ABSTRACT

Title of Document: DYNAMIC BAYESIAN INFERENCE NETWORKS AND HIDDEN MARKOV MODELS FOR MODELING LEARNING PROGRESSIONS OVER MULTIPLE TIME POINTS

Younyoung Choi, Doctor of Philosophy, 2012

Directed By: Professor, Robert J. Mislevy,
Department of Measurement, Statistics and Evaluation

The current study examines the performance of a Bayesian Inference Network (BIN) for modeling Learning Progressions (LP) as a longitudinal design approach. Recently, Learning Progressions, defined by measurable pathways that a student may follow in building their knowledge and gaining expertise over time (National Research Council, 2007; Shin, Stevens, Short & Krajcik, 2009), have captured attention in mathematics and science education (Learning Progressions in Science Conference, 2009). While substantive, psychological, instructional, and task developmental aspects has been proposed in the LP framework, few assessment design frameworks have been designed to link the theory embodied in a progression, tasks that provide evidence about a student's level on that progression, and psychometric models that can link them. Specially, few psychometric models have been proposed to characterize the relationship between student performance and levels on learning progressions in a longitudinal design approach. This dissertation introduces an approach to modeling LPs over multiple time points using Bayesian Inference Networks, referred to as dynamic Bayesian Inference Networks (DBINs). The DBINs are a framework for modeling LPs over time by integrating the theory embodying LPs, assessment design, and interpretation of student performances. The technical aspects of this dissertation cover the fundamental concepts of the graphical model for constructing a DBIN. It is shown that this modeling strategy for change over multiple time points is equivalent to a hidden Markov model. An expectation-maximization (EM) algorithm is presented for estimating the parameters in the model. Two simulation studies are conducted that focus on the construction of a simple DBIN model and an expanded DBIN model with a covariate. The extension that incorporates a covariate for students is useful for studying the effect of instructional treatments, students' background, and motivation on a student's LP. An application illustrates the ideas with real data from the domain of beginning computer network engineering drawn from work in the Cisco Networking Academy.

DYNAMIC BAYESIAN INFERENCE NETWORKS AND HIDDEN MARKOV MODELS FOR MODELING LEARNING PROGRESSIONS OVER MULTIPLE TIME POINTS

by

Younyoung Choi

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2012

Advisory Committee:
Professor Robert J. Mislevy, Chair
Professor George Macready
Professor Patricia Campbell
Professor Hong Jiao
Professor Jeffrey Harring

DEDICATION

To dear mother and father, whose love has made this all worthwhile, and whose patience and support have made this all possible.

ACKNOWLEDGEMENTS

A number of people and parities have made this work possible. This dissertation research was supported in part by Cisco Systems and the Cisco Networking Academy. I would like to express my sincere gratitude to the department of Measurement, Statistics, and Evaluation and those who supported me during my graduate school life. To CISCO academy, Martin Benson, Kristina Chapple, and Kristen DiCerbo, who gave me insightful comments and valuable information for the application study.

To Dr. Daisy Rutstein and Jihyun Kim, who supported and encouraged me as my mentors for my graduate life and dissertation journey.

To Dr. Jen Gray, who helped with my last edit for publishing my dissertation.

To Dr. Jiao, who gave me warm advice and insightful suggestions on my dissertation.

To Dr. Macready, Dr. Campbell, and Dr. Harring for their thoughtful suggestions.

To Dr. Hancock, whose amazing statistical courses and support our EDMS department.

To Dr. Dayton, for giving me the once in a lifetime opportunity to study in this great program.

Most of all, my deepest gratitude goes to Dr. Robert J. Mislevy, for his patience, encouragement, vision have allowed me to continue my dissertation work, for his supports throughout my doctoral study, for his wisdom and guidance about my attitude of study.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Learning Progressions (LPs) are defined by measurable pathways that a student may follow in the process of building their knowledge and gaining expertise over time (National Research Council, 2007; Shin, Stevens, Short & Krajcik, 2009). LPs are hypotheses about how a student's understanding of knowledge, skills, and abilities in a targeted area develops over time (Corcoran, Mosher, & Rogat, 2009). Generally, LPs consist of several levels or units, each of which represents a given state of knowledge, skill, and abilities (KSAs) required for a student to achieve mastery at that level. As an example, Gotwals, Songer, and Bullard (2009) developed a LP about complex inquiry reasoning for building evidence-based explanations in biodiversity and ecology (see Table 1). The LP consists of five levels which contain a given state of KSAs required for a student to be at that level in the domain of biodiversity and ecology. The five levels are hierarchically structured so that level 5 requires higher KSAs than level 4, level 4 requires higher KSAs than level 3, and so on.

Table 1

An example of a learning progression regarding biodiversity content adapted from Gotwals, Songer, and Bullard (2009)

Level	Description
Level0[1]	No systematic understanding of biodiversity
Level1	Student understands that a habitat is a place that provides food, water, shelter, and space for living things.
Level2	Student understands that animals have different features that they use to survive in different habitats. Student understands that there are observable internal and external differences. Student understands that some of these differences are used to distinguish major groups.
Level3	Student understands that richness and abundance are two different measures of the amount of animal life in a habitat or area.
Level4	Student understands that biodiversity is a measure of the number and variety of different organism in a particular area (habitat, ecosystem, or biome).
Level5	Student understands that an area has high biodiversity if it has both high richness and high abundance.

Note. The description of level 0[1] has been modified from the original LP so as to describe a progression of students' understanding.

Since LPs provide useful information for improving student learning, they have recently captured the attention of professionals in mathematics and science education (Learning Progressions in Science conference, 2009). The major objectives in the study of LPs are to provide (1) information regarding the state of a student with respect to the level of understanding of a given concept and (2) diagnostic information regarding the strength and weakness of a student's understanding along a curriculum (Gotwals, Songer, & Bullard, 2009; Schwarz, et al., 2009; Shin, Stevens, Short & Krajcik, 2009). To provide such information about the learning states of a student (i.e., the current, past, and future levels of a student on LPs), the first step is to develop tasks for gathering student responses that provide evidence about students' KSAs relative to their levels on LPs. More specifically, once key task features that can evoke evidence about student states have been identified by drawing on research, the information can be used for constructing the tasks that can elicit student responses containing evidence about student KSAs. In the measurement community, there has been a shift in the development of assessments to the incorporating the cognitive aspects about structure and acquisition of knowledge into the assessment development system in order to accurately represent the knowledge, skills, and abilities with respect to the purpose of the assessment (Mislevy, 1994; Nichols, Chipman, & Brennan, 1995). This movement allows assessments to produce diagnostic feedback based on the expected ways in which students understand and solve tasks in addition to providing an overview of each student's ability level (Leighton & Gierl, 2007; Mislevy, 1994; Nichols, Chipman, & Brennan, 1995).

Furthermore, the emphasis on the connection between cognitive psychology and measurement contributes to providing meaningful information for instructional uses. Linn (1986) stated that traditional standardized assessments have very little instructional uses in terms of what should be done to improve a student's level of achievement. Put another way, overall test scores from transitional standardized assessments provide relatively less information about the nature of a student's weaknesses, strengths, and errors than cognitively developed assessments. Huff and Goodman (2007) found that a large percentage of teachers wished they had more individualized diagnostic information from these assessments. The National Research Council (NRC) (2001) reported that formative and timely feedback is important to students in their development. In order to address these issues, Mislevy (1995) called for the importance of creating assessments that were able to provide meaningful information regarding students by collaborating cognition and instruction. This call, in part, motivated the introduction of formative assessment (Black & Wiliam, 1998; Wiliam, 2007). The notion of formative assessment was initially based on concept of "mastery learning", in which students do not progress to the next learning objective until they have mastered the current one (Bloom, Hastings, & Madaus, 1971). Recently, the use of formative assessment was expanded to identify a gap between actual student levels and desired levels of performance and to provide information for reducing student weaknesses. For this purpose, Wiliam and Black (1998) stated that an assessment must produce evidence of student levels and elicit performance associated with KSAs at that level. Consequently, they suggest that a combination of cognitive theory of learning, assessment design, measurement models, and curriculum provides the most beneficial information for student learning.

Expressing a similar viewpoint, the NRC proposed the Assessment Triangle containing three vertices: cognition, observation, and interpretation (Pellegrino, Chudowsky, & Glaser, 2001). The Assessment Triangle emphasizes the theoretical and empirical connections among theory, task design, and analytic methods in order to create valid assessment and provide reliable inferences. This notion is also applied to the study of LPs. One of the major challenges in the study of LPs is to develop a suitable framework of linking among *theory* embodied in LPs, *tasks* that provide observable evidence about a student's capability relative to those LPs, and *analytic models* that interpret student performance (Learning Progressions in Science Conference, 2009).

Evidence Centered Design is an assessment design framework that provides guidance for generating tasks that evoke evidence about students' KSAs, and for coherently connecting between theory embodied in an application and task design, and for choosing analytic models that characterize the relationship among them (Mislevy, 2003). This dissertation addresses how the Evidence Centered Design approach helps to solve the challenges in the study of LPs.

Once tasks have been developed, another major issue in the study of LPs is modeling the relation that links student performance on assessment tasks to their levels on the LPs (West, et al., 2009). One of the roles of psychometric models is to characterize the relation between student performance and levels on the LPs. Previously, measurement of proficiency change in accordance with development theory, cognitive psychology, and learning science has been a significant issue in educational and psychological research such as Piaget's (1950) stages of cognitive development, Siegler's (1981) multiple strategies in proportional reasoning ability of children, and Rock and Pollack-Ohls's (1987) math learning as a dynamic latent variable consisting of a series of discrete stages.

Following this trend, various approaches in psychometric models have been proposed for addressing the measurement of proficiency change with different perspectives in terms of (1) focusing on group differences or individual differences, (2) considering proficiency as either a quantitative growth or a qualitative growth, and (3) a sampling issue of a static modeling approach based on cross-sectional design or a dynamic modeling approach by repeated measurement with same students. The modeling approach in the study of LPs differs from the previous transitional development theory-based research of proficiency change in that the modeling LPs requires a highly integrated approach among (1) theory embodied to LPs, (2) developing assessments to elicit student responses relative to an LP, and (3) interpreting student performance relative to their levels on LPs. As such, the study of LPs not only offers the opportunity to explore student progressions in their knowledge and practices over time across a variety of contexts such as classroom based environments and standardized assessment environment, but also provides useful information for designing effective instructional materials that help students develop meaningful engagement in the practices and content over time (Schwarz, et al., 2009).

The particular analytic methods for measuring proficiency change investigated in this dissertation need to be designed with the following distinguishing features that are best matched to learning progression research: (1) observations of student responses to assessment tasks are categorical variables, (2) latent variables are discrete variables with levels representing LPs of qualitative growth, (3) the qualitative growth focuses on proficiency change within the same individual over time in a longitudinal fashion (a dynamic modeling approach), (4) the role of the measurement model formalizes the characteristics of the underlying latent variables of which the observations are indictors, (5) theory provides information about the nature and structure of expected change, and (6) theory and task design provide a theoretical framework for creating and modeling observable evidence.

Some psychometric models matched to these statistical characteristics have been proposed in Latent Class Analysis, Rule Space model (Tatsuoka, 1983), Cognitive Diagnosis models (Leighton & Gierl, 2007), and hidden Markov models (Wiggins, 1955; Collins & Wugalter, 1992). This dissertation introduces Bayesian Inference Networks (BINs) over multiple time points, referred to as dynamic Bayesian Inference Network (DBINs). The BIN is one of the statistical modeling frameworks that have capabilities of modeling proficiency change by integrating substantive theory, designing assessment tasks, and the interpretation of student performances on the assessment (West et al., 2010). Statistically, the BINs offer efficient statistical estimation methods to handle computational challenges arising in longitudinal analyses (Almond, Mislevy, Steinberg, Williamson, & Yan, in progress).

This dissertation contains four parts. The first part (chapter 2) addresses the assessment design framework using the ECD approach. The second part (chapters 3 and 4) describes how BINs can be used to model learning progressions over multiple time points. Specifically, it addresses the questions of how the current, past, and future levels of a student on LPs are inferred. Consequently, these two parts explain the issue of how the BINs can model LPs over time by connecting defined LPs, assessment design, and the interpretation of student performances. The third part (chapter 6 and 7) examines two simulation studies for evaluating the performances of DBINs in the context of an LPs study. The evaluation focuses on how different constraints on the relation between observables and LPs and the relation of the LPs between two consecutive measurement points effect parameter recovery in estimation using Netica software. The research questions related to the two simulation studies are as follows:

Study 1: The first simulation study focuses on a simple DBIN model. The research questions for the first simulation study are as follows:

(1) How well can parameter estimates of conditional probabilities for observable variables be recovered?

(2) How well can parameter estimates of transition probabilities between two latent variables be recovered?

(3) How well can the distribution of students indicating student classification of levels at the first measurement occasion and the second measurement occasion be recovered? This will indicate the classification of students into the levels on an LP.

The factors that will be varied in the simulation are (1) sample size, (2) task size, (3) distributions of the students on the LP at the first measurement occasion, (4) types of transition probability tables, and (5) types of conditional probability tables of tasks. Bias, Root Mean Squared Difference (RMSD), and Standard Deviation of Estimate (SDE) are used for evaluating the parameter recovery.

Study 2: The second simulation study incorporates a covariate for students into the DBIN model (i.e., incorporating a covariate into a transition probability matrix), which would be useful for studying the effect of instructional treatments, students' background, and motivation on student learning progressions. The research questions of the second simulation study are as follows:

(1) How well can parameter estimates of conditional probabilities for observable variables with respect to different values of a covariate be recovered?,

(2) How well can parameter estimates of transition probabilities between two latent variables with respect to different values of a covariate be recovered?, and

(3) How well can the distribution of students indicating student classification in terms of their levels at the first measurement occasion and the second measurement occasion be recovered with respect to the values of a covariate?

The factors that will be varied in the simulation are (1) sample size, (2) task size, (3) types of transition probability tables, (4) types of conditional probability tables of tasks, and (5) proportions of group membership on a covariate. Bias, Root Mean Squared Difference (RMSD), and Standard Deviation of Estimate (SDE) are used for evaluating the parameter recovery.

The last part (chapter 8) carries out an analysis with DBINs using real data from the domain of beginning computer network engineering drawn from work in the Cisco Networking Academy in order to confirm the results of the simulation studies.

As technical aspects of a BIN, the fundamental concepts of the graphical models for constructing a BIN are described. Belief updating is presented from the approach of the junction tree method. Parameter estimation in a BIN (often called "learning" in the BIN and expert systems literature) is presented from the approach of the expectation-maximization (EM) algorithm for Bayesian modal estimates. This part will also show

how this BIN's modeling strategy for proficiency change over multiple time points is equivalent to hidden Markov models (HMMs) (chapters 3, 4, and 6)

CHAPTER 2: LITERATURE REVIEW

This chapter provides a review of assessment design framework to integrate the substantive context of an application, psychometric modeling, and task design. In particular, the sections in this chapter review (1) current perspectives of learning progressions, (2) task designs to elicit observable evidence to the learning progressions, and (3) psychometric models that can be used to identify and accumulate evidence for managing the uncertainty of the relation between learning progressions and student performances.

Assessment Triangle

The National Research Council (NRC) (2001) defined an assessment triangle with three vertices: cognition, observation, and interpretation (Figure 1) that must work coherently in order to develop valid assessments. Cognition is defined as the theory of proficiency and performance that is embodied in the application. Observation is defined as the tasks or situations used to elicit student performance regarding what one desires to measure. The observation activities are related to the design of assessment tasks. Interpretation concerns the mapping of the observations onto cognition. The activity of choosing an appropriate measurement model is related to the interpretation.

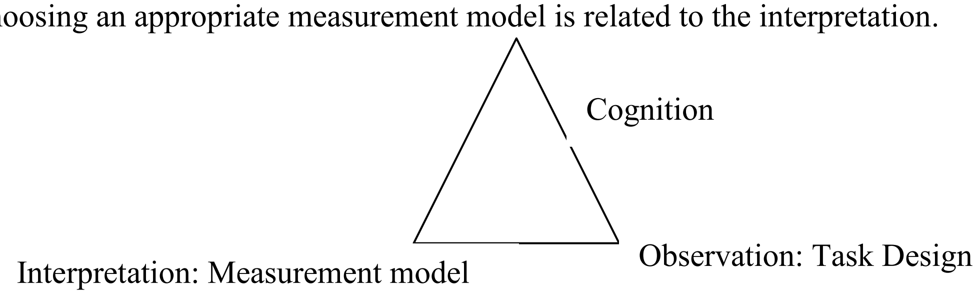


Figure 1**.** Assessment Triangle

One of the major challenges in the study of LPs is to develop a suitable framework of linking among theory embodied in LPs, tasks that provide observable evidence about a student's capability relative to that LPs, and analytic models that interpret student performance (Learning Progressions in Science conference, 2009). The notion of the assessment triangle corresponds to the Evidence Centered Design framework (Mislevy, Almond, & Lukas, 2003). The challenge in the study of LPs can be addressed through the assessment design framework using the ECD approach.

Assessment Design Framework

Assessments developed under the trait psychology perspective have been reliable indicators of the general state of students KSAs when the purpose of measurement was to compare students' abilities and to select students. However, such an overall score probably does not provide sufficient information for the purposes of (1) measuring complex aspects of KSAs and evaluating student learning progressions, (2) understanding distinguishable systematic patterns associated with different characteristics of groups, task features, and ways to solve the tasks, and (3) providing diagnostic information connected with curriculum and instruction. For these purposes, more information is needed for designing assessments and interpreting student performances. The findings from cognitive research have been discussed as improving validity in educational assessment for these purposes by embracing the principles in defining abilities, designing assessment, constructing items/tasks, defining principles for automated scoring, modeling psychometric models to analyze observations, and interpreting the results (Embretson, 2000; Mislevy, 1995; Nichols, Chipman, & Brennas, 1995).

Assessment necessitates containing tasks that reflect aspects of targeted KSAs and their measured structures. At this point, understanding the structure of knowledge, acquisition, and other attributes is essential. Advances in cognitive research provides (1) representations of the structure of knowledge and (2) distinguishable features from expert and novice in perception, procedures, and acquisition, how students are progressing, and different types of learning (Leighton & Gierl, 2007; Mislevy, 1995; Nichols, Chipman, & Brenna, 1995). This information helps to specify the complex aspects of KSAs that are supposed to be measured in assessment, the task key features for distinguishing students, and the rationales for identifying and accumulating evidence from complex data (Mislevy, 2003). In addition to defining the structure of complex KSAs and identifying tasks features, cognitive research helps psychometric models to be meaningfully structured for accumulating evidence by specifying the relation of linking student performance on assessment tasks to theory (Mislevy, 2003). Furthermore, the connection among theory, task design, and analytic method provides information about how students are progressing and where they are having difficulties solving the tasks. The information is useful in the selection of instructional strategies such as re-teaching, utilizing alternative instructional approaches, altering the difficulty level of tasks or assignments, or offering more opportunities for practice (Shute et al., 2009).

The Evidence Centered Design (ECD) framework (Mislevy, Almond, & Lukas, 2003) as a structured assessment design framework guides the incorporation of findings from cognitive research into assessment design, so that (1) all tasks have been generated to provide the opportunity to obtain evidence about the targeted KSAs, (2) the scoring systems are designed to capture the features of student work that serve as evidence about the KSAs, and (3) the characteristics of student in terms of the targeted KSAs are summarized based on evidence (Mislevy, 2003). Since this dissertation introduces the psychometric model given a condition where substantive theory provides a theoretical framework for creating tasks and accumulating evidence, the following section reviews the ECD as an assessment design framework and explains what is meant to explicitly incorporate cognitive theory into the assessment.

Evidence Centered Design (ECD) Framework

The ECD framework is a general assessment framework which supports the notion that an assessment is built upon evidential argument. For an assessment to be considered as an evidential argument, it consists of a series of descriptive models that addresses the following three questions: (1) what complex of knowledge, skills, or other attributes should be assessed?, (2) what behaviors or performances should reveal those constructs, and what are the connections?, and (3) what tasks or situations should elicit those behaviors? (Mislevy, Steinberg, & Almond, 2003). Based on these questions, the ECD framework provides guidance when developing assessments for various purposes as evidentiary argument. Figure 2 shows each stage in the framework (Mislevy, Steinberg, & Almond, 2003).

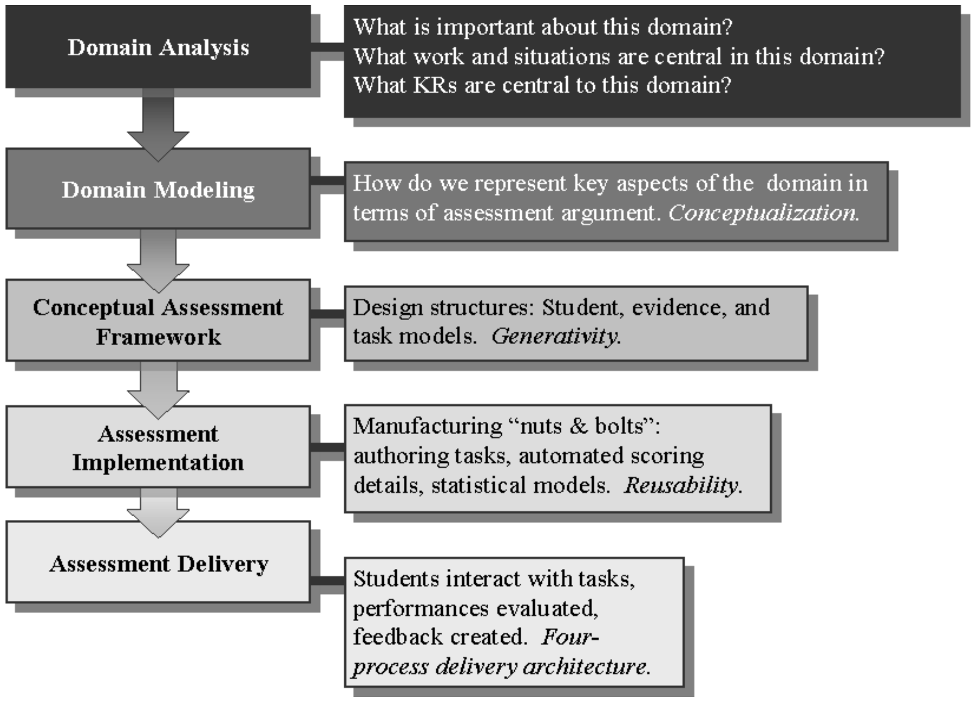


Figure 2. Layers of the Evidence Centered Design Framework

Specially, the Domain Modeling and Conceptual Assessment Framework are closely associated with the facet of incorporating substantive theory into psychometric modeling. In the Domain Modeling, information from analyses of which complex KSAs are central to a domain is organized to form the assessment arguments. Through a tool called Design Pattern, assessment designers can create the substance and structure of an assessment argument. The Conceptual Assessment Framework (CAF) provides technical specifications for operational elements, which explain how the information gathered and organized in domain modeling can coherently serve as evidential arguments while operating the assessment. The CAF specifies five models: (1) the student models, (2) the task models, (3) the evidence models, (4) the assembly model, and (5) the presentation model. Figure 3 shows the CAF model.

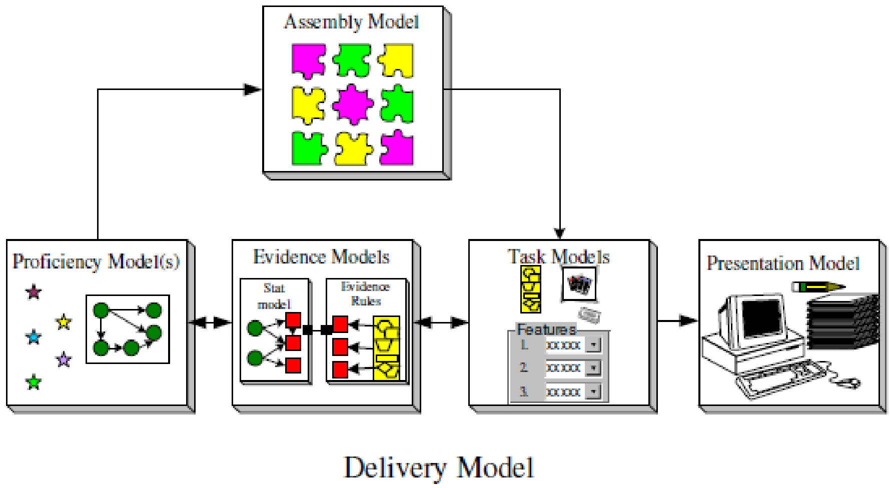


Figure 3. The conceptual assessment framework

The five models address the five questions below (Mislevy, Steinberg, & Almond, 2003).

- What complex of knowledge, skills, or other attributes should be assessed?

  In the student model, aspects of knowledge, skills, or abilities and their configuration are supposed to be specifically addressed. Later, the configuration in the student model will be used as a representation of the variables in a BIN. Since various structures and different levels of complexity in the student model can be constructed, this raises an issue of determining of which set of the student model variables is minimally sufficient to differentiate student performances in terms of the purpose of an assessment. With regard to this, psychological perspectives can offer the rationales involved in constructing the student model variables because different psychological perspectives suggest different notions of knowledge, acquisition, and learning processes. For this purpose, statistical methods such as model fit statistics can also be utilized to determine an adequate structure for the student model. From the two procedures, one can construct a student model that can afford to capture sufficient evidence for the purpose of an assessment.

For LP research, the student model can be built by the findings of developmental theory, learning science, and cognitive psychology. The student model variables can be specified by the aspects of KSAs associated with levels, progresses, stages of learning progressions, and diagnostic information. For example, a student model variable can represent either a LP of a particular domain or a level of a particular LP. The student model contains information about how many levels are useful in defining a LP, what specifications of levels of a LP are defined, and providing information on what evidence is needed to evaluate the LPs of a student. Therefore, the structure of the student model for assessing LPs would be more complex than a traditional assessment with the same observables. This structure can be verified and confirmed by statistical model comparison methods. In the case of the first example of the LPs in chapter 1, the student model can consist of either (1) one student model variable with five classes corresponding to the LP with five levels or (2) five student model variables corresponding to each level of the LP. In any case, the model shows that five levels are defined for assessing the LP in the domain of biodiversity and ecology and contain information regarding what KSAs are required for students at that level. In addition to this information, the structure of the student model variables is also specified at this model. For this LP, a hierarchical relationship among the student model variables could be one of the adequate structures. The student model is connected with the task model through the evidence model, explaining how each observable depends on the student model variables.

- What behaviors or performances should reveal those constructs? How they are connected?

  The evidence model defines how evidence from observables can be identified, accumulated, and linked to the student model variables. It explains the nexus of observables and expectations defined in the student model. The evidence model contains two components: Evidence Rules and Statistical model. The Evidence Rules specify the rules to identify evidence from the work products that a student produced from a particular task. The measurement model explains how evidence is accumulated and synthesized across tasks in terms of student model variables. Various psychometric models such as classical test theory, item response theory models, and cognitive diagnostic models are involved in this part depending on the purpose of an assessment; therefore, one of the issues here is to choose the suitable psychometric model for the purpose of an assessment.

  For the study of LPs, the evidence model provides (1) information about how student performances are modeled and interpreted relative to the level of an LP, (2) information about the criterion for comparing observed and expected LPs, and (3) information about feedback within and across task level. Consequently, the evidence model provides inferential reasoning from observables of tasks and expectations in the student model. For example, in the case of the first LP example in chapter 1, a student response pattern of tasks is used for inferring which levels the student has reached. This dissertation will focus on addressing this area. The DBIN is investigated as one of the suitable psychometric models for modeling LPs.

- What tasks or situations should elicit those behaviors?

The task model provides a set of specifications for the situations, environments, and contexts to elicit student performances to obtain evidence needed for the evidence model. The task model contains presentation material, work product, and task model variables. The presentation material describes the material which is presented to the student. The work products are student performances and responses to tasks. The task model variables are specifications of aspects/features of tasks which are more likely to evoke the desired evidence. They can be varied depending on the targeted KSAs and degrees of difficulty.

For assessing LPs, the task model provides information for developing tasks to elicit student performances relative to the levels of a learning progression. Specifically, it contains the following information: (1) the key features of tasks are important to elicit student's understanding with respect to the targeted KSAs at a particular level of a LP, (2) the key features of tasks which are more likely to distinguish student performances into different levels of a LP, (3) the key features which make a task more or less difficult, (4) other characteristics/contexts of a task that effect its difficulty, and (5) the aspects and features that inform the quality of tasks for assessing LPs. In the case of the first LP example in chapter 1, task designers and domain experts identify what key task features can produce different response patterns among different levels of students. For example, the key features that are able to distinguish between students who understand the concept of biodiversity and those who do not are identified and incorporated into designing tasks to elicit different response patterns.

- How much do we need to measure?

  The assembly model describes how the three models above work together to form a balanced assessment properly reflecting what is needed to be measured.

  For assessing LPs, the assembly model describes how the three models are combined for inferring a student learning progression in a given assessment situation. For instance, the number of tasks (i.e., task size) with respect to the different levels on a LP and the task type are determined to construct an optimized assessment.

- How does assessment look?

> The presentation model describes how a task is provided to students. There are many different means for delivering an assessment such as paper and pencil format, computer and web-based format, and simulation and game-based format. The requirements for presenting assessments differ depending on the format.

In the assessment framework through the ECD model, the substantive theory is explicitly reflected in the development of assessment and psychometric models. The next section discusses the substantive evidence of LPs. The section to follow addresses how tasks are designed for eliciting LPs and what psychometric models are suitable for this purpose.

Substantive Research on Learning Progressions

LPs are descriptions of increasingly sophisticated ways of thinking about or understanding a topic (National Research Council, 2007). The differences from other developmental approaches are that the LPs can be nonlinear progressions, have the possibility to provide diagnostic information about student' progress connected with instructions, and curriculum is closely linked to assessment tasks, curriculum, and instruction (Schwarz et al., 2009). Therefore, the LPs are more integrated concepts for defining aspects of KSAs in the LPs, identifying what levels are addressed by a specific LP, determining how they are to be assessed, how they can provide diagnostic feedback, and identifying how they are linked to instructions and curriculum. The research of LPs offers the opportunity to explore how students build their KSAs over time, what evidence is needed not only for assessing students' learning, but also for evaluating and refining the defined learning progressions, curricula, and instruction (Schwarz et al, 2009).

Specifically in assessment, the LPs are also characterized as measurable and testable pathways that a student may follow in building his other knowledge and gaining expertise over time (National Research Council, 2007; Shin, Stevens, Short & Krajcik, 2009). Although there are many possible pathways where students may progress, common expected natures of paths exist and can be defined (West, et al., 2009). These legitimated pathways are used as grounded concepts of assessing LPs (West, et al., 2009).

LPs research is gaining popularity in the science education and mathematics community (Learning Progressions Science Conference, 2009). For instance, Schwarz et al (2009) developed related to the construction and use of a scientific model in science. The Berkeley Evaluation and Assessment Research (BEAR) Assessment System has been studied in developing LPs in Living by Chemistry (Clasesgens, Scalise, Wilson ,& Stacy, 2009) and Carbon Cycle (Mohan, Chen, & Anderson, in press). Draney (2009) presented LPs in the domains of Living by Chemistry and Carbon Cycle. The study described an integrated Assessment System that provides meaningful interpretations of student performances relative to LPs linked to the cognitive and developmental goals of a curriculum (Draney, 2009). Alonzo and Steedle (2009) have developed LPs in the science content domains of earth science, life science, and physical science. Briggs and Alonzo (2009) have developed an LP for the conceptual understanding of Earth and the Solar System and an associated set of items. The LP describes students' developing understanding of a target idea in earth science according to National Science Education

Standards documents (Briggs & Alonzo, 2009). Table 2 is the example of the LP of Earth and the Solar System.

Table 2

An example of a learning progression of Earth and Solar System adapted from Briggs and Alonzo, 2009

Level	Description
Level0[1]	No systematic understanding of earth and solar system
Level1	Student does not recognize the systematic nature of the appearance of objects in the sky. Student may not recognize that the Earth is spherical.
Level2	Student recognizes that the sun appears to move across the sky every day and the observable shape of the Moon changes every 28 days.
Level3	Student knows that the Earth orbits the Sun, the Moon orbits the Earth, and the Earth rotates on its axis.
Level4	Student is able to coordinate apparent and actual motion of objects in the sky.
Level5	Student is able to put the motions of the Earth and Moon into a complete description of motion in the Solar System.

Note. The description of level 0 has been modified from the original LP so as to describe a progression of students' understanding.

West, et al. (2009) developed a learning progression of IP (Internet Protocol) addressing skills in the field of computer networking.

As the development of assessments, curriculum, and instruction associated with LPs are of interest in various disciplines, challenges arise in many areas, including (1) designing a coherent assessment system, (2) inferring student learning progression levels based on the responses to assessment tasks, and (3) interpreting the difference between expected and observed students' progress mapped to the conceptually defined learning progression. More specifically, a number of inferential challenges of modeling LPs have arisen (Learning Progressions Science Conference, 2009): (1) deciding what methodologies can be used for the inference about students' learning progression levels based on student performance on a set of assessment tasks, (2) determining how students' inconsistent patterns can be explained and modeled, (3) how observed student responses could be compared to expected student responses, (4) understanding how the substance of learning progressions and assessment tasks could be refined by the implications of differences between observed and expected responses, (5) what technologies could be used for explaining student movement along the learning progression with a non-linear sequence of change rather than a simple linear path by longitudinal accounts of student learning beyond the cross-sectional research (Briggs & Alonzo, 2009), and (6) establishing how to model complex LPs to explain different subgroups that have different learning progressions in terms of exogenous variables.

Wilson (2009) did research on assessment structures that one could build to undergird a learning progression through construct map. The research focused on measurement perspective exploring whether to use a traditional form of a uni- or multidimensional model, or to include elements of structural equation modeling, or even more complex ones such as the Structured Constructs model as well as the task design. The majority of this dissertation focuses on psychometric modeling using BIN over multiple time points to address these inferential challenges. Specially, this dissertation will address the challenges of (1), (5), and (6) above. This chapter described conceptual developments of LPs. The next section will illustrate task design that provides evidence about the level of the LP a student may have research.

Task Design of Assessing the Learning Progressions and Validity

Assessment must be thoughtfully designed so as to obtain evidence about the targeted knowledge, skills, and abilities (NRC, 2001). Evidence Centered Design provides guidance for designing assessment tasks to elicit student performances for obtaining evidence about the aspects of KSAs that need to be measured. Particularly, the previous section discussed the idea that the task model in the Conceptual Assessment Framework serves to identify the key features of tasks that allow for distinguishing student performances and constructing the tasks that provide evidence with respect to targeted aspects of KSAs. Through the assessment design framework, tasks are generated to reflect the targeted aspects of KSAs by means of incorporating the identified task features that evoke evidence about the KSAs or the targeted strategies. Therefore, the change in one of the task features can require students to use different KSAs. It is possible to generate isomorphic items by incorporating different features to require students to use the same KSAs.

In a similar view, Embretson (1998) has discussed integrating the principles of cognitive psychology into the assessment design through the Cognitive Design System. Embretson's (1998) Cognitive Design System shared a similar lens to the Conceptual Framework Assessment (Mislevy, 2003) through the task model, the student model, and the evidence model. Embretson (1998) applied task features in terms of different rules in solving the tasks of Raven progression matrix, which identified different rules in solving matrix problems through cognitive research (Carpenter, et al. (1990). They found that the different rules in solving the tasks cause individual differences of working memory capacity and abstraction capacity. Specifically, the rules are as follows (Embretson, 1998): (a) Identify relations: An element is the same across the row or column entries, (b) Pairwise progression relations: An element changes systematically from entry to entry (i.e., size increases across the rows), (c) Figure addition or subtraction relations: The first two entries in the row or column sum to the last entry, (d) Distribution of three relations: An object or attribute appears just once in each row and column, and (e) Distribution of two relations: Distribution of three relations have null values (i.e., one matching element is missing). Based on the findings of Carpenter, et al. (1990), Embretson (1998) identified task features that examinees are required to use different rules in solving a task. She found that using tasks generated by different features can distinguish examinees who use different rules for solving tasks as well as those who have different levels of working memory and abstraction capacity. Embretson (1998) illustrated one example in her study (Figure 4). This task requires the highest level rule, in which examinees need to understand the distribution of three relations and determine which is missing an element in solving the task. The first illustration is missing a diamond shape with two horizontal parallel bars, determined by the previous patterns of two rows and columns. The second one has a missing element of a square with a horizontal bar given the previous patterns. The two tasks require the examinee to use the same rule in completing the tasks while their surface features differ. Also, the two tasks measure the same level of working memory capacity and abstraction capacity.

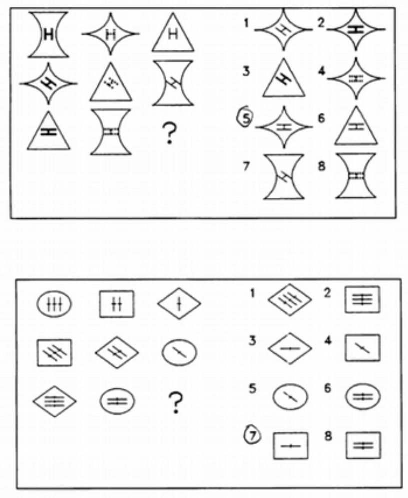


Figure 4. An example of a task in Raven Progressive Matrix

As another example, Figure 5 shows that the task required the use of Rule C. In order to solve both tasks, examinees are required to understand that the first two entries in the row or column sum to the last entry. This rule requires examinees to use a lower level of working memory capacity and abstraction capacity than the first example task. Furthermore, the examples show that it is possible to generate some isomorphic tasks which require the same rules, but appear with different features. In contrast, the tasks can be generated to measure the different rules, but appear with similar features. Through these principles, assessment tasks can be generated by systematically varying and expanding the task stimulus features.

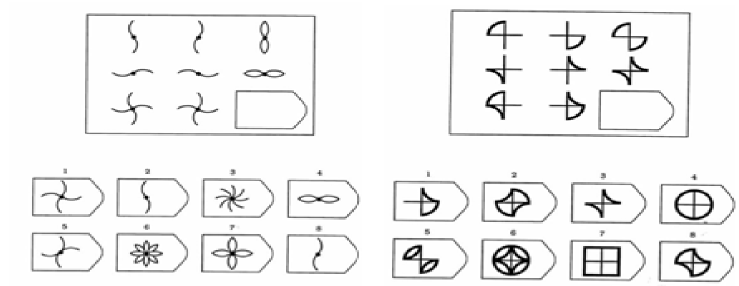


Figure 5. An example of a task in Raven Progressive Matrix

In the study of LPs, by designing assessment tasks that target different levels associated with different aspects of targeted KSAs, it becomes possible (1) to obtain the level of the LPs a student may have attained, (2) to draw conclusions about the value, sequence, and structure of a student's learning, and (3) to gather empirical evidence to guide the development and refinement of the hypothesized LPs associated with assessment and curriculum.

Gotwals, Songer, and Bullard (2009) presented a set of designed tasks that were linked to the LP of inquiry reasoning in order to gather evidence of how students use their content knowledge to formulate scientific explanations associated with a range of ecology, classification, and biodiversity domains. By using the key task features associated with the LP, they can determine which level a student may have attained and what a student knows about the domain of the LPs. For example, the Task in Figure 6 has a scenario for assessing the concept of biodiversity. Given the scenario, two tasks are generated relative to different levels of the LP of biodiversity. Both of the tasks ask the student to provide an answer and the rationale. The first question (Question A) relating to the lower level of the LP, asks students to identify which zone has the highest richness.

The answer and its evidence are straightforward in the task because the Zone B clearly has the highest animal richness. Whereas, the other task (Question B) related to a higher level of the LP is to ask students to identify which zone has the highest biodiversity, given the same scenario. While the answer is the same as the previous task (Zone B), providing appropriate and sufficient evidence of the answer is not as straightforward because students need to understand different concepts between the richness of animals and the abundance of animals (Gotwals, Songer, & Bullard, 2009).

This table shows school yard animal data collected using CyberTracker. Use the table to help you answer the question.

School Yard Animal Data

Animal Name	Zone A	Zone B	Zone C
Pillbugs	1	3	4
Ants	4	6	10
Robins	0	2	0
Squirrels	0	2	2
Pigeons	1	1	0

Question A	Question B
Which zone has the highest richness, given the same scenario?	Which zone has the highest biodiversity, given the same scenario?
Make a Claim: Write a sentence that answers the scientific question. Answer:	Make a Claim: Write a sentence that answers the scientific question. Answer:
Give your Reasoning: Write the scientific concept or definition that you thought about to make your claim Answer:	Give your Reasoning: Write the scientific concept or definition that you thought about to make your claim Answer:

Figure 6. An example of a task taken from Gotwals, Songer, and Bullard (2009)

In another example, West, et al. (2009) identified the features of tasks relative to different levels of the LP in IP Addressing Skills. Two tasks in Figure 7 require different level of IP Addressing skills to obtain a correct answer. The two tasks look similar on the surface, but the stem of Task A is /24, while that of Task B is /28. This change requires students to use a more advanced IP addressing skill (West, et al., 2009).

Task A	Task B
It is necessary to block all traffic from an entire subnet with a standard access control list. What IP address and wildcard mask should be used in the access control list to block only hosts from the subnet on which the host 192.168.16.43/24 resides?	It is necessary to block all traffic from an entire subnet with a standard access control list. What IP address and wildcard mask should be used in the access control list to block only hosts from the subnet on which the host 192.168.16.43/28 resides?
A.192.168.16.0 0.0.0.15	A.192.168.16.0 0.0.0.15
B.192.168.16.0 0.0.0.31	B.192.168.16.0 0.0.0.31
C.192.168.16.16 0.0.0.31	C.192.168.16.16 0.0.0.31
D.192.168.16.32 0.0.0.15	**D.192.168.16.32 0.0.0.15
E.192.168.16.32 0.0.0.16	E.192.168.16.32 0.0.0.16
**F.192.168.16.0 0.0.0.255	F.192.168.16.0 0.0.0.255

Figure 7. An example of a task taken from West, et al (2009)

If the purposes of an assessment are to provide evidence about the level of an LP where a student may have reached, the learning trajectory of a student over time, and diagnostic feedback relative to a student LP beyond a general proficiency of a student, it is important that assessment tasks are designed to cover students with a variety of ability levels. Appropriate and sufficient evidence with respect to all levels in LPs can be obtained by incorporating task features, given that the task features are associated with the key aspects that students require certain levels of understanding to complete the task.

The previous two sections explained the theory embodied LPs and task design. The next section emphasizes the role of analytic models that characterize the connections between theoretical aspects and empirical evidence from student performances based on assessment tasks. Statistical models such as latent class models, Rule Space model, cognitive diagnostic models, and hidden Markov models are useful tools for reasoning from observed change patterns to expected change patterns on student performances. The next section discusses several psychometric models that are useful for analyzing data for LP research.

Psychometric Modeling

Many psychometric models have been proposed for measuring change over time in latent variables. The psychometric models can be distinguished by conceptual differences between (1) quantitative growth and qualitative growth, (2) static and dynamic latent variables, and (3) a cross-sectional sample design approach and a longitudinal sample design approach. Proficiency change as a continuous variable is often expressed as quantitative growth, modeled by means of latent growth curve approaches (e.g., McArdle & Hamagami, 1991; Willett & Sayer, 1994, 1996). In this case, the quantitative growth can be defined in terms of an increase or decrease in the amount of knowledge or ability. In contrast, movement between stages or stage sequential change is often described by qualitative growth (Collins & Flaherty, 2002). A typical example of the qualitative growth is Piaget's model based on the cognitive development of children. The qualitative growth focuses on the critical pinpoints that represent a qualitatively different way of thinking and doing (Collins & Flaherty, 2002). Collins (1996) defined the conceptual distinction between static and dynamic latent variables. The static latent variables are not expected to change over time, or, put another way, the change is not of interest, rather group differences at a particular time are of interest in static modeling of latent variables. On the other hand, dynamic latent variables examine change in systematic over time (Collins & Flaherty, 2002). Depending on the interests and research questions, the same variable can be often seen as either static or dynamic. The distinction between them is often drawn by what kinds of the sample design approaches are used such as a cross sectional sample design and a longitudinal sample design (Collins & Flaherty, 2002). The modeling strategies and the selection of a suitable model can be determined by (1) a careful consideration of substantive theory, (2) what are observed from the assessment tasks, (3) the purpose of an assessment, and (4) the desired level of precision at which student characteristic.

This dissertation will focus on psychometric models that are best matched to LP research. Specifically, since proficiency change aligned with theory embodied learning progressions over time is of interest, the models considered here have three key properties. (1) Observations are student responses to assessment tasks, so observables are categorical variables. (2) An LP is operationalized as a latent variable with several latent classes representing qualitatively different levels in the learning progression, and (3) psychometric models make inferences about latent level change on an LP over time when the variable of interest is unobservable, but task design and theory provide a theoretical framework for creating and modeling observable evidence as well as information about nature and structure of expected change. The next subsections review key terminologies of latent class models, Rule Space models, diagnostic classification models, and hidden Markov models.

Latent Class Models

Latent class models are statistical methods often used to identify homogenous subgroups and to differentiate heterogeneous subgroups by their responses to dichotomous or polychromous items. Through the analysis of the models, students are identified as homogenous groups with respect to membership in a latent class, while they are identified as heterogeneous groups between latent classes. The latent class models are also referred to as finite mixture models (McLachlan & Peel, 2000) or unrestricted/unconstrained latent class models. In latent class models, observations are used to estimate the probability of class membership for the latent class variables and the probability of responses to an item given the latent class membership. The models assume that the probability of a particular response on any one item is independent of the probability of any given response on any other item after conditioning on latent class membership; this is known as the local independence assumption (Lazarsfeld & Henry, 1968). The general forms for the latent class models are as follows:

Assuming that there are C classes with levels $c = 1, \ldots, C$, the probability of latent class membership has two constraints:

$$\sum_c \pi_c = 1 \quad , \quad 0 < \pi_c < 1 \quad \text{for all c} \tag{1}$$

Then, conditional probability of responses to item *i* for student j given latent class c is as follows:

$$P(X_{ij} = 1 \mid C = c) = P_{i|c} \quad and \quad P(X_{ij} = 0 \mid C = c) = 1 - P_{i|c} \tag{2}$$

Assuming local independence within each latent class, the conditional probability of response pattern given latent class c is as follows:

$$P(X_j \mid C = c) = \prod_i p_{i|c}{}^{x_{ij}} (1 - p_{i|c})^{1-x_{ij}} \tag{3}$$

Therefore, marginal probability of response pattern X*j* is as follows:

$$P(X_j) = \sum_c \pi_c \prod_i p_{i|c}{}^{x_{ij}} (1 - p_{i|c})^{1-x_{ij}} \tag{4}$$

The latent class models have three general constraints: (1) the parameters are non-negative, (2) the mixing proportions should sum to 1, and (3) the conditional probabilities also should sum to 1 for each item within each latent class. In terms of learning progressions research, since a learning progression is a categorical latent variable, the latent class models can be directly applied by considering each latent class as a level on a learning progression. There are different modeling strategies by combining the latent class models with other models. One of the modeling strategies is to explicitly incorporate variables that characterize substantive features of tasks. Specifically, cognitive diagnostic modeling, often referred to as cognitive diagnostic models (Leighton & Gierl, 2007), or diagnostic classification models (Rupp, Templin, & Henson, 2010), has taken the strategy that makes them possible to provide diagnostic information.

Another modeling strategy is combining Markov chain model with latent class model, referred to as the hidden Markov models, latent Markov models (Wiggins, 1955), or latent transition model (Collins & Wugalter, 1992; Graham et al.; Hansen, 1991). These models are helpful for analyzing a series of latent class models over multiple time points. The next section will discuss the key terminology of a unified model (Henson, Templin, & Willse, 2009) of diagnostic classification models, Rule Space model, and hidden Markov models.

Diagnostic Classification Models

Diagnostic classification models are statistical models that were developed to classify students in terms of their mastery states on each attribute (DiBello, Roussos, Stout, & Junker, 2007; Rupp & Templin, 2008; Rupp, Templin, & Henson, 2010). The Diagnostic Classification models contain multiple attributes. The term attribute refers to latent aspects of knowledge, skills, and abilities that are supposed to be measured in an assessment. Student mastery states on the attributes of interests are estimated based on students' observed response patterns. A composite of the student mastery states on the attributes is referred to as an attribute profile (Rupp et al., 2010). Therefore, the attribute profile is a pattern used for providing diagnostic feedback. Many models such as Deterministic inputs, noisy and gate (DINA: e.g., Junker & Sijtsma, 2001; Templin & Henson, 2006), Deterministic inputs, noisy or gate (DINO: e.g., Junker & Sijtsma, 2001; Templin & Henson, 2006), and reparameterized unified model in DCMs have been proposed. These models differ depending on what variables are of interest and which condensation rules are used for modeling attributes; however, a central concept of modeling is linking findings from cognitive psychology (Rupp & Templin, 2008). Since there is more than one attribute involved and tasks can depend on multiple attributes, their relations are represented by a loading structure, often called a Q matrix (Tatsuoka, 1990). The Q matrix contains the targeted attributes and specification of which attributes are measured by which task(s) based on substantive theory. In the case of an LP, the Q matrix can be constructed as hierarchy attribute relationship along different levels in an LP. Table 3 shows an example of the Q matrices with 8 tasks for the study of LPs.

Table 3

Exemplary Q-matrix in the case of a learning progression with four levels

Task	Attribute1 Required KSAs at level 1	Attribute 2 Required KSAs at level 2	Attribute 3 Required KSAs at level 3	Attribute 4 Required KSAs at level 4
1	1	0	0	0
2	1	0	0	0
3	0	1	0	0
4	0	1	0	0
5	0	0	1	0
6	0	0	1	0
7	0	0	0	1
8	0	0	0	1

Rows indicate tasks, columns correspond to attributes, and values indicate which attributes are measured by which tasks. Task 1 and Task 2 measure the KSAs required to be at level 1 in the LP. In other words, aside from slips, mastery of the KSAs is required to correctly answer Task 1 or Task 2. To construct a Q matrix, many sources may be used. In the case of LPs research, cognitive developmental theory, learning science and objectives about learning in curriculum can be sources for specifying Q matrix (Buck & Tatsuoka, 1998).

The general model for DCMs is as follows (Rupp et al, 2010):

$$P(X_r = x_r) = \sum_C \pi_c \prod_I p_{ic}^{x_{ir}} (1 - p_{ic})^{1 - x_{ir}} \qquad (5)$$

Different DCMs provide different parameterizations of p_{ic} based on the relation between tasks and attributes, and among attributions (Rupp et al., 2010). A unified model referred to as the *log-linear cognitive diagnosis model* (LCDM) framework (Henson, Templin, & Willse, 2009) can capture the different DCMs such as DINA, DINO, and RUM. The LCDM is as follows (Rupp et al, 2010):

$$P\left(Y_{ij} = 1 \mid \alpha_i, q_j\right) = \frac{\exp\left[\lambda_{0j} + \lambda_j^{'} h\left(\alpha_i, q_j\right)\right]}{1 + \exp\left[\lambda_{0j} + \lambda_j^{'} h\left(\alpha_i, q_j\right)\right]} \qquad (6)$$

where *i* and *j* denote student and task, respectively; λ_{0j} is an intercept and $\lambda_j^{`}$ represents a vector of coefficient indicating the effects of attribute mastery on the response probability for item *j*; and $h(\alpha_i, q_j)$ is a set of linear combinations of α_i and q_j. The intercept can be interpreted as a *guessing parameter* and λ_{ju} parameters represent the main effects of each attribute *u* on the response probability for item *j*, and the λ_{juv} parameters represent the two-way interaction effects of the combination of the mastery states of attributes *u* and *v* on the response probability for item *j*. Depending on the number of attributes included in the assessment, the LCMD can have different number of main effects, two-way, and three-way interactions effects. For example, suppose that there is a task associated with two attributes. Various structures between tasks and attributes that affect the probability of correctly response to the task can be modeled. The first case could be a situation where a student can get correct answer to the task when both attributes have been mastered. In this case, the response probability for this task is modeled computed by reducing the LCMD only with interaction effects as follows (Rupp et al, 2010):

$$P(Y_{j1} = 1 | \alpha_i, q_j) = \frac{\exp(\lambda_{10} + \lambda_{112}\alpha_1\alpha_2)}{1 + \exp(\lambda_{10} + \lambda_{112}\alpha_1\alpha_2)} \quad (7)$$

This model is referred to as the DINA model because the DINA model reflects a case where the mastery of attributes cannot compensate for the lack of mastery of any other attribute(s) (de la Torre, 2008; Junker & Sijtsma, 2001). The second case could be a situation where a student can have a correct answer to the task when one of the attributes has been mastered, which is referred to as the DINO model. The DINO model can be modeled by the three LCMD models as follows (Rupp et al, 2010):

$$P(Y_{i1} = 1 | \alpha_i, q_j) = \frac{\exp(\lambda_{10} + \lambda_{11}\alpha_1)}{1 + \exp(\lambda_{10} + \lambda_{11}\alpha_1)} = \frac{\exp(\lambda_{10} + \lambda_{12}\alpha_2)}{1 + \exp(\lambda_{10} + \lambda_{12}\alpha_2)} = \frac{\exp(\lambda_{10} + \lambda_{11}\alpha_1 + \lambda_{12}\alpha_2 + \lambda_{112}\alpha_1\alpha_2)}{1 + \exp(\lambda_{10} + \lambda_{11}\alpha_1 + \lambda_{12}\alpha_2 + \lambda_{112}\alpha_1\alpha_2)} \quad (8)$$

Since the DINO model reflects the assumption that mastery of subset of attribute(s) can compensate for the lack of mastery of other attribute(s), the LCMD models can be modeled with only main effects, only interaction effects, or both of them (Rupp et al, 2010). Lastly, the compensatory RUM can be also specified by the LCMD model by considering a situation where the probability of getting a correct response to the task increase as the number of attributes mastered increases. The compensatory RUM model in the LCMD framework is as follows (Rupp et al, 2010):

$$P(Y_{i1}=1|\alpha_i,q_j)=\frac{\exp(\lambda_{10}+\lambda_{11}\alpha_1+\lambda_{12}\alpha_2)}{1+\exp(\lambda_{10}+\lambda_{11}\alpha_1+\lambda_{12}\alpha_2)} \tag{9}$$

In the case of an LP research, depending on the structures among LPs and relationship between LPs and tasks, DCMs can be modeled as special cases of DINA, DINO, or RUM.

Rule Space Model

Rule Space model is a statistical method for classifying students into one or more pre-specified attribute mastery profiles (K. Tatsuoka & M. Tatsuoka, 1990; K. Tatsuoka, 1993). The attributes are associated with different cognitive skills, processing skills, strategies, and knowledge components in order to successfully complete tasks. The model is used to assess whether a student has mastered the cognitive skills or attributes required to solve tasks, to diagnose a student's misconception, and to provide meaningful information to guide instruction. For these purposes, the Rule Space model explicitly incorporates cognitive theory into designing tasks and classifying student responses. Mainly, the Rule Space model addresses two issues: (1) the identification of task features and task design by incorporating the task features and (2) statistical analysis for classifying student responses into the patterns.

An adjacency matrix, reachability matrix, and incidence matrix are constructed in the task design phase. The adjacency matrix expresses the direct relation between attributes. The reachability matrix specifies the indirect as well as the direct relation among attributes. Hence, the reachability matrix is directly used for constructing the incidence matrix, often referred to as Q matrix, which represents each task by the attributes being assessed. Since the initial Q matrix constructed by the reachability matrix represents all combinations of the relations between tasks and attributes, the Q matrix sometimes can be reduced or constrained depending on a particular relation among attributes. In general, the curriculum specialists, domain experts, and cognitive researchers provide the specification of the attributes in a specific content area. In the statistical classification phase, under the assumption that the relations among attributes are true, ideal item response patterns can be produced by using the Q matrix.

Furthermore, the ability continuum based on one dimensional IRT analysis is derived using the ideal item response patterns. As such, it can be expected that a student having high value of θ would have an ideal item response pattern with many 1s and few 0s, while a student having low value of θ would have an ideal item response pattern with many 0s and few 1s as well as it can be reported by values located on the IRT scale. On the other hand, the observed student response patterns are often subject to fluctuations. That is, there are some cases in which students of high ability get easy items incorrect or students of low ability get hard items correct. These inconsistent patterns are called response unusualness, referred to as ζ in the Rule space. Both the ideal response patterns and the actual student response patterns are then plotted on a two-dimensional Cartesian coordinate system, called the Rule Space, characterized by the θ (the ability continuum derived from an item response analysis) and ζ (response unusualness) (Tatsuoka, 1995). In order to classify a student response pattern into one of the ideal response patterns, the Mahalanobis distance is computed between the ideal response patterns and the observed student response pattern. The observed student response pattern is classified to the ideal response pattern that embraces the student's point in the smallest value of the Mahalanobis distance.

In the case of LP research, relations among the attributes could be structured as a liner hierarchy order based on learning paths. Figure 8 indicates a possible hierarchy attributes for an LP study. In Figure 8, attribute 1 is prerequisite to attribute 2, attribute 2 is prerequisite to attribute 3, and attribute 3 is prerequisite to attribute 4.

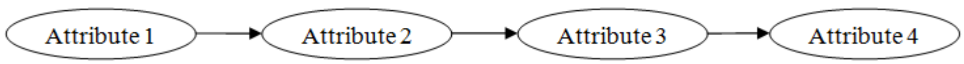


Figure 8. A hierarchical relation

This hierarchical relation can be expressed in each row of the adjacency matrix

Table 4

Exemplary Adjacency matrix in the case of a learning progression with four levels

	A1* Required at L1*	A2* Required at L2*	A3* Required at L3*	A4* Required at L4*
A1* Required at L1*	0	1	0	0
A2* Required at L2*	0	0	1	0
A3* Required at L3*	0	0	0	1
A4* Required at L4*	0	0	0	0

Note. A1* indicates attribute 1. A2* indicates attribute 2. A3* indicates attribute 3. A4* indicates attribute 4. L1* indicates level1, L2* indicates level2, L3* indicates level3, and L4* indicates level4.

The linear hierarchy structure is one of the attribute structures that are addressed in Attribute Hierarchy Method (AHM: Leighton, Gierl, & Hunka, 2004). The AHM is a variation of Rule Space model and a psychometric method for classifying examinees' test item responses into a set of structured attribute patterns associated with different components from a cognitive model of task performance. The structured attribute patterns are varied in terms of different types of hierarchical structures (Figure 9).

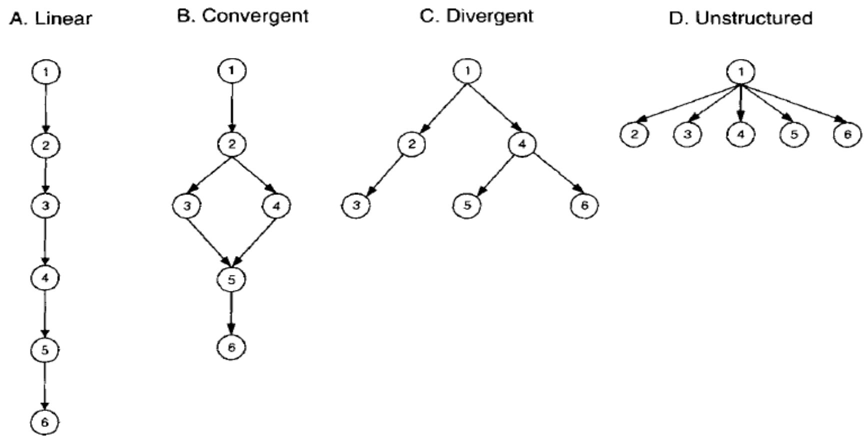


Figure 9. Four hierarchical structures using six attributes

The AHM addresses the issue that the cognitive attributes that describe the problem solving process are not isolated pieces, but rather working in a certain hierarchical related mode . Similar to the Rule Space model, the AHM has two stages including the structure identification and the statistical classification analysis. Once the hierarchical structures are identified, the fit of the hierarchy is evaluated relative to the actual student response data from the random sample, and then the attribute probabilities are computed in order to provide examinees with specific information about their attribute patterns. Therefore, as in other cognitive diagnosis models, the validity of the results of statistical classification of Rule Space model and the AHM depends on how correctly the structure of attributes is identified, how well each task is generated based on the Q matrix, and the amount of error in the student's responses (Birenbaum, Kelly, & Tatsuoka, 1993).

Most research on DCMs and the Rule Space model has focused on the classification at a given time point, the movements from one attribute at one point in time to others at the next point in time are not of interest. In other words, transition proportions of skills, levels, and strategies between consecutive measurement time points are not explained. Consequently, DCMs and Rule Space model cannot be directly applied for modeling learning progressions over multiple time points. The models would need to be extended by connecting a series of cross-sectional latent class models or combining multiple adjacent matrices over multiple time points. In other words, multiple latent class models along consecutive measurements may be linked so as to capture as closely as possible change over time. Markov chain models can describe transition proportions of latent classes between consecutive time points. The explanation of Markov chain models is provided in the next section.

Markov Chain Models

Multiple latent class models can be linked to make the statements about what happens from the first measurement to the next measurement. The core statistical models for the study of change in qualitative status over time are Markov chain models. The Markov chain models have been applied in situations such as attitude change, learning, cognitive development, and epidemiology (Langeheine & van de Pol, 2002). Variations of the Markov chain models (i.e., hidden Markov models, mixed hidden Markov models, and mixed hidden Markov models with several groups) have been proposed (Langeheine & van de Pol, 2002). The models concern modeling change over time in observed categorical variables by using transition probabilities. The transition probabilities of the all models can be constrained to be the same for all time point changes. Central to these models is the Markov property. The first-order Markov property assumes that only the status on the most recent occasion is important for predicting the present status (Langeheine & van de Pol, 2002). A simple Markov model for three time points is specified as follows:

$$P_{ijk} = \pi_{i^1} \ \tau_{j^2|i^1} \tau_{k^3|j^2} \tag{10}$$

where π_{i^1} indicates the observed initial marginal distribution at Time 1; $\tau_{j^2|i^1}$ is the observed transition probability for a transition from Time 1 to Time 2; and is the observed transition probability for a transition from Time 2 to Time 3.

In the context of LPs research, we can consider a situation where student levels on an LP are measured at three discrete time points. Since the Markov chain model assumes all variables are observable, an LP with some levels is an observed categorical variable in the Markov chain model. The subscripts, *i, j*, and *k*, are the manifest levels in a learning progression for Times 1, 2, and 3; hence, if four levels are specified in a learning progression, *i* = 1, 2, 3, and 4, *j* = 1, 2, 3, and 4, and *k* = 1, 2, 3, and 4 from Time 1 to Time 3, respectively. π_{i^1} is the observed probability of students at Time 1 who are at levels 1,2,3, and 4, which correspond to the initial marginal distribution of the levels. $\tau_{j^2|i^1}$ and $\tau_{k^3|j^2}$ are the transition probabilities between two consecutive time points. $\tau_{j^2|i^1}$ represents the observed transition probabilities from Time 1 to Time 2 for those in level *j* at Time 2, given that they were in level *i* at the Time1; hence, the transition probabilities contain information of movements from a certain level at Time 1 to other levels at Time 2. Similarly, the $\tau_{j^2|i^1}$ and $\tau_{k^3|j^2}$ indicate the transition probabilities from Time 2 to Time 3 for those in level *k* at Time 3, given that they were in level *j* at the Time 2.

Since the data of interest in the Markov chain model are manifest observed responses, the Markov model assumes that the responses are measured without error. This may not be a reasonable assumption in psychological and educational research. Extensions of the Markov chain model by incorporating the notion of latent class analysis have been proposed in order to account for measurement error (Langeheine & van de Pol, 2002). In these extensions, the classification of students is based on a hypothesized latent structure. The appropriateness of the hypothesized latent structure can be examined through the measures of model fit to the observed responses. In the context of LPs research, students' observable responses to assessment tasks are not perfectly reliable measures of students' latent levels on LPs. Instead, the observable responses serve as the indicators to make inferences of students' learning progressions. For example, there could be some situations where a student with a high level on an LP gets an incorrect answer to an easy task, while a student with a low level of an LP gets a correct answer to a hard task. The incorporation of the notion of latent class analysis to Markov chain models addresses these inconsistent patterns, often referred to as measurement error.

Hidden Markov Model

The hidden Markov model combines certain features of latent class model and those of a simple Markov chain model. The model is also referred to as latent Markov models as proposed by Wiggins (1955) or a Latent Transition Analysis (Collins & Wugalter, 1992). Wiggins (1955) proposed the latent Markov Model. This model has been applied for identifying unobservable latent state change such as strategies, levels, and skills by analyzing observable student responses at each point in time. For three time points, the model is specified as follows:

$$P_{ijk} = \sum_{c}^{A}\sum_{b}^{B}\sum_{c}^{C}\pi_{a^1}\ \rho_{i^1|a}\ \tau_{b^2|a^1}\ \rho_{j^2|b}\ \tau_{c^3|b^2}\ \rho_{k^3|c} \tag{11}$$

where P_{ijk} is the model expected probability that student may belong to *i*, *j*, and *k* categories in the variable of interest at time points 1, 2, and 3; π_{a^1} is the latent initial marginal distribution with respect to latent states at Time1; $\rho_{i^1|a}$ refers to the response probability associated with category *i* given latent state *a* at Time 1; $\tau_{b^2|a^1}$ is the transition probability for transitions from Time 1 to Time 2.

The subscripts, *a*, *b*, and *c*, indicate latent states at Time 1, Time 2, and Time 3, respectively. The subscripts, *i*, *j*, and *k*, refer to responses to a categorical variable. The response probabilities in this model serve to take measurement error into account. In the case of LPs study, an LP with some levels is considered as a latent variable with some classes. The subscripts, *a*, *b*, and *c*, are the latent levels in the LP for Times 1, 2, and 3. π_{a^1} refers to the latent initial marginal states in the learning progression at Time 1. $\rho_{i^1|a}$

indicates conditional probabilities of response, given the latent levels in the LP at Time 1. $\tau_{b^2|a^1}$ refers to the transition probabilities from the latent levels at Time 1 to others at Time 2.

Latent transition analysis (Collins & Wugalter, 1992) captures the same notions of the hidden Markov model. There is no fundamental difference between the latent transition analysis and the latent Markov model. The difference is that the latent transition analysis incorporates multiple indicators at each time point into the model and expresses all Time t points while the hidden Markov model incorporates a single indicator at each time points.

The main issue of the family of Markov chain models is modeling the transition probabilities between at consecutive time points (Rost, 1989). Theory-based information can be incorporated into the transition probability matrix by imposing some constraints. For example, in the LPs research, forward movements where students always learn or stay can be considered in order to explain transitions over time. The transition probability matrix for the forward movements can be modeled by imposing a constraint that all transition probabilities in backward movements to be zeros. Different patterns in transitions matrix depending on substantive theory can be modeled by imposing some restrictions such as (1) sets of transition probabilities must to be particular values, or (2) must be equal to each other.

Hidden Mixed Markov Model

Manifest variables (e.g., different instructions, interventions, and individuals' demographic background) and latent exogenous variables (e.g., attitude, intelligence, and other abilities) can impact the status change over time (Langeheine & van de Pol, 2002). The latent Markov model can be extended to explain the impact of manifest or latent exogenous variables on change (Van de Pol & Langeheine, 1990). In the context of LP study, a manifest/latent variable (e.g., instructions, curriculum, and attitude) can influence student learning progressions over time. The extended model can explain the effectiveness of the manifest/latent variable on a student learning progression over multiple time points. As compared with the Hidden Markov model, the Hidden Mixed Markov model contains an additional person parameter for group membership in the form of a manifest/latent exogenous variable. Also, all other sets of parameters are conditional on the group membership (Collins & Wugalter, 1992; Chung, Walls, & Park, 2007; Langeheine & van de Pol, 2002). For three time points, the model can be specified as follows:

$$P_{ijk} = \sum_{g}^{G}\sum_{c}^{A}\sum_{b}^{B}\sum_{c}^{C}\delta_g \pi_{a^1|g}\ \rho_{i^1|ag}\ \ \tau_{b^2|a^1g}\ \ \rho_{j^2|bg}\ \ \tau_{c^3|b^2g}\ \ \rho_{k^3|cg} \tag{12}$$

where P_{ijk} is the model expected probability that student may belong to i, j, and k categories in the variable of interest at time points 1, 2, and 3; δ_g is the proportion of group membership in a manifest/latent variable. The model shows that all other parameters are conditional on the group membership.

Similar to the Hidden Markov model, the subscripts, *a*, *b*, and *c*, indicate the latent states at Time 1, Time 2, and Time 3. The subscripts, *i*, *j*, and *k*, refer to student responses in a categorical variable. All responses probabilities and transition probabilities are conditional on the group membership of the manifest/latent variable as well as the latent classes at each time point. In the case of LPs research, a learning progression with some levels is considered as a latent variable with some classes. The subscripts, *a*, *b*, and *c*, are the latent levels in a learning progression for Times 1, 2, and 3. $\pi_{a^1|g}$ refers to the latent initial marginal states in a learning progression at Time 1, given the group membership of the manifest/latent variable. All conditional responses probabilities, $\rho_{i^1|ag}$, $\rho_{j^2|bg}$, and $\rho_{k^3|cg}$ are conditional on the latent levels in the learning progression at each time point and the group membership of the manifest/latent variable. Similarly, transition probabilities and $\tau_{c^3|b^2g}$ are conditional on the latent levels in the LP at previous time point and group membership of the manifest/latent variable. The parameters being estimated here are the latent initial probabilities, conditional probabilities of responses, and transition probabilities. Later, in the Bayes nets, the parameters being estimated are probabilities in Bernoulli and categorical distributions, specially, probabilities in tables of marginal and conditional probabilities.

CHAPTER 3: A BAYESIAN INFERENCE NETWORKS

This chapter introduces a Bayesian Inference Network as a psychometric modeling method for measuring student status in an LP. The BINs are different from the psychometric models that were previously mentioned, in that the BIN is a probability based statistical modeling framework, instead of a specific statistical model (Rutstein & Mislevy, in press); hence it is a very flexible statistical modeling method. A BIN represents the probabilistic relations among variables by means of probability theory and graphical models (Almond et al, in progress). Because the BINs are a flexible modeling framework, modeling BINs comes with more decisions: definitions of variables, relations among them, and reasoning from observations of the variables.

Depending on sampling design, there may be two modeling approaches under the BINs for the case of LPs research. A static modeling approach using a cross sectional sampling design focuses on inferences of student levels on LPs at a given measurement time point. Therefore, this approach could provide an interesting view of group differences with different abilities under the ordered latent variable representing the LP. For example, students are provided with an assessment that consists of the tasks that measure different levels on LPs. Some of the tasks were constructed by using the distinct features that can allow students to use KSAs at level 1, some of the tasks constructed by using the distinct features that can allow students to use KSAs at level 2, and so on. Given the one administered assessment, the levels of students under the ordered LP can be measured.

In contrast, a dynamic modeling approach using repeated measurement is aimed to make inferences of level change through an LP over multiple time points. The explanation of BINs starts with the static modeling approach. The sections in this chapter review the (1) fundamental notion of probability based reasoning and (2) key terminology and concepts of BINs. The next chapter will move toward to the dynamic modeling approach with BIN for measuring a student's level of change in an LP over multiple time points, referred to as DBINs.

Probability Based Reasoning

It is difficult or sometimes even impossible to construct a model covering all aspects of real world situations. Rather, a model can be constructed as a simplified representation focused on certain key aspects of real world situations (Ingham & Gilbert, 1991). Modeling real situations is a process of building a coherent system by extracting distinct features of the real world situations and constructing their relations (Ingham & Gilbert, 1991). Mislevy (2009) described model based reasoning in terms of how reconceiving a real world situation can be constructed through a model. Specifically, it explains not only how a representational system of complex real world situations is constructed with distinct entities and their relations, but also how the system deals with uncertainty in explaining, predicting, and inferring for real world situations. Since the model is a simplified representational system, there may not be an exact correspondence between the real world entities and the idealizations in the model (West et al, 2010). That is, a student's performance across different tasks may provide inconsistent patterns compared to the idealized patterns from a model. Probability theory is a prevailing method for dealing with the inconsistent patterns (Kjaerulff & Madsen, 2007). BINs are a probability based statistical modeling framework for reasoning and making decisions with uncertain and inconsistent patterns. In the case of LPs, student performances on assessment tasks are a particular real world situation. A model is built to explain student levels and level change on LPs, given an expected performance. Modeling LPs is a process of mapping between student performances and LPs. There may not be an exact correspondence between student responses to assessment tasks and the expected responses from the model. For example, there may be some students who have reached a high level in a LP, yet may get incorrect answers to a task that requires only skills on a low level of the LP, or some students who have reached only a low level in a LP, yet may get correct answers to a task that requires skills on a high level of the LP. BIN is a probability based reasoning framework that allows us to manage these problems of uncertainty and inconsistent patterns. In addition to this, it technically provides a compact representation and an efficient method for gathering evidence from data (Kjaerulff & Madsen, 2008).

There are some requirements in order for BINs to be reasonably modeled, (Kjaerulff & Madsen, 2008). First, variables and their possible values in a BIN must be well defined. Secondly, information about the structure of the variables must be available, so that the structure can be identifiable. Thirdly, there must be uncertainty associated with problem such as measurement error. Therefore, modeling LPs using BINs can be more valuable when integrating the work of the development of LPs, task design, and the interpretation of student performances relative to their levels on the LPs. The following section will consider basic terminology, fundamental concepts, and a graphical representation of BINs.

Bayesian Inference Networks: Graph Theory and Graphical Model

Bayesian Inference Networks combine probability theory and graph theory to represent the probabilistic relations among variables under uncertainty. To facilitate an efficient representation of complex situations with many variables, the BINs use a graphical representation to represent dependence and independence relations among the variables (Kjaerulff & Madsen, 2008). The graphical representation is based on a finite acyclic directed graph (DAG). The acyclic directed graphs are directed graphs containing no directed cycles (Almond et al, in progress).

The graphical model consists of three main concepts: (1) nodes representing unobservable or observable variables, (2) edges representing relations among variables, and (3) a joint probability distribution over all the variables in the network, implied by conditional probability distributions indicated by the edges. A graph is a pair $\mathcal{G} = (\mathcal{A}, E)$, where $\mathcal{A}$ is a set of nodes (variables) and E is a set of edges in which one edge is a line between two vertices (Almond et al, in progress). Each variable, $\mathcal{A}_i$, is associated with a finite set of possible values $\{a_{i,1}, a_{i,2}, \ldots, a_{i,n-1}, a_{i,n}\}$. An edge is expressed by the two variables it connects ($\mathcal{A}_1$, $\mathcal{A}_2$). The meaning of the directed groups is that the edges are directed, usually expressed as arrows.

In the directed graph ($\mathcal{G} = (\mathcal{A}, E)$), there is a dependent relationship among the variables and additional terminologies are incorporated to express this dependent relationship. The sets of variables with an arrow pointing from themselves to another set of variables ($\mathcal{A}$) are called *parents* of $\mathcal{A}$. They are denoted pa ($\mathcal{A}$| $\mathcal{G}$) or simply pa($\mathcal{A}$). The variable $\mathcal{A}$ with an edge toward it are the children of $\mathcal{A}$. If a directed graph contain no directed cycles, the graph is referred to as an acyclic directed graph. The acyclic directed graph is a key graphical model of BINs.

The graphical representation of BINs can be expressed by three probability distributions: marginal probability, conditional probability, and joint probability. A direct dependency among variables represented by an arrow in a directed graph is expressed by a conditional distribution. The states of variables that do not have any parents in a directed graph are expressed by a marginal distribution. The joint product of the probability distributions of all variables in a directed graph is a joint probability distribution for the full set of variables. The formal notation of the conditional probability distribution associated with each variable given all of its parents' variables is as follows (Almond et al, in progress):

$$P(A_i = a_i \mid pa(A_i)) \tag{13}$$

Since the joint probability distribution of a set of finite valued variables (A_i, . . . , A_n) is represented recursively in terms of the product of conditional distributions, the formal notation of a joint distribution associated with BINs is as follows (Almond et al, in progress):

$$P(A_i = a_i, ..., A_n = a_n) = \prod P(A_i = a_i \mid pa(A_i)) \tag{14}$$

If there are no parents (i.e., pa(A) is empty), then the conditional probability is regarded as a marginal probability. These marginal, conditional, and joint probability distributions are the formal relationship of BINs to probability theory.

To understand the concepts, consider an example of an acyclic directed graph with three variables (Figure 10).

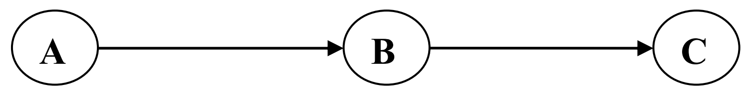


Figure 10. An Acyclic Directed Graph

The acyclic directed graphs are represented by a joint probability distribution over three variables, A, B, and C, which can be decomposed into a product of conditional probability distributions. The conditional dependences of the variables correspond to the acyclic directed graphs. The factorization is as follows:

$$P(A,B,C) = P(C|A,B)\ P(B|A)\ P(A) \qquad (15)$$

P(C|A,B) is a conditional distribution of variable C, given the variable of A and B. P(B|A) is a conditional distribution, given the variable A. P(A) is a marginal distribution. These probability distributions correspond to the directed graphical model (Figure 10). For example, P(C|A,B) is the probability distribution of variable C that is conditionally on its parents of variable A and B in the directed graph. Because there is no direct edge from A to C, P(C|A,B) simplifies to P(C|B). That is, A and C are conditionally independent given B.

Once all of the interrelationships are expressed in terms of the recursive representation of the joint distribution of variables, it is possible to calculate the updated states of any variables by the effect of new information about another set of variables through Bayes' rule (Almond, Mislevy, Steinberg, Williamson, and Yan, in progress). For example, suppose that there are two variables X and Y, P(X,Y). The variable X is given the variable Y. Bayes Theorem is obtained as follows:

$$P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X), \text{ therefore, } P(Y|X) = P(X) / [P(X|Y)P(Y)]$$

When X=x is observed, P(Y|x) can be calculated by Bayes Theorem:

$$\mathrm{P}(Y \mid X = x) \propto \mathrm{P}(Y)\,\mathrm{P}(X = x \mid Y) \qquad (16)$$

Two probability expressions for Y are involved in this expression. The first is the prior distribution, P(Y), expressing the initial belief about Y before any observations have been made. The second is the posterior distribution, P(Y|x), expressing updating belief about Y based on the observation X=x. In Equation 16, the posterior distribution, P(Y|X=x), is proportional to the likelihood, P(X=x|Y), multiplied the prior distribution, P(Y).

As the number of variables increases updating the full joint distribution using the Bayes theorem becomes prohibitive due to the increased number of parameters (Almond et al, in progress). Efficient calculation methods in BIN have been proposed (Lauritzen & Spiegelhalter, 1988). Lauritzen and Spiegelhalter (1988) and Jensen, Olesen, and Andersen (1990) developed updating methods for BINs based on the concept of the message passing in a tree structure, referred to as a junction tree (Kjaerulff & Madsen, 2008). The process of updating BINs through the junction tree algorithm is explained in the Appendix A.

Bayesian Inference Network: Estimation

In addition to the belief updating through the junction tree algorithm, a second mathematical part of BINs is estimating probability matrices from data through Maximum Likelihood (ML) method or Bayesian estimation method. If we could observe all observations for all variables, latent as well as observable, learning (estimation) would be easy by the ML method or Bayesian estimation method.

The ML method finds the maximum likelihood estimates (MLEs) of the parameter values which maximize the likelihood of the data. The likelihood function is follows as:

$$L = P(D \mid \Theta) = P(d_1 \mid \Theta) \cdot \ldots \cdot P(d_n \mid \Theta) \qquad (17)$$

where D consists of *n* independent cases: d_1, d_2, …, d_n; Θ indicates a set of the parameters in the probability matrices. (In this dissertation, the probabilities are estimated directly, so the parameters Θ are the conditional probabilities. It is possible, however, to model the probabilities more parsimoniously as parametric functions (Almond et al., in progress), in which case the parameters of these functions are the parameters to be estimated and serve the role of Θ in this equation.)

The likelihood can be re-expressed by taking its logarithm, to produce the log likelihood function is as follows:

$$\log(L) = \log(P(D \mid \Theta)) = \log(P(d_1 \mid \Theta)) + \ldots + \log(P(d_n \mid \Theta)) \qquad (18)$$

The ML method finds the parameter values which maximize the log likelihood function of the data, which are the same values that maximize the likelihood itself.

The Bayesian estimation incorporates a prior on the parameter. A common point estimate, called Maximum A Posterior (MAP), are estimates which find the parameter values that yields the maximizing value for the realized data. If there is prior information about the variables before the estimation starts from data, the prior knowledge can be considered as part of the data and combined with the new information. Based on Bayes rule, combining prior information is expressed as follows:

$$P(\Theta \mid D) \propto P(D \mid \Theta) P(\Theta) \tag{19}$$

There are two terms in this equation. The first term, $P(D \mid \Theta)$, is determined by data. The second term, $P(\Theta)$, expresses prior information. If the logarithm of the equation is obtained, the equation 20 is expressed as follows:

$$\log(P(D \mid \Theta)) + \log(P(\Theta)) \tag{20}$$

The log likelihood function, $\log(P(D \mid \Theta))$, consists of *n* independent cases: d_1,d_2, . . ,d_n, so the $\log(P(D \mid \Theta))$ is expressed as follows:

$$\log(P(D \mid \Theta)) = \log(P(d_1 \mid \Theta)) + \log(P(d_2 \mid \Theta)) + ... + \log(P(d_n \mid \Theta)). \tag{21}$$

Each of the term $\log(P(d_i \mid \Theta)$ is obtained by data or cases. The term of $\log(P(\Theta))$ is obtained from prior information. Finally, $P(\Theta \mid D)$ are computed by combining the $(P(d_i \mid \Theta))$ from data and $P(\Theta)$ from prior information. There are many ways to determine the term, $P(\Theta)$, such as prior knowledge and previous experience from data analysis or domain experts. As noted above, estimating the probability matrices would be straightforward if the observations could be directly observed. However, in educational and psychological settings, latent variables are involved in measurement models. In other words, by their nature, the values of the latent variables can never be observed. There are three techniques for dealing with the latent variables that are commonly used in BINs software programs: Expectation and Maximization (EM) algorithm, gradient ascent, and Markov Chain Monte Carlo Estimation (MCMC). This dissertation uses the EM algorithm in the Bayesian estimation paradigm (Dempster, Laird & Rubin, 1977). The details regarding these two conceptions with respect to DBINs will be provided in Section 5.2.

Identification Issues in Estimation

Estimation in latent class analysis is subject to model identification issues on estimation, one of which is the label switching problem (Dai, 2009). The issue of the label switching problem is not avoided under Bayesian estimation. Some approaches to deal with it in the context of latent class analysis have been suggested such as artificial identification constraints (Diebolt & Robert, 1994), relabeling of algorithms to perform a k-means type clustering of the MSMS samples (Celeux, 1998), label invariant loss functions (Celeux, Hurn, & Robert, 2000), and considering parameters as known prior information (Chung, Loken, & Schafer, 2004). Specifically, the method suggested by Chung, Loken, & Schafer (2004) is a relatively simple solution to tackle the label switching issue. This dissertation followed Chung's method.

Label switching is not a significant issue in the application study using real data because the labels switched can be fixed through prior information, thus making it possible that the results can be directly interpreted. However, it is a significant issue in the simulation study using many replications for evaluating parameter recovery or in the application study if there is any prior information about latent classes. In BINs, there are some feasible solutions to deal with the label switching issue. The possible ways to fix the label switching issue in BINs are (1) considering the latent variables as variables with many missing values (i.e, supplying a small number of parameters as known values) (Chung, Loken, & Schafer, 2004), (2) reconstructing the BINs by incorporating a constraint variable node into the BINs, and (3) incorporating prior information when estimating parameters.

This study uses the third method that incorporates prior information when estimating parameters. If there is prior information or experiences about the variables before the learning starts from data, the prior knowledge / previous experience can be considered as part of the data and combined with the new information to construct BINs. The incorporation of prior information into the parameters from data is analytically explained using Bayesian estimation: two terms, (1) $P(D \mid \Theta)$, which is determined by data, and (2) $P(\Theta)$, which is prior information. The use of a prior to avoid label switching is tantamount to adding a small amount of data to the actual data, in order to bias the resulting solution to one particular labeling among all those that would be consistent with the likelihood.

The Netica software program has a function that can incorporate the prior probabilities into a variable in the BINs before EM estimation starts (Netica-C API manual, 2006). When estimating $P(\Theta \mid D)$, furthermore, different degrees of the weights can be applied to initial prior information (Netica -C API manual, 2006). This dissertation used this function to deal with the label switching issue. For instance, the prior probability table for a variable can be set as below. Then, the degree of experience can be chosen as a weight of the prior probability table using the function of *SetNodeExperience*. The weight is known as the number of cases. This procedure would predispose the solution toward one of a number of possible labeling.

(1) Setting prior probabilities of variable A given variable B

	A				
		A1	**A2**	**A3**	**A4**
	B1	0.1	0.1	0.5	0.3
B	**B2**	0	0.1	0.2	0.7
	B3	0	0	0.5	0.5
	B4	0	0	0	1

(2) Setting a degree of experience for the variable A using the function below.

SetNodeExperience_bn (A parent_states, 1.0)

The BINs can be used for modeling student performances in education setting. The next section will illustrate a simple example of BINs in education setting.

Bayesian Inference Networks: An Example in Educational Setting

Mislevy (e.g., 1999, 2002, and 2003) has constructed BINs for modeling student performance in educational settings. This section illustrates a simple example that uses BINs for modeling student performance in an education setting. The example has a proficiency variable representing students' mastery states in a particular domain and three tasks designed to measure the proficiency of the proficiency variable. A BIN can be constructed by using the plausible hypothesized conditional probability for each task and the marginal probability of the proficiency variable. Figure 11 displays an initial BIN representation of students' performance on the assessment with the three tasks in Netica (Norsys Software Corporation, 2008).

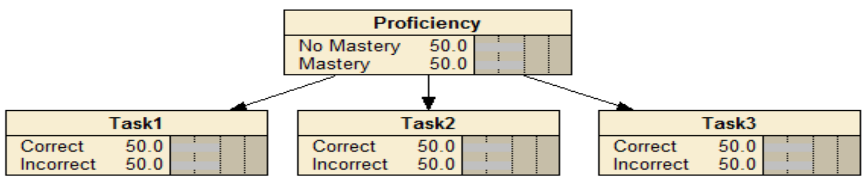


Figure 11. An initial BIN for a simple example with three tasks measuring proficiency in a domain.

The structure of the BIN in this example echoes traditional Item Response Theory (IRT) assumptions: Unidimensionality (i.e., there is only one ability variable for students) and Local Independence (i.e., the responses to any two items are independent given the students' ability in a domain). In the BIN, there is a direct arrow from the proficiency variable to each of the tasks, while there are not any direct arrows between the tasks. It states that the probabilities of whether or not a student correctly answers each task are dependent on the student's status in the proficiency variable, while they are independent of the responses to other tasks. The difference between IRT and BINs is that a student's ability is represented on a continuous continuum in IRT, while it is represented as a discrete latent variable with some states in BINs.

In this example, the proficiency variable has two values, representing the mastery and non-mastery states of each student. The task variable has two values for correct and incorrect responses. The proficiency variable and task variables have their own probability tables. For the proficiency variable, equal probabilities are considered as being able to take the values of mastery and non-mastery when there is no information regarding students' proficiency and any information regarding the task has not been observed. For the task variables, hypothesized conditional probabilities reflecting task characteristics associated with the states of the proficiency variable are considered. The initial marginal probability table for the proficiency variable is listed in Table 5. The conditional probability tables for the task variables are listed in Table 6.

Table 5

Initial marginal Probabilities of two states in the proficiency variable

Proficiency Variable		
Status	Low	Medium
Probability	0.5	0.5

Table 6

Conditional probabilities of answering correctly to three tasks given the student's states on the proficiency variable.

		Task1		Task2		Task3	
		Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Profic-iency	no mastery	0.3	0.7	0.2	0.8	0.1	0.9
	mastery	0.8	0.2	0. 6	0.4	0.4	0.6

The hypothesized conditional probabilities are set to reflect the task characteristics. In this example, tasks are of increasing difficulty: Task 3 is more difficult than Task 2 and Task 2 is more difficult than Task 1. The nature of these tasks corresponds to the pattern of the conditional probabilities of the three tasks that the probability of getting correct answer decreases from Task 1 to Task 3. The conditional probabilities can be determined either by reflecting on the nature of the tasks from domain experts or by estimating from actual student responses to the tasks. The notation of joint distribution associated with the BIN in this example can be written as follows:

$$P(Y_{1i} = y_{1i}, \ldots, Y_{3i} = y_{3i}, X = x_j) = \prod P(Y_i = y_i \mid X_j) P(X = x_j) \tag{22}$$

where X_j is the mastery states of a student and Y_{ni} is the outcome of a task.

Once a student's response to Task1 has been observed, that information is propagated through the network via Bayes' theorem to yield the posterior probability distribution of the student's states. Furthermore, the probabilities of answering correctly to Task 2 and 3 are updated based on the updated probabilities of a student's status on the proficiency variable. Figure 12 shows that all probabilities of the proficiency variable and tasks are updated.

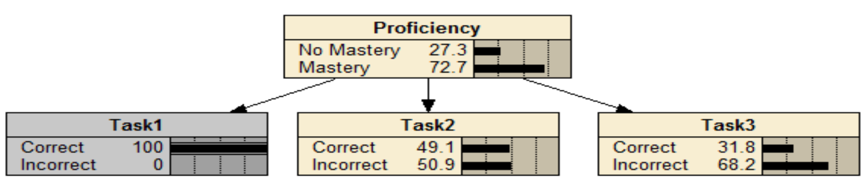


Figure 12. The same BIN as the Figure 3.3.1, but the response to Task 1 has been made.

From knowing that a student gets a correct answer to Task 1, it can be inferred that the student is more likely to have mastered the proficiency. Consequently, it is shown that the probabilities of answering correctly Task 2 and 3 increases. This updating can be written as follows based on Bayes rule:

$$P(X_j \mid Y_1 = 1) = \frac{P(Y_1 = 1 \mid X_j)P(X_j)}{\sum_j P(X_j)P(Y_1 = 1 \mid X_j)} \tag{23}$$

This is often written as:

$$P(X_j \mid Y_1 = 1) \propto \mathrm{P}(Y_1 = 1 \mid X_j)P(X_i) \tag{24}$$

Equation 24 states that the posterior distribution of the proficiency variable is proportional to the product of the likelihood of the proficiency variable given the response to Task 1 and the prior distribution of the proficiency variable. The likelihood of the proficiency variable can be found once the response to the task has been observed. Then, the posterior distribution is obtained by multiplying the prior distribution of the proficiency variable and the likelihood. This procedure can be found in Table 7.

Table 7

Computation of the posterior distribution of the proficiency variable

		Prior Probability	Likeli-hood	Likeli-hood * Prior	Normali-zation Constant	Posterior Probability
Proficiency	Mastery	0.5	0.8	0.40	0.55	0.727
	Non-Mastery	0.5	0.3	0.15	0.55	0.273

Suppose that a student's response pattern for all tasks has been obtained. Figure 13 shows a situation in which all observations of the tasks have been made. Once a student's response pattern of all tasks has been observed, the probabilities of the student's mastery statues on the proficiency variable are updated. Considering a situation in which a student has the response pattern [1, 1, 0] for each task, the probability that the student has mastered the proficiency is 0.842 and the probability that the student has not mastered the proficiency is 0.158. From the posterior distribution of the proficiency variable, it can be inferred that the student is more likely in the mastery status, with a probability of 0.842.

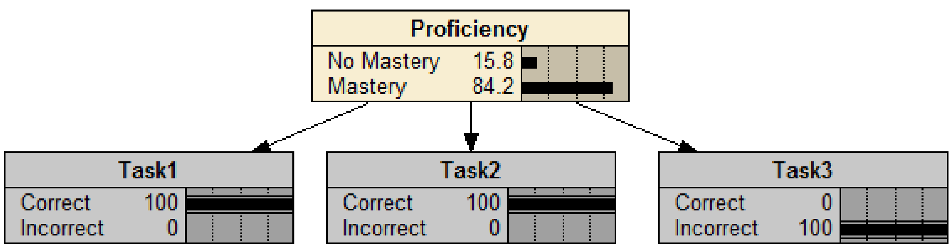


Figure 13. The same BIN, but the observations of the task 1, 2 and 3 have been made**.**

The BIN in an education setting is useful for understanding both task characteristics and student characteristics. With respect to a student, the BIN is useful for making inferences about student's status on latent variable(s) of interest. With respect to a task, the BIN is useful for examining the quality of a task by comparing the expected conditional probability and the data driven conditional probability. The next example models LPs using BINs with a static approach.

A Sample Static Approach for Modeling Learning Progressions

The static approach for modeling LPs is based on a cross-sectional sample design in which group differences are of interest. The following example describes a modeling of a LP in a static model approach based on a cross-sectional sample design through BINs. The building of BINs for LPs starts by developing the structure of LPs through the statistical method (e.g., model comparison for determining number of latent classes) and substantive theory grounded in the findings of contemporary research in cognition, developmental education, and the learning sciences. For a simple example, it is assumed that there is a latent variable representing a LP with four different levels. In terms of tasks, there are sixteen observable variables. It is assumed that all tasks have been designed with respect to a set of knowledge, skills, and abilities that students would be expected to possess at each status. Figure 14 displays a BIN representation of the LP with the four levels and sixteen tasks constructed in Netica.

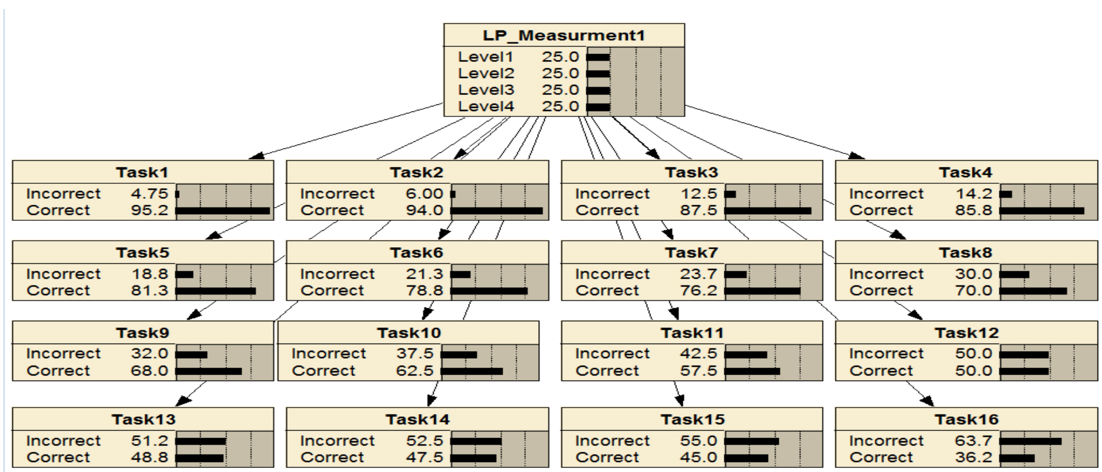


Figure 14. A BIN representation of LPs with a static approach

The latent variable representing an LP here is called LPs_Measurement1, which has an initial prior distribution where there are equal probabilities of being at each level. Since it is assumed that each task has been designed to evoke evidence about one or more targeted levels on the LP by means of key task features that differentiate students on the four levels, each task has a different conditional probability structure with respect to each level. In this BIN example, 16 tasks were used with the following assumptions; Tasks 1 through 4 were designed to have the particular task features associated with Level 1, Task 5 through Task 8 were designed to have the particular task features associated with Level 2, Task 9 through Task 12 were generated to have the particular task features associated with Level 3, and Task 13 through Task 16 were designed to have particular task features for Level 4.

It is noted that the conditional probabilities of responses to each task correspond to what tasks are designed because the level of task difficulty is ordered across tasks. As an example, Table 8 indicates a hypothetical conditional probability table for Task 8. It specifies conditional probabilities for a hypothetical observed variable at different levels of the latent variable (LP_Measurement1). As mentioned above, this task is designed for differentiating between Level 1 and Level 2; hence there is a significant gap between the probability of correct response given Level 1 and one given Level 2.

Table 8

Conditional Probability Table for Task 8

		Task 8	
		Incorrect	Correct
LP_Measurement1	Level1	0.6	0.4
	Level2	0.3	0.7
	Level3	0.2	0.8
	Level4	0.1	0.9

As another example, Table 9 shows a conditional distribution table for Task 16. It is shown that only students at Level 4 are relatively higher probability to correct response to Task 16 than those at other levels.

Table 9

Conditional Probability Table for Task 16

		Task 8	
		Incorrect	Correct
LP_Measurement1	Level1	0.90	0.10
	Level2	0.80	0.20
	Level3	0.55	0.35
	Level4	0.30	0.70

These hypothesized conditional probabilities can be compared with the observed conditional probabilities obtained from students' actual responses to each task in order to examine how well the task has been designed to classify students in terms of the levels and evaluate the BIN structure.

Once a student's responses have been observed, that information is propagated through the network via Bayes' theorem. The posterior distribution in the LP_Measurement 1 variable can be obtained by combining the initial prior distribution with likelihood of the LP_Measurement 1 given the student's response. From the posterior distribution, it can be inferred where the student is most likely in one among the four levels. Figure 15 shows the BINs for a student who has completed six tasks.

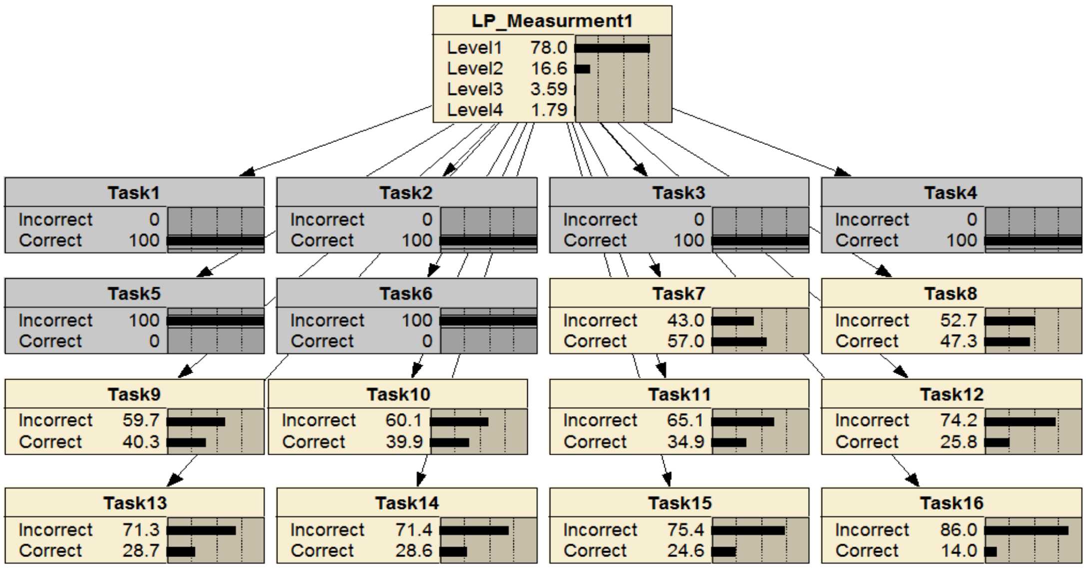


Figure 15. A BIN representation of LPs with a static approach. The 6 tasks has been observed

The first four tasks were correctly answered and the next two were incorrectly answered. Given this response pattern, the posterior probabilities of being on levels are .76, .16, .035, and .01 respectively. From this information, it can be inferred that the student with this response pattern is most likely in Level 1. Figure 16 shows the BIN for a student who has completed sixteen tasks. The student correctly answered all tasks. The response pattern is [1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1]. Given this response pattern, the posterior probabilities of being in the levels are .0002, .0013, .0785, and .92 respectively. From this information, it can be inferred that the student with this response pattern is most likely in level 4.

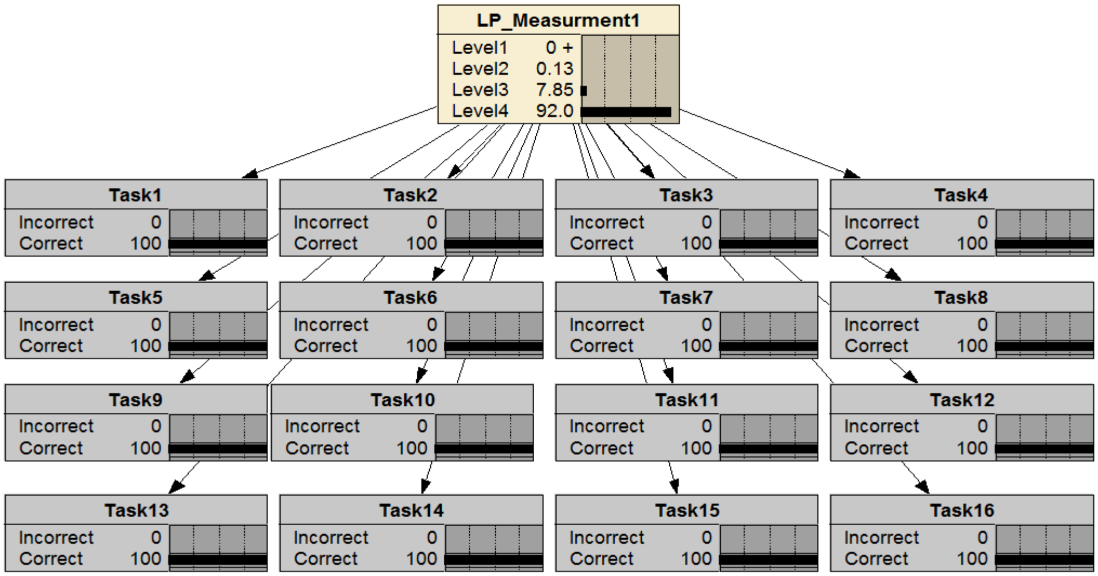


Figure 16. A BIN representation of LPs with a static approach. The 16 tasks have been observed.

Again, the BIN is useful (1) for assessing the quality of tasks with respect to measuring a student level on LP and (2) for classifying students in terms of the levels in the LP. The static BINs described above work well for modeling LP at a given point in time (Rutstein & Mislevy, in press; West, et al, 2009), but educators are also interested in analyzing proficiency change over time. A longitudinal sample design has the benefit of allowing change to be examined within the same individuals over time, although it also poses challenges in educational applications (e.g., practice effect and item drift). In other words, the observations can provide evidence of students' past and future states as well as their current states using longitudinal design approach. The next chapter will explain BINs with a dynamic approach and illustrate the ideas with some examples.

CHAPTER 4: DYNAMIC BAYESIAN INFERENCE NETWORK

This chapter introduces the Dynamic Bayesian Inference Network (DBINs) as one of the psychometric modeling methods for measuring LPs over multiple time points. In particular, the sections in this chapter include (1) fundamental concepts and formal notations for DBINs, (2) an example of DBINs for modeling LPs, and (3) an example of DBINs of modeling LPs with a covariate.

Dynamic Bayesian Inference Networks: Fundamental Concepts

Dynamic Bayesian Inference Networks (DBINs) are a way to extend a static BINs to model probability distributions over multiple time points (Murphy, 2002). Hence, a common approach to representing DBINs is combining several static BINs for a desired number of time slices. In this respect, the DBIN is also referred to as a time-sliced BIN (Kjaerulff & Madson, 2008). In the static approach, the probabilistic network is restricted to represent the state of a system at a certain point in time. In contrast, the DBINs consider the problem of monitoring the state of a dynamic process over a specific period of time. In the previous example, the BINs consider student states of an LP at a certain measurement occasion; however, it may also be of interest to monitor the state change over multiple measurement occasions. The DBINs can be used to make inferences about the previous states, current states, and future states of a system over a specific period of time. The mathematical procedures for inference and updating procedures in DBINs are the same as for the static approach.

The setting of interest in DBIN is the situation in which there are underlying hidden states of the phenomena that generate the observations, and in which the hidden states evolve over time (Murphy, 2002). As a simple example for understanding how to extend a static BIN to the structure of DBIN, Figure 10 considers the acyclic directed graph. Assume that a given BIN is an appropriate model for representing a phenomenon of interest at a certain point in time. A DBIN can be constructed based on the static network by copying the nodes of the static network and linking appropriately across time points. Figure 17 is the DBIN based on extending the BIN in Figure 10, in which each time slice consists of the structure shown in the original acyclic directed graph, while they are linked across multiple time slices.

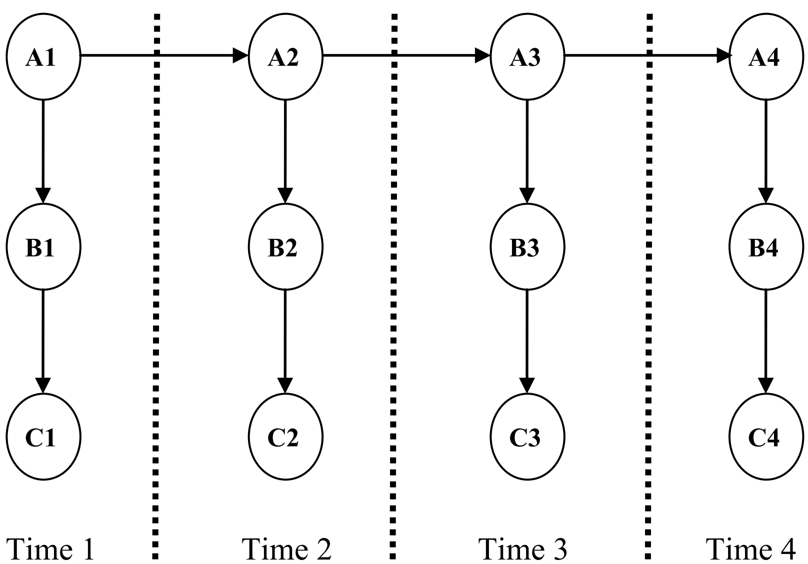


Figure 17. A representation of dynamic Bayesian Network

For the case of a BIN for modeling LPs, the DBIN contains a prior of the hidden state, $P(X_1)$, a transition function of the hidden states over multiple time points,

$P(X_t| X_{1:t-1})$, and an observation function given the hidden state, $P(Y_t| X_t)$.

There are two techniques for parameter estimation commonly used in commercial software programs: gradient ascent and EM (expectation maximization). This dissertation uses the EM to evaluate learning. The technical details will be provided in chapter 5.2. Another aspect of DBINs supports the monitoring of observations concerning the change of the system over a specific period of time. Once an observation has been made on a subset of the variables in the network at a certain point in time, researchers are able to make inferences about the remaining unobserved variables in the network at any given time points. In other words, the DBINs reflect the states at previous and future points in time as well as the current state because the states at the current point in time will impact the state in the future and are impacted by the state in the past (Kjaerulff & Madson, 2008). There are three main inferences that can be performed using DBIN (Kjaerulff & Madson, 2008; Murphy, 2002):

Smoothing: the process of monitoring states at previous time t-1 given evidence at time t,

Filtering: the process of monitoring states at the current time given evidence at the current time t,

Prediction: the process of monitoring states at future time t+1 given evidence at time t.

Regarding the inferences and learning of parameters, this dissertation assumes three properties.

First, the links of time slices are defined by the conditional probability of the variables at a current Time t given the variables at previous Time t-1. This property is called the first-order Markov property, under which the variables at Time t+1 are d-separated from the variables at Time t-1 given the variables at Time t (Kjaerulff & Madson, 2008). In other words, the states at Time t+1 only depend on the states at time t. Under the assumption, the transition probability, $P(X_t \mid X_{1:t-1})$, can be denoted as $P(X_t \mid X_{t-1})$. The formal notation for linking variables over multiple time points under the assumption of the first-order Markov property is denoted as follows:

$$\begin{aligned}\Pr(X_{t+1},\ldots,X_1) &= \Pr(X_{t+1} \mid X_t)\times\Pr(X_t \mid X_{t-1})\times\ldots\times\Pr(X_2 \mid X_1)\Pr(X_1)\\ &= \prod \Pr(X_t \mid X_{t-1})\end{aligned} \tag{25}$$

Second, observations are structured under the assumptions of conditional independence of observations and the first-order Markov property. The assumption of conditional independence indicates that $P(Y_t)$ is conditionally independent of $P(Y_{t`})$, given the $X_t$, where $t \neq t`$. Under the assumption of conditional independence, $P(Y_t \mid Y_{1:t-1}, X_t)$ simplifies to $P(Y_t \mid X_t)$. The assumption of the first-order Markov property indicates that $P(X_{t+1})$ is conditionally independent to $P(X_{t-1})$, given the X_t. Third, the term of dynamic suggests a modeling system that refers to state change over time, not networks or structures change over time.

Under the three assumptions, the formal notation of DBIN at Time t can be expressed with respect to a graphical model (Murphy,2002).

$$P(A_t \mid A_{t-1}) = \prod P(A_t^i \mid pa(A_t^i)) \tag{26}$$

where $A_t = (X_t, Y_t)$. The variable A consists of the latent variable (X) and the observation (Y).

The joint probability distribution is as follows:

$$P(A_{1:t}) = \prod_{t=1}^{T} \prod_{i=1}^{N} P(A_t^i \mid pa(A_t^i)) \tag{27}$$

For a simple example in an educational setting, the example of the BINs with the static approach above (Figure 13) can be extended to a representation of DBINs. Suppose that the same students have been repeatedly measured three times with three tasks. To construct a DBIN for this situation, the first step is to build three static BINs corresponding to each measurement occasion, and then, the proficiency variable of each static BIN is linked to each other across time points. Figure 18 shows the DBIN for this situation.

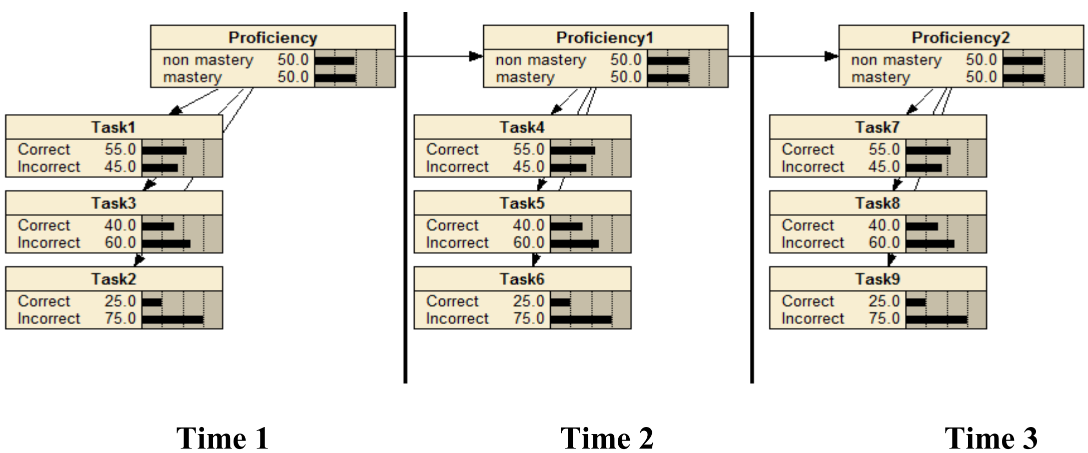


Figure 18. A representation of DBIN extended the BINs with static approach

Notice that there are three identical structures associated with the three time slices. The DBINs contain two kinds of reasoning: (1) reasoning about a student's current, past, and future proficiency, (2) reasoning about conditional probabilities of the tasks given the student's proficiency states. Once the responses to the three tasks have been made at the first measurement, the probabilities of which states the students are likely to at the first measurement occasion are estimated. In addition to reasoning about the current state, the states that the student will be more likely to be at mastery or non-mastery at future points in time can be inferred. As another example, the next section will introduce an example of the DBINs for modeling LPs that extends the example of the BINs with the static approach discussed in the previous chapter.

Bayesian Inference Networks: A Simple Example of a Dynamic Approach for Modeling Learning Progressions

In addition to a cross-sectional approach for modeling LPs, the longitudinal modeling of a student's LP over multiple time points can be investigated. In an educational setting, such testing situations can be considered where the same students are repeatedly measured at more than one point during a period of instruction (e.g., a course in a semester or an intervention). The tasks used are designed by incorporating task features in which students can be differentiated in terms of their levels of understanding and achievement that are theoretically grounded in cognitive developmental theory. In such situations, the investigation of the patterns of student level change across measurement occasions can offer diagnostic information that is customized to reflect individual learning and provide an informative evaluation of the effectiveness of instruction. For a simple example for modeling LPs through the DBINs, suppose that four measurements are designed. At each time point, there is a latent variable representing a LP and the observables that depend on them in probability. It is assumed that four levels are identified in the LP by domain experts. Each measurement consists of sixteen tasks across time points, which means that they have the same task characteristics reflecting the same set of knowledge, skills, and abilities over four measurement time points. Figure 19 shows an example of modeling LPs with a DBIN. This example shows an initial status of a DBIN with a latent variable and sixteen tasks with four measurement occasions when no observations have been made yet.

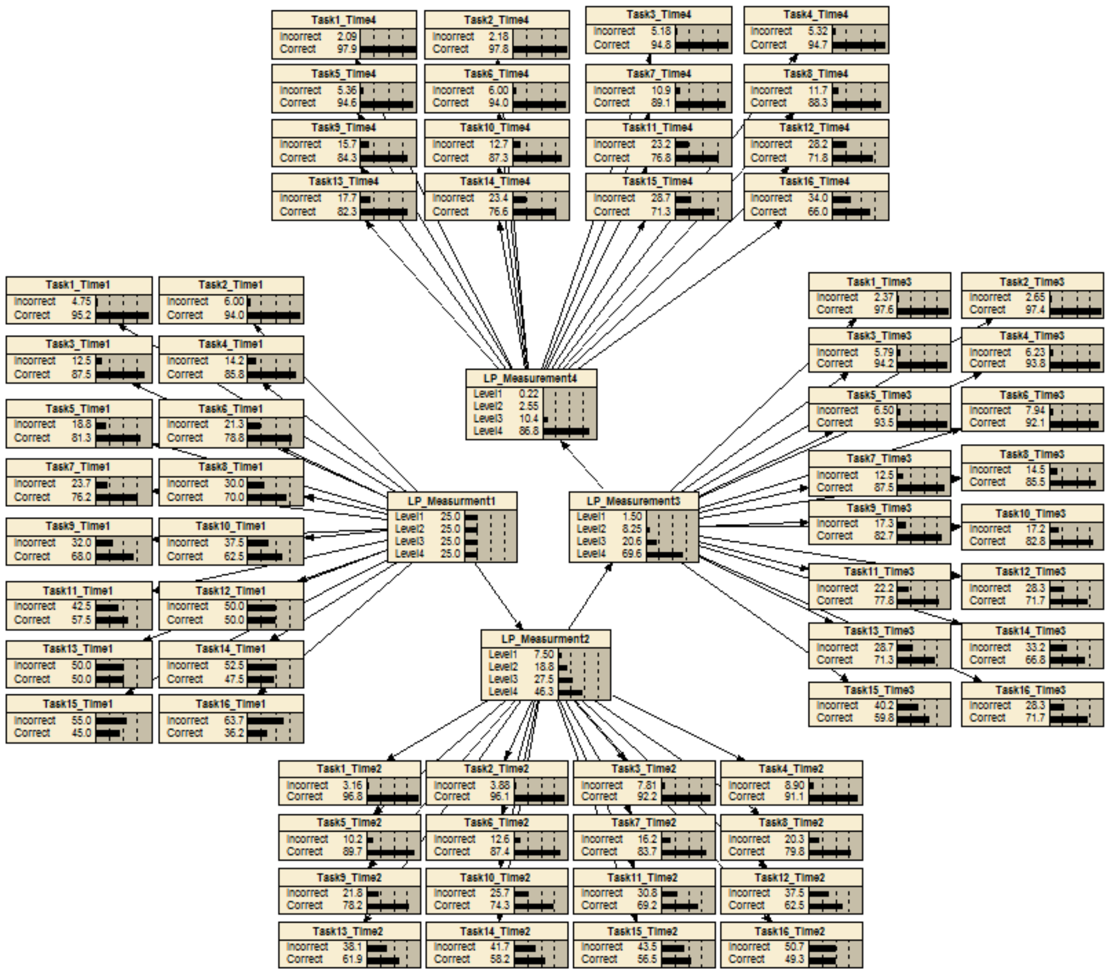


Figure 19. An initial representation of DBIN for modeling LPs

The DBIN contains two parts; (1) four latent variables for LPs where the latent variable at each measurement occasion is connected to the latent variable at the previous measurement occasion and (2) sixteen observables at each time point linked to the latent variable for that time point. Reasoning about status change over time can be investigated by focusing on transition probability tables for the four latent variables in the DBIN. To understand inference about student level change, Figure 19 shows the network containing the four latent variables.

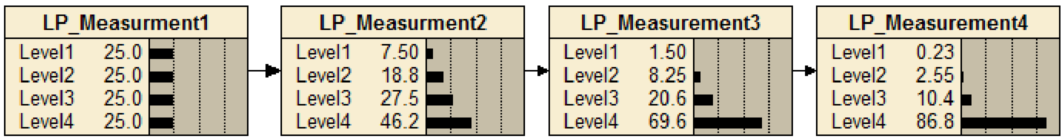


Figure 20. A DBIN representation of the four latent variables without tasks

The LP_Measurement 1 indicates the LP at the first measurement occasion, the LP_Measurement 2 indicates the LP at the second measurement occasion, and so on. In the network, each variable is connected to its previous variable. Two types of probability tables are involved: an initial probability table at the first measurement and the transition probability tables between two consecutive time points. The initial probability table and three transition probability tables are shown in table 10, 11, 12, and 13 below.

Table 10

Initial marginal Probabilities of three states in the proficiency variable

	LP_Measurement1			
Status	Level1	Level2	Level3	Level4
Probability	0.25	0.25	0.25	0.25

Table 11

Transition Probability Table for LP_Measurement2 given LP_Measurement1

		LP_Measurement2			
		Level1	Level2	Level3	Level4
LP_ Measurement1	Level1	0.30	0.55	0.10	0.05
	Level2	0	0.20	0.60	0.20
	Level3	0	0	0.30	0.60
	Level4	0	0	0	1

Table 12

Conditional Probability Table for LP_Measurement3 given LP_Measurement2

		LP_Measurement3			
		Level1	Level2	Level3	Level4
LP_ Measurement2	Level1	0.20	0.60	0.15	0.05
	Level2	0	0.20	0.60	0.20
	Level3	0	0	0.30	0.70
	Level4	0	0	0	1

Table 13

Conditional Probability Table for LP_Measurement3 given LP_Measurement3

		LP_Measurement3			
		Level1	Level2	Level3	Level4
LP_ Measurement3	Level1	0.15	0.60	0.20	0.05
	Level2	0	0.20	0.60	0.20
	Level3	0	0	0.25	0.75
	Level4	0	0	0	1

For this example, the probabilities are hypothetically set in order to illustrate the structure of a DBIN. They could be estimated by observations or determined by theory or domain expert opinion, as discussed in Section 3.2 and 5.2. Table 10 contains the hypothesized initial probabilities at the first measurement where the probability of being at each level is 0.25. Tables 11, 12, and 13 are the three hypothesized conditional probability tables of LP_Measurement2, LP_Measurement3, and LP_Measurement3 given all possible values of the previous levels on the LP, referred to as the transition probability tables. The variables in the DBIN representation in Figure 19 show the marginal probability tables of each variable. The computation of the marginal probabilities is as follows, where subscript c indicates the latent level and the superscript number indicates time:

Measurement Point 1:

$P(\theta^1{}_c) = P(\theta^1{}_{c=1}) + P(\theta^1{}_{c=2}) + P(\theta^1{}_{c=3}) + P(\theta^1{}_{c=3})$

Measurement Point 2:

$P(\theta^2{}_{c=1}) = P(\theta^2{}_{c=1} \mid \theta^1{}_{c=1})\, P(\theta^1{}_{c=1})$

$P(\theta^2{}_{c=2}) = P(\theta^2{}_{c=2} \mid \theta^1{}_{c=1})\, P(\theta^1{}_{c=1}) + P(\theta^2{}_{c=2} \mid \theta^1{}_{c=2})\, P(\theta^1{}_{c=2})$

$P(\theta^2{}_{c=3}) = P(\theta^2{}_{c=3} \mid \theta^1{}_{c=1})\, P(\theta^1{}_{c=1}) + P(\theta^2{}_{c=3} \mid \theta^1{}_{c=2})\, P(\theta^1{}_{c=2}) + P(\theta^2{}_{c=3} \mid \theta^1{}_{c=3})\, P(\theta^1{}_{c=3})$

Measurement Point 3:

$P(\theta^3{}_{c=1}) = P(\theta^3{}_{c=1} \mid \theta^2{}_{c=1})\, P(\theta^2{}_{c=1})$

$P(\theta^3{}_{c=2}) = P(\theta^3{}_{c=2} \mid \theta^2{}_{c=1})\, P(\theta^2{}_{c=1}) + P(\theta^3{}_{c=2} \mid \theta^2{}_{c=2})\, P(\theta^2{}_{c=2})$

$P(\theta^3{}_{c=3}) = P(\theta^3{}_{c=3} \mid \theta^2{}_{c=1})\, P(\theta^2{}_{c=1}) + P(\theta^3{}_{c=3} \mid \theta^2{}_{c=2})\, P(\theta^2{}_{c=2}) + P(\theta^3{}_{c=3} \mid \theta^2{}_{c=3})\, P(\theta^2{}_{c=3})$

Measurement Point 4:

$P(\theta^3{}_{c=1}) = P(\theta^3{}_{c=1} \mid \theta^3{}_{c=1})\, P(\theta^3{}_{c=1})$

$P(\theta^3{}_{c=2}) = P(\theta^3{}_{c=2} \mid \theta^3{}_{c=1})\, P(\theta^3{}_{c=1}) + P(\theta^3{}_{c=2} \mid \theta^3{}_{c=2})\, P(\theta^3{}_{c=2})$

$P(\theta^3{}_{c=3}) = P(\theta^3{}_{c=3} \mid \theta^3{}_{c=1})\, P(\theta^3{}_{c=1}) + P(\theta^3{}_{c=3} \mid \theta^3{}_{c=2})\, P(\theta^3{}_{c=2}) + P(\theta^3{}_{c=3} \mid \theta^3{}_{c=3})\, P(\theta^3{}_{c=3})$

$P(\theta^3{}_{c=3}) = P(\theta^3{}_{c=3} \mid \theta^3{}_{c=1})\, P(\theta^3{}_{c=1}) + P(\theta^3{}_{c=3} \mid \theta^3{}_{c=2})\, P(\theta^3{}_{c=2}) + P(\theta^3{}_{c=3} \mid \theta^3{}_{c=3})\, P(\theta^3{}_{c=3}) +$
$P(\theta^3{}_{c=3} \mid \theta^3{}_{c=3})\, P(\theta^3{}_{c=3})$

$P(\theta^1{}_{c=1})$, $P(\theta^1{}_{c=2})$, $P(\theta^1{}_{c=3})$, and $P(\theta^1{}_{c=3})$ are the initial marginal probabilities of the LP at the first measurement; $P(\theta^2{}_{c=1})$, $P(\theta^2{}_{c=2})$, $P(\theta^2{}_{c=3})$ and $P(\theta^2{}_{c=3})$ are the marginal probabilities of each level at the second measurement occasion, and so on. $P(\theta^{t+1}{}_c \mid \theta^t{}_c)$ is the transition probability. For example, $P(\theta^2{}_{c=2} \mid \theta^1{}_{c=1})$ indicates the probability of moving from Level 1 at the first measurement occasion to Level 2 at the second measurement occasion; as another example, $P(\theta^3_{c=3} \mid \theta^2_{c=2})$ is the probability of moving from Level 2 at the second measurement occasion to Level 3 at the third measurement occasion.

In this example, the hypothesized transition probabilities are restricted in such a way that all the probabilities of reverse changes are zero. This constraint reflects an LP considering only forward movements over time. Other types of transition probability patterns can be considered depending on different substantive theory embodied LPs. The different patterns can be modeled by (1) constraining sets of transition probabilities to be equal to zero, (2) restricting them to be a particular value, or (3) fixing them to be equal to each other. Table 14, 15, and 16 show the different types of the transition matrix patterns including a forward movement, an adjacent movement, and all possible movements.

Table 14

Forward movements LPs model between two measurement points

		LP_Measurement2				
		Level1	Level2	Level3	Level4	Marginal
LP_Measure-ment1	Level 1	$P(\theta^2_{c=1} \mid \theta^1_{c=1})$	$P(\theta^2_{c=2} \mid \theta^1_{c=1})$	$P(\theta^2_{c=3} \mid \theta^1_{c=1})$	$P(\theta^2_{c=4} \mid \theta^1_{c=1})$	1
	Level 2	0	$P(\theta^2_{c=2} \mid \theta^1_{c=2})$	$P(\theta^2_{c=3} \mid \theta^1_{c=2})$	$P(\theta^2_{c=4} \mid \theta^1_{c=2})$	1
	Level 3	0	0	$P(\theta^2_{c=3} \mid \theta^1_{c=3})$	$P(\theta^2_{c=4} \mid \theta^1_{c=3})$	1
	Level 4	0	0	0	1	1

Table 15

Adjacent movements LPs model between two measurement points

		LP_Measurement2				
		Level1	Level2	Level3	Level4	Marginal
LP_Measur-ement1	Level 1	$P(\theta^2_{c=1} \mid \theta^1_{c=1})$	$P(\theta^2_{c=2} \mid \theta^1_{c=1})$	0	0	1
	Level 2	$P(\theta^2_{c=1} \mid \theta^1_{c=2})$	$P(\theta^2_{c=2} \mid \theta^1_{c=2})$	$P(\theta^2_{c=3} \mid \theta^1_{c=2})$	0	1
	Level 3	0	$P(\theta^2_{c=2} \mid \theta^1_{c=3})$	$P(\theta^2_{c=3} \mid \theta^1_{c=3})$	$P(\theta^2_{c=4} \mid \theta^1_{c=3})$	1
	Level 4	0	0	$P(\theta^2_{c=3} \mid \theta^1_{c=4})$	$P(\theta^2_{c=4} \mid \theta^1_{c=4})$	1

Table 16

All possible movements LPs model between two measurement points

		LP_Measurement2				
		Level1	Level2	Level3	Level4	Marginal
LP_Measur-ement1	Level 1	$P(\theta^2_{c=1} \mid \theta^1_{c=1})$	$P(\theta^2_{c=2} \mid \theta^1_{c=1})$	$P(\theta^2_{c=3} \mid \theta^1_{c=1})$	$P(\theta^2_{c=4}\ \theta^1_{c=1})$	1
	Level 2	$P(\theta^2_{c=1} \mid \theta^1_{c=2})$	$P(\theta^2_{c=2} \mid \theta^1_{c=2})$	$P(\theta^2_{c=3} \mid \theta^1_{c=2})$	$P(\theta^2_{c=4} \mid \theta^1_{c=2})$	1
	Level 3	$P(\theta^2_{c=1} \mid \theta^1_{c=3})$	$P(\theta^2_{c=2} \mid \theta^1_{c=3})$	$P(\theta^2_{c=3} \mid \theta^1_{c=3})$	$P(\theta^2_{c=4} \mid \theta^1_{c=3})$	1
	Level 4	$P(\theta^2_{c=1} \mid \theta^1_{c=4})$	$P(\theta^2_{c=2} \mid \theta^1_{c=4})$	$P(\theta^2_{c=3} \mid \theta^1_{c=4})$	$P(\theta^2_{c=4} \mid \theta^1_{c=4})$	1

To understand how the transition function works for the purpose of investigating state change over time, one could consider a situation where student status at the first measurement occasion is known. This information is propagated through the network by Bayes theorem. The posterior distribution of the next three variables given the student's states at the first measurement occasion can be updated by using the transition function.

Figure 21 shows the posterior distributions of three variables given a student latent Level 1 at the first measurement occasion. It can be inferred that the student is most likely at Level 2 at the second measurement occasion with .55, at Level 3 with .405 at the third measurement occasion, and at Level 4 with .61 at the fourth measurement occasion, respectively. Figure 22 depicts a situation where a student was at level 2 at the first measurement occasion. The probabilities of the student being at Level 3 at the second measurement, at Level 4 at the third measurement, and at Level 4 at the fourth measurement are .6,.66,and .89, respectively.

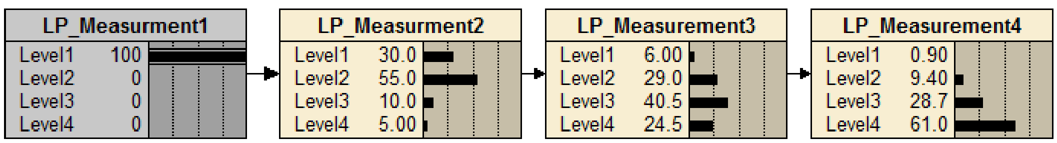


Figure 21. A DBIN representation of the four latent variables given a student latent Level 1 at the first measurement occasion

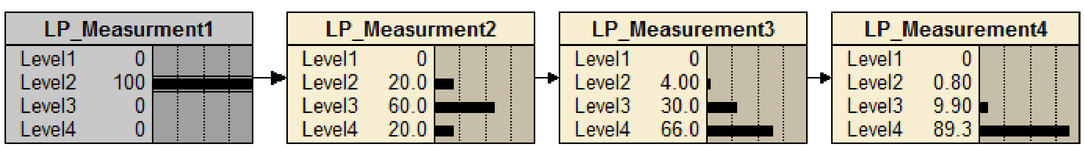


Figure 22. A DBIN representation of the four latent variables without task given a student latent Level 2 at the first measurement occasion

The next step considers the DBIN with tasks. Once a student's responses have been observed at any given time point, that information is propagated through the network via Bayes' theorem. The posterior distributions of the latent variable at that time point as well as the latent variables at previous and future time points are obtained.

Figure 23 shows a situation in which observations, [1,1,1,0,0,0,0,0,0,0,0,0,0,0,0,0], have been made at the first measurement occasion. Based on these observations, the posterior distributions of the four latent variables are updated. It is shows that the student is more likely to be at Level 1 at the first with 0.99, at Level 2 at the second with 0.55, at Level 3 at the third with .30, and at Level 4 at the fourth with 0.61 when the student has the particular response pattern given 16 tasks at the first measurement occasion.

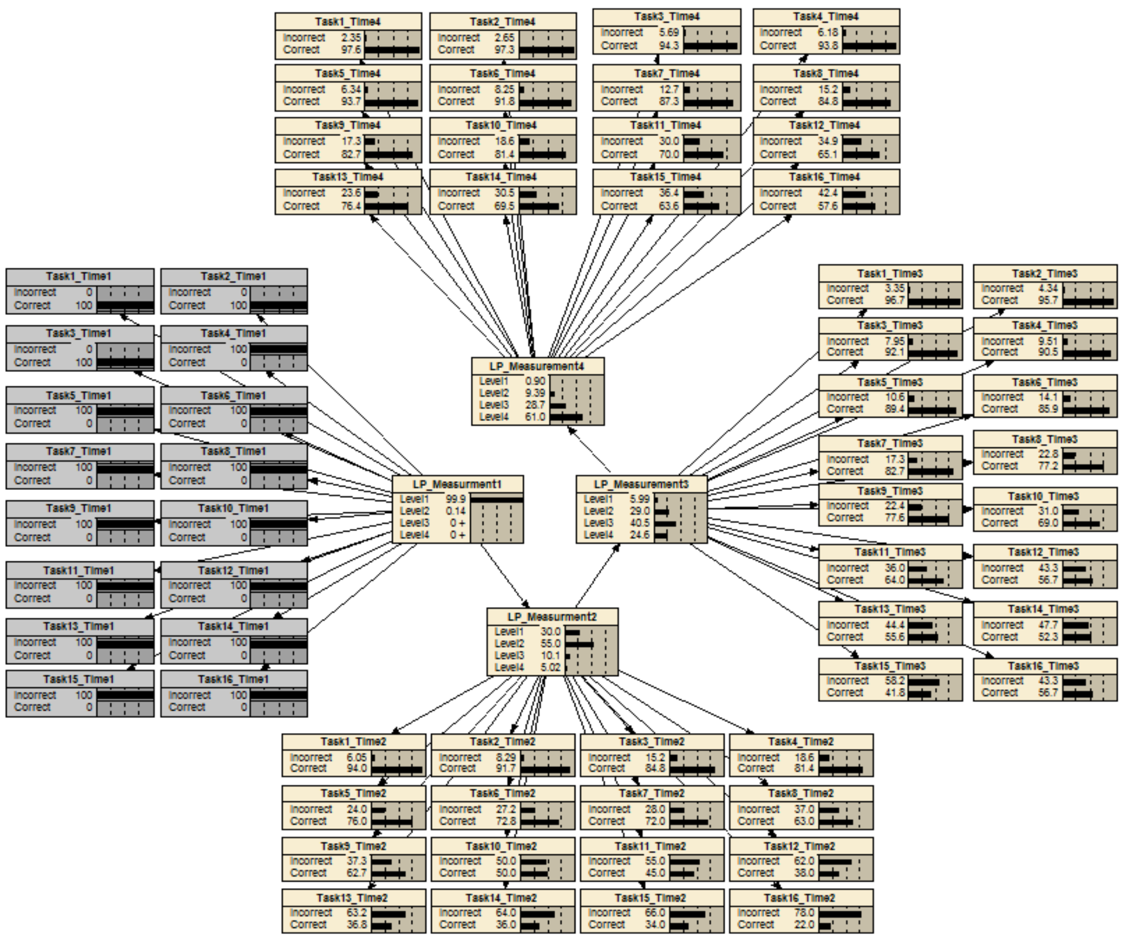


Figure 23. A representation of DBIN when the student has the particular response pattern given 16 tasks at the first measurement occasion

Figure 24 shows a situation in which the observations of all 32 tasks have been made, [1,1,1,1,0,0,0,0,0,0,0,0,0,0,0,0] at the first measurement, [1,1,1,1,1,1,1,1,0,0,0,0,0,0,0,0] at the second measurement, [1,1,1,1,1,1,1,1,1,1,1,1,0,0,0,0] at the third measurement, and [1,1,1,1,1,1,1,1,1,1,1,0] at the fourth measurement. It is shown that the posterior distributions of the four latent variables are updated when a student has the responses pattern given 32 tasks at the four measurement occasions. It can be inferred that the probability of the student being at level 1 at the first occasion is .99, the probability of the student being at level 2 is.81, the probability of the student being at level 3 is.77, and the probability of the student being at level 3 is .86 given their particular response patterns in each of the 32 tasks. In other words, from the posterior distributions, it can be inferred that the student with this particular responses pattern is most likely at Level 1 at the first measurement occasion, at Level 2 at the second measurement occasion, at Level 3 at the third measurement occasion, and at Level 4 at the fourth measurement occasion.

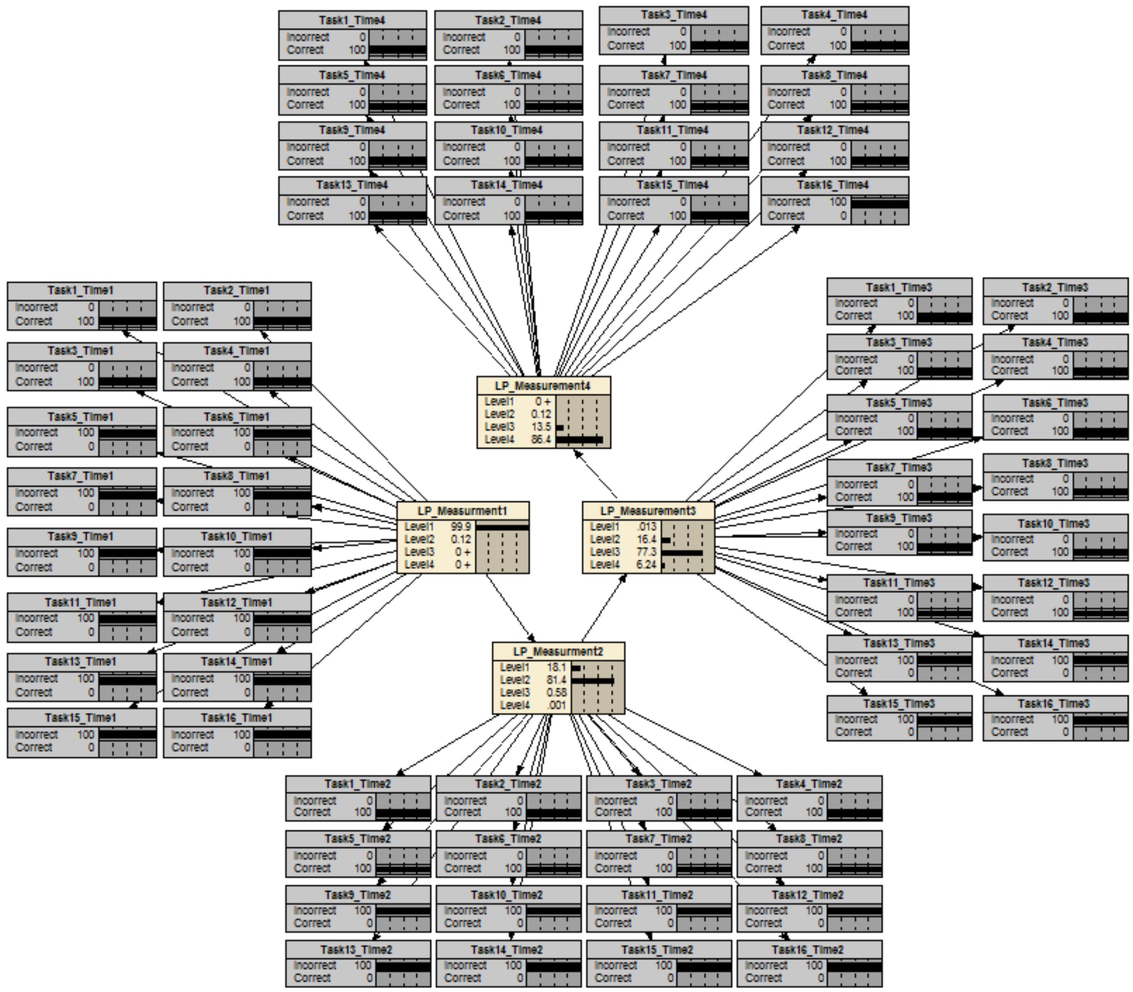


Figure 24. A representation of DBIN with the observation of all 32 tasks.

In addition to making inferences about students' status changes over multiple time points, the DBIN can be also used for examining the quality of tasks to see if the task is appropriately located at the expected level on LPs. A comparison between the expected conditional probability table and the observed conditional probability table of each task provides evidence of task quality. If there is a sufficient amount of difference between them, one should consider checking whether the task has been appropriately designed.

The BINs can be easily extended to more complex models by incorporating covariate or more measurement occasions. The next example considers how to extend a DBIN to more complex models by incorporating a covariate.

A Dynamic Model Approach with Covariate

A manifest variable (e.g., different instructions, interventions, or individuals' demographic backgrounds) or a latent exogenous variable (e.g., attitude, intelligence, or social economic status) can impact students' status change. It can be investigated by constructing more complex DBINs that incorporate any such variables into a transition probability. This section demonstrates how a DBIN can be extended for inferring status change when a covariate for students is incorporated. For the case of LPs research, different instruction can differently influence student LPs with respect to different levels. For example, suppose that a DBIN is constructed with two measurement occasions. The latent variable representing an LP at each measurement occasion has four levels. Additionally, the DBIN contains a variable indicating two sets of instruction connected to the latent variable at the second measurement. Figure 25 shows the initial status of the DBIN with two latent variables, four tasks at each measurement, and an observed instruction-related variable, in the state where no evidence of the mode of instruction that has been entered.

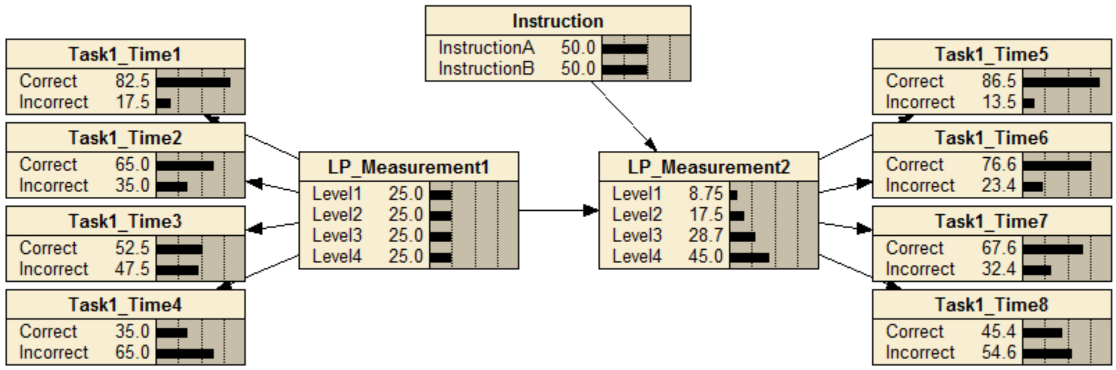


Figure 25. A representation of DBIN with a covariate

Figure 26 illustrates changes in student levels on the LP from one time point to the next after instruction A. In contrast, Figure 27 demonstrates a change in student levels on the LP from one time point to the next if a student has received instruction B. Comparing the marginal probabilities of the LP_Measurement 2 variable in Figure 26 and Figure 27 shows that the instruction A was more effective for students at Level 2 while the instruction B was more effective for students at Level 3. Thus, when instruction A was used, most of the students at Level 2 changed to Level 3, while most of the students at Level 3 have changed to Level 4 when instruction B was used.

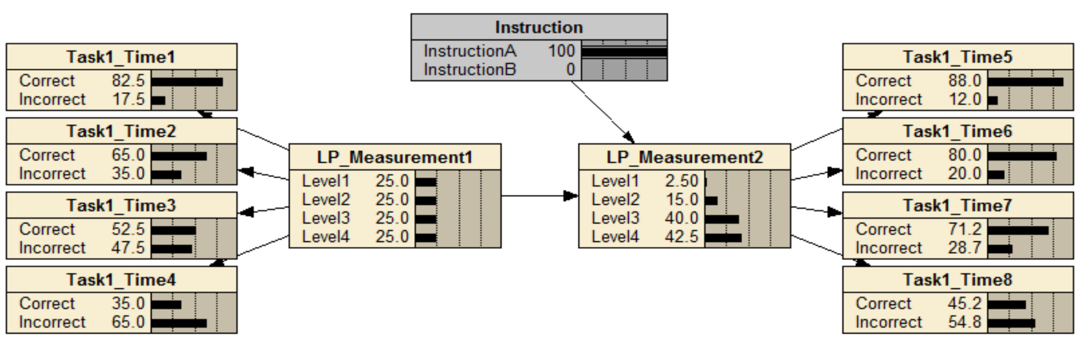


Figure 26. A representation of DBIN with a covariate when instruction A was used

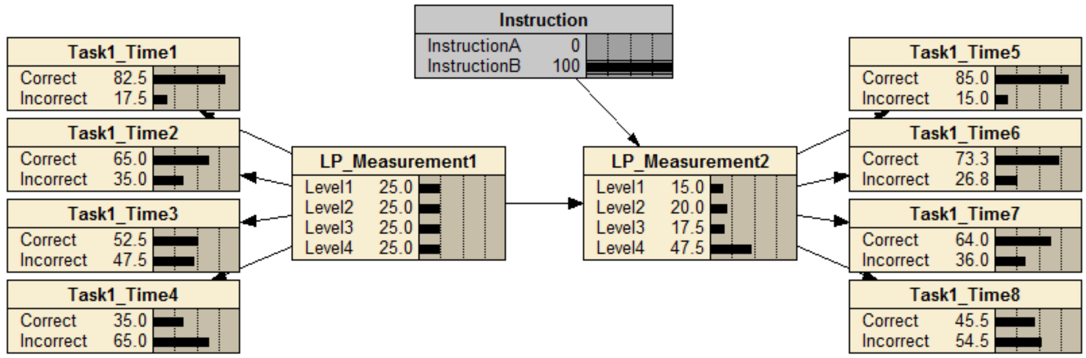


Figure 27. A representation of DBIN with a covariate when instruction B was used

The detailed movements with respect to the instruction variable can be investigated by estimating transition probability tables, to the extent that the instruction variable has different effects on the transition probabilities. Table 17 and Table 18 are hypothesized transition probabilities used for constructing the DBINs. Notice that the transition probabilities are dependent on the type of instruction received. The different probabilities yield different marginal probabilities of the LP_Measurement2 variable.

Table 17

Transition Probability Table for LP_Measurement2 given LP_Measurement1 with the instruction A

		LP_Measurement2			
		Level1	Level2	Level3	Level3
LP_Measure-ment1	Level1	0.1	0.5	0.3	0.1
	Level2	0	0.1	0.6	0.3
	Level3	0	0	0.7	0.3
	Level4	0	0	0	1

Table 18

Transition Probability Table for LP_Measurement2 given LP_Measurement1 with the instruction B

		LP_Measurement2			
		Level1	Level2	Level3	Level4
LP_Measure-ment1	Level1	0.60	0.20	0.15	0.05
	Level2	0	0.60	0.25	0.15
	Level3	0	0	0.30	0.70
	Level4	0	0	0	1

Therefore, if the transition probabilities can be estimated with respect to a covariate in LP, it can be used to evaluate the effectiveness of an instruction that may have differential effectiveness in terms of different latent levels (Graham et al, 1991).

CHAPTER 5 DBINS, HMMS, AND EMALGORITHM

HMMs can be expressed in the representation of a DBIN. The elements of DBINs correspond to a standard algebraic expression of HMMs. In addition, the structural relationships between the elements of DBINs correspond to those in HMMs through the concept of Markov property and conditional independence. This chapter compares two models to describe how their statistical properties are related to the same concept. In addition to the comparison, this chapter explains the EM algorithm as the estimation method for the parameters of DBINs.

Correspondence between DBINs and HMMs

An HMM is comprised of a Markov chain and observables (Cappé et al, 2005). A Markov chain is a sequence of discrete random variables with the Markov property. By the term hidden, a Markov chain is latent, denoted by X_t containing *n* possible states, X_t = {1,…,n}. What are directly observable are other sets of observables called indicators linked to the Markov chain, denoted by Y_t containing n possible states, Y_t = {1,…,n}. There are parameters in three distributions being estimated: (1) The initial state distribution, $\pi(i)$ representing a multinomial distribution, (2) the transition model, $A(i,j)$ representing a conditional multinomial distribution, and (3) the observation model, $P(Y_t \mid X_t)$ (Murphy, 2002).

The initial state distribution, $\pi(i)$, corresponds to the initial probability distribution, $P(X_1 = i)$ in DBINs. The transition model, $A(i,j)$, is the transition probability distribution, $P(X_t = j \mid X_{t-1} = i)$ in DBINs. The observation model, $P(Y_t \mid X_t)$, corresponds to the conditional probability distribution of observables, $P(Y_t \mid X_t)$ in DBINs. The formal probabilistic notation of a hidden Markov chain in the HMMs is denoted as follows:

$$\begin{aligned} P(X_{t+1},...,X_1) &= P(X_{t+1}|X_t)\times P(X_t|X_{t-1})\times ...\times P(X_2|X_1)P(X_1) \\ &= \prod P(X_t|X_{t-1}) \end{aligned} \tag{28}$$

Under the assumptions of conditional independence and the first-order Markov property defines that the observations $\{Y_n\}$ are independent given the states of a hidden Markov chain $\{X_n\}$ at a given time point:

$$P(Y_{n+1},...,Y_1|X_1,...,X_n) = \prod P(Y_n|X_n) \tag{29}$$

Additionally, an extended probability-based model can be expressed by incorporating a covariate, shown below:

$$\Pr(X_n,...,X_1,|\,G_n,...,G_1) = \prod \Pr(X_n|X_{n-1},G_n) \tag{30}$$

where G_n is of the value of a covariate.

EM algorithm

DBINs have parameters in initial probability distribution, $P(X_t)$, transition probability distribution, $P(X_t|X_{t-1})$, and conditional probability matrix, $P(Y_t|X_t)$. If it was possible to observe observations for the variables in DBINs, estimation could be done by using the maximum likelihood (ML) method or the Bayesian estimation method (Maximum a posterior (MAP)). The ML method finds the maximum likelihood estimates (MLEs) of the parameters of each probability distribution (i.e., the parameter values which maximize the likelihood of the data). The log likelihood in DBINs from the equation 31 is follows as:

$$\log L = \log \prod_{t=1}^{T} \prod_{i=1}^{N} P(A_t^i \mid pa(A_t^i)) = \sum_{t=1}^{T} \sum_{i=1}^{N} \log P(A_t^i \mid pa(A_t^i)) \tag{31}$$

The log-likelihood function decomposes into a series of terms per node (Murphy, 2002). Specifically, BINs have categorical variables, so the distributions of the variables take the form of either Bernoulli distributions or multinomial distributions. If there are just two categories for the outcomes of a variable, the random variable follows a Bernoulli distribution. If there are more than two categories in variable, the random variable follows a multinomial distribution. The Bernoulli distribution is as follows:

$$p(r \mid n, \pi) \propto \pi^r (1-\pi)^{n-r} \tag{32}$$

where π is the probability that an event of success will occur; 1- π is the probability of the occurrence of a failure.

Once *n* trials occur and *r* successes are observed, Equation 33 is interpreted as a likelihood function, L(π | *r, n*). From the likelihood function, the maximum likelihood estimate (MLE) of π is obtained as the value that maximizes the likelihood. If there are more than two categories in a variable, the random variable follows a multinomial distribution.

$$P(R_1 = r_1, \ldots, R_k = r_k \mid \pi_1, \ldots, \pi_k) \propto \prod_{k=1}^{K} \pi_k^{r_k} \tag{33}$$

where $\pi_1,...,\pi_k$ is the probability that a success or occurrence event in category *k* from *n* independent samples of the categorical variable. r_k is the count of the number of observations in category k, so $\sum_{k=1}^{K} r_k = n$.

The equation 33 is interpreted as a likelihood function, L(**π** |***r****, n*). From the likelihood function, the MLEs of **π** are obtained as the values that maximize the likelihood.

The Bayesian estimation method incorporates a prior on the parameter to the likelihood. A common point estimate in this case is called a maximum a posterior (MAP) estimates; that is, what parameter value yields the maximizing value for the realized data. If a conjugate prior for the Bernoulli and multinomial distribution is used in the Bayesian estimation method, this has the advantage of eliminating concerns about the normalizing constant. The conjugate prior for the Bernoulli distribution is beta distribution. Therefore, the posterior distribution about π after combining the beta prior distribution and the likelihood through Bayes theorem is as follows:

$$p(\pi \mid a,b) \times L(\pi \mid r,n) \propto beta(\pi \mid a,b) \times L(\pi \mid r,n) \propto Beta(\pi \mid a+r, b+(n-r)) \tag{34}$$

The conjugate prior for the multinomial distribution is a Dirichlet distribution as follows:

$$p(\boldsymbol{\pi} \mid a_1,...,a_k) \propto {\pi_1}^{a_1-1} \times \cdots \times {\pi_k}^{a_k-1} \tag{35}$$

Therefore, the posterior distribution about **π** can be thought of as the results of combining the Dirichlet prior distribution and the likelihood through the Bayes theorem as follows:

$$p(\boldsymbol{\pi} \mid a_1,\ldots,a_k) \times L(\boldsymbol{\pi} \mid \mathbf{r}, n) \propto Dirch(a_1 + r_1,\ldots,a_k + r_k) \tag{36}$$

If there were observations of latent variables, the estimation of the parameters in the distribution of the latent variables would be simple. However, the values representing the levels of LPs are not directly observed in LPs research. Expectation and Maximization (EM) algorithm, gradient ascent, and Markov chain Monte Carlo Estimation (MCMC) are commonly used in BINs software programs in order to estimate the values of the parameters of the distributions of the latent variables. This dissertation uses the EM algorithm to estimate parameters of DBINs (Dempster, Laird, & Rubin, 1977).

The basic idea of EM algorithm is to estimate parameters in iterative cycles. The first step is to start with an initial guess of parameters, and then compute the expected sufficient statistics in the E (expectation) step. The second step is the M step. The M step estimates parameters using the expected sufficient statistics as if they were actually sufficient statistics computed from the data. This procedure finds the values to maximize the expected complete data log-likelihood. The EM procedure is repeated until convergence criterion is met. At iteration *k,* the expectation of the likelihood in DBINs is as follows:

$$\mathrm{Q}(\Theta \mid \Theta^{\mathrm{k}}) = E[\log \prod_{t=1}^{T} \prod_{i=1}^{N} P(A_t^i \mid pa(A_t^i)] \tag{37}$$

In the Bayesian estimation method, this is additionally multiplied by the priors. In the E step, the expectation of the complete data log likelihood is calculated. Then, the next step is the M (maximization) step. The M step is performed with the expected values from the

E step. This procedure finds the values to maximize the expected complete data log-likelihood.

$$\Theta^{K+1} = \max_{\Theta} Q(\Theta \mid \Theta^{K}) \tag{38}$$

The cycles of the E step and M step continues until the criterion of convergence is met. For instance, the parameters of the transition probability table are computed as follows. The E step expected values of the count of each category given the provisional parameters:

$$E[P(X_{t-1}=i, X_t=j \mid \Theta)] \tag{39}$$

The $E[P(X_{t-1} = i, X_t = j \mid \Theta^{K})$ is called the expected sufficient statistic (ESS) for the transition matrix. Since the prior of the transition probability is a Dirichlet distribution, the Dirichlet prior distribution is combined with the likelihood.

Then, the M step finds the values to maximize the expected complete data log-likelihood. The cycle of the E step and M step continues until convergence.

CHAPTER 6: SIMULATION DATA STUDY 1: A SIMPLE DBIN

This chapter investigates a simulation data study for evaluating the performance of a simple DBIN in the context of an LPs study. In particular, the sections in this chapter contain the overview of the simulation study, an explanation of the method of data generation, a description of simulation conditions, and a discussion of the results. The evaluation of the performances of a simple DBIN focuses on how different constraints on (1) the relation between observables (tasks) and LPs and (2) the relation of the LPs between two consecutive measurement points affect parameter recovery in estimation using Netica software (Norsys Software Corp, 2008).

Overview

The first simulation study focused on the construction of a simple DBIN model. In the case examined, there were two measurement occasions and each measurement has multiple observable variables measuring one LP. Figure 28 displays the model that was examined in this study.

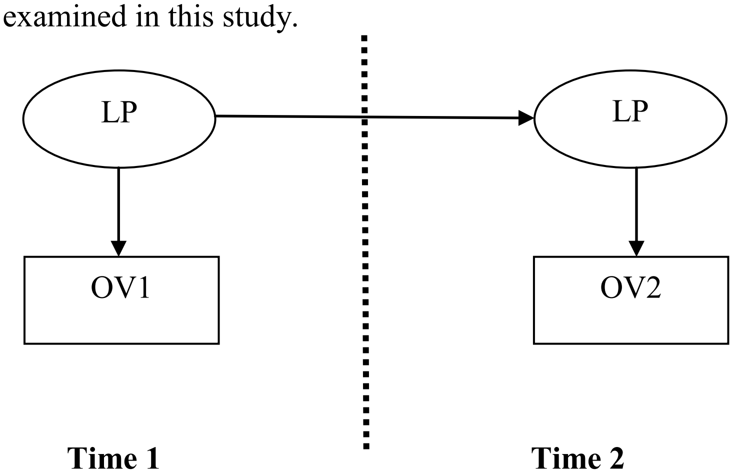


Figure 28. A model for the first simulation data study

The LP assumed that there were four levels. Based on a literature review, four levels is a common number of levels used in practice (Learning Progression in Science Conference, 2009). Further research may be applied to LPs with more than four levels. Each level represents the aspects of knowledge, skills, and ability for students required at the level on an LP. Table 19 includes the parameters that needed to be estimated for this model.

Table 19

Parameters that need to be estimated for the first simulation study

Parameters		
Latent Variable at the first measurement occasion	Transition probability matrix	Conditional probability of all tasks, given levels
$P(LP_{t=1}=1)$	$P(LP_{t=2}=1 \| LP_{t=1}=1)$	At time 1,
$P(LP_{t=1}=2)$	$P(LP_{t=2}=2 \| LP_{t=1}=1)$	$P(OV_j \| LP_i)$
$P(LP_{t=1}=3)$	$P(LP_{t=2}=3 \| LP_{t=1}=1)$	For each task
$P(LP_{t=1}=4)$	$P(LP_{t=2}=4 \| LP_{t=1}=1)$	
	$P(LP_{t=2}=2 \| LP_{t=1}=2)$	At time 2,
	$P(LP_{t=2}=3 \| LP_{t=1}=2)$	$P(OV_j \| LP_i)$
	$P(LP_{t=2}=4 \| LP_{t=1}=2)$	For each task
	$P(LP_{t=2}=3 \| LP_{t=1}=3)$	
	$P(LP_{t=2}=4 \| LP_{t=1}=3)$	j = task
	$P(LP_{t=2}=4 \| LP_{t=1}=4)$	i= level

The research questions for this first simulation study were:

(1) How well can parameter estimates of conditional probabilities for observable variables be recovered?

(2) How well can parameter estimates of transition probabilities between two latent variables be recovered?

(3) How well can the proportions of students at the first measurement (indicating student classifications of levels at the first measurement) and the proportions (marginal probabilities) of an LP at the second measurement occasion (indicating student classifications of levels at the second measurement) be recovered?

These research questions were addressed by computing (1) Root Mean Squared Difference (RMSD), (2) bias, and (3) Standard Error (SE).

Data Generation and Simulation Conditions

Data was simulated using R. The proportions of each level on the LP at the first time point, the transition probabilities, and the conditional probabilities of correctly responding to each task given each level of the LP were considered for generating response data.

The fixed factors of this simulation study were included in the structure of the DBIN. The structure of the DBIN considered here includes two measurement occasions and four levels in each LP.

The factors varied in this simulation study were (1) sample size, (2) the number of tasks, (3) distributions of the students on LPs at the first measurement, (4) the types of transition probability tables, and (5) the types of conditional probability tables of the tasks.

Three cases were considered for the distributions of students on LPs at the first measurement occasion. Table 20 displays the different distributions of students. The first case represents an equal probability of students being at each of the four levels. The second case represents the students who are mostly of high ability. The third case represents the students who are mostly of low ability.

Table 20

Distribution of student on LP at the first measurement

Case	Description	Level 1	Level 2	Level 3	Level 4
1(Equal)	Equal probability at all levels representing general population	0.25	0.25	0.25	0.25
2 (High)	High ability students	0.05	0.10	0.25	**0.60**
3(Low)	Low ability students	0.15	**0.60**	0.15	0.10

Two cases for the conditional probability table of observables were considered. The rationale of the probabilities chosen here is taken from the literature of cognitive diagnosis or mastery testing using probabilities for true and false positive probabilities used in some studies (Leighton & Gierl, 2007). The first case represents that the tasks are well designed for classifying students into their levels. The case uses the probability of .2 of answering the task incorrectly if students are at a lower level than the levels that task requires. The probability of .85 is used for answering the task correctly if students are at the level or at higher levels than the task requires. The value .2 corresponds to the probability of getting correct multiple choice tasks with 5 options. On the other hand, the second case represents that tasks are relatively poorly designed for classifying students into their levels, therefore, the probabilities of answering the task incorrectly are chosen with the higher number while the probabilities of answering the task correctly are chosen with the lower number than the first case. The probability of .35 of answering the task incorrectly is used for students who are at a lower level than the levels that task requires. The probability of .70 of answering the task correctly is used for students who are at the level or at higher levels than the task requires. Table 21 displays the first case of the conditional probability table with nine tasks. The probabilities indicate the probabilities of answering the task correctly given each level. This basic structure will be used in order to duplicate more tasks.

Table 21

The first case of conditional probability table

	Task								
	Level1			Level2			Level3		
Level of Student	1	2	3	1	2	3	1	2	3
Level 1	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Level 2	0.85	0.85	0.85	0.20	0.20	0.20	0.20	0.20	0.20
Level 3	0.85	0.85	0.85	0.85	0.85	0.85	0.20	0.20	0.20
Level 4	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85

Table 22 displays the second case of the conditional probability table. The differences in values of the conditional probabilities between the two levels are smaller than the first case in Table 21. This structure is duplicated for generating 30 tasks.

Table 22

The second case of conditional probability table

	Task								
	Level1			Level2			Level3		
Level of Student	1	2	3	1	2	3	1	2	3
Level 1	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
Level 2	0.70	0.70	0.70	0.35	0.35	0.35	0.35	0.35	0.35
Level 3	0.70	0.70	0.70	0.70	0.70	0.70	0.35	0.35	0.35
Level 4	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70

For the types of transition probability tables, three cases were considered for the transition probability tables. In this study, based on developmental theory, only the forward directions in transitions were considered by setting two constraints into the transition probabilities matrix. The two constraints are: 1) the sums of all probabilities at each row in a transition matrix are constrained to one and 2) the probabilities in backward directions such as moving from high levels to low levels in the transition matrix are constrained as zeros. Table 23 shows the three cases in the transition probability table.

Table 23

Nine cases of transition probability table

Transition Probability	Description Equal Transition	Large Transition	Small Transition
P(1|1)	0.25	0.10	**0.60**
P(2|1)	0.25	0.10	**0.20**
P(3|1)	0.25	**0.50**	0.10
P(4|1)	0.25	**0.30**	0.10
P(2|2)	0.33	0.10	**0.70**
P(3|2)	0.33	0.20	0.20
P(4|2)	0.33	**0.70**	0.10
P(3|3)	0.50	0.10	**0.90**
P(4|3)	0.50	**0.90**	0.10
P(4|4)	1.00	1.00	1.00

The simulation conditions in the transition probability table reflect the following situations:

- The first case has the equal transition probability for each cell. This indicates that students at all levels have the same probability of staying at the same level or moving to higher levels with same probability.
- The second case suggests that students have a high probability of moving to higher levels, while the transition probabilities that students stay at the same

level are small. For example, P(2|0) is 0.5, P(3|1) is 0.8, and P(3|2) is 0.9; but, P(0|0), P(1|1), and P(2|2) are only value 0.1.

- The third case represents low probabilities that students will transition to higher levels, while the transition probabilities that students will stay at the same level are relatively large. For example, P(2|0) is 0.1, P(3|1) is 0.1, and P(3|2) is 0.1; but, P(0|0) is 0.5, P(1|1) is 0.7, and P(2|2) are 0.9.

Different sample sizes and task sizes were also considered. In their review of literature, Harwell and Stone (1996) found that a sample size of 100 or less sample size is generally considered a small sample, a sample size of around 500 is considered a medium sample, and a sample size of 1000 is considered a large sample. Estimating conditional probabilities for items in Bayes nets is analogous to estimating item parameters for the same tests. This study estimates a small number of conditional probabilities for item response conditional on a latent proficiency variable. The only difference from the study in IRT literature is that the latent ability is discrete rather than continuous. Therefore, following the literature, two sample sizes of 100 and 1000 are considered in this study. The 100 sample size is considered a small sample size. The 1000 sample size is considered a large number of students. The purpose of comparing different sample sizes is to investigate how large a sample size may be needed to provide reliable results.

The task sizes used in this study are 9 and 30. The 9 task size is considered a small task size. If there are too few observables, the model will not be identified. Almond, et al (2008) used three observables per level in order to keep the model simple but identified. Following the literature, three tasks for each level in the latent variable representing a LP were used as the minimum task size in this simulation study. Each task in this study was assumed to be dichotomously scored and was designed to measure only one level of the LP. The relation between the task and the latent variable was set at a medium relationship by using .8 as the probability of a correct response given that the student has the skill level required by the task. In order to consider a medium or more than medium task size, a task size of 30 was additionally examined in this simulation study.

With the combination of (1) different distributions of students at the first measurement, (2) different transition probability tables, (3) different conditional probability tables, (4) sample size, and (5) task size, the total number of cells for this study is 72 conditions. 100 replications were chosen in each cell based on the previous simulation studies (Harwell and Stone, 1996). The simulation conditions are summarized in Table 24.

Table 24

Simulation conditions of the first study

Simulation Condition		# of Case
Distribution of student at the first measurement	Equal, high, and Low	3
Conditional probability table	Case 1 and 2	2
Transition probability table	Case 1, 2, and 3	3
Sample size	50 and 1000	2
Task size	9 and 30	2
Total		72
Replication		100

Estimation

Once the task responses were generated, an EM estimation implemented in the Netica C API (Norsys Software Corp, 2008) was used to estimate the parameters. R code for generating responses and evaluating parameter recovery were written. One disadvantage of the Netica application is that it does not have any function that automatically runs many replications. Therefore, the Netica C API was used for this simulation study. The Netica C API has the same capabilities as the Netica application such as building, modifying, learning, and inferring networks. However, it is a complete library of C-callable functions for working with BINs. Equations can also be embedded into programs written in any language as long as the language can call C functions. The visual studio (2010) was used to call Netica C API in this study. The syntax for implementing the DBINs in this simulation study has been written in C language. The syntax of the Netica C API can be found in Appendix B.

In order to implement an EM algorithm, two criteria needed to be set to stop iterations. Two criteria were default in Netica: One was the maximum number of iteration steps and the other was the minimum change in data log likelihood between consecutive iterations. This simulation study used the defaults: (1) 1000 for the maximum number of iterations and (2) 1.0e-5 for the minimum change in data log likelihood between two iterations. The iteration was terminated when either of the two conditions was met. All replications in all cells were convergent before 1000 iterations in this simulation study.

A label switching issue occurred. In order to handle this label switching issue in the simulation study, the method described above of incorporating prior information when estimating parameters was used. The Netica contains a function that is able to incorporate the prior probability table of each variable in the BINs before EM learning starts (Netica C API manual, 2006). The different degrees of weights can be applied to initial prior information. The value of 1 was used as the weight of prior, which is equivalent to the amount of information contained in a data set with sample size of 1 (Netica C API manual, 2006). The detailed information is in chapter 3.

Results

Parameter recovery in terms of the different conditions was examined by comparing estimates with true parameter values used for response data. 100 replications were run. Three criteria were used to evaluate the overall accuracy of the method in each condition: (1) Root Mean Squared Difference (RMSD), (2) Bias, and (3) Standard Deviation of Estimates (SDE), often called Standard Error (SE).

Root Mean Squared Difference

$$RMSD = \sqrt{\frac{\sum_i (\hat{\theta}_i - \theta)^2}{I}} \tag{40}$$

where I is the number of replication, $\hat{\theta}_i$ is the estimate, and θ is the true parameter.

Bias

$$Bias = \bar{\hat{\theta}} - \theta \tag{41}$$

where $\bar{\hat{\theta}}$ is the mean of the estimates and θ is the true parameter..

Standard error

$$SD = \sqrt{\frac{\sum_i (\hat{\theta}_i - \bar{\hat{\theta}})}{I}} \tag{42}$$

where I is the number of replication, $\hat{\theta}_i$ is the estimate and $\bar{\hat{\theta}}$ is the mean of the estimates

Gifford and Swaminathan (1990) stated that the Root Mean Squared Difference (RMSD) for any particular parameter across replications can be separated into bias in estimation and the variance of the estimates across replications. Harwell (1996) reported that the three criteria are interrelated in that the squared RMSD is equal to the sum of the squared bias and the squared SDE. In other words, RMSD can be seen as a measure of the total error of parameter estimation, and is composed of a systematic error element (Bias) and a random error element (SE). Harwell (1996) said that smaller values of this index suggest that the estimates are fairly stable and reliable, while larger values indicate that the estimates may be unreliable. The three criteria were computed for each simulation condition. The results chapter contains three parts. The first part shows the values of bias, RMSD, and SDE of the parameters of all probability tables in terms of simulation conditions. The second part displays graphs that show a comparison of the different simulation conditions in terms of the parameters of all probability tables. The graphs provide information about which simulation conditions influenced the parameter of estimation the most. The third part shows the results of ANOVA analysis with all 72 simulation conditions. In the first part, the values of bias, RMSD, and SDE were organized by the simulation conditions of the sample size, task size, types of initial probability distributions, transition probability tables, and conditional probability tables.

Sample Sizes

Table 25

Bias for the student distribution at the first measurement in terms of the different sample sizes

Sample Size = 100				Sample Size = 1000			
Bias				Bias			
Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
0.01047	0.00491	-0.01906	0.00368	-0.00451	0.00209	0.00000	0.00241

Table 26

Bias for the transition probability in terms of the different sample sizes

Bias				
Sample Size = 100				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.0385	0.0336	-0.0332	-0.0390
Level 2	0.0062	-0.0447	0.0273	0.0112
Level 3	0.0029	0.0082	-0.0590	0.0479
Level 4	0.0014	0.0071	0.0081	-0.0167
Sample Size = 100				
Level 1	-0.0070	0.0044	0.0052	-0.0026
Level 2	0.0000	-0.0071	-0.0018	0.0089
Level 3	0.0014	0.0001	-0.0037	0.0022
Level 4	0.0004	0.0000	0.0012	-0.0016

Table 27

Bias for conditional probability tables of the first type of the tasks in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00524	-0.00524	-0.00189	0.00189
Level 2	0.00019	-0.00019	-0.00451	0.00451
Level 3	0.00658	-0.00658	0.00136	-0.00136
Level 4	0.00158	-0.00158	0.00218	-0.00218

Table 28

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01002	-0.01002	0.00329	-0.00329
Level 2	0.02391	-0.02391	0.00025	-0.00025
Level 3	-0.01055	0.01055	-0.00568	0.00568
Level 4	0.00189	-0.00189	0.00302	-0.00302

Table 29

Bias for conditional probability tables of the third type of the tasks at the first measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00130	-0.00130	0.00131	-0.00131
Level 2	0.00336	-0.00336	-0.00223	0.00223
Level 3	-0.00632	0.00632	-0.00339	0.00339
Level 4	-0.01355	0.01355	-0.00513	0.00513

Table 30

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.04947	-0.04947	0.00219	-0.00219
Level 2	0.00305	-0.00305	-0.00217	0.00217
Level 3	-0.00475	0.00475	-0.01182	0.01182
Level 4	0.00023	-0.00023	-0.00007	0.00007

Table 31

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.02029	-0.02029	0.00570	-0.00570
Level 2	0.04714	-0.04714	-0.00374	0.00374
Level 3	0.00543	-0.00543	-0.00188	0.00188
Level 4	0.00592	-0.00592	-0.00018	0.00018

Table 32

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00941	-0.00941	0.00289	-0.00289
Level 2	0.00586	-0.00586	-0.00550	0.00550
Level 3	0.01551	-0.01551	0.00223	-0.00223
Level 4	-0.00588	0.00588	-0.00091	0.00091

Table 33

Average of RMSDs for the different sample sizes

Condition	Sample Size = 100	Sample Size = 1000
	RMSD	RMSD
Parameters		
DS* at Time1	0.064614	0.038974
TPT*	0.095234	0.040602
Average of CPTT*	0.076190	0.038426

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 34

Average of SDE for the different sample sizes

Condition	Sample Size = 100	Sample Size = 1000
	SD	SD
Parameters		
DS* at Time1	0.004252	0.001283
TPT*	0.006366	0.002052
Average of CPTT*	0.017728	0.006381

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Tables above show the bias of the initial probability distribution parameters (DS) in terms of the two different sample sizes. Table 26 shows the bias of the transition probability table parameters (TPT) in terms of the sample of 100 and the sample of 1000. Table 27, Table 28, and Table 29 indicate the bias of the conditional probability table parameters of the tasks used at the first measurement. Table 30, Table 31, and Table 32 show the bias of the conditional probability table parameters of the tasks implemented at the second measurement. The bias values decreased as the sample size increased, that is, the bias values were dramatically lower with a sample size of 1000 than a sample size of 100. Table 33 shows the average of the RMSDs of the parameters of DS, TPT, and CPT. The RMSDs were lower as the number of sample size increased. Among the parameters of DS, TPT, and CPT, it was observed that TPT seems to be the most affected by the sample size. A similar pattern was also observed for SDE values. In terms of two different samples sizes, SDEs with 1000 samples were lower than those with 100 samples.

To summarize, the results showed that more samples reduced bias and variance of the estimates for all parameters of the probability distribution tables under the different conditions considered in this study. Among the parameters of DS, CPT, and TPT, TPT seems to be the most affected by different sample sizes compared the others. TPT seems to be more biased compared to the others and CPT seems to have more error variance than others. A following section will use analysis of variance to test for the statistical significance of these prima facie effects.

Task Sizes

Table 35

Bias for the student distribution at the first measurement in terms of the different task sizes

Task Size = 9				Task Size = 30			
Bias				Bias			
Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
0.00646	0.00729	-0.0161	0.00236	-0.0005	-0.0003	-0.0029	0.00374

Table 36

Bias for the transition probability in terms of the different task sizes

Bias				
Task Size = 9				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.02603	0.01791	-0.02745	-0.01649
Level 2	0.00521	-0.02790	0.00204	0.02066
Level 3	0.00399	0.00796	-0.02030	0.00833
Level 4	0.00182	0.00636	0.00346	-0.01160
Task Size = 30				
Level 1	0.00544	0.02016	-0.00051	-0.02509
Level 2	0.00100	-0.02392	0.02348	-0.00056
Level 3	0.00030	0.00034	-0.04237	0.04172
Level 4	0.00000	0.00077	0.00585	-0.00662

Table 37

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different task sizes

	Task Size = 9		Task Size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00123	-0.00123	0.00212	-0.00212
Level 2	-0.00079	0.00079	-0.00353	0.00353
Level 3	0.00907	-0.00907	-0.00112	0.00112
Level 4	0.00203	-0.00203	0.00173	-0.00173

Table 38

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different task sizes

	Task Size = 9		Task Size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01563	-0.01563	-0.00231	0.00231
Level 2	0.02134	-0.02134	0.00283	-0.00283
Level 3	-0.01374	0.01374	-0.00249	0.00249
Level 4	0.00335	-0.00335	0.00155	-0.00155

Table 39

Bias for conditional probability tables of the third type of the tasks at the first measurement in terms of the different task sizes

	Task Size = 9		Task Size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00384	-0.00384	-0.00123	0.00123
Level 2	-0.00031	0.00031	0.00144	-0.00144
Level 3	-0.00822	0.00822	-0.00149	0.00149
Level 4	-0.01311	0.01311	-0.00557	0.00557

Table 40

Bias for conditional probability tables of the first type of the tasks at the second measurement in terms of the different task sizes

	Task Size = 9		Task Size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.04141	-0.04141	0.01025	-0.01025
Level 2	-0.00021	0.00021	0.00109	-0.00109
Level 3	-0.01724	0.01724	0.00067	-0.00067
Level 4	0.00003	-0.00003	0.00013	-0.00013

Table 41

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different task sizes

	Task Size = 9		Task Size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.02259	-0.02259	0.00340	-0.00340
Level 2	0.03717	-0.03717	0.00622	-0.00622
Level 3	0.00237	-0.00237	0.00118	-0.00118
Level 4	0.00570	-0.00570	0.00004	-0.00004

Table 42

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different task sizes

	Task Size = 9		Task Size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00923	-0.00923	0.00307	-0.00307
Level 2	0.00161	-0.00161	-0.00125	0.00125
Level 3	0.01440	-0.01440	0.00334	-0.00334
Level 4	-0.00560	0.00560	-0.00119	0.00119

Table 43

Average of RMSD for the different task sizes

Condition	Task Size = 9	Task Size =30
	RMSD	RMSD
Parameters		
DS* at Time1	0.073203	0.018310
TPT*	0.093266	0.044939
Average of CPTT*	0.079744	0.030369

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 44

Average of SDE for the different task sizes

Condition	Task Size = 9	Task Size =30
	SD	SD
Parameters		
DS* at Time1	0.003301	0.002234
TPT*	0.004878	0.003540
Average of CPTT*	0.011927	0.012182

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 35 shows the bias of the DS in terms of the two different task sizes. Table 36 shows the bias of the TPT in terms of a task size of 9 and a task size of 30. Table 37, Table 38, and Table 39 indicate the bias of the CPT parameters of the tasks used at the first measurement. Table 40, Table 41, and Table 42 show the bias of the CPT parameters of the tasks implemented at the second measurement. The bias values decreased as the task size increased. Compared to the condition of sample size, the size of the sample resulted in a larger difference than the task size in the parameters of TPT. However, there were similar differences between two conditions of sample size and task size in the parameters of DS and CPT. The reduced amount of bias values of the CPT parameters was the largest among other parameters (DS and TPT) when the number of tasks increased. It implied that the precision of estimates in the CPT were more sensitive to the task size implemented than the other parameters (DS and TPT). Table 42 shows the comparison of the average of RMSDs between two different task sizes of 9 and 30. It was observed that RMSD decreased as the task size increased.

In summary, the results showed that the precision of estimates increased as more tasks were used. Among the parameters of DS, CPT, and TPT, DS and CPT seem to be more effected by different task sizes. TPT seems to be more biased compared to the others and CPT seems to have more error variance than others. The variances of estimation were less reduced as task size increased compared to samples size increased. The statistical significance of the effects will be addressed in a following section using analysis of variance.

Initial Probability Distributions

Table 45

Bias for the student distribution at the first measurement in terms of the different initial probability distributions

	Bias			
Condition	Level 1	Level 2	Level 3	Level 4
Equal distribution	-0.00006	-0.01106	0.00695	0.00417
Negatively skewed distribution	0.00407	0.01377	-0.03108	0.01324
Positively skewed distribution	0.00492	0.00780	-0.00446	-0.00826

Table 46

Bias for the transition probability in terms of the different initial probability distributions

Bias				
Equal distribution				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.00938	0.02372	-0.01675	-0.01636
Level 2	0.00252	-0.01064	0.00441	0.00372
Level 3	0.00583	0.00919	-0.04080	0.02578
Level 4	0.00192	0.00113	0.00490	-0.00796
Negatively skewed distribution				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.00749	0.02815	-0.01995	-0.01568
Level 2	0.00592	-0.05278	0.00953	0.03733
Level 3	0.00061	0.00073	-0.00257	0.00123
Level 4	0.00011	0.00273	0.00891	-0.01176
Positively skewed distribution				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.03034	0.00523	-0.00523	-0.03033
Level 2	0.00088	-0.01432	0.02434	-0.01091
Level 3	0.00000	0.00254	-0.05061	0.04808
Level 4	0.00070	0.00684	0.00014	-0.00768

Table 47

Bias for conditional probability tables of the first type of the tasks in terms of the different initial probability distributions

	Equal distribution		Negatively skewed distribution		Positively skewed distribution	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.00059	-0.00059	0.00267	-0.00267	0.00177	-0.00177
Level 2	-0.00761	0.00761	0.00633	-0.00633	-0.00520	0.00520
Level 3	0.01043	-0.01043	-0.00477	0.00477	0.00626	-0.00626
Level 4	0.00038	-0.00038	0.00413	-0.00413	0.00113	-0.00113

Table 48

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different initial probability distributions

	Equal distribution		Negatively skewed distribution		Positively skewed distribution	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.01559	-0.01559	0.00036	-0.00036	0.00402	-0.00402
Level 2	-0.00636	0.00636	0.03135	-0.03135	0.01126	-0.01126
Level 3	-0.01575	0.01575	-0.00264	0.00264	-0.00595	0.00595
Level 4	0.00965	-0.00965	0.00124	-0.00124	-0.00353	0.00353

Table 49

Bias for conditional probability tables of the third type of the tasks at the first measurement in terms of the different initial probability distributions

	Equal distribution		Negatively skewed distribution		Positively skewed distribution	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00136	0.00136	0.00721	-0.00721	-0.00192	0.00192
Level 2	0.00133	-0.00133	-0.00582	0.00582	0.00617	-0.00617
Level 3	-0.01099	0.01099	-0.00294	0.00294	-0.00064	0.00064
Level 4	-0.00664	0.00664	-0.01007	0.01007	-0.01130	0.01130

Table 50

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different initial probability distributions

	Equal distribution		Negatively skewed distribution		Positively skewed distribution	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.01865	-0.01865	0.02079	-0.02079	0.03805	-0.03805
Level 2	-0.00539	0.00539	0.01189	-0.01189	-0.00518	0.00518
Level 3	-0.00690	0.00690	-0.00988	0.00988	-0.00808	0.00808
Level 4	0.00096	-0.00096	-0.00011	0.00011	-0.00060	0.00060

Table 51

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different initial probability distributions

	Equal distribution		Negatively skewed distribution		Positively skewed distribution	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.00684	-0.00684	0.00801	-0.00801	0.02413	-0.02413
Level 2	0.01965	-0.01965	0.02054	-0.02054	0.02491	-0.02491
Level 3	-0.00504	0.00504	0.00287	-0.00287	0.00749	-0.00749
Level 4	0.00430	-0.00430	0.00315	-0.00315	0.00116	-0.00116

Table 52

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different initial probability distributions

	Equal distribution		Negatively skewed distribution		Positively skewed distribution	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.01940	-0.01940	-0.00177	0.00177	0.00082	-0.00082
Level 2	-0.00101	0.00101	-0.00050	0.00050	0.00205	-0.00205
Level 3	0.00645	-0.00645	0.01090	-0.01090	0.00926	-0.00926

Table 53

Average of RMSD for the different types of initial probability distributions

Condition	Equal Distribution	High Distribution	Low Distribution
	RMSD	RMSD	RMSD
Parameters			
DS* at Time1	0.054772	0.050501	0.048445
TPT*	0.077904	0.078355	0.069776
Average of CPTT*	0.059979	0.061253	0.059772

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 54

Average of SDE for the different types of initial probability distributions

Condition	Equal Distribution	High Distribution	Low Distribution
	SD	SD	SD
Parameter			
DS* at Time1	0.002749	0.002993	0.002559
TPT*	0.004199	0.004703	0.003724
Average of CPTT*	0.011237	0.013054	0.011872

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 45 shows the bias values of the DS in terms of the different initial probability distributions. It seems that the negatively skewed distribution has higher bias values at the higher levels compared to the equal distribution and the positively skewed distribution. Table 46 shows the bias of the TPT. There do not seem to be any patterns exhibited. Table 47, Table 48, and Table 49 indicate the bias of the CPT parameters of the tasks used at the first measurement. Table 50, Table 51, and Table 52 show the bias of the CPT parameters of the tasks implemented at the second measurement. Similarly, the bias values of the CPT parameters do not seem to have any distinct patterns in terms of the different initial probability distributions. Table 53 and Table 54 show the RMSDs and SDEs computed in terms of the different types of initial probability distributions. The RMSDs were similar to each other across three different types of initial probability distributions. This suggests that the different types of initial probability tables did not influence the precision of estimates. Table 54 shows the SDE. Similarly, there is no distinct pattern across the different types of initial probability tables. The SDEs of the parameters of the conditional probability tables were relatively higher than other parameters, meaning that the estimates of the conditional probability table of each task exhibited more fluctuation than other parameters.

To summarize, the results showed that the different types of initial probability distributions did not influence the error variances. DS, CPT, and TPT had similar bias values and variances of estimation across the different types of initial probability distributions.

Conditional Probability Tables

Table 55

Bias for the student distribution at the first measurement in terms of the different types of conditional probability tables

Conditional Probability Table 1				Conditional Probability Table 2			
Bias				Bias			
Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
-0.00121	0.00281	-0.00402	0.00242	0.00716	0.00419	-0.01503	0.00368

Table 56

Bias for the transition probability in terms of the different types of conditional probability tables

Bias				
Conditional Probability Table 1				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.00156	0.03562	-0.01734	-0.01984
Level 2	0.00345	-0.02489	0.03125	-0.00980
Level 3	0.00032	0.00390	-0.03214	0.02792
Level 4	0.00010	0.00004	0.00226	-0.00241
Conditional Probability Table 2				
Level 1	0.02991	0.00245	-0.01061	-0.02175
Level 2	0.00276	-0.02693	-0.00573	0.02990
Level 3	0.00398	0.00440	-0.03051	0.02214
Level 4	0.00172	0.00709	0.00704	-0.01585

Table 57

Bias for conditional probability tables of the first type of the tasks at the first measurement at the different types of conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00137	0.00137	0.00472	-0.00472
Level 2	-0.00111	0.00111	-0.00321	0.00321
Level 3	0.00076	-0.00076	0.00719	-0.00719
Level 4	0.00042	-0.00042	0.00334	-0.00334

Table 58

Bias for conditional probability tables of the second type of the tasks at the first measurement at the different types of conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00171	-0.00171	0.01161	-0.01161
Level 2	0.00234	-0.00234	0.02182	-0.02182
Level 3	-0.00004	0.00004	-0.01619	0.01619
Level 4	-0.00092	0.00092	0.00583	-0.00583

Table 59

Bias for conditional probability tables of the third type of the tasks at the first measurement at the different types of conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00065	-0.00065	0.00197	-0.00197
Level 2	-0.00069	0.00069	0.00181	-0.00181
Level 3	-0.00355	0.00355	-0.00616	0.00616
Level 4	-0.00836	0.00836	-0.01031	0.01031

Table 60
Bias for conditional probability tables of the first type of the tasks at the second measurement at the different types of conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01955	-0.01955	0.03211	-0.03211
Level 2	0.00197	-0.00197	-0.00109	0.00109
Level 3	0.00050	-0.00050	-0.01707	0.01707
Level 4	-0.00062	0.00062	0.00078	-0.00078

Table 61
Bias for conditional probability tables of the second type of the tasks at the second measurement at the different types of conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00275	-0.00275	0.02324	-0.02324
Level 2	0.01849	-0.01849	0.02491	-0.02491
Level 3	0.00270	-0.00270	0.00085	-0.00085
Level 4	0.00079	-0.00079	0.00495	-0.00495

Table 62
Bias for conditional probability tables of the third type of the tasks at the second measurement at the different types of conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00334	-0.00334	0.00896	-0.00896
Level 2	-0.00213	0.00213	0.00249	-0.00249
Level 3	0.01624	-0.01624	0.00150	-0.00150
Level 4	0.00329	-0.00329	-0.01008	0.01008

Table 63

Average of RMSD for the different types of conditional probability tables

Condition	Conditional Probability Table 1	Conditional Probability Table 2
	RMSD	RMSD
Parameter		
DS* at Time1	0.038068	0.065151
TPT*	0.045469	0.093009
Average of CPTT*	0.043223	0.073575

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 64

Average of SDE for the different types of conditional probability tables

Condition	Conditional Probability Table 1	Conditional Probability Table 2
	SD	SD
Parameter		
DS* at Time1	0.002020	0.003515
TPT*	0.003569	0.004849
Average of CPTT*	0.010836	0.013273

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 55 shows the bias of the DS in terms of the different initial probability distributions. The bias values were slightly smaller in the first conditional distribution table than the second conditional distribution table. Table 56 shows the bias of the TPT.

There was almost no difference between the two conditions. Unlike the DS and TPT, the bias values of the CPT parameters show some differences between the two types of conditional distribution tables. Table 57, Table 58, and Table 59 show the bias of the CPT parameters of the tasks implemented at the first measurement. Table 60, Table 61, and Table 62 show the bias of the CPT parameters of the tasks implemented at the second measurement. As shown in the tables, the bias values were higher when using the second conditional distribution table than using the first conditional distribution table. This implies that the type of conditional distribution table influences the precision of the CPT parameters.

Table 63 and Table 64 show the RMSDs and SDEs. A similar pattern of results was observed for RMSDs and SDEs. It was observed that RMSDs of the first type of conditional probability table were smaller than the second type of conditional probability table. This implies that the precision of estimates increased when the tasks were well designed for classifying students with different levels. The SDEs of the first conditional probability table were smaller than those of the second conditional probability table. The results suggest that the estimates were more reliable when the task that was well designed for classifying students with different levels was implemented.

In summary, the results showed that when the conditional probabilities of each task with respect to each level were distinct (i.e., tasks could be considered to be relatively well designed for the purpose of classifying students with different levels), the error variances of estimates were reduced. Among the parameters of DS, CPT, and TPT, TPT and CPT seem to be more effected by different types of conditional probability tables of each task. TPT and CPT seem to be more biased compared to DS. CPT seemed to have more error variance than DS and TPT. The statistical significance of these effects will be examined in a following section.

Transition Probability Tables

Table 65

Bias for the student distribution at the first measurement in terms of the different transition probability distributions

	Bias			
Condition	Level 1	Level 2	Level 3	Level 4
Transition Probability 1	0.00179	-0.00540	-0.00064	0.00425
Transition Probability 2	0.00041	0.00394	-0.00951	0.00515
Transition Probability 3	0.00673	0.01197	-0.01844	-0.00026

Table 66

Bias for the transition probability in terms of the different transition probability distributions

Bias				
Transition Probability 1				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.02686	-0.00347	-0.01524	-0.00815
Level 2	0.00343	-0.03771	0.03910	-0.00483
Level 3	0.00017	0.00629	-0.06982	0.06336
Level 4	0.00152	0.00683	0.00246	-0.01081
Transition Probability 2				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.02949	0.03070	-0.01426	-0.04593
Level 2	0.00221	0.01261	-0.01164	-0.00318
Level 3	0.00225	0.00034	-0.00589	0.00330
Level 4	0.00056	0.00001	0.00680	-0.00737
Transition Probability 3				
	Level 1	Level 2	Level 3	Level 4
Level 1	-0.00915	0.02987	-0.01243	-0.00829
Level 2	0.00367	-0.05264	0.01081	0.03815
Level 3	0.00401	0.00583	-0.01827	0.00843
Level 4	0.00066	0.00386	0.00469	-0.00921

Table 67

Bias for conditional probability tables of the first type of the tasks in terms of the different transition probability distributions

	Transition Probability 1		Transition Probability 2		Transition Probability 3	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.00356	-0.00356	0.00099	-0.00099	0.00048	-0.00048
Level 2	-0.00017	0.00017	-0.00103	0.00103	-0.00528	0.00528
Level 3	0.00259	-0.00259	0.00023	-0.00023	0.00910	-0.00910
Level 4	0.00305	-0.00305	0.00495	-0.00495	-0.00237	0.00237

Table 68

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different transition probability distributions

	Transition Probability 1		Transition Probability 2		Transition Probability 3	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.00038	-0.00038	0.01070	-0.01070	0.00889	-0.00889
Level 2	0.01783	-0.01783	0.00310	-0.00310	0.01532	-0.01532
Level 3	-0.00358	0.00358	-0.01373	0.01373	-0.00703	0.00703
Level 4	-0.01056	0.01056	0.00862	-0.00862	0.00930	-0.00930

Table 69

Bias for conditional probability tables of the third type of the tasks at the first measurement in terms of the different transition probability distributions

	Transition Probability 1		Transition Probability 2		Transition Probability 3	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00141	0.00141	0.00234	-0.00234	0.00299	-0.00299
Level 2	0.00006	-0.00006	-0.00181	0.00181	0.00344	-0.00344
Level 3	0.00104	-0.00104	-0.00815	0.00815	-0.00746	0.00746
Level 4	-0.00930	0.00930	-0.00815	0.00815	-0.01056	0.01056

Table 70

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different transition probability distributions

	Transition Probability 1		Transition Probability 2		Transition Probability 3	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.02391	-0.02391	0.04776	-0.04776	0.00582	-0.00582
Level 2	0.00916	-0.00916	-0.00467	0.00467	-0.00316	0.00316
Level 3	-0.00716	0.00716	-0.00279	0.00279	-0.01490	0.01490
Level 4	-0.00307	0.00307	0.00043	-0.00043	0.00289	-0.00289

Table 71

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different transition probability distributions

	Transition Probability 1		Transition Probability 2		Transition Probability 3	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.00128	-0.00128	0.03157	-0.03157	0.00614	-0.00614
Level 2	0.00437	-0.00437	0.06120	-0.06120	-0.00047	0.00047
Level 3	0.00174	-0.00174	0.00566	-0.00566	-0.00208	0.00208
Level 4	0.00044	-0.00044	0.00088	-0.00088	0.00729	-0.00729

Table 72

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different transition probability distributions

	Transition Probability 1		Transition Probability 2		Transition Probability 3	
	Bias		Bias		Bias	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Level 1	0.00351	-0.00351	0.01305	-0.01305	0.00189	-0.00189
Level 2	-0.00438	0.00438	-0.00127	0.00127	0.00619	-0.00619
Level 3	0.00352	-0.00352	0.02625	-0.02625	-0.00316	0.00316
Level 4	-0.00312	0.00312	0.00399	-0.00399	-0.01106	0.01106

Table 73

Average of RMSD for the different types of transition probability tables

Condition	TD1	TD2	TD3
	RMSD	RMSD	RMSD
Parameter			
DS* at Time1	0.054818	0.048109	0.047131
TPT*	0.073189	0.069639	0.069639
Average of CPTT*	0.055768	0.075002	0.046763

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 74

Average of SDE for the different types of transition probability tables

Condition	Equal Distribution	High Distribution	Low Distribution
	SDE	SDE	SDE
Parameter			
DS* at Time1	0.003135	0.002503	0.002665
TPT*	0.005476	0.003942	0.003209
Average of CPTT*	0.012126	0.013333	0.010703

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 65 shows the bias of the DS in terms of the different transition probability tables. It was observed that bias values were similar across the different transition probability tables in DS, TPT, and CPT. Distinct patterns were not found in their bias values with respect to the different transition probability tables. Table 73 shows the RMSDs of the initial probability distribution parameters, transitional probability table parameters, and conditional probability table parameters in terms of the three different types of transition probability tables. It was also reported that RMSDs were very similar to each other across the different types of transition probability tables. This implies that three types of transition probability tables did not influence the level of estimate precision.

A similar pattern of results was also observed for SDE values (Table 74). The SDE values were similar to each other across three types of transition probability tables.

To summarize, regardless of what types of transition probability tables were used, the results seem to indicate that the precision and reliability of estimates were similar to each other in DS and TPT. However, CPT seemed to be more effected by different types of transitional probability tables than the others. The statistical significance of these effects will be examined in a following section.

In order to understand the effect of the different conditions on the precision and reliability of estimates, the graphical representations of RMSDs were drawn, and are presented in the next part of this dissertation.

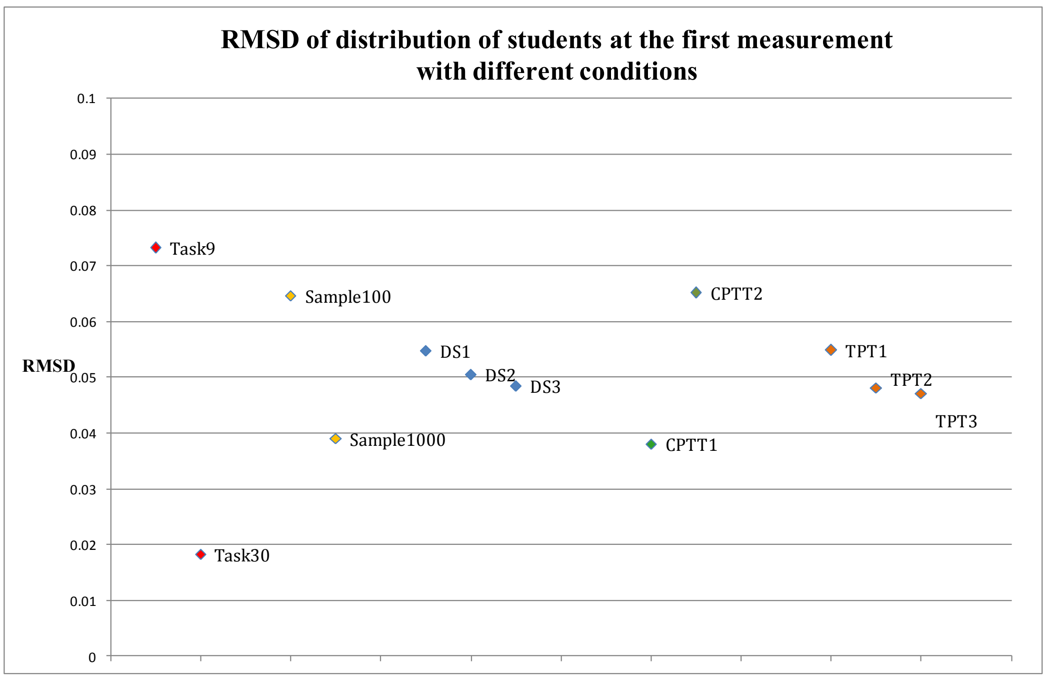


Figure 29. RMSDs of the DS at the first measurement with different conditions

The three graphs (Figure 29, Figure 30, and Figure 31) display the RMSDs of all the conditions with respect to the parameters of DS, TPT, and CPT of each task. For the parameters of the distribution of students at the first measurement (Figure 29), the simulation conditions of task size, sample size, and conditional probability table of each task seem to influence the precision of estimates more than other factors. As shown in

Figure 29, the plots of three types of initial distribution tables and the plots of three types of transition probability tables were clustered to each other more than other factors, meaning that their RMSDs were similar to each other. Therefore, the effect of the different types of the transition probability tables and the initial probability tables seems to be small.

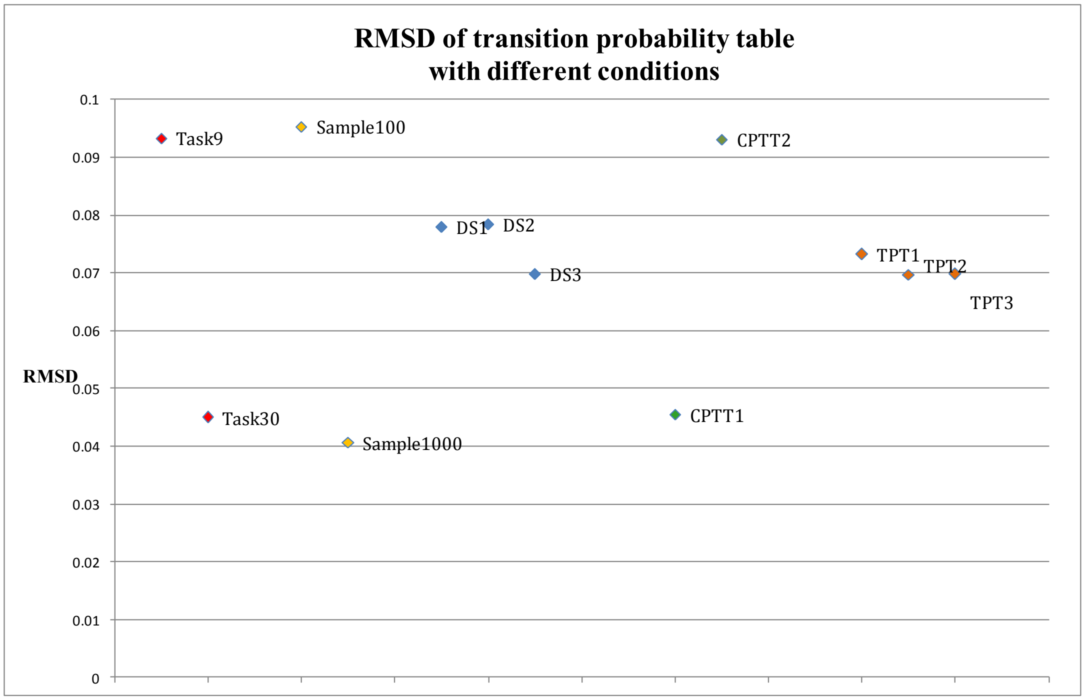


Figure 30. RMSDs of the TPT with different conditions

A similar pattern of results was also observed for the transition probability tables. The RMSDs of the different conditions in the task size, the sample size, and the conditional probability tables seem to have more effect on the precision of estimates than the other factors (i.e., the distribution of students at the first measurement and the transition probability table). As shown in Figure 30, two plots representing the different task sizes, sample sizes, and two types of conditional probability tables were relatively located far from each other, meaning that there were some differences in RMSDs between two conditions; whereas, the three conditions of two factors (i.e., the distribution of students and transition probability tables) were clustered to each other, implying that the precision of estimates seem to be similar no matter what type of transition probability tables were used and no matter how students were initially distributed at the first measurement.

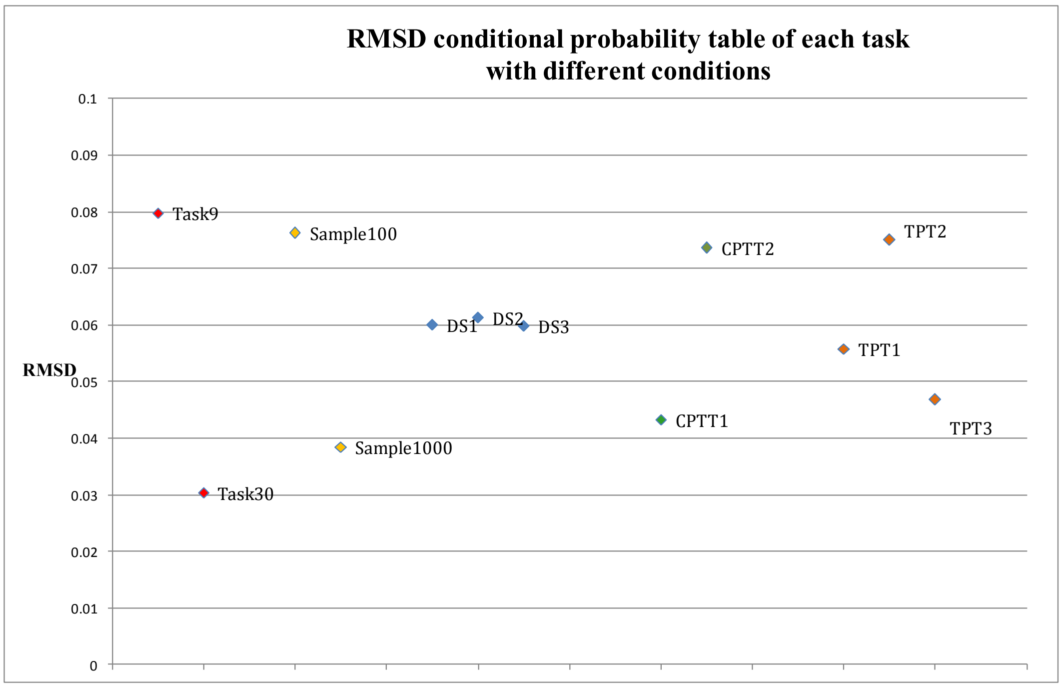


Figure 31. RMSDs of the CPT with different conditions

A slightly different pattern of results was observed for the conditional probability table of each task (see Figure 31). As with other probability tables, RMSDs of the different conditions in three factors (task size, sample size, and conditional probability table of each task) seemed to affect the precision of estimates more than the other factors. However, the three conditions of the transition probability tables also resulted in different values of RMSDs. Like the other parameters, the similar values of RMSDs were reported no matter how students were distributed at the first measurement.

Since 72 conditions were investigated in this study, the simple descriptive statistics and the graphical representations do not seem to be an efficient way of detecting important effects and of estimating the magnitude of effects. Harwell (1991) noted that simply reporting the results by descriptive analyses increases the chance that important effects will go undetected and that the magnitude of effects will be misestimated. Therefore, the results of this simulation study are summarized by both descriptive and inferential analyses in order to provide meaningful evidence concerning the research questions (Harwell, 1991; Harwell & Stone, 1996).

ANOVA (Analysis of Variance) was used for the inferential analyses of this study. Because the purpose of the simulation study was to evaluate the parameter estimation of the success of parameters recovery, the RMSD was used as the dependent variable (e.g., Harwell & Janosky, 1991; Kim, Cohen, Baker, Subkoviak, & Leonard, 1994; Stone, 1992). The simulation conditions served as the independent variables. The main effect for each independent variable and the interaction effects among them were examined. The results were summarized in terms of each parameter table (i.e., the distribution of students at the first measurement, transition probability table, and conditional probability table of each task). Before conducting the ANOVA, the dependent variable was investigated to see if there was a lack of model fit. It was found that the distributions of RMSD values of all parameters were very positively skewed, implying that the normality assumption was violated. Therefore, the RMSD was transformed using a log transformation, so that it had an approximate normal distribution.

A three-way ANOVA with the independent variables was fit to the transformed RMSD values for the parameters of the distribution of students at the first measurement with a main effect model followed by an interaction model by two variables and the three variables at a time. Because there were no significant second-order interaction effects (i.e., three variables at a time), the second-order interactions were not included in the model. The results are reported in Table 75. The magnitude of significant effects was estimated using η^2 (see Table 75).

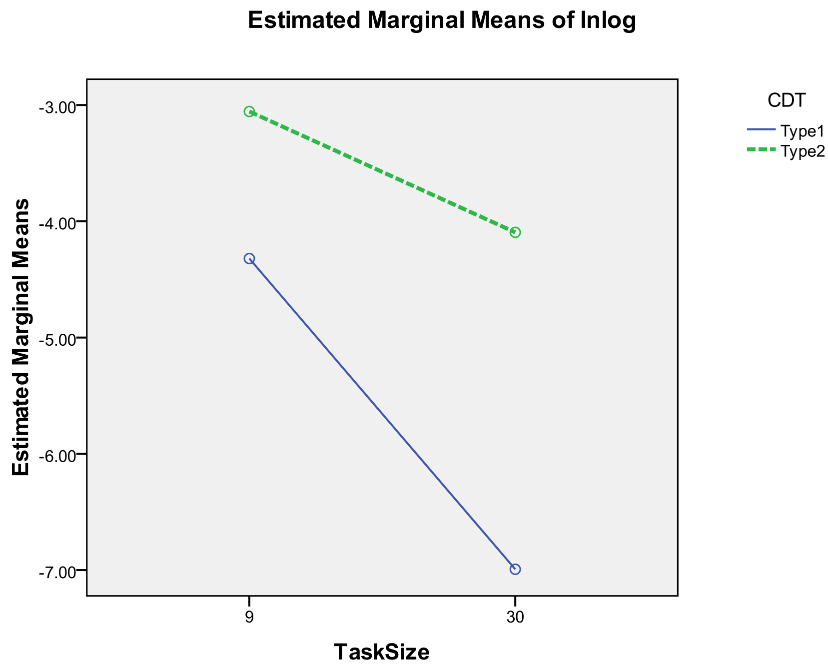
Table 75

ANOVA results for RMSD values for the DSs at the first measurement

Source	*DF*	*F*	*P*	η^2
SampleSize	1	61.497	.000	.147
TaskSize	1	118.344	.000	.284
CPT	1	149.628	.000	.359
TPT	2	1.872	.164	.009
IPT	2	.264	.769	.001
SampleSize x TaskSize	1	.000	.999	.000
SampleSize x CPT	1	.885	.351	.002
SampleSize x TPT	2	.486	.618	.002
TaskSize x CPT	1	22.917	.000	.055
TaskSize x TPT	2	.806	.452	.004
CPT x TPT	2	.214	.808	.001
Error	55			
a. R Squared = .868 (Adjusted R Squared = .830)				

Note. CPT indicates the conditional distribution table. TPT is the transition probability table. IPT indicates the initial probability table.

Using an alpha level of 0.05, the ANOVA revealed three significant main effects and an interaction effect. The condition of sample size was found to be statistically significant, $F(1, 55) = 61.497$, $p < 0.00$. The condition of task size was observed to be statistically significant, $F(1, 55) = 118.344$, $p < 0.00$. The condition of conditional probability table was found to be statistically significant, $F(1, 55) = 149.628$, $p < 0.00$. The results described above suggested that there were influences of sample size, task size, and types of conditional probability table on the accuracy of the estimates. In addition to the main effects, the task size x CPT interaction effect was found to be statistically significant, $F(1, 55) = 22.917$, $p < 0.00$. It provided evidence that the mean difference among the levels of the factor (sample sizes) is not constant across the types of conditional probability tables. In order words, there is a joint effect of the conditions of sample sizes and CPT. Using η^2 as the measure of effect size, the different types of the conditional tables has the largest effect size, which accounted for 36% of the total variability in the RMSD of the DS parameters. It means that more discriminating tasks lead to better estimates of parameters in the model than less discriminating tasks. It implies for practitioners to make effects to try to make high quality tasks. If there are the situations that lower discriminating tasks are avoidable, making longer tests to get good estimates of the model parameters would be an alternative solution. The task size made more difference than the sample size in the DS parameters.



Note. lnlog indicates the log transformed RMSD of DS at the first measurement.

Figure 32. Profile Plot of task size and CPT for DS

To provide assistance in graphically understanding the interaction effect, the profile plot of task size and CPT is displayed (Figure 32). On the X axis are the levels of task size (9 and 30 tasks), and the Y axis provides the cell means on the dependent variable. The plot shows that the influence of task size was greater when using well designed tasks than when using poorly designed tasks. In other words, the quality of the task was an important factor that influenced the accuracy of the estimate as well as the number of tasks implemented.

A three-way ANOVA with the independent variables was also fit to the RMSD values for the parameters of the transition probability table with a main effect model followed by an interaction model by two variables and the three variables at a time. There were no significant second-order interaction effects (i.e., three variables at a time). Therefore, the second-order interactions were not included in the model. The results are reported in Table 76.

Table 76

ANOVA results for RMSD values for TPTs

Source	*DF*	*F*	*P*	η^2
SampleSize	1	195.491	.000	.405
TaskSize	1	75.348	.000	.156
CPT	1	97.565	.000	.202
TPT	2	4.635	.014	.019
IPT	2	.260	.772	.001
SampleSize * TaskSize	1	16.847	.000	.035
SampleSize * CPT	1	25.651	.000	.053
SampleSize * TPT	2	.593	.556	.002
TaskSize * CPT	1	.048	.827	.000
TaskSize * TPT	2	1.539	.224	.006
CPT * TPT	2	.434	.650	.002
Error	55			

a. R Squared = .886 (Adjusted R Squared = .853)

Using an alpha level of 0.05, this test revealed four significant main effects and two interaction effects. For the main effects, the conditions of sample size, task size, CPT, and TPT were found to be statistically significant, $F(1, 55) = 195.491$, $p < 0.00$, $F(1, 55) = 75.348$, $p < 0.00$, $F(1, 55) = 97.565$, $p < 0.00$, and $F(1, 55) = 4.635$, $p < 0.014$

respectively. The results suggested that there were influences of sample size, task size, and types of conditional probability table on the accuracy of estimate. The means of 100 samples and 1000 samples were statistically significantly different averaging over all other factors. The means of 9 tasks and 30 tasks were significantly different averaging over all other factors. Also, the means of two types of CPT significantly differed averaging over other factors. Lastly, the means of three types of TPT were significantly different averaging over other factors. Using η^2 as the measure of effect size, the condition of sample size had the largest effect size, which accounted for 41% of the total variability in the RMSD of the TPT parameters.

Since the factor of TPT had three levels, it required a follow-up test in order to determine which types of TPT were significantly different. The post-hoc Tukey HSD test was conducted. The results indicated that the means of type 1 and type 2 were statistically significantly different and the means of type 1 and type 2 were significantly different, whereas the means of type 2 and type 3 did not differ (Table 77).

Table 77

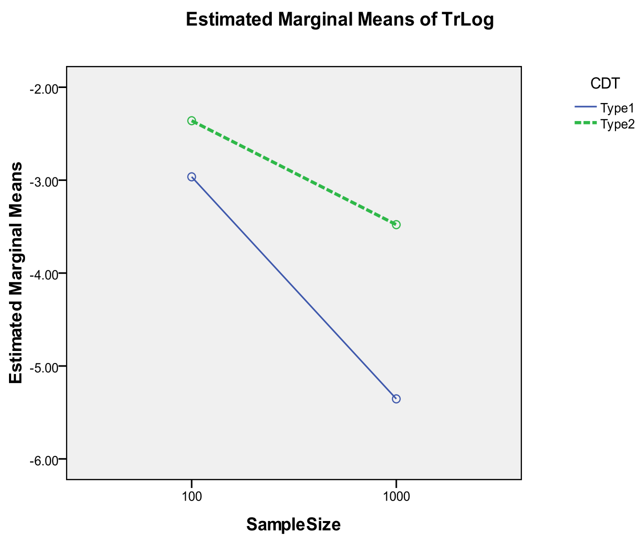
Tukey HSD of TPT

		Mean Difference	Std. Error	*P*
Type1	Type2	.3902	.1534	.036
Type1	Type3	.4792	.1534	.008
Type2	Type3	.0890	.1534	.831

The error term is Mean Square (Error) = .282.

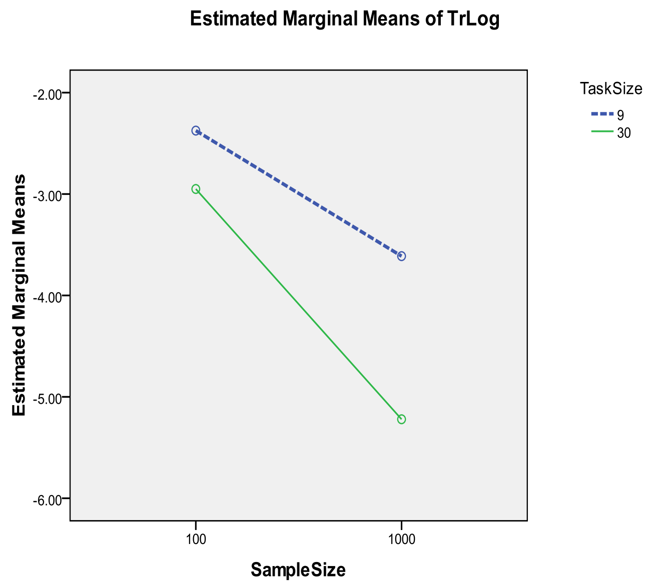
In addition to the main effects, Sample size x Task size interaction effect and Sample size x CPT interaction effect were found to be statistically significant, $F(1, 55) = 16.847$, p <

0.00, $F(1, 55) = 25.651$, $p < 0.00$ respectively. This provided evidence that the mean difference among the levels of the factor (sample size) was not constant across the types of conditional probability tables. Also, the mean difference among the levels of sample size was not equal across the levels of task size. In other words, there was a combined effect of the conditions of sample size and CPT and a joint effect to the conditions of sample size and task size. To understand the interaction effect, the profile plots were drawn (Figure 29 and Figure 30).



Note. Trlog indicates the log transformed RMSD of TPT

Figure 33.Profile plot of sample size and CPT for TPT



Note. Trlog indicates the log transformed RMSD of TPT

Figure 34. Profile plot of task size and sample size for TPT

A three-way ANOVA with the independent variables was also fit to the RMSD values for the parameters of the conditional probability table of each task with a main effect model followed by two variables at a time interaction models and three variables at a time interaction models. No significant second-order interaction effects were found (i.e., three variables at a time). Therefore, the second-order interactions were not included in the model. The results are reported in Table 78. The magnitude of significant effects was estimated using η^2 (see Table 78).

Table 78

ANOVA results for RMSD values for CPTs

Source	*DF*	*F*	*P*	η^2
SampleSize	1	158.010	.000	.318
TaskSize	1	128.525	.000	.259
CPT	1	101.660	.000	.205
TPT	2	16.646	.000	.067
IPT	2	.780	.463	.003
SampleSize * TaskSize	1	.384	.538	.001
SampleSize * CPT	1	6.767	.012	.014
SampleSize * TPT	2	.075	.928	.000
TaskSize * CPT	1	2.711	.105	.005
TaskSize * TPT	2	.678	.512	.003
CPT * TPT	2	.704	.499	.003
Error	55			

a. R Squared = .889 (Adjusted R Squared = .857)

Using an alpha level of 0.05, the ANOVA revealed that the factor of sample size was found to be statistically significant, $F(1, 55) = 158.010$, $p < 0.00$, the factor of task size was observed to be statistically significant, $F(1, 55) = 128.525$, $p < 0.00$, the factor of conditional probability table was found to be statistically significant, $F(1, 55) = 101.660$, $p < 0.00$, and the factor of TPT was observed to be statistically significant, $F(1, 55) = 16.646$, $p < 0.00$. The results suggested that there were influences of sample size, task size, types of conditional probability table, and the types of transition probability table on the accuracy of estimates. Using η^2 as the measure of effect size, the sample size had the largest effect size, which accounted for 32% of the total variability in the RMSD of the CPT parameters.

Since the factor of TPT had three levels, the post-hoc Tukey HSD test was conducted in order to determine which types of TPT were significantly different. The results indicate that the means of all pairs were statistically significantly different, d = -.3250, $p < 0.007$, d = .3071, $p < 0.01$, and d = -.6321, $p < 0.00$ respectively.

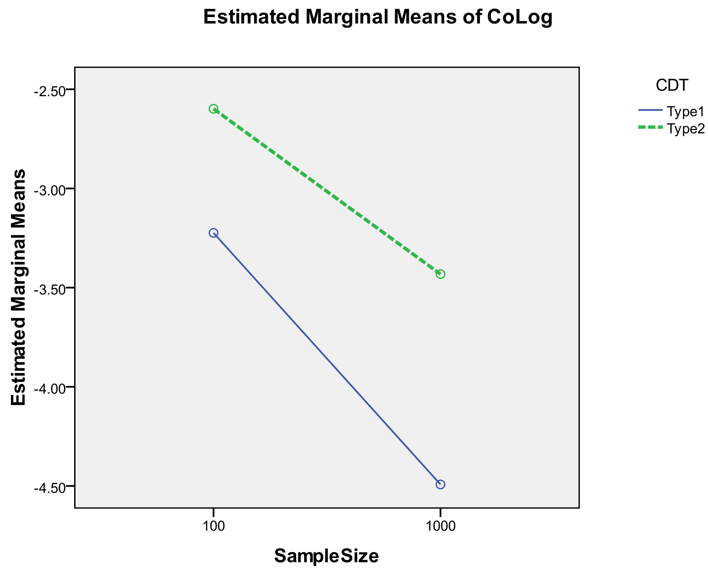
Table 79

Tukey HSD test

		Mean Difference	Std. Error	*P*
Type1	Type2	-.3250	.10220	.007
Type1	Type3	.3071	.10220	.011
Type2	Type3	-.6321	.10220	.000

The error term is Mean Square (Error) = .282.

Furthermore, the sample size and CPT interaction effect was found to be statistically significant, $F(1,55) = 6.767$, $p < 0.012$. This provided evidence that there was a joint effect of sample size and CPT. For graphically understanding the interaction effect, the profile plot of sample size and CPT was displayed (Figure 31). On the X axis are the levels of sample size (100 and 1000 tasks), and the Y axis provides the cell means on the dependent variable. The plot shows that the effect of sample size on the accuracy of estimate increases when using well designed tasks as opposed to using poorly designed tasks.



Note. Colog indicates the log transformed RMSD of CPTs.

Figure 35. Profile plot of sample size and CPT for CPTs

Discussion

This first simulation study evaluated the simple DBIN model with the different conditions of sample size, task size, types of conditional probability tables, types of initial probability tables, and types of transition probability tables. In summary, for the parameters of the initial distribution table, it was observed that the different conditions of sample size, task size, and types of conditional probability table statistically influence the accuracy of estimates. As sample size and task size increased, the accuracy of estimates increased. When tasks that were well designed for classifying different levels were used, the estimates were more stable and reliable. A similar pattern was found for the parameters of the transition probability table. There were statistically significantly influences of the different conditions of sample size, task size, conditional probability tables, and transition probability tables on the accuracy of estimates. Additionally, the different types of transition probability tables were found to be a factor that influenced the accuracy of the estimates. According to the follow-up test, the equal transition has significantly lower RMSD values than the unequal transitions (i.e., large transition and small transition).

A similar pattern of results was observed for the conditional probability table parameters. Three factors of sample size, task size, and the types of conditional probability table were also found to be significant. The task size of 30 had significantly lower RMSD than the task size of 9 and the sample size of 1000 had significantly lower RMSD than the sample size of 100. Furthermore, the task design for classifying student levels was found to be an important factor that affected the precision of estimates. In addition, the different transitions also influenced the accuracy of estimates of the conditional probability table parameters.

The next chapter describes the second simulation study with a more complex DBIN model incorporating a covariate with the similar simulation conditions. The accuracy of estimates will be compared between a simple DBIN model and a complex DBIN model.

CHAPTER 7: SIMULATION DATA STUDY 2: A DBIN WITH A COVARIATE

A manifest variable (e.g., different instruction, interventions, and individual demographic background) or a latent exogenous variable (e.g., attitude, intelligence, and social economic status) may differently impact student change with respect to different levels. This impact can be investigated by constructing more complex DBINs by incorporating covariate variables as parents of transition probabilities. This chapter investigates a simulation data study that incorporates a covariate for students to the simple DBIN model. Although it is possible to extend the model to adding continuous covariates (Clogg & Goodman, 1985), this simulation study considered a discrete covariate. The Bayes net framework focuses on discrete variables and Netica C API can be currently utilized only for the discrete variables. However, it can be incorporated in by discretizing the continuous variable, which is an option for example in the Netica.

The sections in this chapter contain the overview, data generation methods, simulation conditions, and preliminary results. The evaluation of the performance of the extended DBIN focuses on how different constraints on (1) the relationship between the observables and LPs and (2) the relationship of the LPs between two consecutive measurement points with respect to a covariate effects parameter recovery in estimation using Netica C API.

Overview

The second simulation study focuses on the construction of a DBIN with a covariate. The case examined is one in which there are two measurement occasions and each measurement has multiple observable variables measuring one LP. In addition to this simple model, a covariate variable is involved in the transition probability matrix. Figure 37 displays the model that is examined in this study.

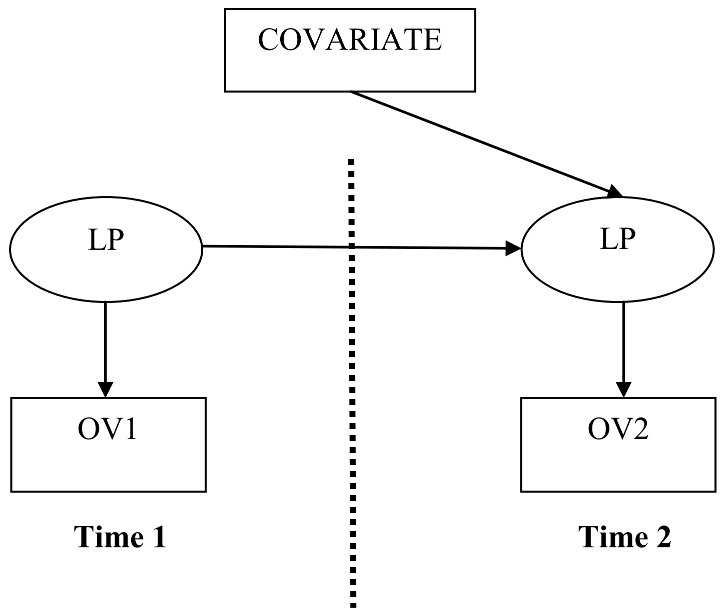


Figure 36. A model for the second simulation data study

As in the first simulation study, the latent variable representing an LP at each measurement occasion has four levels. Additionally, the DBINs contain a variable indicating two different types of instruction connected to the latent variable (indicating the learning progression) at the second measurement. Table 80 includes the parameters that need to be estimated for this model.

Table 80

Parameters that need to be estimated for the second simulation study

Latent Variable at the first measurement occasion	Transition probability matrix given a covariate =1	Transition Probability matrix given a covariate =2	Observable Variables
$P(LP_{t=1}=1)$	$P(LP_{t=2}=1 \| LP_{t=1}=1, C=1)$	$P(LP_{t=2}=1 \| LP_{t=1}=1, C=2)$	$P(OV_j \| LP_i)$
$P(LP_{t=1}=2)$	$P(LP_{t=2}=2 \| LP_{t=1}=1, C=1)$	$P(LP_{t=2}=2 \| LP_{t=1}=1, C=2)$	For each
$P(LP_{t=1}=3)$	$P(LP_{t=2}=3 \| LP_{t=1}=1, C=1)$	$P(LP_{t=2}=3 \| LP_{t=1}=1, C=2)$	task
$P(LP_{t=1}=4)$	$P(LP_{t=2}=4 \| LP_{t=1}=1, C=1)$	$P(LP_{t=2}=4 \| LP_{t=1}=1, C=2)$	$P(OV_j \| LP_i)$
	$P(LP_{t=2}=2 \| LP_{t=1}=2, C=1)$	$P(LP_{t=2}=2 \| LP_{t=1}=2, C=2)$	For each
	$P(LP_{t=2}=3 \| LP_{t=1}=2, C=1)$	$P(LP_{t=2}=3 \| LP_{t=1}=2, C=2)$	task
	$P(LP_{t=2}=4 \| LP_{t=1}=2, C=1)$	$P(LP_{t=2}=4 \| LP_{t=1}=2, C=2)$	
	$P(LP_{t=2}=3 \| LP_{t=1}=3, C=1)$	$P(LP_{t=2}=3 \| LP_{t=1}=3, C=2)$	j = task
	$P(LP_{t=2}=4 \| LP_{t=1}=3, C=1)$	$P(LP_{t=2}=4 \| LP_{t=1}=3, C=2)$	i= level
	$P(LP_{t=2}=4 \| LP_{t=1}=4, C=1)$	$P(LP_{t=2}=4 \| LP_{t=1}=4, C=2)$	

The research questions in this second simulation study were:

(1) How well can the parameters of the conditional probabilities for observable variables, including a covariate, be recovered?

(2) How well can the parameters of the transition probabilities between two latent variables including a covariate be recovered?

(3) How well can the distribution of students indicating student classification of levels at the first measurement occasion be recovered?

(4) How well can the proportions of students at the second measurement indicating student classifications at each level on the LP, including a covariate, be recovered?

These research questions were addressed by computing (1) the Root Mean Squared Difference (RMSD), (2) bias, and (3) standard error.

Data Generation and Simulation Condition

Data was simulated by using R. The transition probabilities with respect to the values of a covariate, conditional probabilities of correctly responding to each task given each state of the LP, and the distribution of students on the LP at the first time point, and group memberships of the covariate were considered for generating response data. Once the item responses were generated, EM estimation implemented in the Netica C API (Norsys Software Corp, 2008) was used to estimate the parameters.

The fixed factors of this simulation study were included in the structure of the DIBN. The structure of the DBIN considered here includes two measurement occasions and four levels on each LP. The factors that vary in this simulation study are (1) sample size, (2) task size, that is, number of tasks, (3) types of transition probability tables, (4) types of conditional probability tables of tasks, and (5) proportions of group membership on a covariate.

One case is considered for the distribution of students on an LP at the first measurement occasion. Table 81 displays the distribution of students. This case is the equal probability of students being at each of the four levels.

Table 81

Distribution of student on LP at the first measurement

Case	Description	Level 1	Level 2	Level 3	Level 4
1(Equal)	Equal probability at all levels representing general population	0.25	0.25	0.25	0.25

Like the first simulation study, two cases for the conditional probability matrix of observables were considered. The first case represents that tasks are well designed for classifying students into their levels. The second case represents that tasks are relatively poorly designed for classifying students into their levels. Table 82 displays the first case of the conditional probability table with 9 tasks. Each of the three tasks was designed to measure each level. This structure is duplicated for 30 tasks.

Table 82

The first case of conditional probability table

	Task								
	Level1			Level2			Level3		
Level of Student	1	2	3	1	2	3	1	2	3
Level 0	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
Level 1	0.85	0.85	0.85	0.20	0.20	0.20	0.20	0.20	0.20
Level 2	0.85	0.85	0.85	0.85	0.85	0.85	0.20	0.20	0.20
Level 3	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85	0.85

Table 83 displays the second case of the conditional probability table. The differences in values of the conditional probabilities between two levels are smaller than the first case. This structure is duplicated for generating 30 tasks.

Table 83

The second case of conditional probability table

	Task								
	Level1			Level2			Level3		
Level of Student	1	2	3	1	2	3	1	2	3
Level 0	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.35
Level 1	0.70	0.70	0.70	0.35	0.35	0.35	0.35	0.35	0.35
Level 2	0.70	0.70	0.70	0.70	0.70	0.70	0.35	0.35	0.35
Level 3	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70	0.70

Four cases are considered for the transition probability tables. The first case included an equal transition probability and no covariate effect. The rest of the three cases had a covariate effect. Table 84 shows the four cases for a transition probability table.

Table 84

Four cases of transition probability table

Case	Case 1		Case 2		Case 3		Case 4	
Covariate	C=1	C=2	C=1	C=2	C=1	C=2	C=1	C=2
P(0|0)	0.25	0.25	**0.1**	0.7	**0.1**	0.7	0.7	0.7
P(1|0)	0.25	0.25	**0.5**	0.1	**0.5**	0.1	0.1	0.1
P(2|0)	0.25	0.25	**0.3**	0.1	**0.3**	0.1	0.1	0.1
P(3|0)	0.25	0.25	**0.1**	0.1	**0.1**	0.1	0.1	0.1
P(1|1)	0.33	0.33	0.8	**0.1**	0.8	0.8	**0.1**	0.8
P(2|1)	0.33	0.33	0.1	**0.6**	0.1	0.1	**0.6**	0.1
P(3|1)	0.33	0.33	0.1	**0.3**	0.1	0.1	**0.3**	0.1
P(2|2)	0.5	0.5	0.9	0.9	0.9	**0.2**	0.8	**0.2**
P(3|2)	0.5	0.5	0.1	0.1	0.1	**0.8**	0.2	**0.8**
P(3|3)	1	1	1	1	1	1	1	1

-

The simulation conditions in the transition probability table reflect the following situations:

- The first case demonstrates that there was no covariate effect on a transition probability. Therefore, the two transition probability tables with respect to each value of the covariate are equal to each other. In addition to the property, the transition probabilities are equal per each cell, so there are no weights to any of the four levels, and thus none of the four levels is favored. In other words, the proportions of moving students are equal across the all levels.
- The second case has a covariate effect: the first value of the covariate has an effect on the level change of students at Level 0 and the second value of the covariate had an effect on the level change of students at Level 1.
- The third case had a covariate effect: the first value of the covariate had an effect on the level change of students at Level 0 and the second value of the covariate has an effect on the level change of students at Level 2.
- The fourth case also had a covariate effect: the first value of the covariate affected the level change of students at Level 1 and the second value of the covariate affected the level change of students at Level 2.

The sample sizes used in this study are 100 and 1000 as same as the first simulation study. The task sizes used in this study are 9 and 30, the same task sizes used in the first simulation study.

Two cases of group memberships on a covariate were considered. The first case represents that group memberships are equality distributed in terms of the covariate, while the second case represents that there is a skewed distribution of group membership on the covariate. Table 85 shows the distributions of students on a covariate.

Table 85

Two cases of the distributions of students

Case	**Description**	**Covariate =1**	**Covariate =2**
1	Equal proportion of group membership	0.5	0.5
2	Skewed proportion of group membership	0.2	0.8

With the combination of different transition probability tables, different conditional probability tables, different transition probability tables, sample size, task size, different distributions of group memberships on a covariate, the total number of cells for this study was 62conditions. The simulation conditions are summarized in Table 86.

Table 86

Simulation conditions of the first study

Simulation Condition		# of Case
Distribution of student at the first measurement	Equal	1
Conditional probability table	Case 1 and 2	2
Transition probability table	Case 1, 2, 3, and 4	4
Sample size	50 and 1000	2
Task size	9 and 30	2
Proportion of group memberships on a covariate	Equal and Skewed	2
Total number of cells		64
Replications per cell		100
Total Simulation Runs		6400

Estimation

Once the task responses were generated, EM estimation implemented in the Netica C API (Norsys Software Corp, 2008) was used to estimate the parameters, just as was done in the first simulation study. R code for generating responses and evaluating parameter recovery was written. The visual studio (2010) was used to call Netica C API in this study. The same criteria of stopping iteration were set as in the first simulation study: (1) 1000 for the maximum number of iterations and (2) 1.0e-5 for the minimum change in data log likelihood between two iterations. The iteration was terminated when either of the two conditions was met. All cells were convergent in this simulation study.

A label switching issue occurred as in the first simulation study. In order to handle this label switching issue in the simulation study, the method of incorporating prior information when estimating parameters was used. The different degrees of weights can be applied to initial prior information. The value of 1 was used as the weight of prior, which is equivalent to the amount of information contained in a data set with sample size of 1 (Netica C API manual, 2006). The detailed information and function about this method was described in chapter 3.

Results

The results of the second simulation study are summarized in the same way as the first simulation study. Parameter recovery in terms of the different conditions was examined by comparing estimates with the true parameter values used for response data generated for the different conditions. One hundred replications were run. Three criteria were used to evaluate the overall accuracy of the method in each condition: (1) Root Mean Squared Difference (RMSD), (2) Bias, and (3) Standard Deviation of Estimates (SDE). The three criteria were computed for each simulation condition. The results are organized by the simulation conditions.

Sample Sizes

Table 87

Bias for the student distribution at the first measurement in terms of the different sample sizes

Sample Size = 100				Sample Size = 1000			
Bias				Bias			
Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
0.01949	-0.01047	0.05208	-0.06111	0.00462	0.01757	0.02956	-0.05175

Table 88

Bias for the transition probability in terms of the different sample sizes

Covariate = 1				
Conditional Probability Table 1				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.00395	0.01246	0.00866	-0.0251
Level 2	0.00647	-0.0027	0.03123	-0.0351
Level 3	0.00199	0.00295	-0.0331	0.02814
Level 4	0.01427	0.0477	0.069	-0.1003
Conditional Probability Table 2				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	-0.0103	0.02126	-0.001	-0.01
Level 2	0.00627	-0.0023	0.02147	-0.0254
Level 3	0.00026	0.00288	-0.0442	0.04109
Level 4	0.01356	0.01703	0.06986	-0.1304

Covariate = 2				
Conditional Probability Table 1				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.01038	-0.0233	0.00209	0.0108
Level 2	0.01069	-0.0657	0.04903	0.00597
Level 3	0.00509	0.0011	-0.0828	0.07664
Level 4	0.02577	0.05056	0.04739	-0.1237
Conditional Probability Table 2				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.00222	-0.0166	0.00487	0.0095
Level 2	0.01306	-0.0532	0.01958	0.02058
Level 3	0.00144	0.00261	-0.1058	0.10179
Level 4	0.01505	0.01912	0.02728	-0.0615

Table 89

Bias for conditional probability tables of the first type of the tasks in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.02564	-0.02564	0.00468	-0.00468
Level 2	-0.01109	0.01109	0.00222	-0.00222
Level 3	-0.00243	0.00243	-0.00191	0.00191
Level 4	0.01353	-0.01353	-0.00119	0.00119

Table 90

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01194	-0.01194	-0.00280	0.00280
Level 2	0.00549	-0.00549	0.01974	-0.01974
Level 3	-0.00185	0.00185	0.00726	-0.00726
Level 4	-0.00781	0.00781	-0.01035	0.01035

Table 91

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00427	-0.00427	0.00002	-0.00002
Level 2	-0.00177	0.00177	-0.00418	0.00418
Level 3	0.06412	-0.06412	0.04727	-0.04727
Level 4	-0.01077	0.01077	0.04145	-0.04145

Table 92

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.03394	-0.03394	0.00827	-0.00827
Level 2	-0.00584	0.00584	-0.00054	0.00054
Level 3	-0.01766	0.01766	-0.01079	0.01079
Level 4	-0.00561	0.00561	0.00052	-0.00052

Table 93

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00603	-0.00603	0.00172	-0.00172
Level 2	0.01913	-0.01913	0.00814	-0.00814
Level 3	0.00090	-0.00090	0.00291	-0.00291
Level 4	-0.01088	0.01088	0.00018	-0.00018

Table 94

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different sample sizes

	Sample Size = 100		Sample Size = 1000	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00952	-0.00952	-0.00236	0.00236
Level 2	0.00625	-0.00625	0.00007	-0.00007
Level 3	0.03786	-0.03786	0.01214	-0.01214
Level 4	0.00517	-0.00517	0.01015	-0.01015

Table 95

Average RMSDs for the different sample sizes

Condition	Sample Size = 100	Sample Size = 1000
	RMSD	RMSD
Parameter		
DS* at Time1	0.077044	0.063177
TPT*	0.101078	0.067127
Average of CPTT*	0.082984	0.045033

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 96

Average SDE for the different sample sizes

Condition	Sample Size = 100	Sample Size = 1000
	SDE	SDE
Parameter		
DS* at Time1	0.004482	0.001591
TPT*	0.009105	0.003229
Average of CPTT*	0.015077	0.005345

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 87 shows the bias of the initial probability distribution parameters (DS) in terms of the two different sample sizes. Table 88 shows the bias values of the transition probability table parameters (TPT) given two different values of the covariate in terms of the sample size of 100 and the sample size of 1000. Table 89, Table 90, and Table 91 indicate the bias of the conditional probability table parameters of the tasks used at the first measurement. Table 92, Table 93, and Table 94 show the bias of the conditional probability table parameters of the tasks implemented at the second measurement. In all cases, the bias values decreased as the sample size increased. Compared to the bias values of the first simulation study, the bias values were higher in the second simulation study. This implies that estimates are less precise for the more complex model than the simple model.

Table 95 shows the RMSDs of the initial probability distribution parameters, transitional probability table parameters, and conditional probability table parameters in terms of two sample sizes. RMSDs were lower for 1000 samples than for 100 samples, which indicate that the precision of the estimate increased as sample size increased. A similar result was also observed for SDE values. In terms of the two different sample sizes, SDEs with 1000 samples were lower than those with 100 samples; therefore, parameter estimates seemed to be more reliable when using more samples.

To summarize, the results show that using a larger sample reduced error variances for all probability distribution table parameters in the different conditions considered in this study. Among the parameters of DS, CPT, and TPT, CPT and TPT seem to be more affected by different sample sizes compared the others (TPT was the parameter set that was the most effected by different sample size in the first simulation study.) Overall, TPT seems to be more biased than the others. Compared to the first simulation study, all values of bias, RMSD, and SDE were larger in the second simulation study. A following section will examine the significance of these effects using analysis of variance.

Task Sizes

Table 97

Bias for the student distribution at the first measurement in terms of the different task sizes

Task Size = 9				Task Size = 30			
Bias				Bias			
Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
0.01791	0.00387	0.03801	-0.05979	0.00621	0.00324	0.04363	-0.05307

Table 98

Bias for the transition probability in terms of the different task sizes

Covariate = 1				
Conditional Probability Table 1				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.0095	-0.0159	0.0042	0.0022
Level 2	0.0041	-0.0401	0.0445	-0.0085
Level 3	0.0036	0	-0.0701	0.0664
Level 4	0.037	0.0629	0.0579	-0.1578
Conditional Probability Table 2				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	-0.004	0.0033	-0.0008	0.0015
Level 2	0.0092	-0.0558	0.0379	0.0087
Level 3	0.0017	0.0036	-0.0971	0.0918
Level 4	0.0001	0.0017	0.0051	-0.007

Covariate = 2				
Conditional Probability Table 1				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.0096	0.0143	0.0063	-0.0302
Level 2	0.0063	0.0009	0.0198	-0.027
Level 3	0.0021	0.0034	-0.0181	0.0125
Level 4	0.0173	0.0253	0.0778	-0.1204
Conditional Probability Table 2				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.004	0.0004	-0.0003	-0.0041
Level 2	0.0006	-0.0093	0.0266	-0.018
Level 3	0.0001	0.0001	-0.0625	0.0623
Level 4	0.0008	0.0125	0.0095	-0.0228

Table 99

Bias for conditional probability tables of the first type of the tasks in terms of the different task sizes

	Task size = 9		Task size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.02053	-0.02053	-0.00614	0.00380
Level 2	-0.00758	0.00758	0.00501	-0.00501
Level 3	0.00180	-0.00180	0.00378	-0.00378
Level 4	0.00734	-0.00734	0.01548	-0.01548

Table 100

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different task sizes

	Task size = 9		Task size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00537	-0.00537	-0.00468	0.00468
Level 2	0.00975	-0.00975	-0.00135	0.00135
Level 3	0.01009	-0.01009	0.00378	-0.00378
Level 4	-0.01680	0.01680	0.01548	-0.01548

Table 101

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different task sizes

	Task size = 9		Task size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00008	-0.00008	0.05095	-0.05095
Level 2	-0.00563	0.00563	0.00100	-0.00100
Level 3	0.06044	-0.06044	0.00378	-0.00378
Level 4	0.02968	-0.02968	0.01548	-0.01548

Table 102

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different task sizes

	Task size = 9		Task size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.02992	-0.02992	-0.00381	0.00381
Level 2	-0.00473	0.00473	-0.00490	0.00490
Level 3	-0.02464	0.02464	0.00378	-0.00378
Level 4	-0.00018	0.00018	0.01548	-0.01548

Table 103

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different task sizes

	Task size = 9		Task size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00289	-0.00289	-0.00270	0.00270
Level 2	0.01953	-0.01953	-0.00402	0.00402
Level 3	0.00651	-0.00651	0.00378	-0.00378
Level 4	-0.00669	0.00669	0.01548	-0.01548

Table 104

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different task sizes

	Task size = 9		Task size = 30	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00342	-0.00342	0.00805	-0.00805
Level 2	0.00315	-0.00315	-0.00192	0.00192
Level 3	0.04194	-0.04194	0.00378	-0.00378
Level 4	0.01724	-0.01724	0.01548	-0.01548

Table 105

Average RMSD for the different task sizes

Condition	Task Size = 9	Task Size = 30
	RMSD	RMSD
Parameter		
DS* at Time1	0.07460	0.06619
TPT*	0.09238	0.08553
Average of CPTT*	0.08280	0.04535

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 106

Average SDE for the different task sizes

Condition	Task Size = 9	Task Size =30
	SDE	SDE
Parameter		
DS* at Time1	0.003162	0.002911
TPT*	0.006303	0.006030
Average of CPTT*	0.011060	0.009632

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 97 shows the bias of the DS in terms of the two different task sizes. Table 98 shows the bias of the TPT in terms of a task size of 9 and a task size of 30. Table 99, Table 100, and Table 101 indicate the bias of the CPT parameters of the tasks used at the first measurement. Table 102, Table 103, and Table 104 show the bias of the CPT parameters of the tasks implemented at the second measurement. The bias values decreased as the task size increased. Compared to the condition of sample size, the condition of task size resulted in similar influence. The reduced amount of bias values of the CPT parameters was the largest among other parameters (DS and TPT) when the number of tasks increased. This implied that the precision of estimates in the CPT was more sensitive to the task size implemented than other parameters (DS and TPT). Table 105 shows the comparison of the average of RMSDs between two different task sizes of 9 and 30. It was observed that RMSD decreased as the task size increased. A similar results pattern was also observed for SDEs.

In summary, the results showed that the precision of estimates increases as more tasks were used. Compared to the simple model, estimates in the more complex model are less precise given the different task sizes. Among the parameters of DS, CPT, and TPT, CPT seemed to be more influenced by different task sizes than the others (DS and CPT seemed to be more influenced by different task sizes in the first simulation study). Overall, TPT seemed to be more biased compared to the others. A following section will examine the significance of these effects.

Covariate Distributions

Table 107

Bias for the student distribution at the first measurement in terms of the different covariate distributions

Equal Distribution				Skewed Distribution			
Bias				Bias			
Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
0.00957	0.01115	0.03229	-0.05300	0.01455	-0.00404	0.04935	-0.05986

Table 108

Bias for the transition probability in terms of the different covariate distributions

Covariate = 1				
Conditional Probability Table 1				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.0034	0.004	-0.0015	-0.0058
Level 2	0.0002	-0.0427	0.0628	-0.0204
Level 3	0.0039	0.001	-0.0988	0.0939
Level 4	0.0221	0.0455	0.0358	-0.1033
Conditional Probability Table 2				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.0022	-0.0166	0.0049	0.0095
Level 2	0.0131	-0.0532	0.0196	0.0206
Level 3	0.0014	0.0026	-0.1058	0.1018
Level 4	0.0151	0.0191	0.0273	-0.0615

Covariate = 1				
Conditional Probability Table 1				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.0239	-0.0066	0.0071	-0.0243
Level 2	0.0007	-0.0061	0.025	-0.0196
Level 3	0.002	0.0006	-0.0488	0.0462
Level 4	0.0045	0.022	0.0487	-0.0752
Conditional Probability Table 2				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	-0.0103	0.0213	-0.001	-0.01
Level 2	0.0063	-0.0023	0.0215	-0.0254
Level 3	0.0003	0.0029	-0.0442	0.0411
Level 4	0.0136	0.047	0.0699	-0.1304

Table 109

Bias for conditional probability tables of the first type of the tasks in terms of the different covariate distributions

	Equal Distribution		Skewed Distribution	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01112	-0.01112	0.01451	-0.01451
Level 2	-0.00100	0.00100	-0.00319	0.00319
Level 3	-0.00044	0.00044	0.00203	-0.00203
Level 4	0.00488	-0.00488	0.01216	-0.01216

Table 110

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different covariate distributions

	Equal Distribution		Skewed Distribution	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00187	-0.00187	0.00259	-0.00259
Level 2	0.01500	-0.01500	0.00554	-0.00554
Level 3	0.00448	0.00044	0.00203	-0.00561
Level 4	-0.00351	0.00351	-0.00996	0.00996

Table 111

Bias for conditional probability tables of the third type of the tasks at the first measurement in terms of the different covariate distributions

	Equal Distribution		Skewed Distribution	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00053	0.00053	0.00013	-0.00013
Level 2	-0.00269	0.00269	-0.00794	0.00794
Level 3	0.05193	0.00044	0.00203	-0.05477
Level 4	0.02287	-0.02287	0.01251	-0.01251

Table 112

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different covariate distributions

	Equal Distribution		Skewed Distribution	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01759	-0.01759	0.01994	-0.01994
Level 2	0.00186	-0.00186	-0.00355	0.00355
Level 3	-0.01236	0.00044	0.00203	0.01140
Level 4	0.00148	-0.00148	-0.00187	0.00187

Table 113

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different covariate distributions

	Equal Distribution		Skewed Distribution	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00363	0.00363	0.00669	-0.00669
Level 2	0.01068	-0.01068	0.01190	-0.01190
Level 3	0.00567	0.00044	0.00203	-0.00283
Level 4	-0.00749	0.00749	0.00147	-0.00147

Table 114

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different covariate distributions

	Equal Distribution		Skewed Distribution	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00226	-0.00226	0.00022	-0.00022
Level 2	0.00224	-0.00224	-0.00061	0.00061
Level 3	0.02512	0.00044	0.00203	-0.02019
Level 4	0.00879	-0.00879	0.01122	-0.01122

Table 115

Average RMSD for the different types of covariate distributions

Condition	Equal Distribution	Skewed Distribution
	RMSD	RMSD
Parameter		
DS* at Time1	0.060164	0.079546
TPT*	0.089519	0.090479
Average of CPTT*	0.067685	0.082984

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 116

Average SDE for the different types of covariate distributions

Condition	Equal Distribution	Skewed Distribution
	SDE	SDE
Parameter		
DS* at Time1	0.003003	0.003070
TPT*	0.005798	0.006536
Average of CPTT*	0.010037	0.010385

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

RMSD, bias, and SDE were computed in terms of the different distributions of a covariate. Table 107 shows the bias of the DS in terms of two types of the covariate.

Table 108 shows the bias of the TPT in terms of an equal distribution and an unequal distribution of a covariate. Table 109, Table 110, and Table 111 indicate the bias of the CPT parameters of the tasks used at the first measurement. Table 112, Table 113, and Table 114 show the bias of the CPT parameters of the tasks implemented at the second measurement. It was observed that the bias values and RMSD values were similar for the two types of covariate distribution. This suggests that the type of covariate distribution did not seem to influence the precision of the estimate. A similar results pattern was also observed for SDEs. To summarize, the results show that using different types of covariate distribution did not seem to affect the error variances of estimates.

Conditional Probability Tables

Table 117

Bias for the student distribution at the first measurement in terms of the different conditional probability tables

Conditional Probability Table 1				Conditional Probability Table 2			
Bias				Bias			
Level 1	Level 2	Level 3	Level 4	Level 1	Level 2	Level 3	Level 4
0.00101	-0.00226	0.00271	-0.00150	0.02307	0.00948	0.07880	-0.11136

Table 118

Bias for the transition probability in terms of the different conditional probability tables

Covariate = 1				
Conditional Probability Table 1				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	-0.0005	0.0014	-0.004	0.0032
Level 2	0.0032	-0.0331	0.0175	0.0128
Level 3	0.0008	0	-0.0223	0.0223
Level 4	0.0001	0.0028	0.0054	-0.0086
Conditional Probability Table 2				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.0061	-0.014	0.0075	0.0005
Level 2	0.01	-0.0618	0.0643	-0.0125
Level 3	0.0046	0.0036	-0.1441	0.1359
Level 4	0.037	0.0617	0.0575	-0.1562

Covariate = 1				
Conditional Probability Table 1				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.0231	-0.0087	0.0023	-0.0172
Level 2	0.0006	0.0142	-0.0029	-0.0124
Level 3	0.0002	0.0034	-0.0125	0.0091
Level 4	0.0002	0.0012	0.0046	-0.0061
Conditional Probability Table 2				
Bias				
	Level 1	Level 2	Level 3	Level 4
Level 1	-0.0102	0.0236	0.0037	-0.0171
Level 2	0.0062	-0.0231	0.0495	-0.0326
Level 3	0.0021	0	-0.0678	0.0657
Level 4	0.0179	0.0366	0.0825	-0.1371

Table 119

Bias for conditional probability tables of the first type of the tasks in terms of the different conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00199	-0.00199	0.02829	-0.02829
Level 2	0.00047	-0.00047	-0.00934	0.00934
Level 3	0.00108	0.00044	0.00203	0.00310
Level 4	-0.00006	0.00006	0.01241	-0.01241

Table 120

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00058	-0.00054	0.00857	-0.00857
Level 2	-0.00168	0.00172	0.02695	-0.02695
Level 3	-0.00090	0.00044	0.00203	-0.00641
Level 4	-0.00125	0.00123	-0.01693	0.01693

Table 121

Bias for conditional probability tables of the first type of the tasks at the third measurement in terms of the different conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00118	-0.00119	0.00305	-0.00305
Level 2	-0.00156	0.00159	-0.00436	0.00436
Level 3	0.00239	0.00044	0.00203	-0.10895
Level 4	0.00196	-0.00193	0.02876	-0.02876

Table 122

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01049	-0.01044	0.03146	-0.03146
Level 2	-0.00139	0.00145	-0.00493	0.00493
Level 3	-0.00117	0.00044	0.00203	0.02728
Level 4	-0.00191	0.00191	-0.00317	0.00317

Table 123

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00013	-0.00013	0.00762	-0.00762
Level 2	0.00530	-0.00530	0.02196	-0.02196
Level 3	0.00154	0.00044	0.00203	-0.00227
Level 4	0.00050	-0.00050	-0.01121	0.01121

Table 124

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different conditional probability tables

	Conditional Probability Table 1		Conditional Probability Table 2	
	Bias		Bias	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00003	0.00003	0.00719	-0.00719
Level 2	0.00330	-0.00330	0.00302	-0.00302
Level 3	0.00281	0.00044	0.00203	-0.04718
Level 4	0.00218	-0.00218	0.01314	-0.01314

Table 125

Average RMSD for the different types of conditional probability tables

Condition	Conditional Probability Table 1	Conditional Probability Table 2
	RMSD	RMSD
Parameter		
DS* at Time1	0.011092	0.099117
TPT*	0.064320	0.109819
Average of CPTT*	0.023166	0.091530

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 126

Average SDE for the different types of conditional probability tables

Condition	Conditional Probability Table 1	Conditional Probability Table 2
	SDE	SDE
Parameter		
DS* at Time1	0.001925	0.004148
TPT*	0.004081	0.008252
Average of CPTT*	0.008460	0.011962

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 117 shows the bias of the DS in terms of the two different task sizes. Table 118 show the bias of the TPT in terms of two types of conditional probability tables. Table 119, Table 120, and Table 121 indicate the bias of the CPT parameters of the tasks used at the first measurement. Table 122, Table 123, and Table 124 show the bias of the

CPT parameters of the tasks implemented at the second measurement. The bias values for the first conditional probability table were smaller than for the second conditional probability table. This implies that the precision of estimates increased when the tasks were well designed for classifying students according to the different levels on a LP. A similar result was also observed for RMSD and SDE. RMSDs for the first type of conditional probability table were smaller than for the second type of conditional probability table. SDEs for the first conditional probability table were smaller than those for second conditional probability table. In summary, the results show that when the conditional probabilities of each task were distinct across levels, so that each task was relatively well designed for the purpose of classifying students according to the different levels, the error variances of the estimates seem to be reduced in this simulation study. All the parameters of DS, CPT, and TPT seemed to be influenced by two types of conditional probability tables. Overall, TPT had more biased values than the other parameters. A following section will examine the significance of these effects.

Transition Probability Tables

Table 127

Bias for the student distribution at the first measurement in terms of the different types of transition probability tables

	Bias			
	Level 1	Level 2	Level 3	Level 4
Transition Probability Table 1	0.02236	0.00863	0.04770	-0.07868
Transition Probability Table 2	0.00212	0.00895	0.02856	-0.03963
Transition Probability Table 3	0.00518	-0.00297	0.05472	-0.05693
Transition Probability Table 4	0.01748	-0.00037	0.03039	-0.04750

Table 128

Bias for the transition probability in terms of the different types of transition probability Tables

Bias				
Covariate 1				
Transition Probability 1				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.01476	0.03073	-0.03365	-0.01184
Level 2	0.01020	-0.06147	0.07125	-0.01998
Level 3	0.00210	0.00004	-0.11390	0.11175
Level 4	0.00800	0.03634	0.01608	-0.06041
Transition Probability 2				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.01715	-0.05727	0.01876	0.02136
Level 2	0.00708	-0.07851	0.06594	0.00549
Level 3	0.00145	0.00322	-0.05303	0.04837
Level 4	0.01558	0.00255	0.04847	-0.06659
Transition Probability 3				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.01969	-0.03528	0.02151	-0.00592
Level 2	0.00661	-0.03553	-0.00249	0.03141
Level 3	0.00604	0.00187	-0.08025	0.07234
Level 4	0.02637	0.06686	0.00317	-0.09640
Transition Probability 4				
	Level 1	Level 2	Level 3	Level 4
Level 1	-0.03809	0.03446	0.00005	0.00358
Level 2	0.00256	-0.01528	0.02824	-0.01552
Level 3	0.00104	0.00202	-0.08202	0.07896
Level 4	0.02286	0.02212	0.05495	-0.09993

Bias				
Covariate 2				
Transition Probability 1				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.01538	0.00920	-0.00991	-0.01467
Level 2	0.00998	-0.00346	0.02653	-0.03304
Level 3	0.00040	0.00577	-0.10604	0.09987
Level 4	0.01654	0.03276	0.01378	-0.06308
Transition Probability 2				
	Level 1	Level 2	Level 3	Level 4
Level 1	-0.01819	0.02802	0.01447	-0.02431
Level 2	0.00001	0.00204	0.04224	-0.04429
Level 3	0.00403	0.00001	-0.04265	0.03861
Level 4	0.00864	0.00300	0.06584	-0.07748
Transition Probability 3				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.02486	-0.01060	0.00151	-0.01577
Level 2	0.00310	0.00218	0.00600	-0.01128
Level 3	0.00002	0.00082	0.00101	-0.00186
Level 4	0.01065	0.03004	0.03618	-0.07686
Transition Probability 4				
	Level 1	Level 2	Level 3	Level 4
Level 1	0.00489	0.00252	0.00562	-0.01303
Level 2	0.00073	-0.01653	0.01708	-0.00128
Level 3	0.00006	0.00032	-0.01261	0.01223
Level 4	0.00028	0.00921	0.05534	-0.06482

Table 129

Bias for conditional probability tables of the first type of the tasks the different types of transition probability tables

	TPT1		TPT2	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.02086	-0.02086	0.01341	-0.01341
Level 2	0.00189	-0.00189	-0.00810	0.00810
Level 3	-0.00052	-0.00073	-0.00838	0.00745
Level 4	0.00941	-0.00941	-0.00281	0.00281
	TPT3		TPT4	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01521	-0.01521	0.01049	-0.01049
Level 2	-0.00848	0.00848	-0.00288	0.00288
Level 3	-0.00138	0.00013	0.00150	-0.00268
Level 4	0.00980	-0.00980	0.00781	-0.00781

Table 130

Bias for conditional probability tables of the second type of the tasks at the first measurement in terms of the different types of transition probability tables

	TPT1		TPT2	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00226	-0.00226	0.00907	-0.00907
Level 2	0.02387	-0.02387	0.01872	-0.01872
Level 3	0.00145	-0.00145	0.00390	-0.00390
Level 4	-0.01838	0.01838	-0.01999	0.01999
	TPT3		TPT4	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00111	-0.00111	0.00551	-0.00551
Level 2	-0.00068	0.00068	0.00805	-0.00805
Level 3	0.00017	-0.00017	0.00499	-0.00499
Level 4	0.00439	-0.00439	-0.00220	0.00220

Table 131

Bias for conditional probability tables of the third type of the tasks at the first measurement in terms of the different types of transition probability tables

	TPT1		TPT2	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.00037	-0.00037	0.00916	-0.00916
Level 2	-0.00917	0.00917	0.00221	-0.00221
Level 3	0.07915	-0.07915	0.04861	-0.04861
Level 4	0.00970	-0.00970	0.00865	-0.00865
	TPT3		TPT4	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00038	0.00038	-0.00053	0.00053
Level 2	-0.00102	0.00102	-0.00368	0.00368
Level 3	0.03761	-0.03761	0.05403	-0.05403
Level 4	0.02364	-0.02364	0.01825	-0.01825

Table 132

Bias for conditional probability tables of the first type of the tasks at the first measurement in terms of the different types of transition probability tables

	TPT1		TPT2	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.05372	-0.05372	0.00287	-0.00287
Level 2	-0.00266	0.00266	-0.00698	0.00698
Level 3	-0.01159	0.01159	-0.02290	0.02290
Level 4	0.00051	-0.00051	-0.00555	0.00555
	TPT3		TPT4	
	Correct	Incorrect	Correct	Incorrect
Level 1	0.01946	-0.01946	0.00789	-0.00789
Level 2	-0.00406	0.00406	0.00088	-0.00088
Level 3	-0.01599	0.01599	-0.00604	0.00604
Level 4	-0.00175	0.00175	-0.00318	0.00318

Table 133

Bias for conditional probability tables of the second type of the tasks at the second measurement in terms of the different types of transition probability tables

	TPT1		TPT2	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00620	0.00620	0.01502	-0.01502
Level 2	0.03127	-0.03127	0.01828	-0.01828
Level 3	0.00911	-0.00911	-0.00349	0.00349
Level 4	0.00464	-0.00464	-0.02356	0.02356
	TPT3		TPT4	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00355	0.00355	0.00963	-0.00963
Level 2	0.01309	-0.01309	-0.00764	0.00764
Level 3	-0.00079	0.00079	0.00263	-0.00263
Level 4	-0.00257	0.00257	0.00007	-0.00007

Table 134

Bias for conditional probability tables of the third type of the tasks at the second measurement in terms of the different types of transition probability tables

	TPT1		TPT2	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00131	0.00131	0.02170	-0.02170
Level 2	0.00822	-0.00822	-0.00489	0.00489
Level 3	0.01680	-0.01680	0.03365	-0.03365
Level 4	0.00996	-0.00996	-0.00304	0.00304
	TPT3		TPT4	
	Correct	Incorrect	Correct	Incorrect
Level 1	-0.00484	0.00484	-0.00116	0.00116
Level 2	0.00833	-0.00833	0.00092	-0.00092
Level 3	0.00900	-0.00900	0.03816	-0.03816
Level 4	0.01209	-0.01209	0.01095	-0.01095

Table 135

Average RMSD for the different types of transition probability tables

Condition	TD1	TD2	TD3	TD4
	RMSD	RMSD	RMSD	RMSD
Parameter				
DS* at Time1	0.078939	0.061694	0.071633	0.068745
TPT*	0.097702	0.093572	0.092467	0.092360
Average of CPTT*	0.073806	0.060949	0.062812	0.068684

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 136

Average SDE for the different types of transition probability tables

Condition	TD1	TD2	TD3	TD4
	SDE	SDE	SDE	SDE
Parameter				
DS* at Time1	0.003682	0.003114	0.002648	0.002701
TPT*	0.007300	0.005084	0.006600	0.005683
Average of CPTT*	0.010859	0.01029	0.009631	0.010065

DS* indicates the distribution of students at the first measurement occasion.
TPT* is the transition probability table.
CPTT* is the conditional probability table of each task.

Table 127 shows the bias of the DS in terms of four different types of transition probability tables. Table 128 shows the bias of the TPT in terms of three different types of transition probability tables. Table 129, Table 130, and Table 131 indicate the bias of the CPT parameters of the tasks used at the first measurement. Table 132, Table 133, and Table 134 show the bias of the CPT parameters of the tasks implemented at the second measurement. It shows that the bias values and RMSD values were very similar to each other across the four different types of transition probability tables. This implies that the four types of transition probability tables do not seem to influence the level of precision of estimates. A similar result was also observed for SDE values. However, the conditional probability table for each task had relatively higher SDEs, implying that the variance of the conditional probability table for each task estimate was greater than for the others. To summarize, regardless of the type of transition probability tables, the results show that the precision and reliability of estimates were similar to each other.

In order to understand the effect of different conditions on the precision and reliability of estimates, the graphical representations of RMSDs across all conditions are presented. The three graphs (Figure 38, Figure 39, and Figure 40) display the RMSDs of all conditions with respect to the initial probability table, transition probability table, and conditional probability table for each task. For the distribution of students at the first measurement (Figure 38), the types of covariate distribution and the conditional probability table for each task seemed to influence the precision of estimates more than other factors

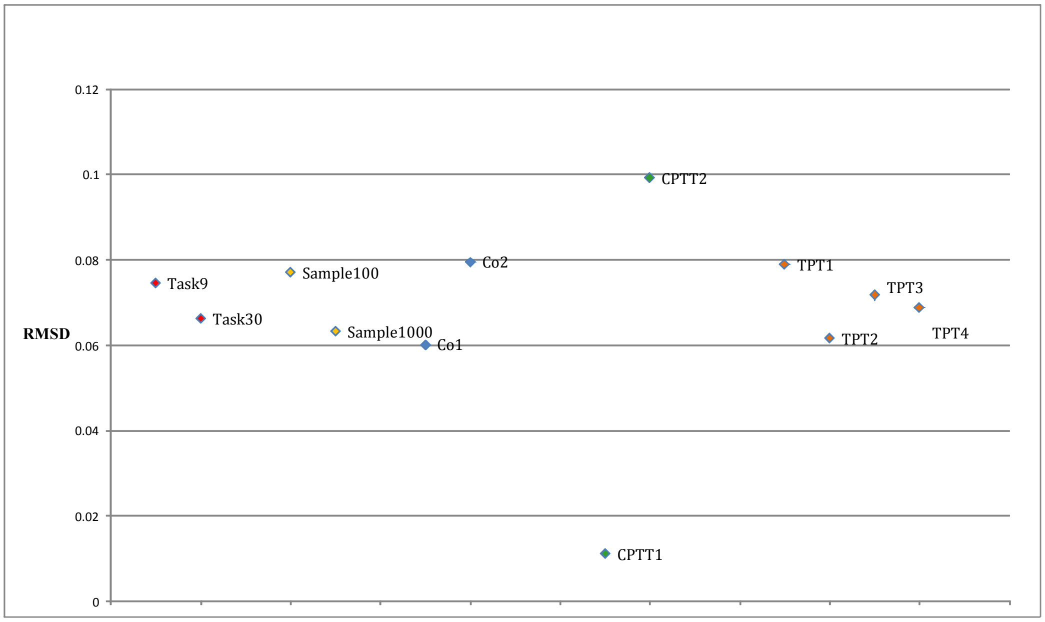


Figure 37. RMSDs of distribution of students at the first measurement with different conditions for initial probability table

As shown in Figure 38, the plots for task size, sample size, and the transition probability table were clustered more closely to each other compared to the conditions of covariate distribution and the conditional distribution table.

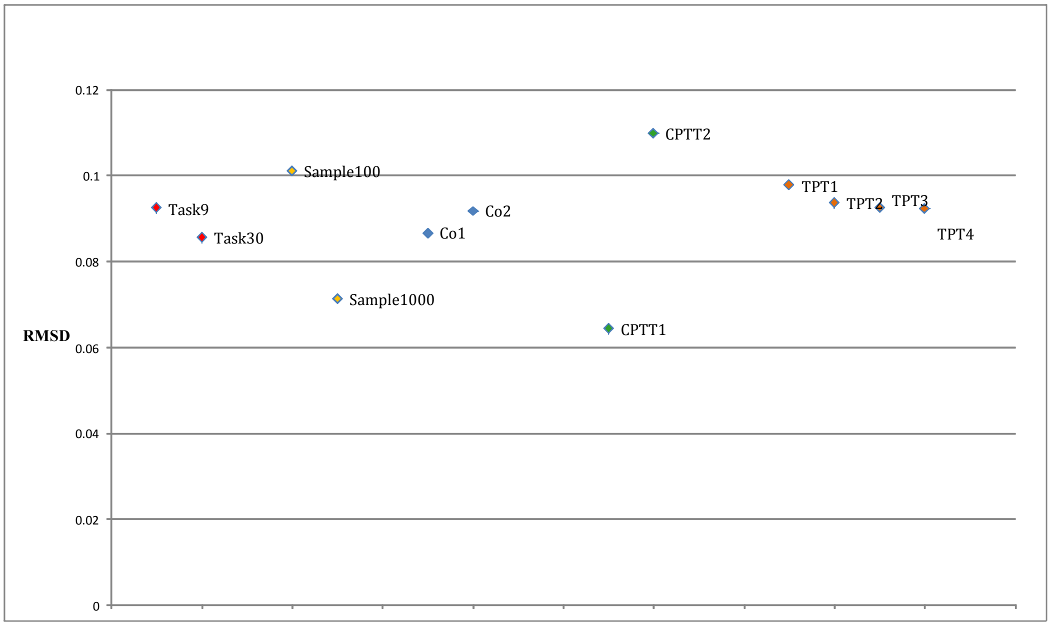


Figure 38. RMSDs of distribution of students at the first measurement with different conditions for transition probability table

Figure 39 shows the RMSDs for the transition probability table. The task size, sample size, and conditional probability table of each task seemed to affect the precision of estimates more than the other factors. As shown in Figure 39, two plots representing different sample size and two types of conditional probability tables were located far from each other, meaning that their RMSDs represent large differences between the two conditions. In contrast, the plots of the task size, covariate distributions, and transition probability tables were clustered close to each other, implying that the precision of the estimates were similar to each other no matter what type of transition probability tables were used and no matter how the covariate was distributed.

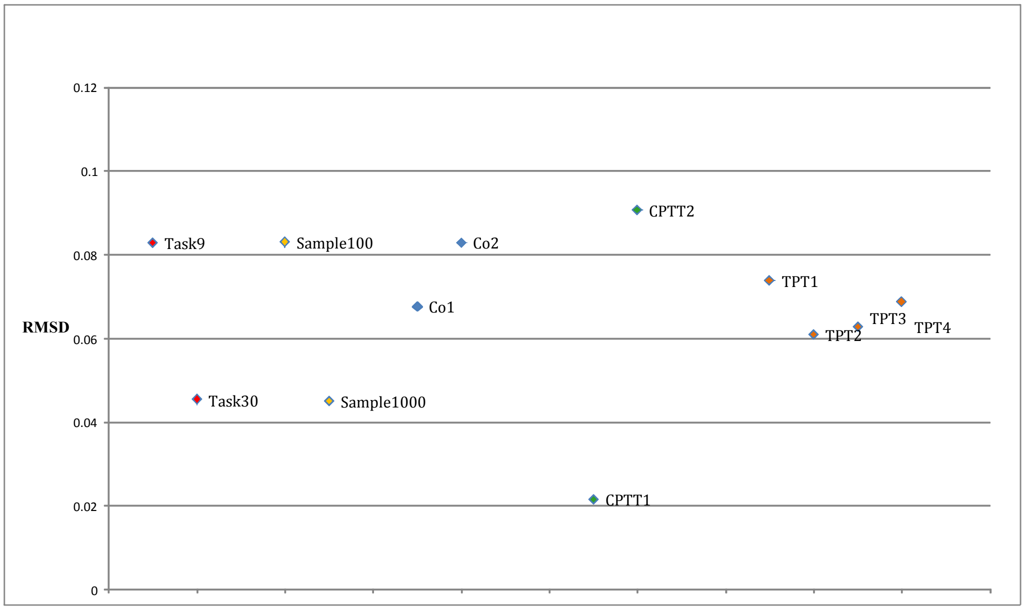


Figure 39. RMSDs of distribution of students at the first measurement with different conditions for conditional probability table

Similar results were observed for the conditional probability table for each task (see Figure 40). Like other probability tables, the task size, sample size, and conditional probability table for each task affected the precision of estimates more than other factors; on the other hand, the conditions for the transition probability table and the covariate distribution yielded similar RMSDs. This implies that the precision of estimates for the conditional probability table seemed to be influenced by different types of transition probability tables and the covariate distribution.

Since 64 conditions were investigated in this study, simple descriptive statistics and graphical representations did not seem to be an efficient way of detecting important effects and estimating the magnitude of these effects. For this reason, the inferential analyses are conducted in the next section.

ANOVA (Analysis of Variance) was used for the inferential analyses of this study. Because the purpose of the simulation study was to evaluate parameter estimation of the success of parameters recovery, the RMSD was used as the dependent variable (e.g., Harwell & Janosky, 1991; Kim, Cohen, Baker, Subkoviak, & Leonard, 1994; Stone, 1992). The simulation conditions served as the independent variables. The main effect for each independent variable and the interaction effects among them were examined. The results are summarized in terms of each parameter table (i.e., the distribution of students at the first measurement, the transition probability table, and the conditional probability table of each task). Before conducting an ANOVA, the dependent variable was investigated to determine model fit. It was found that the distributions of RMSD values for all parameters were positively skewed, implying that the normality assumption was violated. Therefore, the RMSDs were transformed using a log transformation, so that it had an approximately normal distribution.

A three-way ANOVA with the independent variables was fit to the transformed RMSD values for the parameters of the distribution of students at the first measurement with a main effect model followed by two variables at a time interaction models and three variables at a time interaction models. Because no significant second-order interaction effects (i.e., three variables at a time) were observed, the second-order interactions were not included in the model. The results are reported in Table 137. The magnitude of significant effects was estimated using partial eta squared (see Table 137).

Table 137

ANOVA results for the distribution of students at the first measurement (initial probability table)

Source	*Df*	*F*	*P*	η^2
Sample Size	1	24.214	.000	.043
Task Size	1	91.242	.000	.161
CPT	1	347.047	.000	.611
TPT	3	1.246	.303	.007
Covariate	1	1.856	.179	.003
Sample Size * Task Size	1	25.186	.000	.044
Sample Size * CPT	1	.741	.393	.001
Task Size * CPT	1	24.334	.000	.043
Error	53			
a. R Squared = .907 (Adjusted R Squared = .889)				

Note. CPT indicates the conditional distribution table. TPT is the transition probability table.

Using an alpha level of 0.05, the ANOVA revealed three significant main effects and two interaction effects. The condition of sample size was found to be statistically significant: $F(1, 53) = 24.214$, $p < 0.00$. The condition of task size was observed to be statistically significant: $F(1, 53) = 91.242$, $p < 0.00$. The condition of conditional probability table was found to be statistically significant: $F(1, 53) = 347.047$, $p < 0.00$. The results described above suggest that sample size, task size, and types of conditional probability table influenced the accuracy of estimates. In addition to the main effects, the sample size x task size and the task size x CPT interaction effects were found to be statistically significant: $F(1, 53) = 25.186$, $p < 0.00$ and $F(1, 53) = 24.334$, $p < 0.00$. This implies that the mean difference among the levels of the factor (sample sizes) was not constant across the levels of the factor (task size). The mean difference among the levels of the factor (task size) was not equal across the types on the conditional probability table. In order words, there was a joint effect of the conditions of sample sizes and CPT. To graphically understand the interaction effects, the profile plots of sample size x task size and the task size x CPT are displayed. Using η^2 as the measure of effect size, the different types of the conditional probability tables had the largest effect size to explain the RMSD values of the DS parameters. This factor accounted for 61% of the total variability in the RMSD of the DS parameters. This result is the same as the first simulation study.

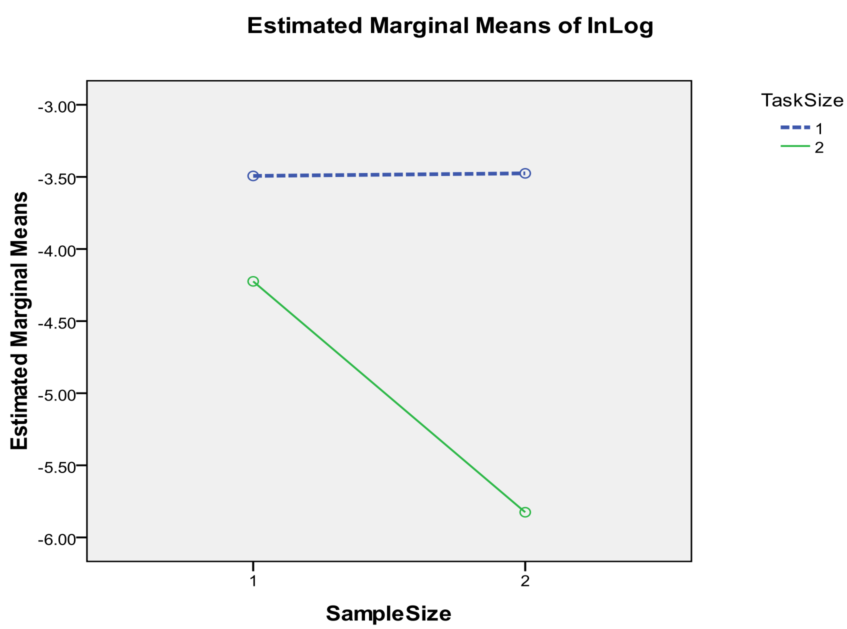
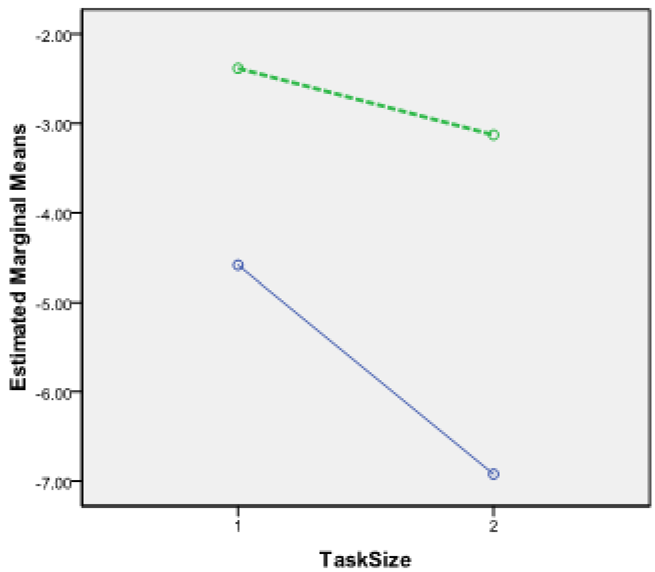


Figure 40 Estimated marginal means of the log transformed RMSD of DS in terms of task size and sample size

This profile plot shows that the RMSD values of 9 tasks did not differ between 100 samples and 1000 samples while the RMSD values of 30 tasks dramatically decreased as sample size increased.



Note. The dotted line is the second conditional probability table and the solid line is the first conditional probability table.

Figure 41**.** Estimated marginal means of the log transformed RMSD of DS in terms of task size and conditional probability table

This profile plot shows that the task size and the task design had a positive relation to the improvement of the accuracy of estimates. When a well-designed task is implemented, the RMSD dramatically decreases. In order words, the quality of task is an important factor that influenced the accuracy of estimates as well as the number of tasks implemented.

A three-way ANOVA with the independent variables was also fitted to the RMSD values for the parameters of the transition probability table with a main effect model followed by two variables at a time interaction models and three variables at a time interaction models. It was also found that there were no significant second-order interaction effects (i.e., three variables at a time). Therefore, the second-order interactions were not included in the model. The results were reported in Table 138.

Table 138

ANOVA results for transition probability table

Source	*Df*	*F*	*P*	η^2
Sample Size	1	67.185	.000	.247
Task Size	1	11.341	.001	.042
CPT	1	123.322	.000	.453
TPT	3	1.292	.287	.014
covariate	1	.321	.573	.001
Sample Size * Task Size	1	8.817	.004	.032
Sample Size * CPT	1	2.981	.090	.011
Task Size * CPT	1	.369	.546	.001
Error	53			
a. R Squared = .805 (Adjusted R Squared = .768)				

Using an alpha level of 0.05, this test was found to have three significant main effects and one interaction effect. For the main effects, the conditions of sample size, task size, and CPT were found to be statistically significant: $F(1, 53) = 67.185$, $p < 0.00$, $F(1, 53) = 11.341$, $p = 0.01$, and $F(1, 53) = 123.322$, $p < 0.01$ respectively. The results suggest that there were significant influences for sample size, task size, and types of conditional probability table on the accuracy of estimate averaging for other factors. In addition to the main effects, the sample size x task size interaction effect was found to be statistically significant: $F(1, 53) = 8.817$, $p = 0.004$. This provides evidence that the mean difference among the levels of the factor (sample size) were not constant across the levels of the factor (task size). In order words, there was a combined effect of sample size and task size. To understand the interaction effect, the profile plot was drawn (Figure 38 and Figure 39). Using η^2 as the measure of effect size, the CPT condition had the largest effect size to explain the RMSD values. It accounted for 45% of the total variability in the RMSDs of the TPT parameters. The sample size had the next largest effect size; 25% of the total variability in the RMSDs was explained by the different sample sizes. In the previous section, the sample size had the largest effect size in explaining the precision of estimates for the TPT parameters in the simple model. Based on the second simulation study, the task design seemed to be a very important factor in yielding precise and reliable estimates as the complexity of the model increases.

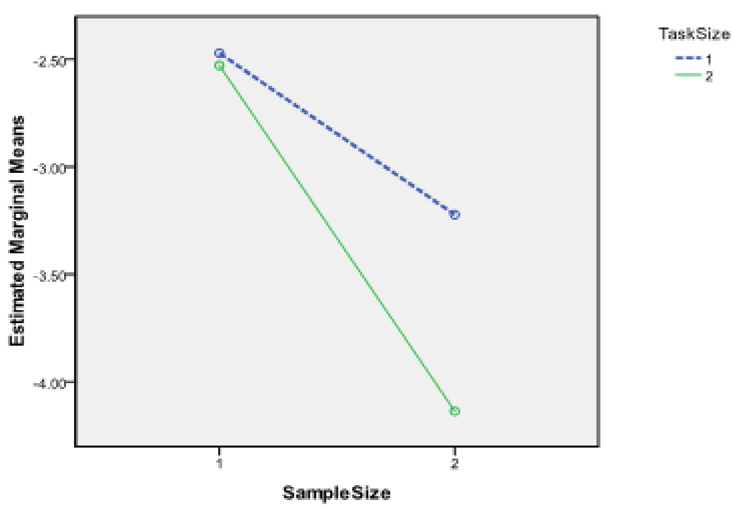


Figure 42 Estimated marginal means of the log transformed RMSD of TPT in terms of task size and sample size

When the sample size was 100, the RMSD values for 9 tasks and 30 tasks were very similar. On the other hand, when the sample size was 1000, the RMSD values were different between 9 tasks and 30 tasks. This means that the effect of task size on the precision of estimates increased when sample size increased.

A three-way ANOVA with the independent variables was also fitted to the transformed RMSD values for the parameters of the conditional probability table for each task with a main effect model followed by two variables at a time interaction models and three variables at a time interaction models. It was also found that there were no significant second-order interaction effects (i.e., three variables at a time). Therefore, the second-order interactions were not included in the model. The results are reported in Table 139.

Table 139

ANOVA results for transition probability table for conditional probability table

Source	*DF*	*F*	*P*	η^2
Sample Size	1	98.031	.000	.238
Task Size	1	51.886	.000	.126
CPT	1	178.887	.000	.434
TPT	3	3.629	.020	.026
covariate	1	1.556	.219	.004
Sample Size * Task Size	1	3.856	.056	.009
Sample Size * CPT	1	.437	.512	.001
Task Size * CPT	1	9.684	.003	.024
Sample Size * TPT	3	1.878	.147	.014
Task Size * TPT	3	.812	.494	.006
CPT * TPT	3	.521	.670	.004
Error	44			

a. R Squared = .893 (Adjusted R Squared = .847)

Using an alpha level of 0.05, the ANOVA revealed that the factors for sample size, task size, CPT, and TPT were found to be statistically significant: $F(1, 44) = 98.031$, $p < 0.00$, $F(1, 44) = 51.886$, $p < 0.01$, $F(1, 44) = 178.887$, $p < 0.01$, and $F(3, 44) = 3.629$, $p < 0.02$, respectively. The results suggest that there were significant influences of sample size, task size, and types of conditional probability table, the types of transition probability table on the accuracy of estimate. Since the factor of TPT has four levels, the post-hoc Tukey HSD test was additionally conducted in order to determine which types of TPT were significantly different. The results indicate that the means of all types with respect to type 1 were statistically significantly different (i.e. type 1 vs. type 2, d = .4176; type 1 vs. type 3, d= .3900; type 1 vs. type 4, d= .4056). It suggests that the mean of RMSD of the equal transition was significantly less than the other types of transition (favored any level on Learning Progression). Using η^2 as the measure of effect size, the different types of the conditional probability tables had the largest effect size to explain the RMSD values. This factor accounted for 44% of the total variability in the RMSD of the CPT parameters. In the previous simulation study, the sample size and task size had the largest effect size in accounting for the total variability of the RMSD values of the CPT parameters. However, the quality of the task design seemed to be the most important factor in yielding precise and reliable estimates as the complexity of the model increased.

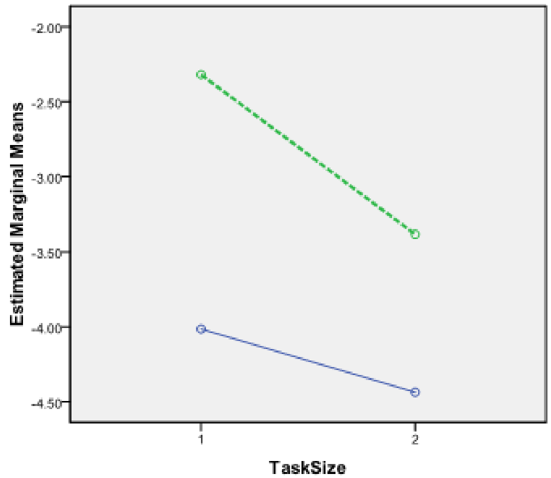
Table 140

Tukey HSD test for TPT

		Mean Difference	Std. Error	*P*
Type1	Type2	.4176	.14485	.030
Type1	Type3	.3900	.14485	.047
Type1	Type4	.4056	.14485	.037
Type2	Type3	-.0277	.14485	.997
Type2	Type4	-.0120	.14485	1.000
Type 3	Type4	.0156	.14485	1.000

The error term is Mean Square(Error) = .168.

Furthermore, task size x CPT interaction effect was found to be statistically significant: $F(1, 44) = 9.684$, $p = 0.03$. This provides evidence that a joint effect of task size and task design affected the precision of estimates. To graphically understand the interaction effect, the profile plot of task size x CPT is displayed.



Note. The dotted line is the second conditional probability table and the solid line is the first conditional probability table.

Figure 43. Estimated marginal means of the log transformed RMSD of CPT in terms of task size and conditional probability table.

Discussion

Overall, it was observed that Bias, RMSD, and SDE were higher in the second simulation study than the first simulation study. That is, the more complex model yielded less precise and reliable estimates. That is because more complex latent variable model needs to estimate more parameters. Although some additional data from the covariate lead more precise estimation, in this study, the increase in model complexity had more influence than the additional information. To summarize the descriptive statistics and inference statistics, the results show that more samples reduced error variances for all probability distribution table parameters within the different conditions considered in this study. Regarding the task size, the results show that more tasks significantly increased the precision of estimates for all probability distribution table parameters. The type of conditional probability tables had a significant effect on the precision of estimates for all parameters of the probability distribution table. It was observed that the mean of RMSD for the first type of conditional probability table was significantly smaller than the second type of conditional probability table. In summary, the results showed that when the conditional probabilities of each task with respect to each level were distinct, so that each task was relatively well designed for the purpose of classifying students according to the different levels, the error variances of estimates were dramatically reduced.

Only the estimates of CPT parameters were influenced by the simulation condition of the different types of transition probability tables. Specifically, the equal transition produced a significantly lower RMSD value than the other unequal transitions (i.e., the transition of favor with level 0 and 1, the transition of favor with level 0 and 2, and the transition of favor with level 1 and 2). On the other hand, no matter what type of transition probability table was used, the results showed that the precision of estimates of DS parameters and TPT parameters were not influenced by the different types of transition probability tables. Regarding the condition of two types of distributions of a covariate, the precisions of estimates of all parameters were not influenced by the different types of covariate distribution. A similar pattern was observed for Bias values. In terms of the SDE values, the SDE values of the CPT parameters were relatively higher than the DS and TPT parameters. This suggests that the estimates of the conditional probability table of each task fluctuated more than other parameters in the complex model. Compared to the simulation study for the simple model, the quality of task design was the most important factor in influencing the accuracy of estimates of all parameters in the complex model.

CHAPTER 8: APPLICATION OF MODEL

Introduction to Data and Analysis Procedure

An application with a real data study was conducted. The data is taken from a course of the Cisco Networking Academy (CNA). The CNA is a global program in which information technology is taught through a blended program of face to face classroom instruction, an online curriculum, and online assessment (West, et al, 2009). Each course contains several chapter exams and a final exam. Students take several chapter exams and a final exam during each course. Therefore, the same students were measured several times over the course of the curriculum. The target populations of the courses are high schools, 2- and 3-year community colleges and technical schools, and 4 year colleges and universities. For this application study, the learning progression of IP Addressing skills was used. The LP of IP addressing skills has been identified by domain experts. The LP originally contained five levels. However, CISCO does not have any tasks for the highest level on the chapter exams at the course level addressed in this study, which means that there are not data with sufficient tasks at the highest level taken by the same students. Therefore, this dissertation collapses level 4 and level 5. In addition, the previous study (West, 2001) showed that the 4-Class model demonstrated the best fit to the data, based on statistical fit in terms of the BIC (Schwarz, 1978) and the bootstrapped likelihood ratio test (McLachlan & Peel, 2000; Nylund, Asparouhov, & Muthén, 2007) conducted in M*plus* (Muthén & Muthén, 1998-2006). The study examined IP addressing LP with 35 items as conditionally independent observable variables, dependent on a single discrete latent variable with values that indicate LP levels. Table 141 is the LP of the IP Addressing skills used for this study. Four levels were in the LP of the IP Addressing skills. Level 1 can be defined for novice students that possibly have pre-course KSAs. Each level contains the descriptions of KSAs.

Table 141

Learning Progression of IP Addressing Skill

IP Addressing Skills Progression	
Level 1 – Novice -Knowledge/Skill (possibly pre-course knowledge and skills)	
1	Student can navigate the operating system to get to the appropriate screen to configure the address.
2	Student knows that four things need to be configured: IP address, subnet mask, default gateway and DNS server.
3	Student can enter and save IP addressing information that has been provided.
4	Student can use a web browser to verify network and or Internet
5	connectivity.
6	Student can verify that the provided information was correctly entered. Student knows that DNS translates names to IP addresses
Level 2 – Basic – Knows Fundamental Concept	
1	Student understands that an IP address corresponds to a source or destination host on the network.
2	Student understands that an IP address has two parts, one indicating the individual unique host and one indicating the network that the host resides on.
3	Student understands how the subnet mask indicates the network and host portions of the address.
4	Student understands the concept of local –vs- remote networks.
5	Student understands the purpose of a default gateway and why it must be specified.
6	Student knows that IP address information can be assigned dynamically.
7	Student is able to create a simple IP addressing scheme based on host or network requirements.
8	Describe the need and features of IPv6 addresses.
Level 3 – Intermediate – Knows More Advanced Concepts	
1	Student understands the difference between physical and logical connectivity.
2	Student understands the difference between Layer 2 and Layer 3 networks.
3	Student understands that a local IP network corresponds to a local IP broadcast domain. (both the terms and the functionality)
4	Student knows how a device uses the subnet mask to determine which addresses are on the local Layer 3 broadcast domain and which addresses are not.

5	Student can use the subnet mask to create an addressing scheme that accommodates design requirements for number of hosts per subnet and number of networks.
6	Student understands why the default gateway IP address must be on the same local broadcast domain as the host.
7	Student understands the ARP process and the role of Layer 2 addresses within a Layer 3 broadcast domain.
8	Student knows how to interpret a network diagram in order to determine the local and remote networks.
9	Student understands how DHCP dynamically assigns IP addresses.
10	Student knows the purpose of private, public, and special reserved addresses such as multicast and loopback, IP address spaces and when to use either one.
11	Student recognizes reserved IPv6 addresses.
Level 4 –Advanced – Can Apply Knowledge and Skills in Context	
1	Student can create an IP addressing scheme for a network using VLSM
2	Student can use a network diagram to find the local network where the configured host is located.
3	Student can use a network diagram to find the other networks attached to the local gateway device.
4	Student can use the PING utility to test connectivity to the gateway and to remote devices.
5	Student can recognize the symptoms that occur when the IP address or subnet mask is incorrect.
6	Student can recognize the symptoms that occur if an incorrect default gateway is configured.
7	Student can recognize the symptoms that occur if an incorrect DNS server (or no DNS server) is specified.
8	Student knows why DNS affects the operation of other applications and protocols, like email or file sharing.
9	Student can use NSlookup output to determine if DNS is functioning correctly.
10	Student can create a DHCP addressing scheme recognizing the importance of excluding addresses.
11	Student is able to convert an IPv4 address to an IPv6 address.

Since the tasks in CNA have enough information about what KSAs are measured through design pattern documents, the levels on LPs were able to be matched to each task.

Once the levels of the LP had been determined for each task by content experts, a DBIN was constructed in Netica.

Twenty-six tasks and a sample size of 1450 students were used for this application study. The DBIN of this LP model after estimating the conditional probability table with the data set is Figure 45.

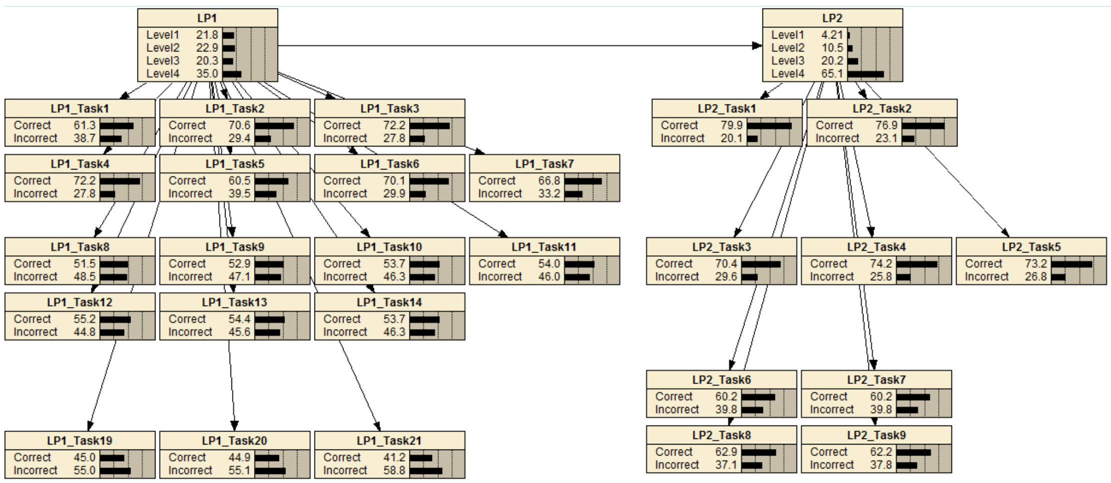


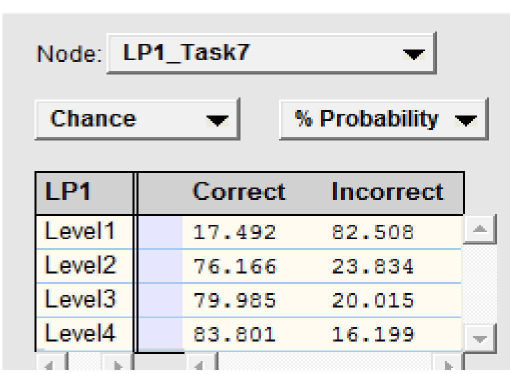
Figure 44 DBIN representation of application study

Estimation

An EM algorithm was used in order to estimate the probability distributions on DBINs. Netica can incorporate the prior information before the estimating starts with data. The prior information can be considered as part of the data by setting a weight to each probability of each variable. This analysis used 1 as the weight of the prior information. The prior probability of each variable was provided by a content expert. The use of the prior information helped to fix the label switching issue and to effect the constraint of no backward movement suggested by the substantive theory of LP research. The value of prior weight is approximately equivalent to the sample size in the nature of its effect on posterior distributions for model parameters; hence the weight has sometimes been called the “equivalent sample size”. The prior values were used as the starting values of EM algorithm. The data analyses were investigated by two aspects regarding task inferences and student inferences.

Task Inferences

A task was classified as being “at the level” if it supported an interpretation that students reaching that level would likely be able to solve or complete the task, whereas students at lower levels would be unlikely to be successful. To classify tasks, the conditional probability table of each task was examined. The results indicated that most of the tasks discriminated between the targeted level and the remaining levels. For example, Figure 46 is the conditional probability table for task 7 at the first measurement provided by Netica. The conditional probability table shows clearly that students at level 2, level 3, and level 4 are likely to successfully solve task 7, whereas students at level 1 are unlikely to successfully solve task 7. Therefore, this task aids in classifying students between level 1 and the higher levels.



Node: LP1_Task7

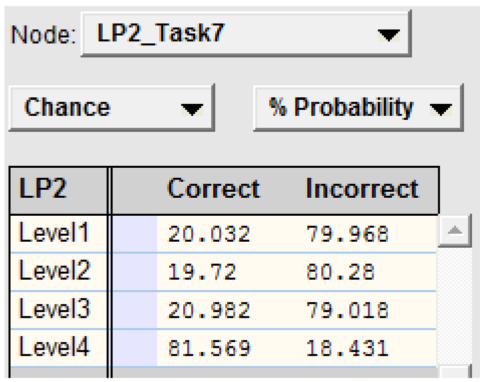
Chance

% Probability

LP1	Correct	Incorrect
Level1	17.492	82.508
Level2	76.166	23.834
Level3	79.985	20.015
Level4	83.801	16.199

Figure 45 Table 7 at the first measurement

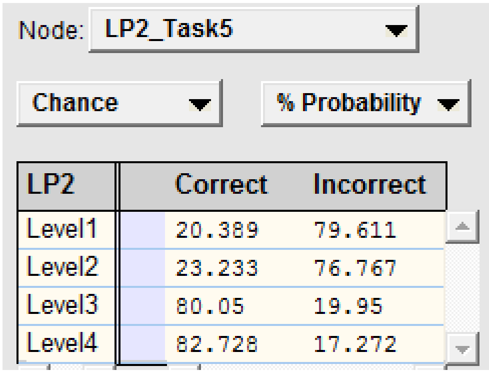
As another example, Figure 47 shows the conditional probability table for task 7 at the second measurement. It is seen clearly that only students at level 4 (the highest) level are likely to solve the task successfully, whereas students at the lower levels are unlikely to have a correct answer to the task. This task aids in distinguishing students at level 4 from the lower levels.



LP2	Correct	Incorrect
Level1	20.032	79.968
Level2	19.72	80.28
Level3	20.982	79.018
Level4	81.569	18.431

Figure 46. Task 7 at the second measurement

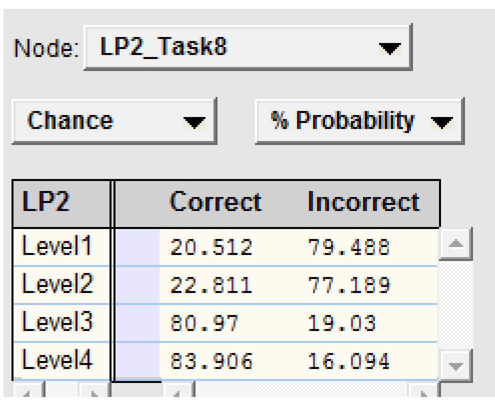
There were a few tasks that were not consistent with the expert based expectations provided by content experts. For instance, Figure 48 displays the conditional probability table for task 5 at the second measurement. It was suggested as being at level 4 by content experts, but it seems to aid in classifying students between at level 3-4 and at level 1-2 based on the data analysis.



LP2	Correct	Incorrect
Level1	20.389	79.611
Level2	23.233	76.767
Level3	80.05	19.95
Level4	82.728	17.272

Figure 47 Task 5 at the second measurement

As another example, Figure 49 shows the conditional probability table for task 8 at the second measurement. This task was suggested as being at level 4 by content expert, but the data analysis showed that this task is useful for classifying students between at level 1-2 and at level 3-4.



LP2	Correct	Incorrect
Level1	20.512	79.488
Level2	22.811	77.189
Level3	80.97	19.03
Level4	83.906	16.094

Figure 48 Task 8 at the second measurement

The last example is task 9 at the second measurement. The content experts identified this task as being level 4, but the data analysis indicated that this task aids in classifying students between at level 3-4 and at level 1-2 (figure 50).

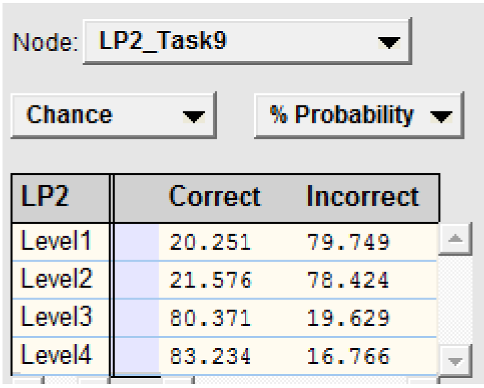


Figure 49 Task 9 at the second measurement

Task 4, 6, and 20 at the first measurement and Task 3 and 6 at the second measurement were more ambiguous patterns in terms of their levels. For instance, the conditional probabilities demonstrated a pattern where students at the lower level have little higher probability of completing the task correctly than students at the higher level.

Across all tasks, eighteen tasks out of 26 exhibited clear and distinct patterns and were consistent with the experts' expectations. They classified between levels as predicted by experts. However, three tasks out of 26 tasks seem to be mismatched with the experts' expectations. That is, they were not located at the expected levels. Initial reviews of these results were passed on to content experts to provide feedback that would help the tasks more sharply target the concepts at their intended levels.

Student Inferences

Once the response pattern had been observed, the conditional probability tables of the LPs also provide information about student levels at two measurements. The information about a student response patterns is propagated through the network via

Bayes theorem to yield posterior distributions of student levels on the LP. The posterior distribution provides the probabilities that a student has reached a specified level. On this basis, it can be inferred that the student is likely to have reached one of the levels. For instance, Figure 46 contains the DBIN for a student who has completed 21 tasks at the first measurement.

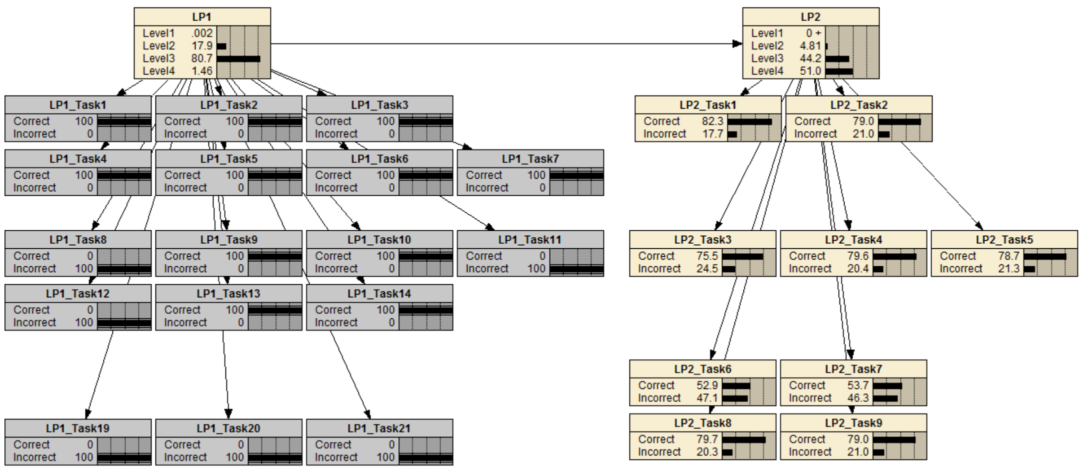
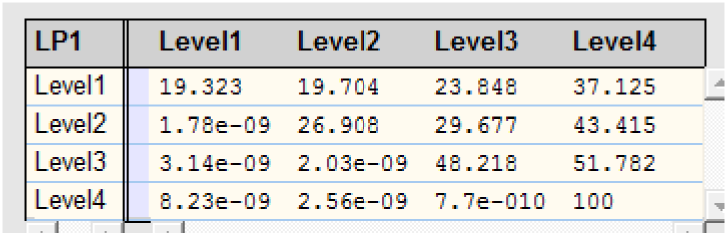


Figure 50 DBIN for a student who has complete 21 tasks at the first measurement

The student has the response pattern of [1,1,1,1,1,1,1,0,1,1,0,0,1,1,0,0,0] at the first measurement. On the basis of this evidence, the posterior distributions for the student's LP1 and LP2 indicate that the student has a probability of being at levels 1-4 of .002, .179, .807, and .146, respectively at the first measurement and a probability of being at levels 1-4 of .000, .048, .442, and .510. On this basis it may be inferred that the student is more likely to be level 3 at the first measurement and is more likely to be level 3 and level 4 at the second measurement. However, there still remains uncertainty. The inclusion of more evidence from the data collection would help to reduce uncertainty regarding the student inference at the second measurement.

Transition Probabilities

In addition to the inference of a student's level change over time, the DBIN offers the probabilities of the transition between two measurements through the transition probability table. Before cleaning abnormal data, the strange and awkward transition from level 4 down to level 1 was observed although the prior distribution was imposed. It means that if the data strongly indicated backwards transitions, they could appear in the posteriors. This finding helped to detect some problematic cases (e.g., students who got all tasks correctly at the first measurement, but got all missing responses or all tasks incorrect at the second measurement). After cleaning the problematic data set, the results were more plausible. Figure 51 is the resulting transition probability table, which shows the probabilities of students having reached each level at the second measurement given their levels at the first measurement. For instance, 19.7% of students at level 1 at the first measurement had moved to level 2 at the second measurement. For the backward transition movements, almost zero probabilities were estimated. That was because not only a constraint of no-backward movements was set using the prior information, but also because the data indicated little or no evidence of this phenomenon. With this information, we can infer the proportions of students that stay at the same level and move to different higher levels between the two consecutive measurements.



LP1	Level1	Level2	Level3	Level4
Level1	19.323	19.704	23.848	37.125
Level2	1.78e-09	26.908	29.677	43.415
Level3	3.14e-09	2.03e-09	48.218	51.782
Level4	8.23e-09	2.56e-09	7.7e-010	100

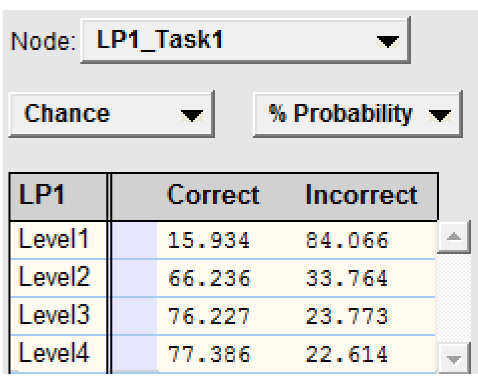
Figure 51Transition probability table

Communicating with Content Experts

The results based on empirical data analysis can serve to aid the development of KSAs that constitute the LP and student inferences. In some cases, the results for tasks were consistent with the expert-based expectation. For other tasks, the results were more ambiguous or suggest an alternative interpretation to that of the experts. The results of the data analysis may be taken back to the content experts for consultation and possible refinements in terms of the definition of the LP, the tasks that assess the aspects of the LP, and the utility of additional tasks for modeling students' progression.

Among three tasks that were not consistent with the expert expectations, the content expert agreed that the level of Task 5 needed to be refined. Task 5 was originally identified as at being level 4 by the content expert, but the data analysis suggested that the task would be useful for classifying students between level 1-2 and level 3-4 (Figure 48).

The content expert commented that the Task 8 and Task 9 fit perfectly into the levels that have been originally identified although the data analysis suggested the different levels. However, he has pointed that Task 8 requires cognitively simple recall process to complete the task, which might make them easier than other tasks at the same levels. In other words, although the tasks measure the higher level of KSAs in terms of content, they required a lower level of cognitive ability to solve the tasks. This may be a possible reason that the data analysis suggested the lower level (level 3) rather than their expectation (level 4). The content expert had a strong belief that task 9 should keep the same level originally identified. The task has relatively higher p-value (0.7) than other tasks in the same level (level 4). Therefore, there may be other factors that influence the level of task difficulty such as difficult distracters and task format. Table 142 shows the summary of the agreement between expectation and data analysis. P-value (i.e., percent-correct) does not provide sufficient information to see if a task is cofrrectly located. For instance, Task 1 at the first measurement seems to be incorrectly located based on the p-value (i.e., it has relatively low p-value). However, the task performed very well for classifying the students between at level 1 and level 2, 3, and 4 (Figure 52).


Node: LP1_Task1

Chance | % Probability

LP1	Correct	Incorrect
Level1	15.934	84.066
Level2	66.236	33.764
Level3	76.227	23.773
Level4	77.386	22.614

Figure 52. Task 1 at the first measurement

Table 142

Consistency of expert expectation and data analysis

Chapter 6			
Task	Expectation	Data Suggestion	P-value
LP1_Task1	Level 2	Level 2	0.560
LP1_Task2	Level 2	Level 2	0.874
LP1_Task3	Level 2	Level 2	0.922
LP1_Task4	Level 2	Ambiguous	0.923
LP1_Task5	Level 2	Level 2	0.509
LP1_Task6	Level 2	Ambiguous	0.874
LP1_Task7	Level 2	Level 2	0.796
LP1_Task8	Level 3	Level 3	0.587
LP1_Task9	Level 3	Level 3	0.655
LP1_Task10	Level 3	Level 3	0.683
LP1_Task11	Level 3	Level 3	0.706
LP1_Task12	Level 3	Level 3	0.732
LP1_Task13	Level 3	Level 3	0.690
LP1_Task14	Level 3	Level 3	0.680
LP1_Task19	Level 4	Level 4	0.757
LP1_Task20	Level 4	Ambiguous	0.760
LP1_Task21	Level 4	Level 4	0.603

Final Exam				
Task	Expectation	Data Suggestion	P-value	Comments
LP2_Task1	Level 2	Level 2	0.745257	
LP2_Task2	Level 2	Level 2	0.596841	
LP2_Task3	Level 3	Ambiguous	0.522898	
LP2_Task4	Level 3	Level 3	0.702869	
LP2_Task5	**Level 4**	**Level 3**	**0.733876**	**Refined as level 3**
LP2_Task6	Level 4	Ambiguous	0.608219	
LP2_Task7	Level 4	Level 4	0.557299	
LP2_Task8	**Level 4**	**Level 3**	**0.754278**	**Cognitively simple**
LP2_Task9	**Level 4**	**Level 3**	**0.703274**	

CHAPTER 9: CONTRIBUTIONS, LIMITATIONS, AND FUTURE STUDY

Contributions of This Study

Formative assessments are increasingly of interest in the field of education where the focus of assessment is moving toward assessing students' learning progress during instruction rather than focusing only on their end of program achievements. The use of formative assessment is being expanded to identify a gap between actual student levels and desired levels of performance and to provide information for reducing student weaknesses. For this purpose, an assessment must produce evidence for revealing student levels and their change over time. The DBIN is a useful statistical modeling method that can make inferences about level change over time when task design and theory provide not only a theoretical framework for creating and modeling observable evidence, but also information about the nature and structure of expected change. It provides real-time updating of estimates for student level during instruction, so that it offers beneficial information to students, instructors, and curriculum developers for enhancing student learning.

Simulation-based assessments, learning systems, and intelligent tutoring systems increasingly have captured attention in education with some potential benefits (VanLehn, 2006). The learning systems have students enter steps leading up to the solution of a problem and it can give feedback and hints on those steps as well as the final answer (Corbett, Koedinger, & Anderson, 1997; Rickel, 1989; VanLehn, 2006). Therefore, learning systems can gather information about student performances on intermediate steps as well as the final answer, so that they can measure not only what a student knows, but also how a student solves problems and what strategies the student used to complete a task. It offers information about which parts or steps are difficult for a student to learn as well as how well the student is doing during a course. In order to obtain such evidence, decisions must be made about monitoring learning and making instructional choices. Bayes nets have been useful on this level (VanLehn, 2006). For instance, a set of knowledge components corresponding to the steps or small pieces of domain knowledge that a student should learn can be built as the nodes of Bayes nets. Then, the probability of mastery of a knowledge component (each node in Bayes nets) can be estimated. The distribution of the probabilities of each knowledge component can reveal which knowledge components have lower probabilities of mastery and higher probabilities of mastery. The knowledge components having lower probabilities can indicate the concepts that students have difficulty understanding. Whereas, the knowledge components having higher probabilities can indicate the concepts that students understand well.

In addition to formative assessment and learning systems, LPs are increasingly of interest in education. The research related to LPs informs the state of a student with respect to their level of understanding of a given concept and diagnostic information regarding the strengths and weaknesses of a student's understanding along a curriculum. Furthermore, the study of LPs offers the opportunity to explore how students build their KSAs over time, and what evidence is needed not only for assessing students' learning, but also for evaluating and refining the defined learning progressions, curricula, and instructions.

However, there has been relatively little work on measurement modeling in the context of LPs, Challenges arise in many areas of LPs research, including (1) designing a coherent assessment system, (2) inferring student learning progression levels based on the responses to assessment tasks, and (3) interpreting the difference between expected and observed students' progress mapped to the conceptually-defined learning progression. More specifically, in terms of inferential challenges of modeling LPs, issues have arisen, including (1) deciding what methodologies can be used for the inference about students' learning progression levels based on student performance on a set of assessment tasks, (2) determining how students' inconsistent patterns can be explained and modeled, (3) determining how observed student responses could be compared to expected student response, and (4) understanding how the substance of learning progressions and assessment tasks could be refined by the implications of differences between observed and expected responses. Bayes nets can be a useful tool for modeling LPs by linking the theory embodied in a progression, tasks that provide evidence about a student's level on that progression, and psychometric models that can characterize the relationship between student performance and levels on learning progressions.

DBINs are a framework for modeling LPs in a longitudinal design approach. This dissertation showed the potential benefits of using BINs for this purpose, focusing on the dynamic case. This dissertation extends the paradigm to changes over time, and additionally includes a covariate structure that can be useful for guiding and evaluating instructional options. The extension that incorporates a covariate for students is useful for studying the effect of instructional treatments, students' background, and motivation on student learning progressions.

The BIN approach has the advantage of building on a structure that can be based on theory and expert opinion, and then data can be accumulated for improving the estimates. Since it is a flexible statistical modeling framework that can build any statistical model in accordance with substantive theory, it can be utilized for a complex performance assessment as well as a simple assessment. In other words, it can be easily extended to complex tasks or multiple aspects of a performance, and also provide real-time updating of estimates for students' proficiency. In the LP research, the work of BINs in light of LPs helps make LPs more useful other than providing student inferences with respect to the level on a LP. Using BINs to model LPs can help lead to efficient and valid task design. The process of identifying initial LPs helps test developers focus on the theory of cognition in the domain and defines the characteristics of individuals at various levels of the LP. BINs confirm these levels and progressions by comparing the results from data analysis, allowing task designers to specify the levels of KSAs at which they are aiming assessment tasks. This helps make task design more principled, well planned, and ultimately more valid. The BINs also help connect curriculum to assessment. For example, curriculum designers can take information from a BIN structure and make decisions about which content areas are more important to emphasize so that students will have a greater probability of mastering future KSAs (DiCerbo & Behrens, 2008).

Although BIN provides a promising means by which to model student performance and their flexibility makes them particularly useful in modeling a various range of assessment situations, there are many decisions that must be made when building a BIN for modeling an assessment. The structure of the relation among variables must be determined. The structure can be determined by communicating with content experts as well as by using statistical methods such as Structural Equation Modeling. In addition, it may need information about the probability distributions of variables in the BIN as prior information. While the probability distributions can be determined from data, the structure can be determined before data is collected. Therefore, it is important to keep in mind the purpose of the assessment in order to determine which relations are important to model. This process is not always straightforward and may require some iterative work before the model can be said to be good enough.

Limitations of the Current Study

This current study was initially designed for modeling a LP over two measurements. The study focused on two inferences: (1) how well a task classifies students with different levels on a LP (quality of a task in terms of a classification) and (2) what would a student's learning path be over two measurements given a LP. The first question can be answered by investigating the conditional probability table of a task. This current study did not address the statistical methods that can evaluate the power of a task as to how well the task discriminates among students with different LP levels, but the conditional probability tables have been inspected to see if there is a distinct pattern. The use of analytic methods such as computing the odds ratio could provide more accurate information about the quality of a task. Regarding the student inference about a path over multiple time points through a LP, this current study was designed only for two measurements. However, it may also be of interest to monitor the level change over more than two measurements. If the situation where students are measured more than two times is considered, it may be of interest to consider the higher-order Markov property other than the first-order Markov property. This current study considered the first-order Markov property in the transition probability table because two measurements were designed. The designing of DBINs with more than two measurements with the higher-order Markov property could reflect a more realistic educational setting.

This current study explored the forward movement transition (i.e., there are no backward movements) in order to coincide with the substantive theory of LPs. This constraint may be very strong, but this is a good candidate to start to learn about DBIN for modeling LP. Other types of movement can be easily designed by imposing constraints. The other types of movements could be built in accordance with the substantive theory or the system of curriculum and instruction during a course. For instance, researchers may be interested in a situation where once a student passed a certain level, the student could move to the adjacent levels at the next measurement. If there is no theory or background information, the appropriate transition model can be determined by statistical model comparison with data.

On the estimation, Bayesian method was used by incorporating prior information. Therefore, the estimates were influenced by prior information as well as data. Netica has a function of what degree prior information influences the estimates. The current study used one degree as a weight of prior information (i.e., theoretically, one degree is equivalent to a sample size of 1). This is very mild prior information. However, the effect of prior information on the estimates has not been carefully investigated in this study.

Through the preliminary analysis and literature review, only two different conditions of sample size and task size were chosen in the simulation studies. However, different sample sizes and task sizes other than the conditions considered could be of interest.

In the second simulation study, a covariate was incorporated into the simple DBIN model. This covariate was considered as a manifest variable, which the observations regarding the covariate have been made before starting the estimation. In other words, group membership of each student in terms of a covariate is known. However, a latent variable such as motivation, attitude, and intelligence could be considered as a covariate. In this case, observable variables posited to depend on it, such as responses to a survey or to an interviewer, would be included in the model as indicators of the latent covariate.

The item equating that this current study used is essentially concurrent calibration. It is assumed that there is no item parameter drift. However, to the extent that there is drift, it introduces a tendency to overestimate examinee's capabilities. This possibility is beyond the scope of the current study.

The current study did not examine all possible conditions in the two simulation studies. This does not reflect all possible situations in real educational setting.

Issues for Future Study

Multidimensionality. This current study was initially designed with one LP variable corresponding to a domain. However, more than one LP reflecting more than one domain may be of interest to be modeled over time. For example, tasks can depend on more than one LP and students might have different learning patterns in terms of multiple LPs representing different domains. More specifically, a task may require a student to be at higher level on one LP while additionally requiring a lower level on another LP. In the case of multiple LPs, different learning paths along the multiple LPs can be modeled over time. Modeling multidimensional LPs for can be investigated in a future study.

Different types of transition movement. The current study considered only the forward transition movement in the two simulation studies because the forward transition movement is the most appropriate structure for representing student movements along a LP. This is a strong constraint but a usual hypothesis and the first natural one to learn about DBIN for modeling LP. However, there can be different types of transition structures such as the backward transition movement and all transition movement. The modeling of the transition probability matrix is very flexible, as effected by imposing different types of restrictions depending on theory (Kaplan, 2008). The transition probability matrix for the forward movements can be modeled by imposing a constraint that all transition probabilities in backward movements are zeros. Different patterns in transition matrices depending on substantive theory can be modeled by imposing some restrictions by (1) constraining sets of transition probabilities to be equal to zero, (2) restricting them to be a particular value, or (3) fixing them to be equal to each other. The scope of this dissertation was the investigation of the forward transition movement, which is the most appropriate structure for representing LPs. It is left to future studies to investigate the selection of the best fit transition movement structure.

Higher-order Markov property. The current study investigated the first-order Markov property to explain transition movements. However, there may be situations that a student's status at more than one previous time point affects the student's current status (higher-order Markov property). The investigation of the higher-order Markov property is left to future study. Moreover, the two simulation studies included in this dissertation only considered two measurement time points. If the higher-order Markov property is investigated, more than two measurement time points would be necessary in the future study.

Psychometric model. Parameters of both transition probability table and conditional probability table can put into parametric forms such as the Samejima-Dibello model (Mislevy et al, 2002). These models increase the complexity of the mathematical structure of the model, but reduce the number of parameters to be estimated and thus can improve stability and accuracy in estimation. Addressing these models as compared to an unconstrained BIN framework would be of interest in the future study.

Prior Information. The current study uses prior information to control the label switching issue, but the effect of different prior information on parameter recovery was not investigated in this dissertation. BINs can have different weights of prior information when estimating parameters. Effects of prior information on parameter recovery can be investigated in future studies.

Appendix A: Bayesian Inference Network: Belief Updating through Junction Tree method

The junction tree method is a popular method that most of the software packages of BINs are used. The belief updating through junction tree algorithm can follow the seven steps below (Almond, Mislevy, Steinberg, Williamson, and Yan, in progress).

(1) Recursive representation of the joint distribution of variables
(2) Acyclic directed graph representation of the probability distribution
(3) Representation as a moralized and triangulated undirected graph
(4) Determination of cliques and clique intersections
(5) Junction tree representation
(6) Potential tables
(7) Calculations with potential tables
(8) Receiving evidence
(9) Updating potential tables

For example, the DAG (figure 10) can be transferred to a junction tree representation

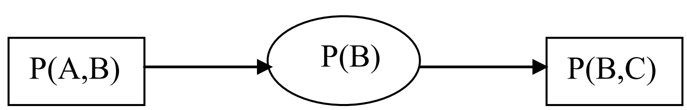


Figure A.1. A junction tree of the acyclic directed graph

The nodes containing both variables, {A,B} and {B,C}, express interrelationships among variables that directly influence one another, called clique nodes. The node for the individual variable, {B}, is the intermediate area where information common to adjacent cliques, called intersection nodes, is found. Each node in the junction tree stores a potential table. The calculation of potential tables can start with the clique nodes. Suppose that there are hypothetical probability distributions of P(B|A), P(C|B), and P(A).

Table A1

Probability distribution of P(A)

Variable A		
Values	A_1=0	A_2=1
Probability	0.5	0.5

Table A2

Probability distribution of P(B|A)

		Variable B	
		B_1=0	B_2=1
Variable A	A_1=0	0.3	0.7
	A_2=1	0.8	0.2

Table A3

Probability distribution of P(C|B)

		Variable C	
		C_1=0	C_2=1
Variable B	B_1=0	0.8	0.2
	B_2=1	0.6	0.4

By combining two probability distributions of P(A) and P(B|A), a potential corresponding to the clique node of {A,B} can be constructed.

$$P(A_1=0, B_1=0) = P(B_1=0|\ A_1=0)P(A_1=0) = 0.3 * 0.5 = 0.15$$
$$P(A_2=1, B_1=0) = P(B_1=0|\ A_2=1)P(A_2=1) = 0.8 * 0.5 = 0.40$$
$$P(A_1=0, B_2=1) = P(B_2=1|\ A_1=0)P(A_1=0) = 0.7 * 0.5 = 0.30$$
$$P(A_2=1, B_2=1) = P(B_1=1|\ A_1=1)P(A_1=1) = 0.2 * 0.5 = 0.10$$

Table A4

A Potential Table of {A,B}

		Variable B		
		B_1=0	B_2=1	Sum
Variable A	A_1=0	0.15	0.35	0.50
	A_2=1	0.40	0.10	0.50
	Sum	0.55	0.45	1

From the potential of {A,B}, the potential of the intersection node of {B} can be calculated by marginalizing out the variable A.

Table A5

A Potential Table of {B}

	Variable B	
Values	B_1=0	B_2=1
Probability	0.55	0.45

The potential of {B} connects to the new clique node of {B,C}. Through the same procedure, the potential of the {B,C} can be computed by combining the probability distributions of P(B) and P(C|B).

$$P(B_1=0, C_1=0) = P(C_1=0|\ B_1=0)P(B_1=0) =0.8 * 0.55 = 0.44$$
$$P(B_2=1, C_1=0) = P(C_1=0|\ B_2=1)P(B_2=1) =0.6 * 0.45 = 0.27$$
$$P(B_1=0, C_2=1) = P(C_2=1|\ B_1=0)P(B_1=0) =0.2 * 0.55 = 0.11$$
$$P(B_2=1, C_2=1) = P(C_1=1|\ B_1=1)P(B_1=1) =0.4 * 0.45 = 0.18$$

Table A6

A Potential Table of {B,C}

		Variable C		
		C_1=0	C_2=1	Sum
Variable B	B_1=0	0.44	0.11	0.55
	B_2=1	0.27	0.18	0.35
	Sum	0.71	0.29	1

Lastly, the potential of {C} can be calculated by marginalizing out the variable B.

Table A7

Potential Table of {C}

Variable C		
Values	C_1=0	C_2=1
Probability	0.71	0.29

The updating belief can be carried out with the junction tree in the same matter, once evidence (observation), {e}, has arrived about variable C. Suppose the observation about variable C is as follows:

Table A8

Evidence for {C}

	Variable C	
Values	C_1=0	C_2=1
Probability	1	0

By combining {e} with the previous potential of {B,C}, the potential of {B,C} is updated, referred as new{B,C|e}.

$$newP(B_1{=}0, C_1{=}0) = oldP(B_1{=}0, C_1{=}0)\ P(C_1{=}0) = 0.44 * 1 = 0.44$$
$$newP(B_2{=}1, C_1{=}0) = oldP(B_2{=}1, C_1{=}0)\ P(C_1{=}0) = 0.27 * 1 = 0.27$$
$$newP(B_1{=}0, C_2{=}1) = oldP(B_1{=}0, C_2{=}1)\ P(C_2{=}1) = 0.11 * 0 = 0$$
$$newP(B_2{=}1, C_2{=}1) = oldP(B_2{=}1, C_2{=}1)\ P(C_2{=}1) = 0.18 * 0 = 0$$

Table A9

A New Potential Table of {B,C|e}

		Variable C		
		C_1=0	C_2=1	Sum
Variable B	B_1=0	0.44*1	0.11*0	0.44
	B_2=1	0.27*1	0.18*0	0.27
	Sum	0.71	0	0.71

By marginalizing out variable C, the new potential of {B} can be computed. The new {B} is an intersection node in the junction tree, connecting two clique nodes. Since there is the old potential of {B}, the adjustment can be obtained by dividing the new potential of {B} by the old one.

Table A10

A Potential Table for {B}

	Variable B	
Values	B_1=0	B_2=1
Probability	0.44/0.55=0.8	0.27/0.45=0.6

Lastly, the potential of {A,B} is updated by combining the adjusted potential of {B} and the old potential of {A,B}.

$$newP(A_1{=}0, B_1{=}0) = oldP(A_1{=}0, B_1{=}0)\ P(B_1{=}0) = 0.15 * 0.8 = 0.12$$
$$newP(A_2{=}1, B_1{=}0) = oldP(A_2{=}1, B_1{=}0)\ P(B_1{=}0) = 0.4 * 0.8 = 0.32$$
$$newP(A_1{=}0, B_2{=}1) = oldP(A_1{=}0, B_2{=}1)\ P(B_2{=}1) = 0.35 * 0.6 = 0.21$$
$$newP(A_2{=}1, B_2{=}1) = oldP(A_2{=}1, B_2{=}1)\ P(B_2{=}1) = 0.1 * 0.6 = 0.06$$

Table A11

A Potential Table of {A,B}

		Variable B		
		B_1=0	B_2=1	Sum
Variable A	A_1=0	0.12	0.21	0.33
	A_2=1	0..32	0.06	0.38
	Sum	0.44	0.27	0.71

The clique node {A,B} is the last node in the Junction tree, so the procedure of updating the BIN is complete. The joint distribution of the tree is P(A,B,C|e). If one wants to interpret the marginal probability distributions for one or more variables, the values of the potential tables are necessary to be normalized to sum to one (Almond, Mislevy, Steinberg, Williamson, & Yan, in progress). For instance, the normalization constant of {A} is 0.71. The marginal probability of potential of {A} is normalized by dividing each probability by the normalization constant. The normalization of the potential table of {A } is as follows:

$$P(A_1=0) = 0.33/0.71 = 0.365$$
$$P(A_2=1) = 0.38/0.71 = 0.535$$

Table A12

A normalized probability table of variable A

	Variable A		
Values	A_1=0	A_2=1	Sum
Probability	0.33	0.38	0.71
Normalized	0.365	0.535	1

Table A12 is the posterior probability table of variable A updated by the observation of variable C. Compared to the prior probability table of variable A (table A1), it is shown that the probability of each state on variable A has changed and learned from observation of variable C.

Appendix B: Netica C-API syntax

```
#include <stdio.h>
#include <stdlib.h>
#include "Netica.h"
#include "NeticaEx.h"

#define CHKERR  {if (GetError_ns (env, ERROR_ERR, NULL))  goto error;}

environ_ns* env;

int main (void){
        net_bn *net = NULL;
        node_bn
        *T1,*T2,*T3,*T4,*T5,*T6,*T7,*T8,*T9,*T10,*T11,*T12,*T13,*T14,*T15,*T1
        6,*T17,*T18,*T19,*T20,*T21,*T22,*T23,*T24,*T25,*T26,*T27,*T28,*T29,*T
        30,
        *Ta1,*Ta2,*Ta3,*Ta4,*Ta5,*Ta6,*Ta7,*Ta8,*Ta9,*Ta10,*Ta11,*Ta12,*Ta13,*
        Ta14,*Ta15,*Ta16,*Ta17,*Ta18,*Ta19,*Ta20,*Ta21,*Ta22,*Ta23,*Ta24,*Ta25
        ,*Ta26,*Ta27,*Ta28,*Ta29,*Ta30,
        *Mone,*Mtwo;
        const nodelist_bn* nodes;
        state_bn parent_states[10];
        stream_ns* casefile = NULL;
        caseset_cs* cases = NULL;
        learner_bn *learner = NULL;
        char mesg[MESG_LEN_ns];
        int res;
        report_ns* err;
        int pn;

        env = NewNeticaEnviron_ns (NULL, NULL, NULL);
        res = InitNetica2_bn (env, mesg);
        printf ("%s\n", mesg);
        if (res < 0)  exit (-1);

        /* Build the net */

        net = NewNet_bn ("Sim1_task30", env);
        CHKERR

        T1 = NewNode_bn ("T1", 2, net);
        T2 = NewNode_bn ("T2", 2, net);
```

```
T3 = NewNode_bn ("T3", 2, net);
T4 = NewNode_bn ("T4", 2, net);
T5 = NewNode_bn ("T5", 2, net);
T6 = NewNode_bn ("T6", 2, net);
T7 = NewNode_bn ("T7", 2, net);
T8 = NewNode_bn ("T8", 2, net);
T9 = NewNode_bn ("T9", 2, net);
T10 = NewNode_bn ("T10", 2, net);
T11 = NewNode_bn ("T11", 2, net);
T12 = NewNode_bn ("T12", 2, net);
T13 = NewNode_bn ("T13", 2, net);
T14 = NewNode_bn ("T14", 2, net);
T15 = NewNode_bn ("T15", 2, net);
T16 = NewNode_bn ("T16", 2, net);
T17 = NewNode_bn ("T17", 2, net);
T18 = NewNode_bn ("T18", 2, net);
T19 = NewNode_bn ("T19", 2, net);
T20 = NewNode_bn ("T20", 2, net);
T21 = NewNode_bn ("T21", 2, net);
T22 = NewNode_bn ("T22", 2, net);
T23 = NewNode_bn ("T23", 2, net);
T24 = NewNode_bn ("T24", 2, net);
T25 = NewNode_bn ("T25", 2, net);
T26 = NewNode_bn ("T26", 2, net);
T27 = NewNode_bn ("T27", 2, net);
T28 = NewNode_bn ("T28", 2, net);
T29 = NewNode_bn ("T29", 2, net);
T30 = NewNode_bn ("T30", 2, net);

Ta1 = NewNode_bn ("Ta1", 2, net);
Ta2 = NewNode_bn ("Ta2", 2, net);
Ta3 = NewNode_bn ("Ta3", 2, net);
Ta4 = NewNode_bn ("Ta4", 2, net);
Ta5 = NewNode_bn ("Ta5", 2, net);
Ta6 = NewNode_bn ("Ta6", 2, net);
Ta7 = NewNode_bn ("Ta7", 2, net);
Ta8 = NewNode_bn ("Ta8", 2, net);
Ta9 = NewNode_bn ("Ta9", 2, net);
Ta10 = NewNode_bn ("Ta10", 2, net);
Ta11 = NewNode_bn ("Ta11", 2, net);
Ta12 = NewNode_bn ("Ta12", 2, net);
Ta13 = NewNode_bn ("Ta13", 2, net);
Ta14 = NewNode_bn ("Ta14", 2, net);
Ta15 = NewNode_bn ("Ta15", 2, net);
```

```
Ta16 = NewNode_bn ("Ta16", 2, net);
Ta17 = NewNode_bn ("Ta17", 2, net);
Ta18 = NewNode_bn ("Ta18", 2, net);
Ta19 = NewNode_bn ("Ta19", 2, net);
Ta20 = NewNode_bn ("Ta20", 2, net);
Ta21 = NewNode_bn ("Ta21", 2, net);
Ta22 = NewNode_bn ("Ta22", 2, net);
Ta23 = NewNode_bn ("Ta23", 2, net);
Ta24 = NewNode_bn ("Ta24", 2, net);
Ta25 = NewNode_bn ("Ta25", 2, net);
Ta26 = NewNode_bn ("Ta26", 2, net);
Ta27 = NewNode_bn ("Ta27", 2, net);
Ta28 = NewNode_bn ("Ta28", 2, net);
Ta29= NewNode_bn ("Ta29", 2, net);
Ta30 = NewNode_bn ("Ta30", 2, net);


Mone = NewNode_bn ("Mone", 4, net);// the latent node;
Mtwo = NewNode_bn ("Mtwo", 4, net);// the latent node;

SetNodeStateName_bn (T1, 0, "correct");
SetNodeStateName_bn (T1, 1, "incorrect");
SetNodeStateName_bn (T2, 0,"correct");
SetNodeStateName_bn (T2, 1,"incorrect");
SetNodeStateName_bn (T3, 0, "correct");
SetNodeStateName_bn (T3, 1, "incorrect");
SetNodeStateName_bn (T4, 0,"correct");
SetNodeStateName_bn (T4, 1,"incorrect");
SetNodeStateName_bn (T5, 0,"correct");
SetNodeStateName_bn (T5, 1,"incorrect");
SetNodeStateName_bn (T6, 0, "correct");
SetNodeStateName_bn (T6, 1, "incorrect");
SetNodeStateName_bn (T7, 0,"correct");
SetNodeStateName_bn (T7, 1,"incorrect");
SetNodeStateName_bn (T8, 0, "correct");
SetNodeStateName_bn (T8, 1, "incorrect");
SetNodeStateName_bn (T9, 0,"correct");
SetNodeStateName_bn (T9, 1,"incorrect");
SetNodeStateName_bn (T10, 0,"correct");
SetNodeStateName_bn (T10, 1,"incorrect");
SetNodeStateName_bn (T11, 0, "correct");
SetNodeStateName_bn (T11, 1, "incorrect");
SetNodeStateName_bn (T12, 0,"correct");
SetNodeStateName_bn (T12, 1,"incorrect");
```

```
SetNodeStateName_bn (T13, 0, "correct");
SetNodeStateName_bn (T13, 1, "incorrect");
SetNodeStateName_bn (T14, 0,"correct");
SetNodeStateName_bn (T14, 1,"incorrect");
SetNodeStateName_bn (T15, 0,"correct");
SetNodeStateName_bn (T15, 1,"incorrect");
SetNodeStateName_bn (T16, 0, "correct");
SetNodeStateName_bn (T16, 1, "incorrect");
SetNodeStateName_bn (T17, 0,"correct");
SetNodeStateName_bn (T17, 1,"incorrect");
SetNodeStateName_bn (T18, 0, "correct");
SetNodeStateName_bn (T18, 1, "incorrect");
SetNodeStateName_bn (T19, 0,"correct");
SetNodeStateName_bn (T19, 1,"incorrect");
SetNodeStateName_bn (T20, 0,"correct");
SetNodeStateName_bn (T20, 1,"incorrect");
SetNodeStateName_bn (T21, 0, "correct");
SetNodeStateName_bn (T21, 1, "incorrect");
SetNodeStateName_bn (T22, 0,"correct");
SetNodeStateName_bn (T22, 1,"incorrect");
SetNodeStateName_bn (T23, 0, "correct");
SetNodeStateName_bn (T23, 1, "incorrect");
SetNodeStateName_bn (T24, 0,"correct");
SetNodeStateName_bn (T24, 1,"incorrect");
SetNodeStateName_bn (T25, 0,"correct");
SetNodeStateName_bn (T25, 1,"incorrect");
SetNodeStateName_bn (T26, 0, "correct");
SetNodeStateName_bn (T26, 1, "incorrect");
SetNodeStateName_bn (T27, 0,"correct");
SetNodeStateName_bn (T27, 1,"incorrect");
SetNodeStateName_bn (T28, 0, "correct");
SetNodeStateName_bn (T28, 1, "incorrect");
SetNodeStateName_bn (T29, 0,"correct");
SetNodeStateName_bn (T29, 1,"incorrect");
SetNodeStateName_bn (T30, 0,"correct");
SetNodeStateName_bn (T30, 1,"incorrect");

SetNodeStateName_bn (Ta1, 0, "correct");
SetNodeStateName_bn (Ta1, 1, "incorrect");
SetNodeStateName_bn (Ta2, 0,"correct");
SetNodeStateName_bn (Ta2, 1,"incorrect");
SetNodeStateName_bn (Ta3, 0, "correct");
SetNodeStateName_bn (Ta3, 1, "incorrect");
SetNodeStateName_bn (Ta4, 0,"correct");
```

```
SetNodeStateName_bn (Ta4, 1,"incorrect");
SetNodeStateName_bn (Ta5, 0,"correct");
SetNodeStateName_bn (Ta5, 1,"incorrect");
SetNodeStateName_bn (Ta6, 0, "correct");
SetNodeStateName_bn (Ta6, 1, "incorrect");
SetNodeStateName_bn (Ta7, 0,"correct");
SetNodeStateName_bn (Ta7, 1,"incorrect");
SetNodeStateName_bn (Ta8, 0, "correct");
SetNodeStateName_bn (Ta8, 1, "incorrect");
SetNodeStateName_bn (Ta9, 0,"correct");
SetNodeStateName_bn (Ta9, 1,"incorrect");
SetNodeStateName_bn (Ta10, 0,"correct");
SetNodeStateName_bn (Ta10, 1,"incorrect");
SetNodeStateName_bn (Ta11, 0, "correct");
SetNodeStateName_bn (Ta11, 1, "incorrect");
SetNodeStateName_bn (Ta12, 0,"correct");
SetNodeStateName_bn (Ta12, 1,"incorrect");
SetNodeStateName_bn (Ta13, 0, "correct");
SetNodeStateName_bn (Ta13, 1, "incorrect");
SetNodeStateName_bn (Ta14, 0,"correct");
SetNodeStateName_bn (Ta14, 1,"incorrect");
SetNodeStateName_bn (Ta15, 0,"correct");
SetNodeStateName_bn (Ta15, 1,"incorrect");
SetNodeStateName_bn (Ta16, 0, "correct");
SetNodeStateName_bn (Ta16, 1, "incorrect");
SetNodeStateName_bn (Ta17, 0,"correct");
SetNodeStateName_bn (Ta17, 1,"incorrect");
SetNodeStateName_bn (Ta18, 0, "correct");
SetNodeStateName_bn (Ta18, 1, "incorrect");
SetNodeStateName_bn (Ta19, 0,"correct");
SetNodeStateName_bn (Ta19, 1,"incorrect");
SetNodeStateName_bn (Ta20, 0,"correct");
SetNodeStateName_bn (Ta20, 1,"incorrect");
SetNodeStateName_bn (Ta21, 0, "correct");
SetNodeStateName_bn (Ta21, 1, "incorrect");
SetNodeStateName_bn (Ta22, 0,"correct");
SetNodeStateName_bn (Ta22, 1,"incorrect");
SetNodeStateName_bn (Ta23, 0, "correct");
SetNodeStateName_bn (Ta23, 1, "incorrect");
SetNodeStateName_bn (Ta24, 0,"correct");
SetNodeStateName_bn (Ta24, 1,"incorrect");
SetNodeStateName_bn (Ta25, 0,"correct");
SetNodeStateName_bn (Ta25, 1,"incorrect");
SetNodeStateName_bn (Ta26, 0, "correct");
```

```
SetNodeStateName_bn (Ta26, 1, "incorrect");
SetNodeStateName_bn (Ta27, 0,"correct");
SetNodeStateName_bn (Ta27, 1,"incorrect");
SetNodeStateName_bn (Ta28, 0, "correct");
SetNodeStateName_bn (Ta28, 1, "incorrect");
SetNodeStateName_bn (Ta29, 0,"correct");
SetNodeStateName_bn (Ta29, 1,"incorrect");
SetNodeStateName_bn (Ta30, 0,"correct");
SetNodeStateName_bn (Ta30, 1,"incorrect");

SetNodeStateName_bn (Mone, 0,"level1");
SetNodeStateName_bn (Mone, 1,"level2");
SetNodeStateName_bn (Mone, 2,"level3");
SetNodeStateName_bn (Mone, 3,"level4");
SetNodeStateName_bn (Mtwo, 0,"level1");
SetNodeStateName_bn (Mtwo, 1,"level2");
SetNodeStateName_bn (Mtwo, 2,"level3");
SetNodeStateName_bn (Mtwo, 3,"level4");

AddLink_bn (Mone,T1);
AddLink_bn (Mone,T2);
AddLink_bn (Mone,T3);
AddLink_bn (Mone,T4);
AddLink_bn (Mone,T5);
AddLink_bn (Mone,T6);
AddLink_bn (Mone,T7);
AddLink_bn (Mone,T8);
AddLink_bn (Mone,T9);
AddLink_bn (Mone,T10);
AddLink_bn (Mone,T11);
AddLink_bn (Mone,T12);
AddLink_bn (Mone,T13);
AddLink_bn (Mone,T14);
AddLink_bn (Mone,T15);
AddLink_bn (Mone,T16);
AddLink_bn (Mone,T17);
AddLink_bn (Mone,T18);
AddLink_bn (Mone,T19);
AddLink_bn (Mone,T20);
AddLink_bn (Mone,T21);
AddLink_bn (Mone,T22);
AddLink_bn (Mone,T23);
AddLink_bn (Mone,T24);
AddLink_bn (Mone,T25);
```

```
AddLink_bn (Mone,T26);
AddLink_bn (Mone,T27);
AddLink_bn (Mone,T28);
AddLink_bn (Mone,T29);
AddLink_bn (Mone,T30);

AddLink_bn (Mtwo,Ta1);
AddLink_bn (Mtwo,Ta2);
AddLink_bn (Mtwo,Ta3);
AddLink_bn (Mtwo,Ta4);
AddLink_bn (Mtwo,Ta5);
AddLink_bn (Mtwo,Ta6);
AddLink_bn (Mtwo,Ta7);
AddLink_bn (Mtwo,Ta8);
AddLink_bn (Mtwo,Ta9);
AddLink_bn (Mtwo,Ta10);
AddLink_bn (Mtwo,Ta11);
AddLink_bn (Mtwo,Ta12);
AddLink_bn (Mtwo,Ta13);
AddLink_bn (Mtwo,Ta14);
AddLink_bn (Mtwo,Ta15);
AddLink_bn (Mtwo,Ta16);
AddLink_bn (Mtwo,Ta17);
AddLink_bn (Mtwo,Ta18);
AddLink_bn (Mtwo,Ta19);
AddLink_bn (Mtwo,Ta20);
AddLink_bn (Mtwo,Ta21);
AddLink_bn (Mtwo,Ta22);
AddLink_bn (Mtwo,Ta23);
AddLink_bn (Mtwo,Ta24);
AddLink_bn (Mtwo,Ta25);
AddLink_bn (Mtwo,Ta26);
AddLink_bn (Mtwo,Ta27);
AddLink_bn (Mtwo,Ta28);
AddLink_bn (Mtwo,Ta29);
AddLink_bn (Mtwo,Ta30);

AddLink_bn (Mone,Mtwo);
CHKERR

SetNodeProbs(T1,"level1", 0.2,0.8);
SetNodeProbs(T1,"level2", 0.85,0.15);
SetNodeProbs(T1,"level3", 0.85,0.15);
SetNodeProbs(T1,"level4", 0.85,0.15);
```

```
SetNodeProbs(T2,"level1", 0.2,0.8);
SetNodeProbs(T2,"level2", 0.85,0.15);
SetNodeProbs(T2,"level3", 0.85,0.15);
SetNodeProbs(T2,"level4", 0.85,0.15);
SetNodeProbs(T3,"level1", 0.2,0.8);
SetNodeProbs(T3,"level2", 0.85,0.15);
SetNodeProbs(T3,"level3", 0.85,0.15);
SetNodeProbs(T3,"level4", 0.85,0.15);
SetNodeProbs(T4,"level1", 0.2,0.8);
SetNodeProbs(T4,"level2", 0.85,0.15);
SetNodeProbs(T4,"level3", 0.85,0.15);
SetNodeProbs(T4,"level4", 0.85,0.15);
SetNodeProbs(T5,"level1", 0.2,0.8);
SetNodeProbs(T5,"level2", 0.85,0.15);
SetNodeProbs(T5,"level3", 0.85,0.15);
SetNodeProbs(T5,"level4", 0.85,0.15);
SetNodeProbs(T6,"level1", 0.2,0.8);
SetNodeProbs(T6,"level2", 0.85,0.15);
SetNodeProbs(T6,"level3", 0.85,0.15);
SetNodeProbs(T6,"level4", 0.85,0.15);
SetNodeProbs(T7,"level1", 0.2,0.8);
SetNodeProbs(T7,"level2", 0.85,0.15);
SetNodeProbs(T7,"level3", 0.85,0.15);
SetNodeProbs(T7,"level4", 0.85,0.15);
SetNodeProbs(T8,"level1", 0.2,0.8);
SetNodeProbs(T8,"level2", 0.85,0.15);
SetNodeProbs(T8,"level3", 0.85,0.15);
SetNodeProbs(T8,"level4", 0.85,0.15);
SetNodeProbs(T9,"level1", 0.2,0.8);
SetNodeProbs(T9,"level2", 0.85,0.15);
SetNodeProbs(T9,"level3", 0.85,0.15);
SetNodeProbs(T9,"level4", 0.85,0.15);
SetNodeProbs(T10,"level1", 0.2,0.8);
SetNodeProbs(T10,"level2", 0.85,0.15);
SetNodeProbs(T10,"level3", 0.85,0.15);
SetNodeProbs(T10,"level4", 0.85,0.15);

SetNodeProbs(T11,"level1", 0.2,0.8);
SetNodeProbs(T11,"level2", 0.2,0.8);
SetNodeProbs(T11,"level3", 0.85,0.15);
SetNodeProbs(T11,"level4", 0.85,0.15);
SetNodeProbs(T12,"level1", 0.2,0.8);
SetNodeProbs(T12,"level2", 0.2,0.8);
SetNodeProbs(T12,"level3", 0.85,0.15);
```

```
SetNodeProbs(T12,"level4", 0.85,0.15);
SetNodeProbs(T13,"level1", 0.2,0.8);
SetNodeProbs(T13,"level2", 0.2,0.8);
SetNodeProbs(T13,"level3", 0.85,0.15);
SetNodeProbs(T13,"level4", 0.85,0.15);
SetNodeProbs(T14,"level1", 0.2,0.8);
SetNodeProbs(T14,"level2", 0.2,0.8);
SetNodeProbs(T14,"level3", 0.85,0.15);
SetNodeProbs(T14,"level4", 0.85,0.15);
SetNodeProbs(T15,"level1", 0.2,0.8);
SetNodeProbs(T15,"level2", 0.2,0.8);
SetNodeProbs(T15,"level3", 0.85,0.15);
SetNodeProbs(T15,"level4", 0.85,0.15);
SetNodeProbs(T16,"level1", 0.2,0.8);
SetNodeProbs(T16,"level2", 0.2,0.8);
SetNodeProbs(T16,"level3", 0.85,0.15);
SetNodeProbs(T16,"level4", 0.85,0.15);
SetNodeProbs(T17,"level1", 0.2,0.8);
SetNodeProbs(T17,"level2", 0.2,0.8);
SetNodeProbs(T17,"level3", 0.85,0.15);
SetNodeProbs(T17,"level4", 0.85,0.15);
SetNodeProbs(T18,"level1", 0.2,0.8);
SetNodeProbs(T18,"level2", 0.2,0.8);
SetNodeProbs(T18,"level3", 0.85,0.15);
SetNodeProbs(T18,"level4", 0.85,0.15);
SetNodeProbs(T19,"level1", 0.2,0.8);
SetNodeProbs(T19,"level2", 0.2,0.8);
SetNodeProbs(T19,"level3", 0.85,0.15);
SetNodeProbs(T19,"level4", 0.85,0.15);
SetNodeProbs(T20,"level1", 0.2,0.8);
SetNodeProbs(T20,"level2", 0.2,0.8);
SetNodeProbs(T20,"level3", 0.85,0.15);
SetNodeProbs(T20,"level4", 0.85,0.15);

SetNodeProbs(T21,"level1", 0.2,0.8);
SetNodeProbs(T21,"level2", 0.2,0.8);
SetNodeProbs(T21,"level3", 0.2,0.8);
SetNodeProbs(T21,"level4", 0.85,0.15);
SetNodeProbs(T22,"level1", 0.2,0.8);
SetNodeProbs(T22,"level2", 0.2,0.8);
SetNodeProbs(T22,"level3", 0.2,0.8);
SetNodeProbs(T22,"level4", 0.85,0.15);
SetNodeProbs(T23,"level1", 0.2,0.8);
SetNodeProbs(T23,"level2", 0.2,0.8);
```

```
SetNodeProbs(T23,"level3", 0.2,0.8);
SetNodeProbs(T23,"level4", 0.85,0.15);
SetNodeProbs(T24,"level1", 0.2,0.8);
SetNodeProbs(T24,"level2", 0.2,0.8);
SetNodeProbs(T24,"level3", 0.2,0.8);
SetNodeProbs(T24,"level4", 0.85,0.15);
SetNodeProbs(T25,"level1", 0.2,0.8);
SetNodeProbs(T25,"level2", 0.2,0.8);
SetNodeProbs(T25,"level3", 0.2,0.8);
SetNodeProbs(T25,"level4", 0.85,0.15);
SetNodeProbs(T26,"level1", 0.2,0.8);
SetNodeProbs(T26,"level2", 0.2,0.8);
SetNodeProbs(T26,"level3", 0.2,0.8);
SetNodeProbs(T26,"level4", 0.85,0.15);
SetNodeProbs(T27,"level1", 0.2,0.8);
SetNodeProbs(T27,"level2", 0.2,0.8);
SetNodeProbs(T27,"level3", 0.2,0.8);
SetNodeProbs(T27,"level4", 0.85,0.15);
SetNodeProbs(T28,"level1", 0.2,0.8);
SetNodeProbs(T28,"level2", 0.2,0.8);
SetNodeProbs(T28,"level3", 0.2,0.8);
SetNodeProbs(T28,"level4", 0.85,0.15);
SetNodeProbs(T29,"level1", 0.2,0.8);
SetNodeProbs(T29,"level2", 0.2,0.8);
SetNodeProbs(T29,"level3", 0.2,0.8);
SetNodeProbs(T29,"level4", 0.85,0.15);
SetNodeProbs(T30,"level1", 0.2,0.8);
SetNodeProbs(T30,"level2", 0.2,0.8);
SetNodeProbs(T30,"level3", 0.2,0.8);
SetNodeProbs(T30,"level4", 0.85,0.15);

SetNodeProbs(Ta1,"level1", 0.2,0.8);
SetNodeProbs(Ta1,"level2", 0.85,0.15);
SetNodeProbs(Ta1,"level3", 0.85,0.15);
SetNodeProbs(Ta1,"level4", 0.85,0.15);
SetNodeProbs(Ta2,"level1", 0.2,0.8);
SetNodeProbs(Ta2,"level2", 0.85,0.15);
SetNodeProbs(Ta2,"level3", 0.85,0.15);
SetNodeProbs(Ta2,"level4", 0.85,0.15);
SetNodeProbs(Ta3,"level1", 0.2,0.8);
SetNodeProbs(Ta3,"level2", 0.85,0.15);
SetNodeProbs(Ta3,"level3", 0.85,0.15);
SetNodeProbs(Ta3,"level4", 0.85,0.15);
SetNodeProbs(Ta4,"level1", 0.2,0.8);
```

```
SetNodeProbs(Ta4,"level2", 0.85,0.15);
SetNodeProbs(Ta4,"level3", 0.85,0.15);
SetNodeProbs(Ta4,"level4", 0.85,0.15);
SetNodeProbs(Ta5,"level1", 0.2,0.8);
SetNodeProbs(Ta5,"level2", 0.85,0.15);
SetNodeProbs(Ta5,"level3", 0.85,0.15);
SetNodeProbs(Ta5,"level4", 0.85,0.15);
SetNodeProbs(Ta6,"level1", 0.2,0.8);
SetNodeProbs(Ta6,"level2", 0.85,0.15);
SetNodeProbs(Ta6,"level3", 0.85,0.15);
SetNodeProbs(Ta6,"level4", 0.85,0.15);
SetNodeProbs(Ta7,"level1", 0.2,0.8);
SetNodeProbs(Ta7,"level2", 0.85,0.15);
SetNodeProbs(Ta7,"level3", 0.85,0.15);
SetNodeProbs(Ta7,"level4", 0.85,0.15);
SetNodeProbs(Ta8,"level1", 0.2,0.8);
SetNodeProbs(Ta8,"level2", 0.85,0.15);
SetNodeProbs(Ta8,"level3", 0.85,0.15);
SetNodeProbs(Ta8,"level4", 0.85,0.15);
SetNodeProbs(Ta9,"level1", 0.2,0.8);
SetNodeProbs(Ta9,"level2", 0.85,0.15);
SetNodeProbs(Ta9,"level3", 0.85,0.15);
SetNodeProbs(Ta9,"level4", 0.85,0.15);
SetNodeProbs(Ta10,"level1", 0.2,0.8);
SetNodeProbs(Ta10,"level2", 0.85,0.15);
SetNodeProbs(Ta10,"level3", 0.85,0.15);
SetNodeProbs(Ta10,"level4", 0.85,0.15);

SetNodeProbs(Ta11,"level1", 0.2,0.8);
SetNodeProbs(Ta11,"level2", 0.2,0.8);
SetNodeProbs(Ta11,"level3", 0.85,0.15);
SetNodeProbs(Ta11,"level4", 0.85,0.15);
SetNodeProbs(Ta12,"level1", 0.2,0.8);
SetNodeProbs(Ta12,"level2", 0.2,0.8);
SetNodeProbs(Ta12,"level3", 0.85,0.15);
SetNodeProbs(Ta12,"level4", 0.85,0.15);
SetNodeProbs(Ta13,"level1", 0.2,0.8);
SetNodeProbs(Ta13,"level2", 0.2,0.8);
SetNodeProbs(Ta13,"level3", 0.85,0.15);
SetNodeProbs(Ta13,"level4", 0.85,0.15);
SetNodeProbs(Ta14,"level1", 0.2,0.8);
SetNodeProbs(Ta14,"level2", 0.2,0.8);
SetNodeProbs(Ta14,"level3", 0.85,0.15);
SetNodeProbs(Ta14,"level4", 0.85,0.15);
```

```
SetNodeProbs(Ta15,"level1", 0.2,0.8);
SetNodeProbs(Ta15,"level2", 0.2,0.8);
SetNodeProbs(Ta15,"level3", 0.85,0.15);
SetNodeProbs(Ta15,"level4", 0.85,0.15);
SetNodeProbs(Ta16,"level1", 0.2,0.8);
SetNodeProbs(Ta16,"level2", 0.2,0.8);
SetNodeProbs(Ta16,"level3", 0.85,0.15);
SetNodeProbs(Ta16,"level4", 0.85,0.15);
SetNodeProbs(Ta17,"level1", 0.2,0.8);
SetNodeProbs(Ta17,"level2", 0.2,0.8);
SetNodeProbs(Ta17,"level3", 0.85,0.15);
SetNodeProbs(Ta17,"level4", 0.85,0.15);
SetNodeProbs(Ta18,"level1", 0.2,0.8);
SetNodeProbs(Ta18,"level2", 0.2,0.8);
SetNodeProbs(Ta18,"level3", 0.85,0.15);
SetNodeProbs(Ta18,"level4", 0.85,0.15);
SetNodeProbs(Ta19,"level1", 0.2,0.8);
SetNodeProbs(Ta19,"level2", 0.2,0.8);
SetNodeProbs(Ta19,"level3", 0.85,0.15);
SetNodeProbs(Ta19,"level4", 0.85,0.15);
SetNodeProbs(Ta20,"level1", 0.2,0.8);
SetNodeProbs(Ta20,"level2", 0.2,0.8);
SetNodeProbs(Ta20,"level3", 0.85,0.15);
SetNodeProbs(Ta20,"level4", 0.85,0.15);

SetNodeProbs(Ta21,"level1", 0.2,0.8);
SetNodeProbs(Ta21,"level2", 0.2,0.8);
SetNodeProbs(Ta21,"level3", 0.2,0.8);
SetNodeProbs(Ta21,"level4", 0.85,0.15);
SetNodeProbs(Ta22,"level1", 0.2,0.8);
SetNodeProbs(Ta22,"level2", 0.2,0.8);
SetNodeProbs(Ta22,"level3", 0.2,0.8);
SetNodeProbs(Ta22,"level4", 0.85,0.15);
SetNodeProbs(Ta23,"level1", 0.2,0.8);
SetNodeProbs(Ta23,"level2", 0.2,0.8);
SetNodeProbs(Ta23,"level3", 0.2,0.8);
SetNodeProbs(Ta23,"level4", 0.85,0.15);
SetNodeProbs(Ta24,"level1", 0.2,0.8);
SetNodeProbs(Ta24,"level2", 0.2,0.8);
SetNodeProbs(Ta24,"level3", 0.2,0.8);
SetNodeProbs(Ta24,"level4", 0.85,0.15);
SetNodeProbs(Ta25,"level1", 0.2,0.8);
SetNodeProbs(Ta25,"level2", 0.2,0.8);
SetNodeProbs(Ta25,"level3", 0.2,0.8);
```

```
        SetNodeProbs(Ta25,"level4", 0.85,0.15);
        SetNodeProbs(Ta26,"level1", 0.2,0.8);
        SetNodeProbs(Ta26,"level2", 0.2,0.8);
        SetNodeProbs(Ta26,"level3", 0.2,0.8);
        SetNodeProbs(Ta26,"level4", 0.85,0.15);
        SetNodeProbs(Ta27,"level1", 0.2,0.8);
        SetNodeProbs(Ta27,"level2", 0.2,0.8);
        SetNodeProbs(Ta27,"level3", 0.2,0.8);
        SetNodeProbs(Ta27,"level4", 0.85,0.15);
        SetNodeProbs(Ta28,"level1", 0.2,0.8);
        SetNodeProbs(Ta28,"level2", 0.2,0.8);
        SetNodeProbs(Ta28,"level3", 0.2,0.8);
        SetNodeProbs(Ta28,"level4", 0.85,0.15);
        SetNodeProbs(Ta29,"level1", 0.2,0.8);
        SetNodeProbs(Ta29,"level2", 0.2,0.8);
        SetNodeProbs(Ta29,"level3", 0.2,0.8);
        SetNodeProbs(Ta29,"level4", 0.85,0.15);
        SetNodeProbs(Ta30,"level1", 0.2,0.8);
        SetNodeProbs(Ta30,"level2", 0.2,0.8);
        SetNodeProbs(Ta30,"level3", 0.2,0.8);
        SetNodeProbs(Ta30,"level4", 0.85,0.15);

        SetNodeProbs(Mone, 0.25, 0.25,0.25,0.25);
        SetNodeProbs(Mtwo, "level1", 0.25, 0.25,0.25,0.25);
        SetNodeProbs(Mtwo, "level2", 0, 1/3 ,1/3 ,1/3);
        SetNodeProbs(Mtwo, "level3", 0,0,0.5,0.5);
        SetNodeProbs(Mtwo, "level4", 0,0,0,1);
        CHKERR

for (pn = 0;  pn < 2;  ++pn)  parent_states[pn] = EVERY_STATE;

SetNodeExperience_bn (Mone,  NULL,  1.0);
SetNodeExperience_bn (Mtwo,  parent_states,  1.0);
SetNodeExperience_bn (T1,  parent_states,  1.0);
SetNodeExperience_bn (T2,  parent_states,  1.0);
SetNodeExperience_bn (T3,  parent_states,  1.0);
SetNodeExperience_bn (T4,  parent_states,  1.0);
SetNodeExperience_bn (T5,  parent_states,  1.0);
SetNodeExperience_bn (T6,  parent_states,  1.0);
SetNodeExperience_bn (T7,  parent_states,  1.0);
SetNodeExperience_bn (T8,  parent_states,  1.0);
SetNodeExperience_bn (T9,  parent_states,  1.0);
SetNodeExperience_bn (T10,  parent_states,  1.0);
SetNodeExperience_bn (T11,  parent_states,  1.0);
```

```
SetNodeExperience_bn (T12, parent_states, 1.0);
SetNodeExperience_bn (T13, parent_states, 1.0);
SetNodeExperience_bn (T14, parent_states, 1.0);
SetNodeExperience_bn (T15, parent_states, 1.0);
SetNodeExperience_bn (T16, parent_states, 1.0);
SetNodeExperience_bn (T17, parent_states, 1.0);
SetNodeExperience_bn (T18, parent_states, 1.0);
SetNodeExperience_bn (T19, parent_states, 1.0);
SetNodeExperience_bn (T20, parent_states, 1.0);
SetNodeExperience_bn (T21, parent_states, 1.0);
SetNodeExperience_bn (T22, parent_states, 1.0);
SetNodeExperience_bn (T23, parent_states, 1.0);
SetNodeExperience_bn (T24, parent_states, 1.0);
SetNodeExperience_bn (T25, parent_states, 1.0);
SetNodeExperience_bn (T26, parent_states, 1.0);
SetNodeExperience_bn (T27, parent_states, 1.0);
SetNodeExperience_bn (T28, parent_states, 1.0);
SetNodeExperience_bn (T29, parent_states, 1.0);
SetNodeExperience_bn (T30, parent_states, 1.0);
SetNodeExperience_bn (Ta1, parent_states, 1.0);
SetNodeExperience_bn (Ta2, parent_states, 1.0);
SetNodeExperience_bn (Ta3, parent_states, 1.0);
SetNodeExperience_bn (Ta4, parent_states, 1.0);
SetNodeExperience_bn (Ta5, parent_states, 1.0);
SetNodeExperience_bn (Ta6, parent_states, 1.0);
SetNodeExperience_bn (Ta7, parent_states, 1.0);
SetNodeExperience_bn (Ta8, parent_states, 1.0);
SetNodeExperience_bn (Ta9, parent_states, 1.0);
SetNodeExperience_bn (Ta10, parent_states, 1.0);
SetNodeExperience_bn (Ta11, parent_states, 1.0);
SetNodeExperience_bn (Ta12, parent_states, 1.0);
SetNodeExperience_bn (Ta13, parent_states, 1.0);
SetNodeExperience_bn (Ta14, parent_states, 1.0);
SetNodeExperience_bn (Ta15, parent_states, 1.0);
SetNodeExperience_bn (Ta16, parent_states, 1.0);
SetNodeExperience_bn (Ta17, parent_states, 1.0);
SetNodeExperience_bn (Ta18, parent_states, 1.0);
SetNodeExperience_bn (Ta19, parent_states, 1.0);
SetNodeExperience_bn (Ta20, parent_states, 1.0);
SetNodeExperience_bn (Ta21, parent_states, 1.0);
SetNodeExperience_bn (Ta22, parent_states, 1.0);
SetNodeExperience_bn (Ta23, parent_states, 1.0);
SetNodeExperience_bn (Ta24, parent_states, 1.0);
SetNodeExperience_bn (Ta25, parent_states, 1.0);
```

```
SetNodeExperience_bn (Ta26,  parent_states,  1.0);
SetNodeExperience_bn (Ta27,  parent_states,  1.0);
SetNodeExperience_bn (Ta28,  parent_states,  1.0);
SetNodeExperience_bn (Ta29,  parent_states,  1.0);
SetNodeExperience_bn (Ta30,  parent_states,  1.0);

CHKERR

        nodes = GetNetNodes_bn (net);

        /* Read the case file into a caseset */

        cases = NewCaseset_cs ("cond1_1", env);
        casefile = NewFileStream_ns ("Data Files\\con1_1.cas", env, NULL);
        AddFileToCaseset_cs (cases, casefile, 1.0, NULL);

        /* Learning the case file into a caseset */
        learner = NewLearner_bn (EM_LEARNING, NULL, env);
        SetLearnerMaxIters_bn (learner, 200); /* terminate at 200 iterations */

        LearnCPTs_bn (learner, nodes, cases, 1.0);

        WriteNet_bn (net,  NewFileStream_ns ("Data Files\\S1_1.dne", env, NULL));
        CHKERR
        CompileNet_bn (net);

end:
        DeleteLearner_bn (learner);
        DeleteStream_ns (casefile);
        DeleteCaseset_cs (cases);
        DeleteNet_bn (net);
        res= CloseNetica_bn (env, mesg);
        printf ("%s\n", mesg);
        printf ("Press <enter> key to quit ", mesg);
        getchar();
        return (res < 0 ? -1 : 0);

error:
        err = GetError_ns (env, ERROR_ERR, NULL);
        fprintf (stderr, "LearnLatent: Error %d %s\n",
              ErrorNumber_ns (err), ErrorMessage_ns (err));
        goto end;
}
```

Appendix C: ANOVA with non-transformed data

Although ANOVA is robust to the violation of normality assumption, the F test is affected by serious light –tailedness, heavy tailedness, or skewed distribution. Based on literature review, the lack of normality assumption can be examined through computing skewness and kurtosis. There are certain rules of thumb in terms of the different sizes of sample. If the sample is small ($n < 100$), then calculate z-scores for skewness and kurtosis and reject as non-normal those variables with either z-score greater than an absolute value of 1.96. If the sample is of medium size ($100 < n < 300$), then calculate z-scores for skew and kurtosis and reject as non-normal those variables with either z-score greater than an absolute value of 3.29. With sample sizes greater than 300, absolute values above 2 are likely to indicate substantial non-normality (De Carlo, 1997; Minium, King, & Bear, 1993). Table C1 and C2 have the values of skewness and kurtosis of RMSDs used in two simulation studies.

Table C1.

The values of skewness and kurtosis in the simulation study 1

	Skewness		Kurtosis	
RMSD	Statistic	Std. Error	Statistic	Std. Error
IPT	2.226	.283	4.913	.559
TPT	1.301	.283	1.770	.559
CPT	1.021	.283	-.037	.559

Table C2.

The values of skewness and kurtosis in the simulation study 2

	Skewness		Kurtosis	
RMSD	Statistic	Std. Error	Statistic	Std. Error
IPT	.918	.299	-.556	.590
TPT	1.586	.299	2.021	.590
CPT	1.249	.299	.686	.590

The values of skewness and kurtosis indicate the violation of the normality assumption. Consequently, the results of ANOVA with the nontransformed RMSDs are different from the results with the transformed RMSDs. The conclusions of the significant variables are not the same between the transformed RMSDs and non transformed RMSDs. When the original data has extreme values of skewness, outliers influence the results. The influence was reduced by the transformation in this study. Furthermore, the original data sets of IPT and TPT were extreme values of kurtosis. Excessive kurtosis tends to effect procedures based on variance and covariances, which lead different results (DeCarlo, 1997). This influence was also reduced by the transformation.

Table C3, C4, and C5 are the ANOVA results using non-transformed variables in the first simulation study.

Table C3.

ANOVA of RMSD values of the parameters of the initial probability table

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.085[a]	16	.005	5.420	.000
Intercept	.067	1	.067	69.047	.000
SampleSize	.013	1	.013	13.648	.001
TaskSize	.024	1	.024	24.477	.000
CDT	.019	1	.019	19.170	.000
TPT	.002	2	.001	1.016	.369
IPT	.001	2	.000	.285	.753
SampleSize * TaskSize	.005	1	.005	4.892	.031
SampleSize * CDT	.014	1	.014	14.247	.000
SampleSize * TPT	.001	2	.001	.704	.499
TaskSize * CDT	.002	1	.002	1.808	.184
TaskSize * TPT	.002	2	.001	1.052	.356
CDT * TPT	.001	2	.001	.679	.511
Error	.054	55	.001		
Total	.205	72			
Corrected Total	.138	71			

a. R Squared = .612 (Adjusted R Squared = .499)

Table C4.

ANOVA of RMSD values of the parameters of the transition probability table

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.134[a]	16	.008	10.341	.000
Intercept	.213	1	.213	263.497	.000
SampleSize	.054	1	.054	66.507	.000
TaskSize	.028	1	.028	34.943	.000
CDT	.034	1	.034	41.853	.000
TPT	.003	2	.001	1.632	.205
IPT	.000	2	.000	.070	.933
SampleSize * TaskSize	.001	1	.001	1.639	.206
SampleSize * CDT	.001	1	.001	.854	.359
SampleSize * TPT	.000	2	.000	.262	.771
TaskSize * CDT	.010	1	.010	12.383	.001
TaskSize * TPT	.000	2	.000	.265	.768
CDT * TPT	.000	2	.000	.306	.737
Error	.044	55	.001		
Total	.386	72			
Corrected Total	.178	71			

a. R Squared = .751 (Adjusted R Squared = .678)

Table C5.

ANOVA of RMSD values of the parameters of the conditional probability table.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.094[a]	16	.006	27.572	.000
Intercept	.159	1	.159	746.451	.000
SampleSize	.025	1	.025	115.245	.000
TaskSize	.032	1	.032	150.758	.000
CDT	.019	1	.019	87.055	.000
TPT	.007	2	.004	17.098	.000
IPT	.000	2	.000	.163	.850
SampleSize * TaskSize	.003	1	.003	14.181	.000
SampleSize * CDT	.000	1	.000	.483	.490
SampleSize * TPT	.001	2	.001	2.864	.066
TaskSize * CDT	.004	1	.004	16.899	.000
TaskSize * TPT	.002	2	.001	5.109	.009
CDT * TPT	.000	2	.000	.042	.959
Error	.012	55	.000		
Total	.262	72			
Corrected Total	.106	71			

a. R Squared = .889 (Adjusted R Squared = .857)

Table C6, C7, and C8 are the ANOVA results using non-transformed variables in the second simulation study

Table C6.

ANOVA of RMSD values of the parameters of the initial probability table.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.274[a]	10	.027	10.125	.000
Intercept	.409	1	.409	151.132	.000
SampleSize	.071	1	.071	26.263	.000
TaskSize	.009	1	.009	3.415	.070
CDT	.163	1	.163	60.141	.000
TPT	.002	3	.001	.250	.861
covariate	.001	1	.001	.256	.615
SampleSize * TaskSize	.001	1	.001	.433	.513
SampleSize * CDT	.019	1	.019	7.146	.010
TaskSize * CDT	.007	1	.007	2.425	.125
Error	.143	53	.003		
Total	.827	64			
Corrected Total	.418	63			

a. R Squared = .656 (Adjusted R Squared = .592)

Table C7.

ANOVA of RMSD values of the parameters of the transition probability table.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.114[a]	19	.006	10.863	.000
Intercept	.145	1	.145	263.544	.000
SampleSize	.020	1	.020	35.690	.000
TaskSize	.017	1	.017	31.282	.000
CDT	.056	1	.056	102.121	.000
TPT	.002	3	.001	1.390	.258
covariate	.000	1	.000	.319	.575
SampleSize * TaskSize	.000	1	.000	.236	.629
SampleSize * CDT	.005	1	.005	8.796	.005
TaskSize * CDT	.009	1	.009	16.095	.000
SampleSize * TPT	.002	3	.001	1.224	.312
TaskSize * TPT	.001	3	.000	.711	.551
CDT * TPT	.001	3	.000	.463	.710
Error	.024	44	.001		
Total	.285	64			
Corrected Total	.138	63			

Table C7.

ANOVA of RMSD values of the parameters of the transition probability table.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.114[a]	19	.006	10.863	.000
Intercept	.145	1	.145	263.544	.000
SampleSize	.020	1	.020	35.690	.000
TaskSize	.017	1	.017	31.282	.000
CDT	.056	1	.056	102.121	.000
TPT	.002	3	.001	1.390	.258
covariate	.000	1	.000	.319	.575
SampleSize * TaskSize	.000	1	.000	.236	.629
SampleSize * CDT	.005	1	.005	8.796	.005
TaskSize * CDT	.009	1	.009	16.095	.000
SampleSize * TPT	.002	3	.001	1.224	.312
TaskSize * TPT	.001	3	.000	.711	.551
CDT * TPT	.001	3	.000	.463	.710
Error	.024	44	.001		
Total	.285	64			
Corrected Total	.138	63			

a. R Squared = .824 (Adjusted R Squared = .748)

Table C8.

ANOVA of RMSD values of the parameters of the conditional probability table.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.114[a]	19	.006	10.863	.000
Intercept	.145	1	.145	263.544	.000
SampleSize	.020	1	.020	35.690	.000
TaskSize	.017	1	.017	31.282	.000
CDT	.056	1	.056	102.121	.000
TPT	.002	3	.001	1.390	.258
covariate	.000	1	.000	.319	.575
SampleSize * TaskSize	.000	1	.000	.236	.629
SampleSize * CDT	.005	1	.005	8.796	.005
TaskSize * CDT	.009	1	.009	16.095	.000
SampleSize * TPT	.002	3	.001	1.224	.312
TaskSize * TPT	.001	3	.000	.711	.551
CDT * TPT	.001	3	.000	.463	.710
Error	.024	44	.001		
Total	.285	64			
Corrected Total	.138	63			

a. R Squared = .824 (Adjusted R Squared = .748)

Appendix D: ANOVA including all 3-way

Appendix D shows the ANOVA results including all 3 ways. The results do not give important insights.

Table D1.

All 3 way ANOVA of the initial probability table in simulation study 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	226.178[a]	46	4.917	9.809	.000
Intercept	1124.656	1	1124.656	2243.552	.000
SampleSize	8.819	1	8.819	17.592	.001
TaskSize	38.918	1	38.918	77.637	.000
CDT	141.625	1	141.625	282.525	.000
TPT	1.006	3	.335	.669	.583
covariate	.522	1	.522	1.042	.322
SampleSize * TaskSize	8.437	1	8.437	16.831	.001
SampleSize * CDT	.165	1	.165	.329	.574
SampleSize * TPT	3.360	3	1.120	2.234	.121
SampleSize * covariate	.014	1	.014	.028	.869
TaskSize * CDT	11.155	1	11.155	22.254	.000
TaskSize * TPT	2.128	3	.709	1.415	.273
TaskSize * covariate	.006	1	.006	.012	.914
CDT * TPT	.851	3	.284	.566	.645
CDT * covariate	.160	1	.160	.319	.579
TPT * covariate	.414	3	.138	.275	.842
SampleSize * TaskSize * CDT	1.616	1	1.616	3.223	.090
SampleSize * CDT * TPT	.223	3	.074	.148	.929
SampleSize * TPT * covariate	.049	3	.016	.033	.992

TaskSize * CDT * TPT	1.214	3	.405	.807	.507
TaskSize * TPT * covariate	.302	3	.101	.201	.894
CDT * TPT * covariate	1.665	3	.555	1.107	.374
SampleSize * TaskSize * TPT	1.923	3	.641	1.279	.313
SampleSize * TaskSize * covariate	.103	1	.103	.205	.656
TaskSize * CDT * covariate	.088	1	.088	.175	.681
Error	8.522	17	.501		
Total	1380.273	64			
Corrected Total	234.700	63			

a. R Squared = .964 (Adjusted R Squared = .865)

Table D2.

All 3 way ANOVA of the transition probability table in simulation study 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	85.290[a]	46	1.854	6.747	.000
Intercept	587.498	1	587.498	2137.905	.000
SampleSize	21.575	1	21.575	78.511	.000
TaskSize	3.371	1	3.371	12.268	.003
CDT	39.244	1	39.244	142.808	.000
TPT	1.232	3	.411	1.494	.252
covariate	.110	1	.110	.402	.535
SampleSize * TaskSize	2.970	1	2.970	10.806	.004
SampleSize * CDT	.976	1	.976	3.552	.077
SampleSize * TPT	1.333	3	.444	1.617	.223
SampleSize * covariate	.095	1	.095	.346	.564
TaskSize * CDT	.157	1	.157	.573	.460
TaskSize * TPT	.766	3	.255	.930	.448
TaskSize * covariate	.628	1	.628	2.285	.149
CDT * TPT	.941	3	.314	1.142	.361
CDT * covariate	.039	1	.039	.143	.710
TPT * covariate	.946	3	.315	1.148	.358
SampleSize * TaskSize * CDT	.000	1	.000	.001	.975
SampleSize * CDT * TPT	1.959	3	.653	2.376	.106
SampleSize * TPT * covariate	.296	3	.099	.359	.783
TaskSize * CDT * TPT	1.116	3	.372	1.354	.290
TaskSize * TPT * covariate	2.037	3	.679	2.471	.097
CDT * TPT * covariate	1.347	3	.449	1.634	.219
SampleSize * TaskSize * TPT	.467	3	.156	.567	.644

SampleSize * TaskSize * covariate	.178	1	.178	.647	.432
TaskSize * CDT * covariate	.156	1	.156	.568	.461
Error	4.672	17	.275		
Total	700.945	64			
Corrected Total	89.961	63			

a. R Squared = .948 (Adjusted R Squared = .808)

Table D3.

All 3 way ANOVA of the conditional probability table in simulation study 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	66.293[a]	46	1.441	8.686	.000
Intercept	776.724	1	776.724	4681.422	.000
SampleSize	15.212	1	15.212	91.688	.000
TaskSize	8.294	1	8.294	49.989	.000
CDT	30.104	1	30.104	181.439	.000
TPT	1.294	3	.431	2.600	.086
covariate	.168	1	.168	1.015	.328
SampleSize * TaskSize	.641	1	.641	3.861	.066
SampleSize * CDT	.030	1	.030	.182	.675
SampleSize * TPT	1.171	3	.390	2.352	.108
SampleSize * covariate	.021	1	.021	.125	.728
TaskSize * CDT	1.591	1	1.591	9.589	.007
TaskSize * TPT	.412	3	.137	.828	.497
TaskSize * covariate	.142	1	.142	.859	.367
CDT * TPT	.406	3	.135	.815	.503
CDT * covariate	.028	1	.028	.167	.688
TPT * covariate	.189	3	.063	.379	.769
SampleSize * TaskSize * CDT	.888	1	.888	4.353	.073
SampleSize * CDT * TPT	.836	3	.279	1.680	.209
SampleSize * TPT * covariate	.111	3	.037	.223	.879
TaskSize * CDT * TPT	.731	3	.244	1.469	.258
TaskSize * TPT * covariate	.447	3	.149	.899	.462
CDT * TPT * covariate	.324	3	.108	.651	.593
SampleSize * TaskSize * TPT	.073	3	.024	.147	.930
SampleSize * TaskSize * covariate	.008	1	.008	.051	.824

TaskSize * CDT * covariate	.277	1	.277	1.668	.214
Error	2.821	17	.166		
Total	869.219	64			
Corrected Total	69.114	63			

a. R Squared = .959 (Adjusted R Squared = .849)

Table D4.

All 3 way ANOVA of the initial probability table in simulation study 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	205.533[a]	51	4.030	7.340	.000
Intercept	1483.462	1	1483.462	2701.809	.000
SampleSize	30.343	1	30.343	55.264	.000
TaskSize	62.902	1	62.902	114.563	.000
CDT	76.850	1	76.850	139.966	.000
TPT	1.717	2	.859	1.564	.234
IPT	.288	2	.144	.262	.772
SampleSize * TaskSize	.058	1	.058	.106	.749
SampleSize * CDT	.361	1	.361	.658	.427
SampleSize * TPT	.481	2	.241	.438	.651
SampleSize * IPT	.212	2	.106	.193	.826
TaskSize * CDT	9.875	1	9.875	17.985	.000
TaskSize * TPT	.603	2	.301	.549	.586
TaskSize * IPT	.352	2	.176	.321	.729
CDT * TPT	.306	2	.153	.278	.760
CDT * IPT	2.429	2	1.215	2.212	.136
TPT * IPT	2.540	4	.635	1.156	.359
SampleSize * TaskSize * CDT	.331	1	.331	.602	.447
SampleSize * CDT * TPT	.073	2	.037	.067	.936
SampleSize * TPT * IPT	2.728	4	.682	1.242	.325
TaskSize * CDT * TPT	1.027	2	.514	.935	.409
TaskSize * TPT * IPT	2.293	4	.573	1.044	.409
CDT * TPT * IPT	.320	4	.080	.146	.963
SampleSize * CDT * IPT	3.655	2	1.828	3.329	.057
SampleSize * TaskSize * IPT	.332	2	.166	.303	.742
TaskSize * CDT * IPT	1.002	2	.501	.913	.417

SampleSize * TaskSize * TPT	.384	2	.192	.350	.709
Error	10.981	20	.549		
Total	1758.632	72			
Corrected Total	216.514	71			

a. R Squared = .949 (Adjusted R Squared = .820)

Table D5.

All 3 way ANOVA of the transition probability table in simulation study 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	129.840[a]	51	2.546	7.976	.000
Intercept	876.493	1	876.493	2745.972	.000
SampleSize	54.049	1	54.049	169.330	.000
TaskSize	21.683	1	21.683	67.930	.000
CDT	26.619	1	26.619	83.395	.000
TPT	2.626	2	1.313	4.113	.032
IPT	.121	2	.061	.190	.829
SampleSize * TaskSize	5.159	1	5.159	16.163	.001
SampleSize * CDT	6.947	1	6.947	21.763	.000
SampleSize * TPT	.333	2	.167	.522	.601
SampleSize * IPT	.232	2	.116	.363	.700
TaskSize * CDT	.000	1	.000	.001	.975
TaskSize * TPT	.776	2	.388	1.215	.318
TaskSize * IPT	.467	2	.233	.731	.494
CDT * TPT	.226	2	.113	.354	.706
CDT * IPT	.161	2	.080	.251	.780
TPT * IPT	2.059	4	.515	1.612	.210
SampleSize * TaskSize * CDT	.526	1	.526	1.647	.214
SampleSize * CDT * TPT	.193	2	.096	.302	.742
SampleSize * TPT * IPT	1.331	4	.333	1.043	.410
TaskSize * CDT * TPT	.247	2	.123	.386	.684
TaskSize * TPT * IPT	1.129	4	.282	.884	.491
CDT * TPT * IPT	.450	4	.113	.352	.839
SampleSize * CDT * IPT	1.263	2	.632	1.978	.164
SampleSize * TaskSize * IPT	.089	2	.044	.139	.871
TaskSize * CDT * IPT	.076	2	.038	.120	.888

SampleSize * TaskSize * TPT	.350	2	.175	.548	.587
Error	6.384	20	.319		
Total	1048.634	72			
Corrected Total	136.223	71			

a. R Squared = .953 (Adjusted R Squared = .834)

Table D6.

All 3 way ANOVA of the conditional probability table in simulation study 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	59.225[a]	51	1.161	7.588	.000
Intercept	823.401	1	823.401	5380.199	.000
SampleSize	19.023	1	19.023	124.299	.000
TaskSize	16.518	1	16.518	107.930	.000
CDT	12.636	1	12.636	82.567	.000
TPT	3.885	2	1.943	12.693	.000
IPT	.201	2	.100	.656	.530
SampleSize * TaskSize	.117	1	.117	.762	.393
SampleSize * CDT	.884	1	.884	5.778	.026
SampleSize * TPT	.009	2	.004	.028	.973
SampleSize * IPT	.035	2	.017	.114	.893
TaskSize * CDT	.487	1	.487	3.183	.090
TaskSize * TPT	.050	2	.025	.165	.849
TaskSize * IPT	.396	2	.198	1.293	.297
CDT * TPT	.205	2	.102	.669	.523
CDT * IPT	.208	2	.104	.680	.518
TPT * IPT	.392	4	.098	.640	.640
SampleSize * TaskSize * CDT	.806	1	.806	4.265	.073
SampleSize * CDT * TPT	.286	2	.143	.933	.410
SampleSize * TPT * IPT	.551	4	.138	.899	.483
TaskSize * CDT * TPT	.159	2	.079	.518	.603
TaskSize * TPT * IPT	.216	4	.054	.352	.839
CDT * TPT * IPT	.409	4	.102	.668	.622
SampleSize * CDT * IPT	.175	2	.087	.572	.574
SampleSize * TaskSize * IPT	.054	2	.027	.176	.840
TaskSize * CDT * IPT	.070	2	.035	.228	.798

SampleSize * TaskSize * TPT	.118	2	.059	.386	.685
Error	3.061	20	.153		
Total	919.545	72			
Corrected Total	62.286	71			

a. R Squared = .951 (Adjusted R Squared = .826)

Appendix E. ANOVA with SDs

Before running ANOVA, I also computed the skewness and kurtosis. It was observed that the distributions of the dependent variables depart from the normal distribution. Therefore, the transformed data were used in ANOVA. Table E1and E2 shows the values of Skewness and Kurtosis.

Table E1.

The values of skewness and kurtosis in the simulation study 1

	Skewness		Kurtosis	
RMSD	Statistic	Std. Error	Statistic	Std. Error
IPT	1.022	.283	.018	.559
TPT	.893	.283	-.458	.559
CPT	.297	.283	-1.236	.559

Table E2.

The values of skewness and kurtosis in the simulation study 2

	Skewness		Kurtosis	
RMSD	Statistic	Std. Error	Statistic	Std. Error
IPT	1.141	.299	1.207	.590
TPT	1.992	.299	5.842	.590
CPT	.371	.299	-1.343	.590

The following tables are the ANOVA results with transformed SDs. The ANOVA results do not provide important additional information other than the results using RMSD.

Table E3.

ANOVA of SDs of the initial probability table in simulation study 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	60.109[a]	16	3.757	33.536	.000
Intercept	2830.246	1	2830.246	25264.422	.000
SampleSize	28.516	1	28.516	254.555	.000
TaskSize	9.634	1	9.634	85.995	.000
CDT	10.724	1	10.724	95.729	.000
TPT	.484	2	.242	2.162	.125
IPT	.331	2	.165	1.476	.237
SampleSize * TaskSize	.582	1	.582	5.194	.027
SampleSize * CDT	.008	1	.008	.075	.786
SampleSize * TPT	.386	2	.193	1.722	.188
TaskSize * CDT	8.205	1	8.205	73.240	.000
TaskSize * TPT	.427	2	.213	1.905	.159
CDT * TPT	.301	2	.151	1.344	.269
Error	6.161	55	.112		
Total	2910.364	72			
Corrected Total	66.270	71			

a. R Squared = .907 (Adjusted R Squared = .880)

Sample size, task size, and CPT were statistically significant. Also, two interaction effects (sample size X task size, Task size X CDP) were statistically significant.

Table E4.

ANOVA of SDs of the transition probability table in simulation study 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	50.532[a]	16	3.158	24.724	.000
Intercept	2440.922	1	2440.922	19108.245	.000
SampleSize	27.023	1	27.023	211.544	.000
TaskSize	4.165	1	4.165	32.602	.000
CDT	4.203	1	4.203	32.902	.000
TPT	2.952	2	1.476	11.556	.000
IPT	.770	2	.385	3.016	.057
SampleSize * TaskSize	.066	1	.066	.518	.475
SampleSize * CDT	.018	1	.018	.140	.709
SampleSize * TPT	.011	2	.006	.044	.957
TaskSize * CDT	11.040	1	11.040	86.426	.000
TaskSize * TPT	.307	2	.154	1.202	.308
CDT * TPT	.322	2	.161	1.261	.291
Error	7.026	55	.128		
Total	2503.478	72			
Corrected Total	57.558	71			

a. R Squared = .878 (Adjusted R Squared = .842)

Sample size, task size, CPT, and TPT were statistically significant. Also, an interaction effects (Task size X CDP) were statistically significant.

Table E5.

ANOVA of SDs of the conditional probability table in simulation study 1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	22.059[a]	16	1.379	51.193	.000
Intercept	1498.999	1	1498.999	55661.585	.000
SampleSize	19.688	1	19.688	731.049	.000
TaskSize	.002	1	.002	.092	.762
CDT	.795	1	.795	29.531	.000
TPT	.727	2	.364	13.506	.000
IPT	.274	2	.137	5.091	.009
SampleSize * TaskSize	.091	1	.091	3.383	.071
SampleSize * CDT	.005	1	.005	.189	.665
SampleSize * TPT	.036	2	.018	.677	.512
TaskSize * CDT	.137	1	.137	5.085	.028
TaskSize * TPT	.107	2	.054	1.989	.147
CDT * TPT	.096	2	.048	1.790	.177
Error	1.481	55	.027		
Total	1528.667	72			
Corrected Total	23.540	71			

a. R Squared = .937 (Adjusted R Squared = .919)

Sample size, CPT, TPT, and IPT were statistically significant. Also, an interaction effects (Task size X CDP) were statistically significant

Table E6.

ANOVA of SDs of the initial probability table in simulation study 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	46.828[a]	10	4.683	20.291	.000
Intercept	2450.733	1	2450.733	10619.426	.000
SampleSize	18.265	1	18.265	79.146	.000
TaskSize	2.276	1	2.276	9.863	.003
CDT	14.575	1	14.575	63.155	.000
TPT	.750	3	.250	1.083	.364
covariate	.132	1	.132	.570	.454
SampleSize * TaskSize	.345	1	.345	1.496	.227
SampleSize * CDT	.261	1	.261	1.130	.293
TaskSize * CDT	9.598	1	9.598	41.589	.000
Error	12.231	53	.231		
Total	2504.656	64			
Corrected Total	59.059	63			

a. R Squared = .793 (Adjusted R **Squared = .754)**

Sample size, task size, CPT, and TPT were statistically significant. Also, an interaction effects (Task size X CDP) were statistically significant

Table E7.

ANOVA of SDs of the transition probability table in simulation study 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	38.733[a]	10	3.873	31.010	.000
Intercept	1877.806	1	1877.806	15033.563	.000
SampleSize	16.044	1	16.044	128.450	.000
TaskSize	1.956	1	1.956	15.657	.000
CDT	12.078	1	12.078	96.696	.000
TPT	.675	3	.225	1.801	.158
covariate	.145	1	.145	1.159	.287
SampleSize * TaskSize	.690	1	.690	5.526	.022
SampleSize * CDT	.001	1	.001	.008	.928
TaskSize * CDT	6.557	1	6.557	52.497	.000
Error	6.620	53	.125		
Total	1920.284	64			
Corrected Total	45.354	63			

a. R Squared = .854 (Adjusted R Squared = .826)

Sample size, task size, and CPT were statistically significant. Also, an interaction effects (sample size X task size, Task size X CDP) were statistically significant.

Table E7.

ANOVA of SDs of the conditional probability table in simulation study 2

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	20.203[a]	19	1.063	16.074	.000
Intercept	1436.746	1	1436.746	21719.403	.000
SampleSize	17.026	1	17.026	257.381	.000
TaskSize	.369	1	.369	5.574	.023
CDT	1.997	1	1.997	30.183	.000
TPT	.257	3	.086	1.293	.289
covariate	.020	1	.020	.310	.581
SampleSize * TaskSize	.085	1	.085	1.291	.262
SampleSize * CDT	.061	1	.061	.915	.344
TaskSize * CDT	.059	1	.059	.888	.351
SampleSize * TPT	.094	3	.031	.474	.702
TaskSize * TPT	.224	3	.075	1.128	.348
CDT * TPT	.099	3	.033	.500	.684
Error	2.911	44	.066		
Total	1471.277	64			
Corrected Total	23.114	63			

a. R Squared = .874 (Adjusted R Squared = .820)

Sample size, CPT, TPT, and IPT were statistically significant. Also, an interaction effects (Task size X CDP) were statistically significant.

References

Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J. D. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement, 44,* 341-359.

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

Almond, R. G., Mulder, J., Hemat, L.A., & Yan, D. (2009). Bayesian network models for local dependence among observable outcome variables. *Journal of Educational and Behavioral Statistics, 34*(4), 491-521.

Almond, R. G., Yan, D., & Hemat, L. (2008). Parameter recovery studies with a diagnostic Bayesian network model. *Behaviormetrika, 35*(2), 159-185.

Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education, 93*, 389-421.

Anderson, E. B. (1985). Longitudinal studies for discrete data based on latent structure models. In D. Magnusson, L. R. Bergman, G. Rudinger, & B. Torestad (Eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 308-322). Cambridge, England: Cambridge University Press.

Birenbaum, M., Kelly, A. E., & Tatsuoka, K. (1993). Diagnosing knowledge states in algebra using the rule-space model. *Journal for Research in Mathematics Education*, *24,* 442-459.

Black, Paul J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*: *Principles*, *Policy*, *and Practice 5 (1)*, 17–73.

Bloom, B. S., Hastings, J. T. & Madaus, G. F. (1971). *Handbook of formative and summative evaluation of student learning.* New York, NY: McGraw-Hill.

Bond, T. G. (1995). Piaget and measurement: Empirical validation of the Piagetian model. *Archives de Psychologie, 63,* 155-185.

Buck, G., & Tatsuoka, K. K. (1998). Application of the Rule Space procedure to language testing: examining attributes of a free response listening test. *Language Testing*, *15*(2), 119-157.

Briggs, D., & Alonzo, A. (2009). *The psychometric modeling of ordered multiple choice item responses for diagnostic assessment with a learning progression.* Paper presented at the Learning Progressions in Science (LeaPS) Conference. Iowa City, IA.

Cappé, O., Moulines, E., & Rydén, T. (2005). *Inference in hidden Markov models*. New York: Springer.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97* 404-431.

Celeux, G., Hurn, M., & Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of American Statistical Association*, *95*, 957-970.

Chung, H., Loken, E., & Schafer, J. L. (2004) Difficulties in drawing inferences with finite-mixture models: A simple example with a simple solution. *The American Statistician, 58*, 152–158.

Chung, H., Walls, T.A., & Park, Y. (2007). A latent transition model with logistic regression. *Psychometrika*, *72(3)*, 413-435.

Clasesgens, J., Scalise, K., Wilson, M., & Stacy, A. (2009). Mapping student understanding in chemistry: The Perspectives of Chemists. *Science Education, 93(1)*, 56-85.

Clogg, C. C., & Goodman, L. A. (1984). Latent structure analysis of a set of multidimensional tables. *Journal of the American Statistical Association, 79*, 762-771.

Collins, L. M. & Wugalter, S. E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research 27*, 131-157.

Collins, L. M., Fidler, P. L., & Wugalter, S. E. (1996). Some practical issues related to estimation of latent class and latent transition parameters. In A. von Eye & C.C. Clogg (Eds.), *Analysis of categorical variables in developmental research*. San Diego, CA: Academic Press, Inc.

Collins, L. M., & Flaherty, B. P. (2002). Latent class models for longitudinal data. In A.

L. McCutcheon & J. A. Hagenaars (Eds.) *Applied latent class analysis*, Cambridge: Cambridge University.

Corcoran, T., Mosher, F., & Rogat, A. (2009). *Learning Progressions in science: An Evidence-based approach to reform.* Philadelphia, Pa: Consortium for Policy Research in Education.

Corrigan, S., Loper, S., Barber, J., Brown, N., & Kulikowich, J. (2009). *The juncture of supply and demand for information: how and when can learning progressions meet the information demands of curriculum developers?* Paper presented at the Learning Progressions in Science (LeaPS) Conference. Iowa City, IA.

Cowell, R. & Dawid A. P. (1992). Fast retraction of evidence in a probabilistic expert system, *Statistics and Computing, 2*, 37–40.

Dai, Y. (2009). *A mixture Rasch model with a covariate: a simulation study via Bayesian Markov Chain Monte Carlo Estimation*. Unpublished doctoral dissertation, University of Maryland, College Park.

De Carlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*, 292-307.

De la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement, 33*.

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

Dempster, A. P., Laird N. M., & D. B. Rubin D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, *Series B*, *39*, 1-38.

DiBello, L. V., Roussos, L.A., & Stout, W. (2007). Review of cognitively diagnostic assessment and a summary of psychometric models. In C.R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (pp. 979-1030). Vo. 26, Psychometrics, Elsevier Science B.V.: The Netherlands.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.). *Cognitively diagnostic assessment*. Hillsdale, NJ: Erlbaum.

Diebolt, J. & Robert, C. P. (1994). Estimation of finite mixture distributions through Bayesian sampling. *Journal of Statistical Society, 56,* 363-375.

Draney, K. (1996). *The polytomous saltus model*: A mixture model approach to the diagnosis of developmental differences. Unpublished doctoral dissertation, University of California, Berkeley.

Draney, K. (2009). *Designing learning progressions with the bear assessment system*. Paper presented at the Learning Progressions in Science (LeaPS) Conference. Iowa City, IA.

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: application to abstract reasoning. *American Psychological Association*, *3(3)*, 380-396.

Embretson, S. E. (2000). Multidimensional measurement from dynamic tests: Abstract reasoning under stress. *Multivariate Behavioral Research*, *35*, 505-542.

Fisher, G. H. (1995b). Linear logistic models for change. In G. H. Fischer 7 I. W. Molenaar (Eds.). *Rasch models*: *Foundations, recent developments and applications* (pp. 15-38). New York: Springer Verlag.

Gruijter, P. N. M., & Van Der Kamp, L. J. T. (Eds.). *Advances in psychological and educational measurement*. New York: Wiley.

Gifford, J. A., & Swaminathan, H. (1990). Bias and the effect of priors in Bayesian estimation of parameters of item response models. *Applied Psychological Measurement*, *14*, 33-43.

Gotwals, A. W., Songer, N. B., & Bullard, L. (2009). *Assessing student's progressing abilities to construct scientific explanations.* Paper presented at the Learning Progressions in Science (LeaPS) Conference. Iowa City, IA.

Harwell, M., Stone, C. A., Hsu, T, & K. L., (1996). Monte Carlo studies in Item Response Theory. *Applied Psychological Measurement*, 20 (2), 101-125.

Huff, K, & Goodman, D. P (2007). The demand for cognitive diagnostic assessment. In J. Leighton & M. Girl (Eds.), *Cognitive diagnostic assessment for education: theory and applications*. New York, NY: Cambridge University Press.

Ingham, A. M. & Gilbert, J. K. (1991). The use of analogue models by students of chemistry at higher education level. *International Journal of Science Education, 13*, 193-202.

Jensen, F. V., Olesen, K. G., & Andersen, S. K. (1990). An algebra of Bayesian belief universes for knowledge based systems. *Networks, 20(5)*, 637–659.

Jensen, F. V., Lauritzen, S. L. & Olesen, K. G. (1990). Bayesian updating in causal probabilistic networks by local computation. *Computational Statistics Quarterly, 4*, 269–282.

Kjaerulff, U. B., & Madsen, A. L. (2007). *Probabilistic networks for practitioners*: *A guide to construction and analysis of Bayesian networks and influence diagrams.* New York, NY: Springer.

Langeheine, R., & Van de Pol, F. (2002). Latent Markov chains. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 304-341). Cambridge, United Kingdom: Cambridge University Press.

Lauritzen, S. L., & Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems (with discussion). *Journal of Royal Statistical Society, Series B, 50(2),* 157–224.

Lauritzen, S. L., & Jensen, F. V. (1996). Local computation with valuations from a commutative semigroup. Technical Report No. R-96-2028, Institute for Electronic Systems, Department of Mathematics and Computer Science, Aalborg University, Aalborg, Denmark.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Leighton, J. & Gierl, M. (Eds). (2007). *Cognitive diagnostic assessment for education: theory and applications*. New York, NY: Cambridge University Press.

Levy, R., & Mislevy, R.J. (2004). Specifying and refining a measurement model for a simulation-based assessment. *CSE Technical Report 619.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Linn, R. L. (1986). Testing and assessment in education: Policy issues. *American Psychologist*, *41*, 1153-1160.

McLachlan, G., & Peel, D. (2000). *Finite mixture models*. New York: John Wiley.

Minium, E.W., King, B.M. & Bear, G. (1993) *Statistical reasoning in psychology and education*, New York: John Wiley and Sons

Mislevy, R. J. (1987). Exploiting auxiliary information about examinees in the estimation of item parameters. *Applied Psychological Measurement*, *11,* 81-91.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychmetrika*, *59*(4), 439-483.

Mislevy, R. J. (1995). Probability-based inference in cognitive diagnosis. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 43-71). Hillsdale, NJ: Erlbaum.

Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. In R.L. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications.* Charlotte, NC: Information Age Publishing

Mislevy, R.J., Almond, R.G., Yan, D., & Steinberg, L.S. (1999). Bayes nets in educational assessment: Where do the numbers come from? In K.B. Laskey & H.Prade (Eds.), *Proceedings of the fifteenth conference on uncertainty in artificial intelligence* (437-446). San Francisco: Morgan Kaufmann.

Mislevy, R. J., Almond, R., Dibello, L., Jenkins, F. Steinberg, L., & Yan, D. (2002). Modeling Conditional Probabilities in Complex Educational Assessments. *CSE Technical Report 580.* Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Mislevy, R. J., Almond, R. G., & Lukas, J. (2003). A brief introduction to "evidence centered assessment design" (ECD). Retrieved June, 2009 from http://education.umd.edu/EDMS/mislevy/papers/index.html.

Mislevy, R.J., & Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User-Modeling and User-Adapted Interaction, 5,* 253-282. Online version available as http://www.cse.ucla.edu/products/Reports/TECH313.pdf.

Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61-90). Mahwah, NJ: Lawrence Erlbaum.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3–62.

Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (2002). Making sense of data from complex assessments. *Applied Measurement in Education*, *15 (4)*, 363-389.

Miselvy, R. J., Yamamoto, K. Y., & Anacker, S. (1992). Toward a test theory for assessing student understanding. In R. A. Lesh & S. Lamon (Eds.), Assessments of authentic performance in school mathematics (pp. 293-318). Washington, DC: American Association for the Advancement of Science.

Mislevy. R. J., & Verhelst, N. D. (1990). Modeling item responses when different subjects employ different solution strategies. *Psychometrika*, *55(2)*, 195-215.

Mislevy. R. J., Wingersky, M. S., Irvine, S. H., & Dann, P. L. (1991). Resolving mixtures of strategies in spatial visualization tasks. *British Journal of Mathematical and Statistical Psychology*, *44,* 265-288. Mislevy, R.J., Wilson , M.R., Ercikan, K., & Chudowsky, N. (2003). Psychometric principles in student assessment. In T. Kellaghan & D. Stufflebeam (Eds.), *International Handbook of Educational Evaluation* (pp. 389-531). Dordrecht, the Netherlands: Kluwer Academic Press. Online version available as http://www.cse.ucla.edu/products/Reports/TR583.pdf.

Muthén, L. K., & Muthén, B. O. (1998-2006). *Mplus User's Guide*. Fourth Editions. Los Angeles, CA: Muthén & Muthén.

Mohan, L., Chen, J., & Anderson, C. W. (in press). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching.*

Murphy, K. R. (2002). *Dynamic Bayesian networks: Representation, inference and learning.* A dissertation of doctor of philosophy.

National Research Council (2001). *Knowing what students know: The science and design of educational assessment.* Committee on the Foundations of Assessment, J. Pellegrino, R. Glaser, & N. Chudowsky (Eds.). Washington DC: National Academy Press.

Nichols, P. D, Chipman, S. F., & Brennan, R. L. (Eds.). (1995). *Cognitively diagnostic asssessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*: *A Multidisciplinary Journal*, *14*, 535-569.

Norsys Software Corp (2007). *Netica manual*. http://www.norsys.com

Pellegrino, J., Chudowsky, N., & Glaser, R., (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* National Research Council's Committee on the Foundations of Assessment. Washington, DC: National Academy Press.

Piaget, J. (1950). *They psychology of intelligence*. London: Lowe & Brydone.

R Development Core Team (2008*). R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. http//www.R-project.org.

Rost, J. (1988). Test theory with qualitative and quantitative latent variables. In R. Langeheine, & Rost, J. (Eds.). *Latent trait and latent class models*. New York: Plenum Press.

Rost. J. (1989). Rasch models and latent class models for measuring change with ordinal variables. In R. Coppi & S. Bolasco (Ed.), *Multivariate data analysis* (pp. 473-483). Amsterdam: Elsevier.

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*, 219-262.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement theory, methods and applications.* New York: The Guilford Press.

Samejima, F. (1983). A latent trait model for differential strategies in cognitive processes (Tch. Rep. No. ONR/RR-83-1. Knoxville, TN: University of Tennessee.

Schwarz, C., Reiser, B., Fortus, D., Shwartz, Y., Acher, A., Davis, B., Kenyon, L., & Hug, B. (2009). *Models: Defining a learning progression for scientific modeling*. Paper presented at the Learning Progressions in Science (LeaPS) Conference. Iowa City, IA.

Shin, N., Stevens, S., Short, H., & Krajcik, J. (2009) *Learning progressions to support coherence curricula in instructional material, instruction, and assessment design*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, June 2009, Iowa City, IA.

Shute, V.J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295-321). Mahwah, NJ: Routledge, Taylor and Francis.

Siegler, R. S., & Campbell, J. (1989). Diagnosing individual differences in strategy choice procedures. In N. Frederiksen, R. Glaser, A. Lesgold, & M. G. Shafto

(Eds.), *Diagnostic monitoring of skill and knowledge acquisition* (pp. 113-120). Hillsdale, NJ: Erlbaum.

Siegler, R. S., & Mcgilly, K. (1989). Strategy choices in children's time-telling. In I. Levin & D. Zakay (Eds.), *Psychological time: A life span perspective*. Elsevier: The Netherlands.

Spiel, C. (1994). Latent trait models for measuring change. In A van Eye & C. C. Clog (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 274-293). Thousand Oaks, CA: Sage Publications.

Spiel, C., Gittler, G., Sirsch, U., & Gluck, J. (1997). Applications of the Rasch model for testing Piaget's theory of cognitive development. In J. Rost & R. Langeheine (Eds.), *Applications of latent trait and latent class models in the social sciences* (pp. 111-117). Munster, Germany: Waxmann.

Stevens, S. Y., Shin, N., & Krajcik, J. (2009). *Towards a model for the development of an empirically tested Learning Progression*. Paper presented at the Learning Progressions in Science (LeaPS) Conference. Iowa City, IA.

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Applied Statistics*, *51*, 337-350.

Tatsuoka, K. K. (1990). *Toward an integration of item-response theory and cognitive error diagnoses*. In Fredrickson, Glaser, R. L., Lesgold, A. M. & Shafto, M. G. (Eds.), *Diagnostic monitoring of skills and knowledge acquisition* (pp. 453-488). Hillsdale, NJ: Erlbaum.

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

West. P., Rutstein, D. W., Mislevy, R. J., Liu, J., Levy, R., DiCerbo, K. E., Crawford, A., Choi, Y., & Behrens, J. T. (2009). *A Bayes net approach to modeling learning progressions and task performances.* Paper presented at the Learning Progressions in Science (LeaPS) Conference. Iowa City, IA.

West, P., Rutstein, D.W., Mislevy, R. J., Liu, J., Levy, R., DiCerbo, K. E., Crawford, A., Choi, Y., Chappel, K., & Behrens, J. T. (2010). *A Bayesian network approach to modeling learning progressions.* CRESST Research Report. Los Angeles: The National Center for Research on Evaluation, Standards, Student Testing (CRESST), Center for Studies in Education, UCLA.

Wiggins L. M. (1955). *Mathematical model for the analysis of multi-wave panels.* Ann Arbor: University Microfilms.

Wilson, M. (2009, June*). The structured constructs, model (SCM): A family of statistical models related to learning progressions*. Paper presented at the Learning Progressions in Science (LeaPS) Conference, Iowa City, IA.

Wilson, M. & Scalise, K. (2006). Assessment to improve learning in higher education: The BEAR Assessment System. *Higher Education*, 52, 635-663.

Wiliam, D. (2007). Keeping learning on track: Formative assessment and the regulation of learning. In Lester, F. K., (Eds.), *The second handbook of mathematics teaching and learning* (pp. 1053–98). Greenwich, Conn.: Information Age Publishing.

Williamson, D. M., Bauer, M., Steinberg, L.S., Mislevy, R.J., Behrens, J.T., & DeMark,

S. (2004). Design rationale for a complex performance assessment. *International Journal of Measurement*, *4*, 303-332.

Williamson, D. M., Mislevy, R. J., & Almond, R. A. (April, 2001). *Model criticism of Bayesian networks with latent variables*. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.

Wilson, M. R. (1989). Saltus: a psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, *105,* 276-289.

Willett, J. B., & Sayer, A. G. (1994). Using covariate structure analysis to detect correlates and predictors of individual change over time. *Psychological Bulletin*, *116(2)*, 363-381.

Willett, J. B., & Sayer, A. G. (1996). Cross-domain analyses of change over time: Combining growth modeling and covariance structure analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 125-157). Mahwah, NJ: Erlbaum.