20	0.4553	0.6171	0.4586	0.4324	0.4474	0.4590	0.4427	0.4488
21	0.3915	0.3784	0.3670	0.3985	0.3961	0.3724	0.3692	0.3764
22	0.4838	0.4267	0.4179	0.4964	0.4429	0.6512	0.5221	0.4297
23	0.5517	0.5580	0.4845	0.4695	0.4592	0.4547	0.5183	0.4879
24	0.4197	0.4182	0.5969	0.4176	0.4315	0.4083	0.4049	0.4198
25	0.4235	0.5325	0.4320	0.4146	0.4322	0.4101	0.4029	0.4060
26	0.3942	0.4031	0.3944	0.4280	0.3979	0.4172	0.4345	0.4403
27	0.4246	0.4433	0.4263	0.4682	0.4586	0.5172	0.4461	0.4484
28	0.4030	0.4066	0.4187	0.4177	0.4277	0.4162	0.4103	0.4525
29	0.4102	0.4634	0.4154	0.4394	0.4070	0.4105	0.4402	0.4245
30	0.4224	0.4146	0.4139	0.4383	0.4401	0.4448	0.4895	0.4404

Table 4.8 includes attribute-level covariate coefficient estimates, and Table 4.9 presents the probability of being active for the corresponding attribute-level coefficients. Similar to the item-level coefficients, the estimated coefficients in bold represent the active coefficients which we decide based on a 0.5 cutoff value on the probability of being active. The binary activeness probabilities suggest that five of the "map skill" coefficients are positively associated with underlying latent attribute membership. More specifically, students who report themselves as "very skilled" in using the maps are more likely to master all three attributes than the students who report themselves in the "not skilled/not very skilled" category. Moreover, students that report themselves as "moderately skilled" has a higher probability of mastering the second and third attributes than those who are in the reference group. In addition, the "gender" coefficient is estimated as an active parameter indicating that male students are more likely to master all three attributes than their female counterparts. In the frequency of using a map, "one to two times in six months" is the only active coefficient with a just over 0.5 activeness probability (i.e., 0.5411). The students using maps "one to two times in six months" are more likely to master the third attribute than the reference group.

Table 4.8: $\boldsymbol{\zeta}$ coefficient estimates for spatial rotation data application for the saturated model.

Attribute	Map Usage					Skill Level		Gender
	1 to 2 times in a month	1 to 2 times a year	1 to 2 times in six months	Daily	Weekly	Moderately	Very	Male
1	-0.0803	0.0560	-0.1180	0.0302	-0.0361	0.1350	**0.8303**	**0.1510**
2	-0.0444	0.0013	-0.0805	-0.0168	-0.0397	**0.1383**	**0.2035**	**0.1397**
3	-0.0601	0.0201	**-0.2292**	-0.0129	-0.1475	**0.3465**	**0.5196**	**0.5189**

The coefficients in bold are estimated as active based on a 0.5 cutoff value on the probability of being an active parameter.

Table 4.9: $\boldsymbol{\zeta}$ binary activeness indicators estimates for spatial rotation data application for the saturated model.

Attribute	Map Usage					Skill Level		Gender
	1 to 2 times in a month	1 to 2 times a year	1 to 2 times in six months	Daily	Weekly	Moderately	Very	Male
1	0.3692	0.3649	0.4082	0.3083	0.2697	0.4898	0.8697	0.5380
2	0.4455	0.4302	0.4782	0.4170	0.4257	0.6019	0.5750	0.5778
3	0.3786	0.3998	0.5411	0.3475	0.4403	0.7466	0.7743	0.8555

4.6 Discussion

Formative assessments are important for providing fine-grained information for students' learning and progress. Exploratory RLCMs remain important tools for broadening the applicability of diagnostic models for developing formative assessments. We considered new methods for including covariates into exploratory RLCMs. Collateral information—here, covariates—about students might benefit educators in several ways. First, covariates may be a source of undesirable interactions between students and items such as DIF; thus, investigating the covariates can uncover and help prevent these undesirable interactions and construct irrelevant variance. Second, covariates might affect the difficulty level of accruing attributes for students. Knowing these effects can help educators to identify and support at-risk students. Covariates can also be used to evaluate educational intervention effects. Third, significant covariates can reduce the uncertainty in the model parameters by providing additional information about students.

Fine-grained information about a student's knowledge profile is another important component in educational environments. Recently, a significant amount of research has been conducted on exploratory RLCMs that can infer the underlying structure between attributes and a binary domain-specific item-skill map. Despite the common interest in exploratory RLCMs and covariates, no study has been conducted on incorporating the covariates in RLCMs. Thus, we proposed three models to include covariates in the exploratory RLCMs—a

saturated model and two of its constrained versions. In the saturated model, we explore the covariates' effects on the probability of correctly answering an item and the probability of mastering an attribute. The $\boldsymbol{\zeta}$ model, one of the constrained models, links covariates with the probability of mastering an attribute, while the $\boldsymbol{\Gamma}$ model links the covariates with the probability of correctly answering an item. Moreover, it may be challenging to distinguish active covariates from inactive ones. Thus, while incorporating covariates into RLCM, we also introduced a variable selection mechanism on the $\boldsymbol{\Gamma}$ and $\boldsymbol{\zeta}$ coefficients to decide the activeness of the covariate coefficients.

The findings suggest that all three models were able to recover the model parameters well. In addition, the latent structure recovery rates in all three models were quite promising—all proposed models recover the entire $\boldsymbol{Q}$ matrix accurately when $N = 4000$. We also investigate the condition when covariates are ignored in the model, the base model, even though the true data generating model included a covariate. Ignoring the covariates increases the mean absolute errors in almost all conditions. Moreover, it yields significantly lower latent structure recovery rates in the base model than in the saturated and gamma models. In the $\boldsymbol{\zeta}$ model, the findings suggest that including the attribute-level covariates in the model improves the recovery of $\boldsymbol{\pi}$, which may be used in practice to classify students. However, we did not find evidence in the current simulation design that omitting attribute-level covariates impacts inference about the latent structure (e.g., $\boldsymbol{Q}$ and $\boldsymbol{B}$). Moreover, in the second simulation study, we found that the variable selection mechanism successfully distinguished active item- and attribute-level covariates from the inactive ones and yields similar promising recovery patterns to the first simulation study. In general, including covariates benefits us not only by providing insights about the relationships between the covariates and the item success and attribute mastery probabilities, but also significantly increases the parameter recovery rates. We fit the saturated model in the real data application, including a variable selection mechanism on item- and attribute-level covariates. The model fit analysis suggests the $K = 3$ as the number of attributes. The findings show that all the

coefficients associated with gender and many coefficients associated with map skill levels are likely to increase a student's master probability on the attributes. Moreover, the item-level coefficients are estimated very close to zero, which suggests little evidence of possible DIF.

In addition to incorporating covariates into RLCMs, we also introduce a novel way to describe the structure of the relationship among the attributes (see subsection "Model for Attribute-Related covariates"). This approach provides great flexibility in structuring the relationships among attributes. In the present study, we assume an unstructured relationship among attributes and estimate the intercept, main-effects, and interaction terms. In the future, researchers may consider employing a regularization technique to infer structure related to relationships among attributes to possibly reveal underlying higher-order structures. Moreover, with the innovation in learning systems, an increasing number of covariates, such as process information from log files or response times, can be recorded during assessments.

Lastly, to increase the availability of the developed method for practitioners, we provide the algorithm throughout an "R" package hosted in GitHub (Yigit, Culpepper, & Balamuta, 2021). We present details on installing the package locally and demonstrating how to analyze a real-life data application in the "Readme" section.

Currently, in this manuscript, we consider only manifest item- and attribute-level covariates. However, the proposed Bayesian framework is quite flexible, and one can easily incorporate latent covariates on both item and attribute levels. For example, a factor model could be used to model latent item-level covariates, $\boldsymbol{X}_i$, such that $p(\boldsymbol{x}_i|\boldsymbol{f}_i, \boldsymbol{\kappa})$ where $\boldsymbol{\kappa}$ denotes the factor model parameters—loading and thresholds. Then, the model for $\boldsymbol{y}_i$ would be $p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_j, \boldsymbol{\beta}_j, \boldsymbol{f}_i)$ where $\boldsymbol{\gamma}_j$ now captures the relationship between $\boldsymbol{f}_i$ and $\boldsymbol{y}_i$. Similarly, one could model the latent attribute-level covariates, $\boldsymbol{Z}_i$ as $p(\boldsymbol{z}_i|\boldsymbol{f}_i, \boldsymbol{\kappa})$ where $\boldsymbol{\kappa}$ again stands for the factor model parameters—loading and thresholds. Under this framework, the examinee's latent profile, $\boldsymbol{\alpha}_i$, would be formulated as $p(\boldsymbol{\alpha}_i|\boldsymbol{\psi}, \boldsymbol{\zeta}, \boldsymbol{f}_i)$, which the $\boldsymbol{\zeta}$ parameters now explain possible associations between $\boldsymbol{f}_i$ and $\boldsymbol{\alpha}_i$.

In conclusion, exploratory RLCMs are important tools for developing and refining for-

mative assessments to accelerate student learning. We addressed the problem of incorporating information about students' context in the form of covariates to strengthen the applicability of these methods for applied research.

Chapter 5

Conclusion

Diagnostic models provide a statistical framework for designing formative assessments by classifying student knowledge profiles according to a collection of fine-grained attributes. They can provide fine-grained information about students' learning profiles to teachers and AI-based learning systems to design student-tailored interventions that accelerate skill development.

Combining a diagnostic model with a learning model will be beneficial to capture the longitudinal perspective of students' learning. A student's learning progress over time can provide valuable information to design the learning materials accordingly. Moreover, a diagnostic model combined with a learning model can project the transition probabilities of attributes from a non-mastery state to a mastery state. A low attribute transition probability may be linked to the difficulty of comprehension of that attribute. Thus, the transition probability of an attribute may create a valuable opportunity for teachers and learning systems to determine which attributes students need to spend more time or practice more to comprehend. For example, AI-based learning systems can keep administering exercises for the attributes until a predefined transition probability is satisfied. Furthermore, teachers can focus and design more practice materials for the attributes which are difficult to be comprehended by students.

Given the benefit of longitudinally tracking students' learning profiles, my second chapter applies an EM-based formulation to estimate students' learning profiles efficiently. Several studies have proposed modeling learning from a longitudinal perspective and using MCMC-based algorithms in the current literature. Although MCMC is a feasible estimation method

in high-dimensional scenarios, it often needs a tremendous number of iterations. Thus, an EM-based algorithm can make the estimation process more computationally efficient, which will make the model more accessible in classroom settings and AI-based learning systems. Currently, in this chapter, the probability of transitioning from mastery to non-mastery state for an attribute is set to zero, which assumes that a student does not forget an attribute once the student mastered the attribute. As a future direction, one can relax this assumption by allowing the transition from the mastery status to the non-mastery status at least after several time points since forgetting can occur after some time even though a student once mastered it.

In addition to the importance of the longitudinal perspectives of learning, the context and ecosystem in which students learn may play an important role in skill and item mastery. Therefore, it is crucial to develop methods for incorporating student covariates into diagnostic models. Existing research is designed to include covariates in confirmatory diagnostic models, which assumes the underlying latent structure (known as $\boldsymbol{Q}$ matrix) is a known component in the model. Since creating a $\boldsymbol{Q}$ matrix is often a time-demanding and challenging task, an alternative can be estimating it as a model parameter. Thus, in the third chapter, several models were proposed for including covariates in exploratory RLCMs. This allows one to both infer the latent structure (i.e., $\boldsymbol{Q}$ matrix) and evaluate the role of covariates on performance and skill mastery. The model parameter posterior distribution is approximated using a Markov chain Monte Carlo (MCMC) algorithm with a Metropolis-within-Gibbs algorithm. Results show that the underlying latent structure and parameter values can be accurately recovered. These models can reveal the possible associations between the covariates and item and attribute mastery probabilities. These associations can be used to identify at-risk students based on students covariates and benefit educators to design student-tailored interventions that accelerate skill development.

Moreover, we proposed a parameter, ψ_{ks}, to model the structural relationships among the attributes, which is used to model the structural attribute probabilities. Currently,

we factorize the attributes so that the latter attributes are dependent on the preceding attributes. However, other factorizations can be applied as well, as long as they are incorporated into the $\boldsymbol{\psi}$ structure by changing the locations of the fixed-zeros. The only restriction is that the number of non-zero elements in the $\boldsymbol{\psi}$ structure has to be less than or equal to $2^K - 1$ due to the identifiability constraint. The $\boldsymbol{\psi}$ structure allows the flexibility to relax the independence assumption among the attributes and make the model more aligned with real-life conditions in educational environments.

Currently, we incorporate manifest covariates in our model. However, there can be many latent covariates in the educational setting that may relate to the item and skill mastery probabilities. Thus, a factor model could be used to model latent item-level covariates, $\boldsymbol{X}_i$, such that $p(\boldsymbol{x}_i|\boldsymbol{f}_i, \boldsymbol{\kappa})$ where $\boldsymbol{\kappa}$ denotes the factor model parameters—loading and thresholds. Then, the model for $\boldsymbol{y}_i$ would be $p(\boldsymbol{y}_i|\boldsymbol{\alpha}_i, \boldsymbol{\gamma}_j, \boldsymbol{\beta}_j, \boldsymbol{f}_i)$ where $\boldsymbol{\gamma}_j$ now captures the relationship between $\boldsymbol{f}_i$ and $\boldsymbol{y}_i$. Similarly, one could model the latent attribute-level covariates, $\boldsymbol{Z}_i$ as $p(\boldsymbol{z}_i|\boldsymbol{f}_i, \boldsymbol{\kappa})$ where $\boldsymbol{\kappa}$ again denotes the factor model parameters—loading and thresholds. Under this framework, the examinee's latent profile, $\boldsymbol{\alpha}_i$, would be formulated as $p(\boldsymbol{\alpha}_i|\boldsymbol{\psi}, \boldsymbol{\zeta}, \boldsymbol{f}_i)$, which the $\boldsymbol{\zeta}$ parameters now explain possible associations between $\boldsymbol{f}_i$ and $\boldsymbol{\alpha}_i$.

After incorporating the covariates into the model, one challenge is distinguishing active covariates, which relate to the outcome variable, from inactive ones. To address this issue, in the fourth chapter, we adopt a hierarchical mixture model on the covariates' prior—a stochastic search variable selection prior (Culpepper, 2019a; George & McCulloch, 1993). This prior is designed to reflect the covariate structure of the design matrix, which yields the covariate's prior distribution to become equivalent to a generalized version of a Zellner (1986) g-prior. In this chapter, results show that the proposed models with the SSVS prior settings result in favorable recovery rates on underlying latent structure and model and examinee parameters. They can also successfully distinguish active covariates from inactive ones.

In real data applications, one common issue with covariates is that covariates can be highly correlated, which yields unstable maximum likelihood coefficient estimates in regres-

sion settings. Ročková and George (2014) conducted a study to address this issue while applying a variable selection algorithm in a regression setting. Their results show that they can mitigate the instability issue in parameter estimates influenced by multicollinearity. In the current design, we investigate the variable selection method performance when the covariates are uncorrelated. One future direction is to examine the current setup with the correlated covariates. We anticipate that the current stochastic search variable selection prior will be robust up to a point when mild correlations are present in the covariates as documented in (George & McCulloch, 1993).

Beyond educational settings, the insights about the association between covariates and attribute possession probabilities may have a valuable application in clinical studies and neuropsychology fields. Covariates may carry important information about the subjects' susceptibility to disease. One possible issue that might arise in these areas is that is the number of covariates is frequently larger than the number of observations. To address this issue, some regularization technique needs to be incorporated into the models discussed in Chapter 4. For a possible future direction, instead of using the SSVS variable selection method, one can apply a regularization technique such as a MCMC based algorithms using the horseshoe prior (Johndrow, Orenstein, & Bhattacharya, 2020).

Bibliography

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, *88*(422), 669–679.

Ayers, E., Rabe-Hesketh, S., & Nugent, R. (2013). Incorporating student covariates in cognitive diagnosis models. *Journal of Classification*, *30*(2), 195–224.

Balamuta, J. J., & Culpepper, S. A. (2021). Exploratory restricted latent class models with monotonicity requirements under pólya-gamma data augmentation. *Psychometrika*. (In Press)

Cauchy, A. (1847). Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, *25*(1847), 536–538.

Chen, Y., & Culpepper, S. A. (2020). A multivariate probit model for learning trajectories: A fine-grained evaluation of an educational intervention. *Applied Psychological Measurement*, *44*(7-8), 515-530.

Chen, Y., Culpepper, S. A., Chen, Y., & Douglas, J. A. (2018). Bayesian estimation of the DINA Q matrix. *Psychometrika*, *83*(1), 89–108.

Chen, Y., Culpepper, S. A., & Liang, F. (2020). A sparse latent class model for cognitive diagnosis. *Psychometrika*, *85*(1), 121-153.

Chen, Y., Culpepper, S. A., Wang, S., & Douglas, J. A. (2018). A hidden Markov model for learning trajectories in cognitive diagnosis with application to spatial rotation skills.

Applied Psychological Measurement, *42*(1), 5–23.

Chen, Y., Li, X., Liu, J., & Ying, Z. (2018). Recommendation system for adaptive learning. *Applied Psychological Measurement*, *42*(1), 24–41.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015a). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*(510), 850–866.

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015b). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*(510), 850-866. Retrieved from https://doi.org/10.1080/01621459.2014.934827 doi: 10.1080/01621459.2014.934827

Chen, Y., Liu, J., Xu, G., & Ying, Z. (2015c). Statistical analysis of Q-matrix based diagnostic classification models. *Journal of the American Statistical Association*, *110*(510), 850–866.

Cheng, R. C. H., Holland, W., & Hughes, N. A. (1996). Selection of input models using bootstrap goodness-of-fit. In *Proceedings of the 1996 winter simulation conference* (pp. 199–206). The University of Kent at Canterbury.

Chiu, C.-Y., & Köhn, H.-F. (2015). Consistency of cluster analysis for cognitive diagnosis the DINO model and the DINA model revisited. *Applied Psychological Measurement*, *39*(6), 465–479.

Colbett, A. T., Anderson, J. R., & OBrien, A. T. (1995). *Student modeling in the ACT programming tutor* (P. Nichols, S. Chipman, & R. Brennan, Eds.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, *4*(4), 253–278.

Culpepper, S. A. (2019a). Estimating the cognitive diagnosis Q matrix with expert knowledge: Application to the fraction-subtraction dataset. *Psychometrika*, *84*(2), 333–357.

Culpepper, S. A. (2019b). An exploratory diagnostic model for ordinal responses with binary

attributes: Identifiability and estimation. *Psychometrika*, *84*(4), 921–940.

Culpepper, S. A., & Balamuta, J. J. (2017). A hierarchical model for accuracy and choice on standardized tests. *Psychometrika*, *82*(3), 820–845.

Culpepper, S. A., & Chen, Y. (2019). Development and application of an exploratory reduced reparameterized unified model. *Journal of Educational and Behavioral Statistics*, *44*(1), 3–24.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179–199.

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

de la Torre, J., van der Ark, L. A., & Rossi, G. (2018). Analysis of clinical data from a cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, *51*(4), 281–296.

Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, *74*(365), 153–160.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*(423), 881–889.

Gonzalez-Brenes, J., Huang, Y., & Brusilovsky, P. (2014). General features in knowledge tracing to model multiple subskills, temporal item response theory, and expert knowledge. In *The 7th International Conference on Educational Data Mining* (pp. 84–91). University of Pittsburgh.

Gonzalez-Brenes, J., & Mostow, J. (2013). What and when do students learn? fully data-driven joint estimation of cognitive and student models. *Educational Data Mining*.

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.

Han, R., Chen, K., & Tan, C. (2020). Curiosity-driven recommendation strategy for adaptive learning via deep reinforcement learning. *British Journal of Mathematical and*

Statistical Psychology. doi: 10.1111/bmsp.12199

Henson, R. A., & Templin, J. L. (2007). Importance of Q-matrix construction and its effects cognitive diagnosis model results. In *Annual Meeting of the National Council on Measurement in Education, Chicago, IL.*

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.

Huang, Z., Liu, Q., Zhai, C., Yin, Y., Chen, E., Gao, W., & Hu, G. (2019). Exploring multi-objective exercise recommendations in online education systems. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 1261–1270).

Iaconangelo, C. J. (2017). *Uses of classification error probabilities in the three-step approach to estimating cognitive diagnosis models* (Unpublished doctoral dissertation). Rutgers, The State University of New Jersey, New Brunswick, NJ.

Johndrow, J., Orenstein, P., & Bhattacharya, A. (2020). Scalable approximate mcmc algorithms for the horseshoe prior. *Journal of Machine Learning Research*, *21*(73).

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.

Kaya, Y., & Leite, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educational and Psychological Measurement*, *77*(3), 369–388.

Li, F. (2008). *A modified higher-order DINA model for detecting differential item functioning and differential attribute functioning* (Unpublished doctoral dissertation). University of Georgia, Athens, GA.

Li, F., Cohen, A., Bottge, B., & Templin, J. L. (2015). A latent transition analysis model for assessing change in cognitive skills. *Educational and Psychological Measurement*, *76*(2), 181–204.

Li, X., & Wang, W.-C. (2015). Assessment of differential item functioning under cognitive diagnosis models: The DINA model example. *Journal of Educational Measurement*, *52*(1), 28–54.

Li, X., Xu, H., Zhang, J., & Chang, H.-H. (2021). Optimal hierarchical learning path design with reinforcement learning. *Applied Psychological Measurement*, *45*(1), 54–70.

Madison, M. J., & Bradshaw, L. P. (2018). Assessing growth in a diagnostic classification model framework. *Psychometrika*, *83*(4), 963–990.

Meredith, W., & Millsap, R. E. (1992). On the misuse of manifest variables in the detection of measurement bias. *Psychometrika*, *57*(2), 289–311.

Minchen, N. D., de la Torre, J., & Liu, Y. (2017). A cognitive diagnosis model for continuous response. *Journal of Educational and Behavioral Statistics*, *42*(6), 651–677.

O'Hara, R. B., & Sillanpää, M. J. (2009). A review of Bayesian variable selection methods: What, how and which. *Bayesian Analysis*, *4*(1), 85–117.

Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a Bayesian networks implementation of knowledge tracing. *International Conference on User Modeling, Adaptation*, 253–278.

Ročková, V., & George, E. I. (2014). Negotiating multicollinearity with spike-and-slab priors. *Metron*, *72*(2), 217–229.

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., Templin, J. L., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York, NY: Guilford Press.

Studer, C. (2012). *Incorporating learning over time into the cognitive assessment framework* (Unpublished doctoral dissertation). Carneign Mellon University, Pittsburg, PA.

Tan, C., Han, R., Ye, R., & Chen, K. (2020). Adaptive learning recommendation strategy based on deep Q-learning. *Applied Psychological Measurement*, *44*(4), 251–266.

Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics*, *10*(1), 55–73.

Templin, J. L. (2020, February 3). *Techniques for diagnostic assessment.* Retrieved from https://patentimages.storage.googleapis.com/0e/22/e4/48f0e0e073d504/US20200273363A1.pdf (US Patent Application # 20200273363)

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *1*(3), 287.

Templin, J. L., Henson, R. A., Templin, S. E., & Roussos, L. (2008). Robustness of hierarchical modeling of skill association in cognitive diagnosis models. *Applied Psychological Measurement*, *32*, 559–574.

Templin, J. L., & Hoffman, L. (2013). Obtaining diagnostic classification model estimates using Mplus. *Educational Measurement*, *32*(2), 37–50.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? the relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, *53*(5), 433–441.

Tjoe, H., & de la Torre, J. (2014). On recognizing proportionality: Does the ability to solve missing value proportional problems presuppose the conception of proportional reasoning? *The Journal of Mathematical Behavior*, *33*, 1–7.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*(5), 1413–1432.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*(2), 287–307.

Wang, S., Yang, Y., Culpepper, S. A., & A., D. J. (2018). Tracking skill acquisition with cognitive diagnosis models: A higher-order, hidden Markov model with covariates. *Journal of Educational and Behavioral Statistics*, *43*(1), 57-87. Retrieved from https://doi.org/10.3102/1076998617719727 doi: 10.3102/1076998617719727

Wang, S., Zhang, S., Douglas, J. A., & Culpepper, S. A. (2018). Using response times to

assess learning progress: A joint model for responses and response times. *Measurement: Interdisciplinary Research and Perspectives*, *16*(1), 45-58. Retrieved from https://doi.org/10.1080/15366367.2018.1435105 doi: 10.1080/15366367.2018.1435105

Wang, S., Zhang, S., & Shen, Y. (2020). A joint modeling framework of responses and response times to assess learning outcomes. *Multivariate Behavioral Research*, *55*(1), 49-68. Retrieved from https://doi.org/10.1080/00273171.2019.1607238 (PMID: 31165632) doi: 10.1080/00273171.2019.1607238

Wang, Z., Guo, L., & Bian, Y. (2014). Comparison of DIF detecting methods in cognitive diagnostic test. *Acta Psychologica Sinica*, *46*(12), 1923–1932.

Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent class models. *Journal of the American Statistical Association*, *113*(523), 1284-1295. Retrieved from https://doi.org/10.1080/01621459.2017.1340889 doi: 10.1080/01621459.2017.1340889

Xu, Y., & Mostow, J. (2012, 7). Comparison of methods to trace multiple subskills: Is LR-DBN best? In (pp. 84–91). Chania, Greece: ERIC.

Ye, S., Fellouris, G., Culpepper, S. A., & Douglas, J. A. (2016a). Sequential detection of learning in cognitive diagnosis. *British Journal of Mathematical and Statistical Psychology*, *69*(2), 139–158.

Ye, S., Fellouris, G., Culpepper, S. A., & Douglas, J. A. (2016b). Sequential detection of learning in cognitive diagnosis. *British Journal of Mathematical and Statistical Psychology*, *69*(2), 139–158.

Yigit, H. D., Culpepper, S. A., & Balamuta, J. J. (2021, 08). *Variable selection for exploratory restricted latent class models with covariates.* Retrieved from https://github.com/hyigit2/egdm_covariates

Yigit, H. D., & Douglas, J. A. (2021). First-order learning models with the GDINA: Estimation with the EM algorithm and applications. *Applied Psychological Measurement*, *45*(3), 143–158.

Yoon, S. Y. (2011). *Psychometric properties of the revised Purdue spatial visualization tests: Visualization of rotations (The Revised PSVT-R)* (Unpublished doctoral dissertation). Purdue University, West Lafayette.

Zellner, A. (1986). *On assessing prior distributions and Bayesian regression analysis with g-prior distributions* (P. Goel & A. Zellner, Eds.). Amsterdam: Elsevier Science.

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

Zhang, S., & Chang, H.-H. (2016). From smart testing to smart learning: How testing technology can assist the new generation of education. *International Journal of Smart Technology and Learning*, *1*(1), 67–92.

Zhang, S., & Chang, H.-H. (2020). A multilevel logistic hidden Markov model for learning under cognitive diagnosis. *Behavior Research Methods*, *52*(1), 408–421.

Zhang, W. (2006). *Detecting differential item functioning using the DINA model* (Unpublished doctoral dissertation). The University of North Carolina at Greensboro, Greensboro, NC.