Analyzing Hierarchical Data with the DINA-HC Approach

Jianzhou Zhang

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2015

ABSTRACT

Analyzing Hierarchical Data with the DINA-HC Approach

Jianzhou Zhang

Cognitive Diagnostic Models (CDMs) are a class of models developed in order to diagnose the cognitive attributes of examinees. They have received increasing attention in recent years because of the need of more specific attribute and item related information. A particular cognitive diagnostic model, namely, the hierarchical deterministic, input, noisy 'and' gate model with convergent attribute hierarchy (DINA-HC) is proposed to handle situations when the attributes have a convergent hierarchy. Su (2013) first introduced the model as the deterministic, input, noisy 'and' gate with hierarchy (DINA-H) and retrofitted The Trends in International Mathematics and Science Study (TIMSS) data utilizing this model with linear and unstructured hierarchies. Leighton, Girl, and Hunka (1999) and Kuhn (2001) introduced four forms of hierarchical structures (Linear, Convergent, Divergent, and Unstructured) by assuming the interrelated competencies of the cognitive skills. Specifically, the convergent hierarchy is one of the four hierarchies (Leighton, Gierl & Hunka, 2004) and it was used to describe the attributes that have a convergent structure. One of the features of this model is that it can incorporate the hierarchical structures of the cognitive skills in the model estimation process (Su, 2013). The advantage of the DINA-HC over the Deterministic, input, noisy 'and' gate (DINA) model (Junker & Sijtsma, 2001) is that it will reduce the number of parameters as well as the latent classes by imposing the particular attribute hierarchy. This model follows the specification of the DINA except that it will pre-specify the attribute profiles by utilizing the convergent attribute hierarchies. Only certain possible attribute pattern will be allowed depending on the particular convergent hierarchy. Properties regarding the DINA-HC and DINA are examined and compared through the simulation and empirical study. Specifically, the attribute profile pattern classification accuracy, the model and item fit are compared between the DINA-HC and DINA under different conditions when the attributes have convergent hierarchies. This study indicates that the DINA-HC provides better model fit, less biased parameter estimates and higher attribute profile classification accuracy than the DINA when the attributes have a convergent hierarchy. The sample size, the number of attributes, and the test length have been shown to have an effect on the parameter estimates. The DINA model has better model fit than the DINA-HC when the attributes are not dependent on each other.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would not have been possible to complete this dissertation without the contributions and support of many people. First and foremost, I would like express my gratitude to my advisor and mentor, Professor Matthew Johnson, for his invaluable advice, patience and encouragement throughout the whole process. I am grateful and thankful to have had Professor Johnson as my advisor during this long journey.

I also would like to thank my dissertation committee members, Professor Young-Sun Lee, Professor Ye Wang, Professor James Corter, and Professor Yang Feng for their questions, comments, and suggestions on the paper. I am especially grateful to my committee chairperson, Professor Lee, for her candid comments and valuable suggestions, to Professor Corter for offering me the opportunity to be his teaching assistant, and to Professors Wang and Feng for their kindness and encouragement.

I would also like to thank my friends, Dr. Jung Yeon Park, Ms. Ruchi Sachdeva, Dr. Rui Xiang, Mr. Huacheng Li, Mr. Zhuangzhuang Han, Ms. Rong Cheng, Ms. Chenmu Xing, Dr. Qi Shen, Ms. Jie Zhou, Mr. Yao Chen, Dr. Yi-Hung Lin, Dr. Luozhou Li, Mr. Long Chen, Dr. Peter Park, Mr. Chao Chen, and Mr. Xiang Yang, for their help and encouragement throughout the study.

Finally I want to thank my family, particularly my sister Jianhang Zhang, for her love and support all these years. I never would finished this dissertation without my family's unconditional love.

This thesis is dedicated to my family who have been a constant source of inspiration

Chapter 1

Introduction

The Item Response Theory (IRT) models are conventional tools used to analyze data sets at the item level, and the Classical Test Theory (CTT) models analyze data at the test level. Both IRT and CTT assume a single overall score to differentiate students' proficiency. Cognitive Diagnostic Models (CDMs) are considered as a popular modeling method to diagnose the cognitive attributes of examinees. They can measure the specific knowledge structures and skills that students need in order to solve test items. CDMs were developed for the purpose of identifying the presence or absence of the multiple fine-grained skills that are required for examinees to produce correct responses. These skills also are called attributes, and they have been described as "Production rules, procedural operations, item types, or, more generally any cognitive tasks" (Tatsuoka, 1990). The advantage of using CDM is that it can incorporate cognitive structures in the psychometric model and thereby allow us to classify examinees into various attribute profiles that indicate their mastery level (Park & Lee, 2014). To be specific, if a student incorrectly answers an algebra question, neither CTT or IRT can identify the missing skills that led the student to the wrong answer. CDM, however, can identify the specific skills that the students lack. For instance, Gierl et al. (2009b) suggested that a simple algebra problem from the SAT test might require multiple attributes such as the conceptual geometric, quadratic equation and abstract geometric series to correctly respond to it. It would be impossible to detect the specific attributes that students need in order to respond to an item correctly by implementing the IRT or CTT methods, so CDMs can potentially provide more detailed information by identifying the presence or absence of specific skills in students.

It has been suggested that certain basic lower level knowledge may be the prerequisite of upper level knowledge (Kuhn, 2001b) and prior knowledge plays an important role on the performance in mathematical, science and other area study (Hudson & Rottmann, 1981). Many mathematical knowledge and science concepts as well as other conceptual domains are not separate segments; there exist certain relationships among those concepts(Kuhn, 2001a). For instance, the mathematical skills are often correlated (Sternberg & Ben-Zeev, 1996). In order to make progress, students need to know how to add two numbers before they master the addition of multiple numbers. In other words, basic concepts such as numbers, proportion, and measurement should be mastered before learning upper level concepts such as geometry and calculus. The strong associations that exist among the concepts would require students to learn basic concepts before they can master the upper level concepts. It's usually difficult to understand complicated network of concepts without prior establishment of simpler concepts. The conceptual knowledge segments can be divided into different parts and each part can be related to each other (Vosniadou & Brewer, 1992). Most of the mathematics and science concepts have connections(Battista, 2004) and they are sometimes dependent on each other, so there will be certain learning points or sequences in the process. For instance, in order to understand algebra, it is important to get an understanding of the concepts of fractions such as decimals and percentage.The hierarchical learning structure is important because it would offer a guide on the specific learning paths students might have when they study various knowledge concepts. It is natural to assume that attribute hierarchy exists among the skills and other concepts, if progress requires the mastery of other attributes.

The popularity of online tests and assessments can provide a large amount of information on students' performances in terms of the specific attributes. The DINA model does not assume an ordering of the attributes and usually there will be no hierarchies involved with the DINA model, so when the DINA model is applied, the parameter estimates will not be accurate (Su, 2013). The hierarchical deterministic, inputs, noisy, "and" gate with convergent hierarchy (DINA-HC) model is a specific type of CDM that was developed in order to handle the hierarchical structure of the educational data. It was introduced by Su (2013) as the modified DINA and DINO with hierarchical configurations (DINA-H and DINO-H) to fit the dependent attributes data. The DINA and DINO models were chosen to be modified because of their simplicity and interpretability. Su (2013) then examined the DINA-H and DINO-H by using two of the four attribute hierarchies (linear and unstructured). The major benefit of using the DINA-HC model is that it can provide less biased estimates than the DINA model for the hierarchical structure data because of the reduced number of parameters and latent class patterns. The hierarchical model with the specific hierarchy is unique because it considers the existing hierarchical nature of attributes, and thus it provides more reliable and detailed information on examinees' learning progressions. However, the study carried out by Su (2013) is limited. The linear attribute hierarchy is the simplest form. For instance, in the Mental Models Theory (MMT) of syllogistic reasoning (Leighton et al., 2004b), from the construction of the initial model to the generation of a conclusion, there are usually a series of steps involved. (1) interpret quantifiers according to the logical criteria (A1); (2) create 1st unique representation of logical quantifiers (A2); (3) create 2nd unique representation of logical quantifiers (A3) (premised on the 1st representation); (4) create 3rd unique representation of logical quantifiers (A4) (premised on 2nd representation); Without the mastering of A1, it is not possible to master the following attributes (A2, A3, and A4). The unstructured hierarchy is an extreme hierarchical structure and does not usually exist in the real settings. The unstructured hierarchy describes a situation where a single attribute (A1) is the prerequisite for all other attributes (A2, A3, A4, and so on), and there is no connection among all the other attributes.

The convergent attribute hierarchy is a more practical hierarchy with a convergent structure, allowing two different paths to be followed from the first to the last attribute. Su (2013) investigated the MSE and model fit between the DINA-H and DINA. In addition, this study further examined the item fit such as the Bias, Absolute Bias Error (a similar alternative to MSE), and the attribute pattern classification accuracy.

Specifically, the research objective is to modify the DINA with convergent attribute hierarchy as the DINA-HC model and to apply both DINA-HC and DINA to the hierarchical attributes data. The results from both models were obtained and compared. In order to achieve this goal, three research questions were prepared:

Question 1: How does the DINA-HC compare with the DINA, in terms of the model fit under different conditions?

Question 2: How do the item parameter estimates (guessing and slipping) for DINA-HC and DINA compare under different conditions?

Question 3: How does the attribute pattern classification accuracy compare between DINA-HC and DINA under different conditions?

The evaluation criteria are discussed in Chapter 3 and the three research questions are addressed through simulation and empirical studies.

This study is an important addition to the existing studies that cover only linear and unstructured attribute hierarchy when estimating the item parameters. Su (2013) investigates the DINA and DINO model with the linear and unstructured attribute hierarchy and suggests that the DINA and DINO model with linear and unstructured attribute hierarchy are preferable when the skills are ordered hierarchically. A confirmatory hypothesis testing (ANOVA) is carried out in the analysis to provide definite answers to the research questions and to help us understand the bias differences between DINA-HC and DINA models.

Furthermore, this study investigates the performance of the DINA-HC and DINA models when the attributes are not dependent on each other. This would provide additional insights into not only the strength of DINA-HC but also the weakness when the data are independent. Another contribution of this paper is to set DINA-HC model as an example, so that the method can be generalized in an equation that can handle different types of attribute hierarchies. Specifically, the alpha matrix (the mastery status of the k^{th} skill by the i^{th} skill pattern) is constrained and pre-specified in the DINA equation. It also is important to note that the bias, the absolute bias over all items, the absolute bias error, and the attribute profile pattern classification accuracy are introduced in the analysis to help us understand the differences between the DINA-HC and the DINA model. Lastly, the The Examination for Certificate of Proficiency in English (ECPE) data are used for the empirical study to examine the performance of DINA with linear hierarchy. Linear hierarchy is one of the four hierarchies and it provides additional insights into the topic of hierarchical DINA. It is been suggested that the Examination for Certificate of Proficiency in English (ECPE) test yields linear attribute hierarchical data (Templin & Bradshaw, 2014). It is used to examine the situation when small number of attributes are present. It is administered yearly and used as a measure of English skills for non-native speakers.

This dissertation includes five chapters. Chapter 2 contains more detailed information on CDMs including their advantages and limitations, the nature of knowledge structures, and the specific hierarchies. Chapter 3 explains the methods involved in this study. Chapter 4 presents the results of simulation and empirical data analyses. Chapter 5 summarizes the results, explain the limitations, and point outs the future research objectives.

Chapter 2

Literature Review

The DINA-HC model is based on the DINA model, which is the most parsimonious and interpretable CDM used to examine the mastery level of the cognitive skills that examinees possess. The central idea of the DINA-HC model is to utilize the DINA model with skill hierarchies. This chapter includes three sections. The first section introduces cognitive diagnostic models and the Q-matrix in greater detail; three popular Cognitive Diagnostics Models (DINA, DINO and GDM) based on attributes and the Q-matrix are described. The second section introduces the hierarchies that exist among various attributes and discuss in detail the four specific attribute hierarchies. The third section reviews the methods for estimating CDMs with attribute hierarchy.

2.1 The Introduction of the Cognitive Diagnostic Models

Cognitive Diagnostic Models (CDMs) intend to assess in detail whether the examinees have mastered the specific skills required to respond to the item. They are superior to the traditional psychometric approaches such as Item Response Theory (IRT) and

Classical Testing Theory (CTT) when it comes to skill assessments (Henson et al., 2009). IRT normally concentrates on the item-level and CTT on the test-level analysis. However, CDMs examine students' mastery of skills on the attribute level, which provides more detailed information about students' specific strengths or weakness (Huebner, 2010). Thus it will be easier for instructors to assess the knowledge level of students and the specific areas that need more attention. Most CDMs are usually developed for the purpose of discovering the latent attributes that examinees possess based on the test items (DeCarlo, 2010). There are several similar models for CDMs in the existing literature including cognitive diagnosis models (Tatsuoka, 1995; de la Torre, 2009a), latent response models (Maris, 1995) and structured item response models (Rupp & Mislevy, 2007). These models have different specific emphases, assumptions, and requirements as compared to the common CDMs. They have different parameters and degree of complexity, but the common purpose of the CDMs is to provide examinees with detailed information and feedback regarding their understanding of the specific skills (attributes). In most cases the specific skills are binary, that is, the examinees either know the attribute or they do not. Knowing of an attribute is usually represented by 1 and not knowing by 0.

Different CDMs may have different assumptions, and thus there will be different ways to categorize those CDMs. One of the most popular methods is to divide them into disjunctive and conjunctive categories (Roussos et al., 2007; Huebner, 2010). Conjunctive models such as DINA (Junker & Sijtsma, 2001), the reduced RUM (DiBello et al., 2006), and the NIDA model (Junker & Sijtsma, 2001) assume that it is necessary to possess all the required attributes to correctly respond to an item unless the examinees respond by guessing. It also assumes that it is not possible to incorrectly respond to an item if all the attributes of this item have been mastered, unless the examinee makes a mistake by slipping. On the other hand, disjunctive models such as the DINO model (Templin et al., 2007) and the General Diagnostic Model (von Davier, 2005) assume that a deficiency in one of the required attributes will be compensated for if the examinees are able to possess other attributes. Without any constraints on the latent class patterns, there will be a maximum of 2^k existing latent class patterns and $2^k - 1$ possible parameters involved. It will be very difficult to estimate all the parameters as the k (skills) increase.

2.1.1 The Deterministic, Input, Noisy 'And' Gate Model

CDMs were developed so that they can be used to examine the cognitive attributes of examinees. They assume that each single item measures a subgroup of the attributes on the test. The Deterministic, input, noisy 'and' gate (DINA) model (Junker & Sijtsma, 2001) is the most parsimonious and interpretable CDM used to determine the mastery level of the cognitive skills that examinees possess. It lays a foundation for most other models in cognitive diagnostic tests (Tatsuoka, 1995, 2002). The probability of correctly responding to an item depends on three factors: the guessing parameter, the slipping parameter, and the latent response variable. The parameter for guessing g_j is the probability of responding to the j^{th} item correctly even if the examinee does not possess all the required attributes. The slipping parameter s_j is the probability of responding to the j^{th} item incorrectly with all the attributes required for the item. The latent response variable ξ_{ij} is a dichotomous variable. If the examinee possesses all the needed attributes to respond correctly and the latent response variable was scored as 1, and if the examinee lacks at least one of the needed attributes and responds incorrectly, then it was scored as 0. The mathematical representations of the guessing and slipping parameters are as follows:

$$s_j = P(X_{ij} = 0|\xi_{ij} = 1) \tag{2.1}$$

$$g_j = P(X_{ij} = 1|\xi_{ij} = 0) \tag{2.2}$$

where ξ matrix is a binary matrix that indicates whether the examinee i has all the needed attributes for the item j, which can be represented as follows:

$$\xi_{ij} = \prod_{k=1}^{n} \alpha_{ik}^{q_{jk}} \tag{2.3}$$

where α_{ik} is the mastery status of the k^{th} skill by the i^{th} skill pattern and q_{jk} is the matrix that expresses whether the j^{th} item needs the k^{th} skill to respond correctly. The final DINA equation can then be represented mathematically as follows:

$$P_j(\alpha_i) = P(X_{ij} = 1|\alpha_i) = g_j^{1-\xi_{ij}}(1 - s_j)^{\xi_{ij}} \tag{2.4}$$

According to this equation, it is necessary to have all the needed skills and not slip in order to correctly respond to an item, and an examinee who lacks at least one of the needed attributes can still respond correctly by guessing. There are 2^k total possible attribute profiles in the DINA model, where k refers to the number of skills that the test measures. The attribute profiles also are known as latent classes. The DINA model categorizes examinees into two classes: one in which the examinees have mastered all the attributes, and the other one in which the examinees lack mastery of at least one of the attributes required by the item. The DINA model is a non-compensatory (conjunctive) model, because it assumes that it is necessary to master all the required skills for an item in order to respond correctly to it unless the examinees respond by guessing. It is important to note that if an examinee is missing one required attribute, this is equivalent to the examinee missing all the attributes. The DINA also assumes that it is not possible to get an item incorrect if the examinee possesses all the required attributes unless it is a mistake made by slipping. As the DINA model is the most parsimonious CDM model, it has only two parameters to be estimated per item: guessing and slipping. However, there are many different CDMs depending upon the model assumptions. Some of the CDMs are Rule Space model (Tatsuoka, 1985, 1990, 2009); the DINA and NIDA models (de la Torre, 2011; Junker & Sijtsma, 2001; Templin, 2006); the DINO and NIDO models (Templin, 2006; Templin & Henson, 2006); the Attribute Hierarchy Method model (Gierl, 2007; Gierl et al., 2009; Leighton et al., 2004a); and the reparameterized unified/fusion model (RUM) (Hartz, 2002).

2.1.2 The Deterministic, Input "Or" Gate Model

The deterministic, input "or" gate (DINO) (Templin, 2006) model is defined in a similar way as DINA and is considered to be the counterpart of the DINA model. Similarly, the probability of correctly responding to an item depends on three factors: the guessing parameter g_j, the slipping parameter s_j, and the latent response variable ω. The guessing parameter for the DINO model is the probability of correctly responding to the item j when the examinee has not mastered any of the needed attributes. The slipping parameter is the probability of responding to the item incorrectly when the examinee has mastered at least one of the needed attributes. The latent class variable ω is a dichotomous variable that is divided into two latent classes. (1) the examinees have mastered at least one of the needed attributes. (2) the examinees have mastered none of the needed attributes specified in the Q-matrix. One of the major differences between DINA and DINO is that it is possible for the DINO model to correctly respond to an item as long as the examinee has mastered at least one of the needed skills. It is important to note that the number of skills and the type of skills are no longer necessary. The mathematical representations of the DINO parameters and the final item response function are as follows:

$$s_j = P(X_{ij} = 0|\omega_{ij} = 1) \tag{2.5}$$

$$g_j = P(X_{ij} = 1|\omega_{ij} = 0) \tag{2.6}$$

$$P_j(\omega_{ij}) = P(X_{ij} = 1|\omega_{ij}) = g_j^{1-\omega_{ij}}(1 - s_j)^{\omega_{ij}} \tag{2.7}$$

where the latent class variable ω_{ij} is a binary disjunctive model and id defined as follows:

$$P_j(\omega_{ij}) = P(X_{ij} = 1|\omega_{ij}) = g_j^{1-\omega_{ij}}(1 - s_j)^{\omega_{ij}} \tag{2.8}$$

The DINA and DINO models both have only two parameters, guessing and slipping, and the latent class responses are divided into two categories: mastery of the skills and non-mastery of the skills. The DINA and DINO models are the simplest CDMs.

2.1.3 The Hierarchical Deterministic, Input, Noisy 'And' Gate Model

The Hierarchical Deterministic, Input, Noisy 'And' Gate Model is a specific type of CDM that was introduced by Su (2013) as the DINA-H to fit hierarchical attributes. Su (2013) investigated the DINA with linear hierarchy (DINA-HL) and unstructured hierarchy (DINA-HU) and concluded that both DINA-HL and DINA-HU outperform the DINA model when the attributes are hierarchical. In terms of attribute hierarchies, there are four basic types: linear, convergent, divergent and unstructured (Leighton et al., 2004b). Those will be discussed in detail in the last section of this chapter. If the DINA model is applied to the hierarchical attribute data without any constraint on the attribute hierarchies, the parameter estimates will not be accurate because of the hierarchical nature of the attributes. The DINA-HC model can reduce the number of possible attribute profiles and decrease the sample size requirements. The study carried out by Su (2013) suggests that the DINA-H model with linear and unstructured hierarchy should be considered instead of the DINA model when the attributes are hierarchical. The DINA-HC model used here applied not only to the hierarchical attributes data but also the non-hierarchical attributes data in the simulation study. The DINA-HC model assumes that the attributes are dependent in a certain way. In other words, some lower level attributes could be the necessary requirements for more complicated attributes (Kuhn, 2001b). When the hierarchical dependence among attributes is pre-specified, the number of possible attribute profiles will be smaller than 2^k (Templin et al., 2010), where 2^k is the number of possible attribute profiles when there are no constraints imposed on the latent classes and k is the number of attributes (skills) that examinees possess.

2.1.4 The General Diagnostic Model

The General Diagnostic Model (von Davier, 2005) is used to define a general class of models for cognitive diagnosis based on a series of models including the Rasch model, item response theory models, skill profile models, and the extension of latent class models. Most of these approaches utilize the Markov Chain Monte Carlo (MCMC) estimation methods, because the maximum likelihood estimations (MLE) are not feasible. The GDMs suggested by von Davier (2005) would outline the parameter estimations by using the MLEs. The concept behind diagnostic models is that various items will have various sets of skills and that experienced professionals are able to build a Q-matrix to solve the items in an assessment. An important feature of GDMs is that they extend the applicability of skill profile models to polytomous items and to skills having more than two proficiency levels (von Davier, 2005). The diagnostic models can be developed and refined based on the GDM framework, which is considered as an item response model, and the probability of the response x, given the respondents v on the item i is shown as follows:

$$P(X = x|i, v) = \frac{\exp(f(\lambda_{xi}, \theta_v))}{1 + \sum_{y=1}^{m_i} \exp(f(\lambda_{xi}, \theta_v))} \tag{2.9}$$

where item responses range from 0 to m_i, the respondents ranges from 1 to N, and item i ranges from 1 to I. The item parameters $\lambda_{xi} = (\beta_{ix},\, q_i,\, \gamma_{xi})$ and a skill vector $\theta_v = (\alpha_{v1}, ..., \alpha_{vk})$, with continuous, ordinal or binary skill variable α_k.

Equation 2.9 is the general form. In the case of the binary data, von Davier (2005, 2008) developed the following linear GMD:

$$P(X = x|i, v) = \frac{\exp(\beta_{ix} + \sum_{k=1}^{k} \gamma_{ixk} b(q_{ik}, a_k))}{1 + \sum_{y=1}^{m_i} \exp(\beta_{ix} + \sum_{k=1}^{k} \gamma_{ixk} b(q_{ik}, a_k))} \tag{2.10}$$

where α_k is the skill vector and b(q,α) =qα for the parsimony.

Equation 2.9 can be used to define both compensatory and non-compensatory models. It has been shown by von Davier (2013) that the DINA (non-compensatory) model can be represented equivalently as a special case of a more general compensatory family of diagnostic models, and the equivalency will hold for all the DINA models with any construction of the Q-matrix. There is no additional structure that is introduced for both DINA and GDM. The equivalency is established by utilizing a mapping of DINA skill space onto an alternative skill space of a DINA-equivalent GDM (von Davier, 2014). The GDM can be seen as a model family that consists of different CDMs where certain restrictions are placed on the parameters. It is worth noting that the family of GDMs can use many kinds of data including not only the binary and polytomous data but also the ordinal and mixed format data.

2.2 The Details of Q-matrix

The Q-matrix is the essential input used during a diagnostic analysis of most CDMs. The proper construction and specification of the Q-matrix will increase the diagnostic power of the CDMs, and the proper specification of a Q-matrix is necessary in order to implement the CDMs (Tatsuoka, 1995). Normally it would be assumed that each item of a test requires some subset of attributes (skills) in order to correctly respond. The particular attributes required by the test items are described by a binary matrix, which is known as the Q-matrix (Tatsuoka, 1983). The columns of the matrix are the attributes (skills), and the rows are the items. When an item requires a certain attribute, the corresponding spot in the column is assigned a 1; otherwise, it is 0. Thus, the responses would only be discrete binary values (1 or 0). The J x K Q-matrix was specified in such a way that K represents the number of skills and J represents the number of items in a given assessment (Tatsuoka, 1995). A detailed Q-matrix can be constructed based on Table 1 drawn from DeCarlo (2010) and taking items 1, 2, 3, 4, 5, and 6 as the examples. There are eight attributes in the table and they are labeled A1, A2 and so on. The Q-matrix is shown in Table 2.1.

Item	A1	A2	A3	A4	A5	A6	A7	A8
$\frac{5}{3} - \frac{3}{4}$	0	0	0	1	0	1	1	0
$\frac{3}{4} - \frac{3}{8}$	0	0	0	1	0	0	1	0
$\frac{5}{6} - \frac{1}{9}$	0	0	0	1	0	0	1	0
$3(\frac{1}{2}) - 2(\frac{3}{2})$	0	1	1	0	1	0	1	0
$4(\frac{3}{5}) - 3(\frac{4}{10})$	0	1	0	1	0	0	1	1
$\frac{6}{7} - \frac{4}{7}$	0	0	0	0	0	0	1	0

Table 2.1: A 6 items to 8 attributes Q-matrix

The Q-matrix presented here includes eight attributes (skills) as follows: (1) convert a whole number; (2) separate a whole number from a fraction; (3) simplify before subtracting; (4) find a common denominator; (5) borrow from whole number part; (6) column borrow to subtract the second numerator from the first; (7) subtract numerators; and (8) reduce answers to the simplest form (Tatsuoka, 1995; DeCarlo, 2010). It is easy to see that mastery of attributes 4, 6, and 7 is required in order to solve item 1; mastery of attributes 4 and 7 is required to solve item 2; and so on. And yet the attributes are not dependent on each other.

2.3 The Cognitive Attribute Hierarchies

The attribute hierarchy method (AHM) utilizes the hierarchies that exist among the attributes and four types of popular hierarchies: linear, convergent, divergent and unstructured were introduced (Leighton et al., 2004a,b; Gierl, 2007; Gierl et al., 2009). This can be a new approach to dealing with hierarchically ordered attributes. If the attributes are hierarchically dependent, then those hierarchies of the attributes can be combined to form more complicated forms of cognitive skills (Kim, 2001). There are four types of popular cognitive attribute hierarchies: linear, convergent, divergent and unstructured (Leighton et al., 2004b). Linear, convergent, and divergent hierarchies are structured. When combining two or more of the four types of hierarchies, the result is even more complex networks of hierarchies (Kim, 2001).

2.3.1 Linear Attribute Hierarchy

In all hierarchies, the first attribute (A1) may be considered hypothetical because it represents all the initial competencies that are prerequisites to the ensuing attributes (Leighton et al., 2004b). In the linear attribute hierarchy, attribute 1 (A1) is a prerequisite for attribute 2 (A2); attribute 2 (A2) is the prerequisite for attribute 3 (A3); and so on. In other words, the examinee cannot master attribute 2 (A2) without mastering attribute 1 (A1), and furthermore the examinee cannot master attribute 6 (A6) without mastering all the previous attributes. Thus There is only one path for reaching a given skill level.

In this study, the linear attribute hierarchies should satisfy the following components.

Component 1: For all the linear attribute hierarchies, there will be a starting point. This starting point will be represented as attribute one (A1).

Component 2: For all the linear attribute hierarchies, there will be an ending point. This ending point will be represented as attribute N (AN), where N equals the number of all possible attributes.

Component 3: The attributes between the starting point and ending point should exhibit a linear relationship, and the mastery of each attribute should require the mastery of all the previous attributes.

The linear attribute hierarchy is the simplest hierarchy. For instance, in the Mental Models Theory (MMT) of syllogistic reasoning (Leighton et al., 2004b), from the construction of the initial model to the generation of a conclusion, there are usually a series of steps involved. (1) interpret quantifiers according to the logical criteria (A1); (2) create 1st unique representation of logical quantifiers (A2); (3) create 2nd unique representation of logical quantifiers (A3) (premised on the 1st representation); (4) create 3rd unique representation of logical quantifiers (A4) (premised on 2nd representation); Without the mastering of A1, it is not possible to master the following attributes (A2, A3, and A4). The hypothetical hierarchy with six attributes can be illustrated in Figure 2.1.

2.3.2 Convergent Attribute Hierarchy

In the convergent attribute hierarchy, the examinee can master attribute 6 (A6) by mastering either attribute 4 (A4), attribute 5 (A5), or both. It also is noted that in order for examinees to master attribute 4 (A4) or 5 (A5), they must also master attribute 1 (A1), attribute 2 (A2), and attribute 3 (A3). In other words, the examinee can reach a given point from multiple paths. In the example illustrated by the following figure, the examinee can reach attribute 6 (A6) either through attribute 4 (A4), attribute 5 (A5), or both.

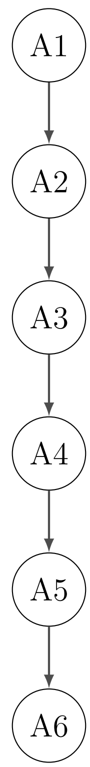

Figure 2.1: A linear hierarchy with 6 attributes

In this study, the convergent attribute hierarchies should satisfy the following components:

Component 1: For all the convergent attribute hierarchies, there will be a starting point. This starting point will be represented as attribute one (A1).

Component 2: For all the convergent attribute hierarchies, there will be a ending point. This ending point will be represented as attribute N (AN), where N equals the number of all possible attributes.

Component 3: The convergence happens when at least two possible paths exist in which one can reach a single point (the convergent attribute).

Component 4: There should be at least one convergence in a particular hierarchy.

This convergent hierarchy can be used to describe the condition wherein different cognitive competencies lead to a single correct end state where the outcome is clear, for example, a multiple choice test that measures the addition of fractions (Leighton et al., 2004b). In order to solve $\frac{5}{6} + \frac{7}{9}$, finding the common multiple is the key. This can be done either by finding the lowest common multiple and using it as the new denominator (A2) or by multiplying the denominator of each fraction and using the multiply as the new denominator (A3). In the case of A2 the lowest common multiple is 18, and we will have $\frac{15}{18}+\frac{14}{18}$; In the case of A3, the multiplication of the denominator of each fraction will result in 54 and we will have $\frac{45}{54} + \frac{42}{54}$. Attribute 4 (A4) is defined as the addition of the fraction. Either through A2, A3, or both, we will get the final answer $\frac{29}{18}$. Here we can say that A2 and A3 converged into A4. The convergent hierarchy with 6 attributes can be shown as follows:

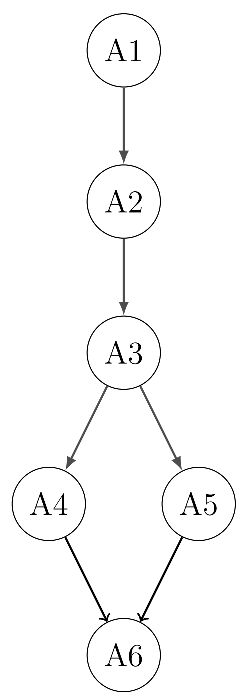

Figure 2.2: A convergent hierarchy with 6 attributes

2.3.3 Divergent Attribute Hierarchy

In terms of the divergent attribute hierarchy, the examinees can master certain level attributes without mastering some of the lower skill level attributes. For instance, the examinee can master attribute 6 (A6) by mastering attribute 4 (A4) and attribute 1 (A1). A1 is the prerequisite of attribute 4 (A4). Similarly, the prerequisites for attribute 5 (A5) are only attribute 1 (A1) and attribute 4 (A4).

In this study, the divergent attribute hierarchies are shown to be able to satisfy the following components:

Component 1: For all the divergent attribute hierarchies, there will be a starting point. This starting point will be represented as attribute one (A1).

Component 2: For all the divergent attribute hierarchies, there will be multiple ending points.

Component 3: The divergence happens when there are at least two possible paths originating from a single point.

Component 4: There should be at least one divergence within a particular hierarchy.

The divergent attribute hierarchy normally is used when there is no absolutely correct answer for an open-ended question and when all the answers taken together can be described as the pool of possible solutions. For instance, students might be asked the reasons that caused the first World War. There might not be an absolutely correct answer, it can be explained from multiple perspectives such as territorial causes, economic conflicts, or political factors. It also is interpreted as the entire ordering of cognitive abilities that are necessary to correctly respond to the problems within a specific domain (Leighton et al., 2004b). Thus there will be different paths diverging from the same point. The divergent hierarchy can be illustrated as shown below.

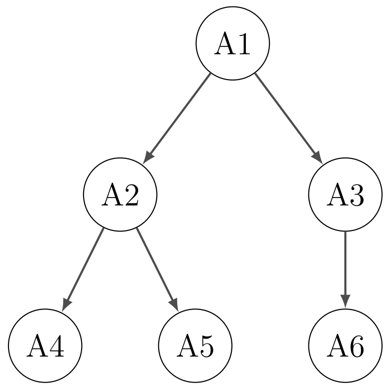

Figure 2.3: A divergent hierarchy with 6 attributes

2.3.4 Unstructured Attribute Hierarchy

In the unstructured attribute hierarchy, attribute 1 (A1) is considered the prerequisite for all the other attributes but there is no specific order among the rest of the attributes. As long as the examinee possesses attribute 1 (A1), the examinee can possess all the other attributes.

In this study, the unstructured attribute hierarchies are shown to be able to satisfy the following components:

Component 1: For all the unstructured attribute hierarchies, there will be a starting point. This starting point will be represented as attribute one (A1).

Component 2: For all the unstructured attribute hierarchies, there will be multiple ending points.

Component 3: All other attributes require only A1 and they are not related with each other .

This hierarchy can be used to describe the case in which many independent outcomes can arise from a single cause. For instance, a low employment rate (A1) can cause a rise in the crime rate (A2), the building of a bridge (A3) and higher inflation (A4). An example of an unstructured hierarchy is illustrated in Figure 2.4.

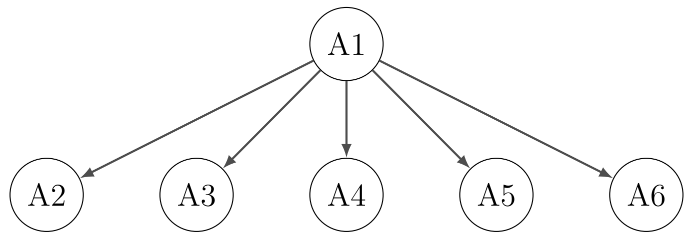

Figure 2.4: An unstructured hierarchy with 6 attributes

It is important to note that for the unstructured, divergent, and convergent attribute hierarchies, there are many different combinations of mastering paths depending upon the number of attributes. In an assessment test there can be various types of attribute hierarchies, and the number of attribute profiles will be different for each assessment (Leighton et al., 2004b).

2.4 Methods for Estimating CDMs with Attribute Hierarchy

The goal is to propose a specific hierarchical structure DINA model that will be able to reduce the number of attribute profiles and make the estimation of parameters and attribute profiles using CDM more efficient by use of the joint MLE and while using the expectation-maximization (EM) algorithms. Gierl et al. (2007) introduced a method (AHM) that utilized the skill hierarchies in the item development stage in order to reduce the number of latent classes. Su (2013) used the EM algorithms in the DINA model estimation conducted by de la Torre (2009) to estimate the item parameters and latent classes. The details are as follows.

The expected number of examinees with attribute profile $\boldsymbol{\alpha_l}$ can be calculated from

$$I_l = \sum_{i=1}^{N} p(\boldsymbol{\alpha_l}|\boldsymbol{X_i}) \tag{2.11}$$

The prior $p(\boldsymbol{\alpha_l})$ is the probability of attribute profile vector $\boldsymbol{\alpha_l}$, and it can be estimated as follows:

$$p(\boldsymbol{\alpha_l}) = \frac{I_l}{N} \tag{2.12}$$

The expected number of examinees with the attribute profile $\boldsymbol{\alpha_l}$ getting item j correct is as follows:

$$R_{jl} = \sum_{l=1}^{L} p(\boldsymbol{\alpha_l}|\boldsymbol{X_i})X_{ij} \tag{2.13}$$

where $p(\boldsymbol{\alpha_l}|X_i)$is the posterior probability that the student i has the attribute profile $\boldsymbol{\alpha_l}$ and l is the number ranging from 1 to L, where L is the number of possible attribute profiles. In the case of DINA, L= 2^k, and in the case of DINA-HC, L is equal to all possible attribute profiles pre-specified for each unique hierarchy.
$R_j^{(x)}$ is the sum of R_{jl}, where l is in the range of 1 to L and can be expressed as follows:

$$R_j^{(x)} = \sum\nolimits_{l:\eta_j(\boldsymbol{\alpha_l})=x} R_{jl} \tag{2.14}$$

$I_j^{(x)}$ is the sum of I_l and can be expressed as follows:

$$I_j^{(x)} = \sum\nolimits_{l:\eta_j(\boldsymbol{\alpha_l})=x} I_l \tag{2.15}$$

When x=0, we have the following guessing estimators:

$$\hat{g}_j = \frac{R_j^{(0)}}{I_j^{(0)}} \tag{2.16}$$

When x=1, we have the following slipping estimators:

$$\hat{s}_j = \frac{I_j^{(1)} - R_j^{(1)}}{I_j^{(1)}} \tag{2.17}$$

where $I_j^{(0)}$ is the expected number of examinees lacking at least one of the required attributes for item j. $R_j^{(0)}$ is the expected number of examinees among $I_j^{(0)}$ that respond to item j correctly. $I_j^{(1)}$ is the expected number of examinees that master the required attributes for item j, and $R_j^{(1)}$ is the expected number of examinees among $I_j^{(1)}$ that have not mastered the required attribute for item j. $I_j^{(0)} + I_j^{(1)} = I_l$ for all items j.

Chapter 3

Methods

This chapter has three sections. The first section introduces the proposed model, the second section presents three research questions, and the third and final section describes the data analysis.

3.1 The HIDINA Model with Attribute Hierarchy

Unlike DINA, the hierarchical DINA will utilize a particular attribute hierarchy to pre-specify the attribute profiles. In the DINA model the number of maximum attribute profiles will be A=2^k, where k is the number of diagnosed attributes for a given assessment. For instance, let us assume that the number of attributes k is 5, and then the number of maximum attribute profiles A will be equal to 32. However, in the HIDINA model the number of attribute profiles will be less because of the constraints imposed by the specific attribute hierarchy. For instance, in the linear attribute hierarchy the number of maximum attribute profiles is only six if there are five attributes.

The attribute profile pattern is illustrated by means of the following table; note that there are only six attribute profile patterns (00000,10000,11000,11100,11110,11111).

Attribute Profile Pattern	A1	A2	A3	A4	A5
1	0	0	0	0	0
2	1	0	0	0	0
3	1	1	0	0	0
4	1	1	1	0	0
5	1	1	1	1	0
6	1	1	1	1	1

Table 3.1: Linear hierarchy attribute profile pattern with 5 attributes

It is important to note that the number of maximum attribute profiles in the hierarchical DINA model depends on the specific attribute hierarchy. Different attribute hierarchies will result in different numbers of attribute profiles. The probability of examinee i getting a correct response on a specific item j on a particular attribute hierarchy is as follows:

$$P_j(\text{constrained-}\boldsymbol{\alpha_i}) = P(X_{ij} = 1|\text{constrained-}\boldsymbol{\alpha_i}) = g_j^{1-\xi_{ij}}(1 - s_j)^{\xi_{ij}} \tag{3.1}$$

The ξ matrix is changed accordingly:

$$\xi_{ij} = \prod_{k=1}^{n} \boldsymbol{\alpha_{ik}}^{q_{jk}} \tag{3.2}$$

where constrained-$\boldsymbol{\alpha_i}$ is the specific skills vector. Let the constrained-$\boldsymbol{\alpha_i} = \boldsymbol{\alpha_{ik}}$ be the examinees' binary skills vector. The mastery of particular skill k by examinee i will be represented by 1 and non-mastery of the particular skill will be represented by 0.

3.2 The Use of Convergent Attribute Hierarchy

The study utilizes the convergent attribute hierarchy in this analysis to examine the DINA-HC model. Studies related to divergent hierarchy can also be examined in future research. The divergent attribute hierarchy describes the situation where multiple paths originate from a single attribute, whereas the convergent attribute hierarchy is the situation where an attribute is a requirement for certain other attributes. In order to reach a certain level of skill attainment, there may be different paths from the beginning to the end point. It's important to note that the convergent hierarchy has never before been studied with DINA.

It is worth mentioning that for the linear attribute hierarchy, each upper level attribute requires the mastery of the previous lower level attribute. If attribute A1 is not present in a particular hierarchy, then all the ensuing attributes will not be mastered by the examinee either. This is the most ideal situation to detect the exact attribute that students do not possess when responding to an item incorrectly. The divergent hierarchy can be used to describe a situation where the answer can consist of multiple components (Leighton et al., 2004b); For instance, for a test item that seeks the reasons for certain events, it will be possible to propose a few reasons that cause certain events. Both convergent and divergent hierarchies have many possible hierarchical structures depending on the specific relationship among the attributes as well as the number of attributes. The unstructured attribute hierarchy refers to the case in which a single attribute can lead to multiple other attributes and when there is no dependent relationship among those attributes. The unstructured attribute hierarchy has the highest number of possible attribute profile patterns as compared to the other three hierarchical structures, under the same condition in which the number of attributes is constant.

Specifically, for a particular attribute hierarchy, the linear attribute hierarchy is usually considered to be the most limited of the four types of hierarchies. It also has the minimum number of attribute profiles, and thus it is considered as the simplest hierarchy to fit the model. The unstructured attribute hierarchy is the least constrained among the four types of hierarchies, so it has the largest number of attribute profiles among the four hierarchies under the condition whereby the number of attributes is the same. Both the linear and the unstructured hierarchies have been studied and evaluated against the DINA model by Su (2013).

In the study, the convergent attribute hierarchy can be used to describe the situation that will result in a single correct end point; for instance, a multiple choice test that measures the subtraction of fractions (Leighton et al., 2004b). The following table illustrates a convergent attribute hierarchy with five attributes:

Attribute Profile Pattern	A1	A2	A3	A4	A5
1	0	0	0	0	0
2	1	0	0	0	0
3	1	1	0	0	0
4	1	1	1	0	0
5	1	1	0	1	0
6	1	1	1	0	1
7	1	1	0	1	1
8	1	1	1	1	1

Table 3.2: Convergent hierarchy attribute profile pattern with 5 attributes

In this case of a convergent attribute hierarchy with five attributes, the mastery of attribute four can be achieved by mastering either attribute 2 or 3. The number of possible attribute profiles for this specific convergent attribute hierarchy is eight and is listed as follows, where the class profiles are 00000, 10000, 11000, 11100, 11010, 11101, 11011, and 11111.

3.3 The Research Questions

The research goal is to compare and contrast the performance of the DINA-HC model in relation to the DINA model when the attributes are independent or hierarchically ordered.

In order to achieve this goal, the following questions are addressed.

1. How does the DINA-HC compare with DINA in terms of the model fit under different conditions?

2. How do the item parameter estimates (guessing and slipping) for DINA-HC and DINA compare under different conditions?

3. How does the attribute pattern classification accuracy compare between DINA-HC and DINA under different conditions?

The evaluation criteria are discussed in the following section. A "parameter recovery" study (French & Dodd, 1998) is carried out, because it is necessary to calibrate the data under different conditions (e.g., number of attributes, number of items, estimating model, and sample size). The CDM parameters from the calibration are compared to the known parameters that were originally generated. If they are close enough, this indicates that they recover the true parameters and that thus the estimation is accurate.

3.4 The Simulation Study

3.4.1 The Simulation Design

The simulation has a 4-factor design for both the hierarchical and the non-hierarchical attributes. Thus it has a 3 (sample size) $\times$ 2 (estimating model) $\times$ 3 (number of attributes) $\times$ 2 (number of test items) design.

Factor One: Sample sizes are set at 500, 2000 and 5000. The sample size values usually are 500 for the small size, and 10000 for a large size in a given assessment (Rupp & Templin, 2008b). It is useful to examine DINA-HC under the most commonly used sample size conditions as well as under small size conditions. There is no specific requirement for the choice of sample size, so it is reasonable to choose 500 for the small sample size, 2000 for the medium sample size, and 5000 for the large sample size.

Factor Two: The estimating models are set as DINA and DINA-HC with 4, 6, and 7 attribute hierarchies as shown in Figure 3.1. The comparison is between DINA-HC and the DINA model.

Factor Three: The number of attributes are set as 4, 6, and 7. It has been noted that most application examples in the various studies of CDMs use 4 to 8 attributes (Rupp & Templin, 2008a). Nevertheless there is no specific requirement for number of attributes. A particular convergent hierarchy with 4 attributes will result in 8 attribute profiles as compared to 16 for the constrained hierarchy. The attribute profile for a particular convergent hierarchy having 6 attributes is 10, as compared to 64 for the unconstrained hierarchy and the attribute profile for a particular convergent hierarchy having 7 attributes is 26 as compared to 128 for the unconstrained hierarchy, all of which may be seen in Figure 3.1. The differences are large enough that the complexity will increase dramatically as the number of attributes increases.

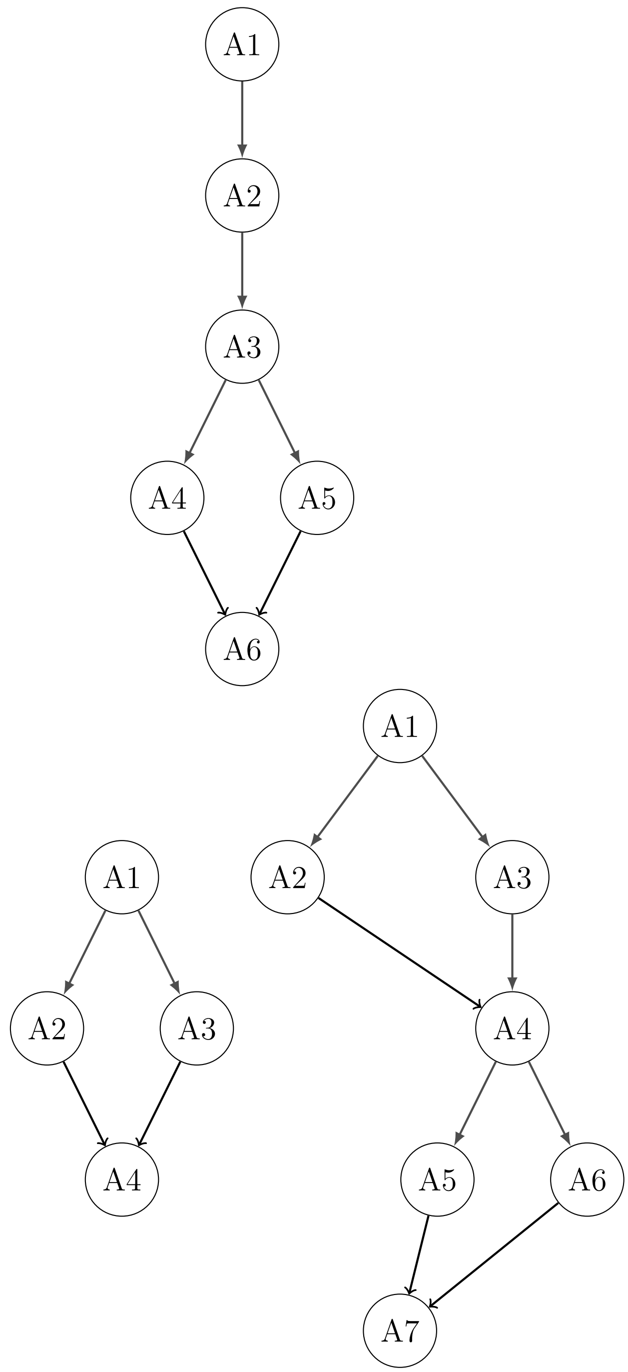

Figure 3.1: Convergent hierarchies with 6, 4, and 7 attributes

The number of attribute profiles are 8, 10, and 26 for the 4, 6, and 7 attributes respectively. Taking the convergent attribute hierarchy with 4 attributes as our example, the attribute profile patterns break down thus:

Attribute Profile Pattern	A1	A2	A3	A4
1	0	0	0	0
2	1	0	0	0
3	1	1	0	0
4	1	0	1	0
5	1	1	1	0
6	1	1	0	1
7	1	0	1	1
8	1	1	1	1

Table 3.3: Possible attribute profiles for 4 attributes

Factor Four: In terms of number of items, there are 20 items and 40 items. The fraction subtraction data of K. K. Tatsuoka (1990) is one of the most widely analyzed CDM datasets. De la Torre (2008, 2009b) suggested using either 15 or 30 items of the fraction subtraction data, while DeCarlo (2011) uses the full set of 20 items to carry out his study. It would be reasonable to choose 20 and 40 for the test length. The simulated Q-matrices with 20 and 40 items can be seen in the appendix.

Before the simulation, CDMs require the establishment of the fixed relationships between attributes and items, so specification of the Q-matrices will be necessary. For the details of the Q-matrices, see the appendix. The default values of both guessing and slipping parameters are set at 0.2, so it would be reasonable to simulate a value close to the default value; in this case, the parameter values are sampled from a uniform distribution on the interval [0.05, 0.35]. The number of possible attribute profiles will no longer be 2^K; instead, it will be pre-specified based on the specific attribute hierarchy.

3.4.2 The Estimating Method

The goal of this simulation study is to examine whether the estimating model fits the simulated data under different conditions.The R package (Team, 2012) was employed to fit the model by using the EM-algorithm to estimate the guessing and slipping parameters. The package "CDM" (Robitzsch et al., 2014) was used. The EM-algorithm procedures based on de la Torre (2009) were followed.

If the skill vectors are known, then it is straightforward to estimate the model parameters. However, the skill vectors are normally not known, and using the joint MLE can lead to an inconsistent estimation (de la Torre, 2009a).

In the DINA model, X_{ij} is the response of student i on the item j. Let α_1, α_2,... α_L denote the allowable attribute patterns in the skill space defined by the specific hierarchy, where L is the number of possible attribute profiles. Taking the convergent attribute hierarchy with the number of 6 attributes based on Figure 2.2 in Chapter 2, we have all the attribute profiles:

Attribute Profile Pattern	A1	A2	A3	A4	A5	A6
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	1	1	0	0	0	0
4	1	1	1	0	0	0
5	1	1	1	1	0	0
6	1	1	1	0	1	0
7	1	1	1	1	1	0
8	1	1	1	1	0	1
9	1	1	1	0	1	1
10	1	1	1	1	1	1

Table 3.4: Possible attribute profiles for 6 attributes

Assuming the observed data X and the attribute profiles $\boldsymbol{\alpha}$, and that the $\boldsymbol{\alpha}$ will be the only difference as compared to the procedures utilized by de la Torre (2009), the Marginalized Maximum Likelihood for the response vector of student i can be expressed as:

$$\text{Likelihood}(\boldsymbol{X}) = \prod_{i=1}^{I} \text{Likelihood}(\boldsymbol{X_i}) = \prod_{i=1}^{I} \sum_{l=1}^{L} \text{Likelihood}(\boldsymbol{X_i}|\boldsymbol{\alpha_l})p(\boldsymbol{\alpha_l}) \tag{3.3}$$

where $p(\boldsymbol{\alpha_l})$is the prior probability for the constrained $\boldsymbol{\alpha_l}$. A guessing and slipping parameter estimation based on the marginalized likelihood can be implemented by using the EM algorithm from section 2.4. Instead of taking the weighted sum of conditional likelihood across all 2^K attribute profiles (de la Torre, 2009b), only the allowable attribute profiles l were used, where l is much smaller than 2^K and depends on the particular attribute hierarchy. The parameter estimation used here is exactly the same as the EM algorithm used by de la Torre (2009), except for the constrained $\boldsymbol{\alpha_l}$.

3.4.3 The Evaluation Methods and Parameter Recovery

In order to address Research Question 1, model fit was checked to see if the two different models fit the different simulated data sets. The model fit can be assessed by using information criteria such as Akaike Information criteria (AIC)(Akaike, 1974) and Bayesian Information criteria (BIC) (Schwarz et al., 1978). The information criteria were defined as follows:

$$\text{AIC} = -2\ln(l) + 2d \tag{3.4}$$

$$\text{BIC} = -2\ln(l) + d\ln(n) \tag{3.5}$$

where l is the maximized likelihood value for the estimated model, d is the number of estimable parameters in the model, and n is the sample size.

The model with a lower value of AIC will be preferred. Similarly, the model with a lower BIC will be preferred, and BIC will penalize free parameters more strongly than does the AIC (Shepherd, 2005). Over the replications, the proportion that each criterion correctly identifies the correct model was calculated.

In order to address Research Question 2, item fit statistics were calculated and compared between the DINA-HC approach and the DINA model.

The bias was used as a measure to evaluate item parameter fit. It is considered to be the difference between the estimator's true value and the estimated value across replications and is expressed as follows.

For the guessing parameter:

$$\text{Bias}(g) = \frac{\sum_{i=1}^{R} (\hat{g}_i - g)}{R} \tag{3.6}$$

where $\hat{g}_j$ is the estimate of the guessing parameter, g is the given true guessing parameter, and R is the number of replications equal to 100 in this case.

For the slipping parameter:

$$\text{Bias}(s) = \frac{\sum_{i=1}^{R} (\hat{s}_i - s)}{R} \tag{3.7}$$

where $\hat{s}_j$ is the estimate of the guessing parameter and s is the given true guessing parameter.

The biases for both the guessing and the slipping parameters of the items are calculated over replications for all possible conditions. The smaller values for the bias of the item parameter estimates indicate more accurate parameter estimates.

The values of bias for both the guessing and the slipping parameters of all the items were compared between the DINA-HC approach and the DINA model under different conditions.

The root mean square error (RMSE) is a measure of precision. It has been used as a standard statistics metric to measure model performance (Chai & Draxler, 2014) and it is defined as follows:

For the guessing parameter:

$$\text{RMSE}(g) = \sqrt{\frac{\sum_{i=1}^{R} (\hat{g}_i - g)^2}{R}} \tag{3.8}$$

For the slipping parameter:

$$\text{RMSE}(s) = \sqrt{\frac{\sum_{i=1}^{R} (\hat{s}_i - s)^2}{R}} \tag{3.9}$$

The RMSE is considered a quadratic scoring guide to measure the average magnitude of the error. It is used to evaluate the models by summarizing the differences between estimated and true values.

The Absolute Bias Error (ABE) is a measure of bias variation over all possible items J. It provides a single value as an indicator of item fit. It is used as an alternative to the Mean Square Error (MSE), because the absolute value will be larger than the square value when the bias is between 0 and 1. The ABE also will be easier to present in a graph and it is defined as follows.

For the guessing parameter:

$$\text{ABE}(g) = \frac{\sum_{j=1}^{J} (\frac{\sum_{i=1}^{R} \left| \hat{g_{ij}} - g \right|}{R})}{J} \tag{3.10}$$

For the slipping parameter:

$$\text{ABE}(s) = \frac{\sum_{j=1}^{J}(\frac{\sum_{i=1}^{R}\left|\hat{s_{ij}} - s\right|}{R})}{J} \tag{3.11}$$

In order to address Research Question 3, the attribute profiles classification accuracy values were calculated and compared between the DINA-HC and the DINA model under different conditions. The generated attribute profiles were matched with the estimated attribute profiles. A larger value of classification accuracy means a better model choice.

3.5 Application to Real Data

In addition to the simulation study, it is important to fit the hierarchical DINA to real data and to compare the performance of the two models. The simulated data have great advantages, such as that the generating values (true values) are known and the model complexity can be controlled. The simulated data are artificial, however, so the impact of model assumptions on model data fit can be investigated only under ideal situation. It can be useful to examine the hierarchical model using real data. The model fit and item fit were compared between the hierarchical DINA with linear hierarchy (DINA-HL) and the DINA model.

3.5.1 The ECPE Data

The Examination for Certificate of Proficiency in English (ECPE) is a test that was developed and scored by the English Language Institute of the University of Michigan and was introduced by Templin and Hoffman (2013) as a tutorial to specify CDMs in Mplus. Administered and scored on a yearly basis, the test is used to measure the advanced English skills of those whose native language is not English. A dataset of 2922 students from a single year's administration was used to analyze a 2003 to 2004 dataset that consists only of a grammar section. The average age of the examinees was 23 years old. Half of the examinees first language is Portuguese (Liu et al., 2009). Three grammar attributes, knowledge of morphosyntactic rules, cohesive rules, and lexical rules, are measured by the ECPE dataset (Templin, 2006). There are 28 ECPE items that represent the three attributes. A Q-matrix has already been proposed (Buck & Tatsuoka, 1998) and is applied to this study. The 28 items were used to measure the three attributes. There are eight items that measure only one attribute, seven items that measure only two attributes, and zero items measure all three attributes. There are 13 items that measure attribute one, that 6 attributes that measure attribute two, and 8 items that measure attribute three. An example of item (Liu et al., 2009) is as follows:

Which of the following is correct?

- I have always ( ) snow.

A); to enjoy

B); enjoyed

C); enjoying

D); to enjoyed

The ECPE data is a linear attribute hierarchical dataset (Templin & Bradshaw, 2014), so the examinees have to master each attribute before proceeding to the next higher level of attribute. In this particular case lexical rules are supposed to be mastered first, followed by mastery of the cohesive rules; the highest level of attribute is the morphosyntactic rules. In the ECPE example, the three attributes would result in 8 attribute profiles (2^3), but since it is a linear attribute hierarchy, there are 4 possible mastery profiles:

Profiles	Morphosyntactic Rules	Cohesive Rules	Lexical Rules
Profile 1	0	0	0
Profile 2	0	0	1
Profile 3	0	1	1
Profile 4	1	1	1

Table 3.5: The attribute profiles for the ECPE data

Both the DINA-HL and DINA models were used to fit the ECPE data. The Model fit and item fit were compared. The Q-matrix describes the mastery status of the attributes by the items. In the case of the ECPE data, there are three attributes (lexical rules, cohesive rules, & morphosyntactic rules) and 28 items in the Q-matrix. The Q-matrix of the ECPE data is shown on the next page.

Item	Morphosyntactic Rules	Cohesive Rules	Lexical Rules
Item 1	1	1	0
Item 2	0	1	0
Item 3	1	0	1
Item 4	0	0	1
Item 5	0	0	1
Item 6	0	0	1
Item 7	1	0	1
Item 8	0	1	0
Item 9	0	0	1
Item 10	1	0	0
Item 11	1	0	1
Item 12	1	0	1
Item 13	1	0	0
Item 14	1	0	0
Item 15	0	0	1
Item 16	1	0	1
Item 17	0	1	1
Item 18	0	0	1
Item 19	0	0	1
Item 20	1	0	1
Item 21	1	0	1
Item 22	0	0	1
Item 23	0	1	0
Item 24	0	1	0
Item 25	1	0	0
Item 26	0	0	1
Item 27	1	0	0
Item 28	0	0	1

Table 3.6: The ECPE Q-matrix

3.6 Summary of the Methods

This chapter has explained the proposed approach using mathematical equations to analyze all hierarchical attributes data. The central idea is to use a specific hierarchy to pre-specify the attribute profiles; thus the number of maximum attribute profiles was greatly reduced and more accurate parameter estimates were obtained. This simulation study uses the convergent attribute hierarchy as one of the four types of attribute hierarchies as it seeks to examine the DINA-HC. This can be an important addition to the existing study. The convergent attribute hierarchy has been chosen because it is not the simplest hierarchy and it has never been studied. The primary goal of this study is to compare the DINA-HC to the DINA model when the attributes are hierarchical. The secondary goal is to compare the the DINA-HC and the DINA model when the attributes are non-hierarchical. Both a simulation study and real data study were used to understand the differences between hierarchical DINA and DINA. The simulation study has a four-factor design (estimating model, number of attributes, sample size, and number of test items) for both hierarchical and non-hierarchical data, and it was studied by using the Expectation Maximization algorithm. The model fit, item fit, and attribute profiles classification accuracy were used as evaluation criteria to examine the performance of the DINA-HC and DINA model under different conditions.

Chapter 4

Results

4.1 Simulation Study Results

This chapter presents the simulation study results. There are a total of 36 conditions for the hierarchical attributes data and non-hierarchical data. The estimation was replicated 100 times. The DINA with the convergent attribute hierarchy (DINA-HC) was compared with the DINA model.

4.1.1 Model Fit for Hierarchical Data

In terms of model fit for the hierarchical attributes using the DINA-HC and DINA model, the average value of AIC and BIC over 100 replications was calculated. BIC tends to penalize parameters more than AIC does, but both criteria were used as the measure of model fit. The number of parameters for each condition is listed in Table 4.1.

AIC and BIC correctly selected the DINA-HC over the DINA in 100% of the replications across all conditions. This indicates that DINA-HC should be preferred when the attributes are convergent. The average values of AIC and BIC over 100 replications for the DINA-HC model are smaller than the DINA model for all conditions according to Table 4.2 and 4.3 and smaller values indicate better model fit, so the DINA-HC model has better model fit than the DINA model when the attribute hierarchy is convergent. Table 4.2 and 4.3 summarize all 36 conditions with both the AIC and BIC values.

	DINA	DINA-HC	DINA	DINA-HC
	J=20	J=40	J=20	J=40
K=4	55	95	47	87
K=6	103	143	49	89
K=7	167	207	65	105

Table 4.1: Number of parameters for each condition

		Model Fit (AIC)					
		DINA-HC	DINA	DINA-HC	DINA	DINA-HC	DINA
		K=4	K=4	K=6	K=6	K=7	K=7
N=500	J=20	10526	10540	11926	12018	12229	12410
	J=40	21376	21390	21165	21266	22258	22449
N=2000	J=20	44746	44760	46131	46225	48735	48916
	J=40	85711	85725	84150	84248	89402	89591
N=5000	J=20	107445	107459	114312	114415	113381	113566
	J=40	219874	219889	207517	207619	216386	216574

Table 4.2: Model fit (AIC) for all conditions: hierarchical data

		Model Fit (BIC)					
		DINA-HC	DINA	DINA-HC	DINA	DINA-HC	DINA
		K=4	K=4	K=6	K=6	K=7	K=7
N=500	J=20	10725	10772	12133	12452	12503	13114
	J=40	21743	21791	21540	21868	22701	23321
N=2000	J=20	45009	45068	46405	46802	49099	49851
	J=40	86198	86257	84648	85049	89990	90750
N=5000	J=20	107751	107817	114631	115086	113805	114654
	J=40	220441	220508	208097	208551	217070	217923

Table 4.3: Model fit (BIC) for all conditions: hierarchical data

4.1.2 Model Fit for Non-hierarchical Data

In terms of model fit for the DINA-HC and DINA model for attributes that are not dependent on each other, the average value of AIC and BIC over 100 replications is calculated. Similarly, AIC and BIC correctly selected the DINA-HC over the DINA in 100% of the replications across all conditions., except that the condition of 7 attributes, 500 sample size, and 20 items favors the DINA-HC model in all 100 replications.

The average values of AIC and BIC over 100 replications for DINA are smaller than the DINA-HC model for all conditions except the condition of 7 attributes, 500 sample size, and 20 items according to Table 4.4 and 4.5, and smaller values indicate better model fit. The DINA model performs better than the DINA-HC model when the attributes are non-hierarchical. Table 4.4 and 4.5 summarize all 36 conditions along with the AIC and BIC values.

		Model Fit (AIC)					
		DINA-HC	DINA	DINA-HC	DINA	DINA-HC	DINA
		K=4	K=4	K=6	K=6	K=7	K=7
N=500	J=20	12260	11973	50073	48941	12638	12529
	J=40	23160	21892	24219	22560	24213	22937
N=2000	J=20	48912	47756	50073	48941	50391	49581
	J=40	92458	87239	96657	89847	96602	91202
N=5000	J=20	122253	119296	125084	122193	125912	123712
	J=40	231019	217962	241583	224598	241287	227622

Table 4.4: Model fit (AIC) for all conditions: non-hierarchical attributes

		DINA-HC	DINA	DINA-HC	DINA	DINA-HC	DINA
		Model Fit (BIC)					
		K=4	K=4	K=6	K=6	K=7	K=7
N=500	J=20	12458	12205	50347	49518	12912	13233
	J=40	23527	22292	24594	23163	24656	23809
N=2000	J=20	49175	48064	50347	49518	50756	50516
	J=40	92945	87771	97156	90648	97190	92361
N=5000	J=20	122560	119655	125404	122864	126336	124800
	J=40	231586	218581	242163	225530	241972	228971

Table 4.5: Model fit (BIC) for all conditions: non-hierarchical attributes

4.1.3 Parameter Estimates

For both the guessing and the slipping parameters, the 108000 bias values based on all the items from 100 replications under 36 conditions are calculated. An ANOVA test is conducted to investigate the performance of four factors (test length, number of attributes, sample size, and estimating model) on the bias and absolute bias under all conditions. The results show that the DINA-HC model provides smaller bias and absolute bias error for the slipping parameter estimates than does the DINA model. In terms of the guessing parameter, the DINA-HC and DINA model also are significantly different.

Guessing Parameter

Specifically, for the guessing parameter, in terms of the main effects, the results show that the sample size is the only significant effect on the bias at 0.05 level. The main effect of the sample size on the bias is statistically significant, given that ($F(2, 107980)=11.88$, $p<0.001$). The main effect of the test lengths on the bias is not statistically significant, given that ($F(1, 107980)=1.12$, $p=0.290$). The main effect of the number of attributes on the bias is not statistically significant such that ($F(2, 107980)=0.73$, $p=0.482$). The main effect of the estimating model on the bias is not statistically significant, given that (F(1, 107980)=2.62, p=0.106).

In terms of the interaction effects, there was significant interaction between the sample size and test lengths, (F(2, 107980)=12.53, p<0.001). There was an significant interaction between the number of attributes and sample size, (F(4, 107980)=3.59, p<0.001). There was a significant interaction between the number of attributes and the estimating model, F(2, 107980)=7.66, p<0.001). There was no significant interaction between the estimation model and the sample size, (F(2, 107980)=2.30, p=0.101). There was no significant interaction between the test lengths and the number of attributes, (F(2, 107980)=2.13, p=0.119). There was no significant interaction between the test lengths and the estimation model, (F(1, 107980)=0.02, p=0.886).

A Tukey post hoc test was conducted to further investigate the differences among the conditions and to compare the three different sample sizes for the main effect. The results show that the sample sizes 2000 and 500 differ significantly with adjusted p<0.001; sample sizes 5000 and 2000 differ significantly with adjusted p=0.020; sample size 5000 is not significantly different from sample size 500 with adjusted p=0.072. All other comparisons were not significant at the 0.05 level of significance.

The absolute bias makes more sense in this case, because the bias can be positive or negative. An ANOVA test also was conducted to examine the effects of four factors on absolute bias. The results are presented in Table 4.6.

The main and interaction effects of the four factors on the absolute bias are all statistically significant. Specifically, the main effect of the estimating model on the absolute bias is statistically significant such that (F(1, 107980)=284.51, p<0.001). There was a significant interaction between the sample size and the estimation model, (F(2, 107980)=8.53, p<0.001). There was a significant interaction between the test lengths and the estimation model, (F(1, 107980)=391.75, p<0.001). There was a

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SampleSize	2	6.44	3.219	5462.745	<.001
TestLength	1	0.66	0,656	1113.056	<.001
NumOfAtt	2	0.79	0.395	670.801	<.001
EstModel	1	0.17	0.168	284.507	<.001
SampleSize:TestLength	2	0.02	0.011	19.407	<.001
SampleSize:NumOfAtt	4	0.25	0.062	104.760	<.001
SampleSize:EstModel	2	0.01	0.005	8.533	<.001
TestLength:NumOfAtt	2	0.28	0.138	234.848	<.001
TestLength:EstModel	1	0.23	0.231	391.747	<.001
NumOfAtt:EstModel	2	0.08	0.038	64.214	<.001
Residuals	107980	63.62	0.001		

Table 4.6: ANOVA Table of Absolute Bias for Guessing Parameter

significant interaction between the number of attributes and the estimation model, (F(2, 107980)=64.21, p<0.001).

Taken together, these results suggest that the two different estimation models (DINA-HC and DINA), the number of attributes, and the test lengths have no effect on the bias value for the guessing parameter. However, the sample size does have an effect on the bias value. All four factors, including the estimation model, have an effect on the absolute bias value of the guessing parameter. The Tukey post hoc test shows that the DINA-HC model differed significantly from the DINA model at 0.05 level of significance; all other comparisons were significant. The DINA-HC model provides smaller absolute bias than does the DINA model.

Slipping Parameter

Similarly, for the slipping parameter, an ANOVA test was conducted on the bias and the absolute bias to examine the effects of all four factors.

In terms of the main effects, there are statistically significant differences of all four factors on the bias at the 0.05 level of significance, with the sample size (F(2, 107993)=19.59, p<0.001), test length (F(1, 107993)=11.81, p=0.001), number of attributes (F(2, 107993)=3.92, p=0.020), and estimating model (F(1, 107993)=5.75, p=0.016) respectively.

In terms of the interaction effects, there was a significant interaction between the sample size and test length: (F(2, 107980)=22.14, p<0.001). There was a significant interaction between the number of attributes and sample size: (F(4, 107980)=3.45, p=0.008). There was a significant interaction between the number of attributes and the estimating model: (F(2, 107980)=3.12, p=0.044). There was a significant interaction between the estimation model and the sample size: (F(2, 107980)=6.65, p=0.001). There was a significant interaction between test length and number of attributes: (F(2, 107980)=13.89, p<0.001). There was a significant interaction between test length and the estimation model: (F(1, 107980)=4.83, p=0.028).

A Tukey post hoc test was conducted to further investigate the difference among the conditions, and the results show that the DINA-HC model differed significantly from the DINA model at a 0.05 level of significance; the number of attributes 4 and 7 differed significantly with adjusted p = 0.030. However the number of attributes 4 and 6, 6 and 7 did not significantly differ, with adjusted p=0.964 and p=0.578 respectively. Most of the interaction comparisons were not significant.

Use of the absolute bias makes more sense here, because the bias will be transformed into a positive absolute value. An ANOVA test also was conducted to examine the effects of four factors on the absolute bias. The results are presented in Table 4.7.

The main and interaction effects of the four factors on the absolute bias are all statistically significant. Specifically, the main effect of the estimating model on the absolute bias is statistically significant, such that (F(1, 107980)=7.69, p=0.006). At the 0.05 level of significance there was a significant interaction between the sample size and the estimation model: (F(2, 107980)=3.54, p=0.029). There was a significant interaction between test length and the estimation model: (F(1, 107980)=10.74,

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
SampleSize	2	5.39	2.694	14284.22	<.001
TestLength	1	0.12	0.120	636.76	<.001
NumOfAtt	2	0.01	0.007	37.68	<.001
EstModel	1	0.00	0.001	7.69	0.0056
SampleSize:TestLength	2	0.06	0.031	162.84	<.001
SampleSize:NumOfAtt	4	0.01	0.003	17.32	<.001
SampleSize:EstModel	2	0.00	0.001	3.54	0.0289
TestLength:NumOfAtt	2	0.06	0.028	146.72	<.001
TestLength:EstModel	1	0.00	0.002	10.74	0.0010
NumOfAtt:EstModel	2	0.00	0.001	4.09	0.0167
Residuals	107980	20.36	0.000		

Table 4.7: ANOVA Table of Absolute Bias for Slipping Parameter

p=0.001). There was a significant interaction between number of attributes and the estimation model: (F(2, 107980)=4.09, p=0.017).

Taken together, these results suggest that the two different estimation models (DINA-HC and DINA), the sample size, the number of attributes, and the test lengths have effects on the bias value for the slipping parameter. All four factors, including the estimation model, have their effects on the absolute bias value for the slipping parameter. The Tukey post hoc test shows that the DINA-HC model differed significantly from the DINA model, with an adjusted p=0.006; all other comparisons were significant. The DINA-HC model provides smaller absolute bias than does the DINA model.

To summarize: the DINA-HC model provides smaller absolute bias value than does the DINA model for both the guessing and the slipping parameters. As the sample size increases, the absolute bias decreases; as the test length increases from 20 to 40, the absolute bias decreases; as the number of attributes increases from 4 to 6, the absolute bias decreases, the number of attributes increases from 6 to 7, the absolute bias increases. When switching from the DINA model to the DINA-HC model, the absolute bias will decrease as well.

Absolute Bias Error

Another important measure used to examine the DINA-HC model over the DINA model is the Absolute Bias Error (ABE). This is an alternative to the Mean Square Error (MSE). The absolute bias error over all the items for guessing parameter and under all conditions is calculated, summarized and plotted in the ensuing sections.

ABE for guessing parameter when J=20

In terms of the absolute bias error over 20 items for the guessing parameter as shown in Figure 4.1, the vertical axis value is set between 0 and 0.025. The horizontal axis represents the three sample sizes 500, 2000, and 5000) in an ascending order. It is clear that the DINA-HC model has smaller ABE than does the DINA model for K=4, K=6, and K=7. It seems that as the number of attributes increases, the absolute bias error increases as well. It also seems that the bias over all items increases as the sample size decreases for the DINA-HC model. It seems there is, however, a specific pattern for the DINA model.

		Absolute Bias Error		
		K=4	K=6	K=7
N=500	DINA-HC	0.00116	0.00464	0.00425
	DINA	0.00244	0.01457	0.02131
N=2000	DINA-HC	0.00114	0.00566	0.00406
	DINA	0.00222	0.01965	0.00869
N=5000	DINA-HC	0.00094	0.00215	0.00296
	DINA	0.00128	0.01576	0.00933

Table 4.8: Absolute bias error value of guessing parameter for J=20

ABE for guessing parameter when J=40

In terms of the absolute bias error over 40 items for the guessing parameter as seen in Figure 4.1, the vertical axis value is set between 0 and 0.006. The horizontal axis represents the three sample sizes 500, 2000, and 50 in an ascending order. It also is clear that the DINA-HC model has smaller ABE than does the DINA model for K=4, K=6, and K=7. The absolute bias error for K=4 is the smallest in this case, but the absolute bias errors for K=6 and K=7 are not very different. It also seems that the bias over all items increases as the sample size decreases. As the number of items increases, the bias for the guessing parameter decreases.

		Absolute Bias Error		
		K=4	K=6	K=7
N=500	DINA-HC	0.00251	0.00314	0.00286
	DINA	0.00221	0.00455	0.00573
N=2000	DINA-HC	0.00099	0.00112	0.00137
	DINA	0.00114	0.00178	0.00237
N=5000	DINA-HC	0.00059	0.00124	0.00082
	DINA	0.00070	0.00182	0.00135

Table 4.9: Absolute bias error value of guessing parameter for J=40

ABE for slipping parameter when J=20

In terms of the absolute bias error over 20 items for the slipping parameter as seen in Figure 4.2, the vertical axis value are is between 0 and 0.012. The slipping parameter has a smaller ABE than the guessing parameter in general. The horizontal axis represents the three sample sizes 500, 2000, and 5000 in an ascending order. It also is clear that the DINA-HC model has a smaller bias than does the DINA model for K=4, K=6, and K=7. The absolute bias error will increase as the number of attributes increases for both models. It also seems that the bias over all items decreases as the sample size increases.

		Absolute Bias Error		
		K=4	K=6	K=7
N=500	DINA-HC	0.00278	0.00217	0.00401
	DINA	0.00288	0.00355	0.01098
N=2000	DINA-HC	0.00123	0.00121	0.00186
	DINA	0.00131	0.00156	0.00319
N=5000	DINA-HC	0.00079	0.00085	0.00081
	DINA	0.00074	0.00197	0.00189

Table 4.10: Absolute bias error value of slipping parameter for J=20

ABE for slipping parameter when J=40

In terms of the absolute bias error over 40 items for the slipping parameter as seen in Figure 4.2, the vertical axis value is set between 0.0005 and 0.0025. The slipping parameter has a much smaller ABE than does the guessing parameter in this case, as compared to the absolute bias error over 20 items. The horizontal axis represents the three sample sizes 500, 2000, and 5000 in an ascending order. It seems that the DINA-HC model has a smaller bias than does the DINA model for K=4, K=6, and K=7. For the DINA model, the absolute bias error will increase as the number of attributes increases; for the DINA-HC model, there is no clear pattern. It also is clear that the bias over all items increases as the sample size decreases.

		Absolute Bias Error		
		K=4	K=6	K=7
N=500	DINA-HC	0.00204	0.00213	0.00197
	DINA	0.00196	0.00212	0.00241
N=2000	DINA-HC	0.00145	0.00114	0.00114
	DINA	0.00149	0.00115	0.00126
N=5000	DINA-HC	0.00072	0.00070	0.00067
	DINA	0.00073	0.00076	0.00080

Table 4.11: Absolute bias error value of slipping parameter for J=40

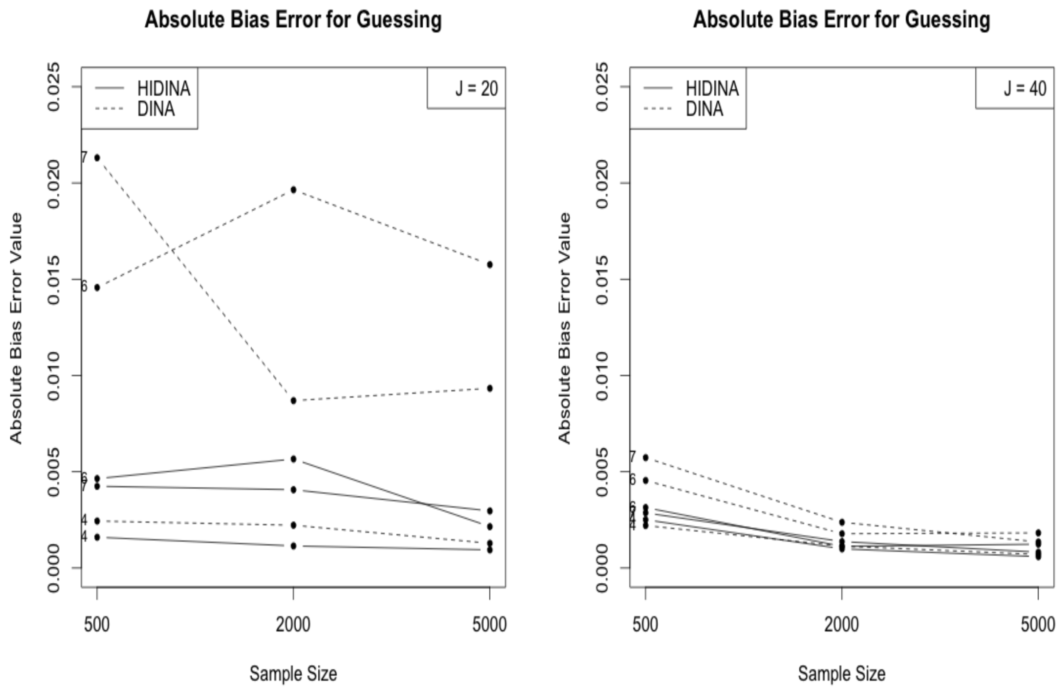

Figure 4.1: Absolute bias error for the guessing parameter

Both Figure 4.1 and 4.2 consist of two sub-figures that are plotting the absolute error values over all the items. These figures are presented in the same page for the convenience of comparing the two.

To summarize: for the ABE, for the guessing parameter, the condition of 5000 sample size, 4 attributes, and 40 items when estimated by the DINA-HC model has the lowest absolute bias error value, 0.00059. The condition of 500 sample size, 7 attributes, and 20 items when estimated by the DINA model has the largest absolute bias error value, 0.0213086. For the slipping parameter, the condition of 5000 sample size, 7 attributes, and 40 items when estimated by the DINA-HC has the lowest absolute bias error value, 0.00067, and the condition of 500 sample size, 7 attributes, and 20 items when estimated by the DINA model has the largest absolute bias error value, 0.01098.

In general, the DINA-HC model has a smaller absolute bias error for the param-

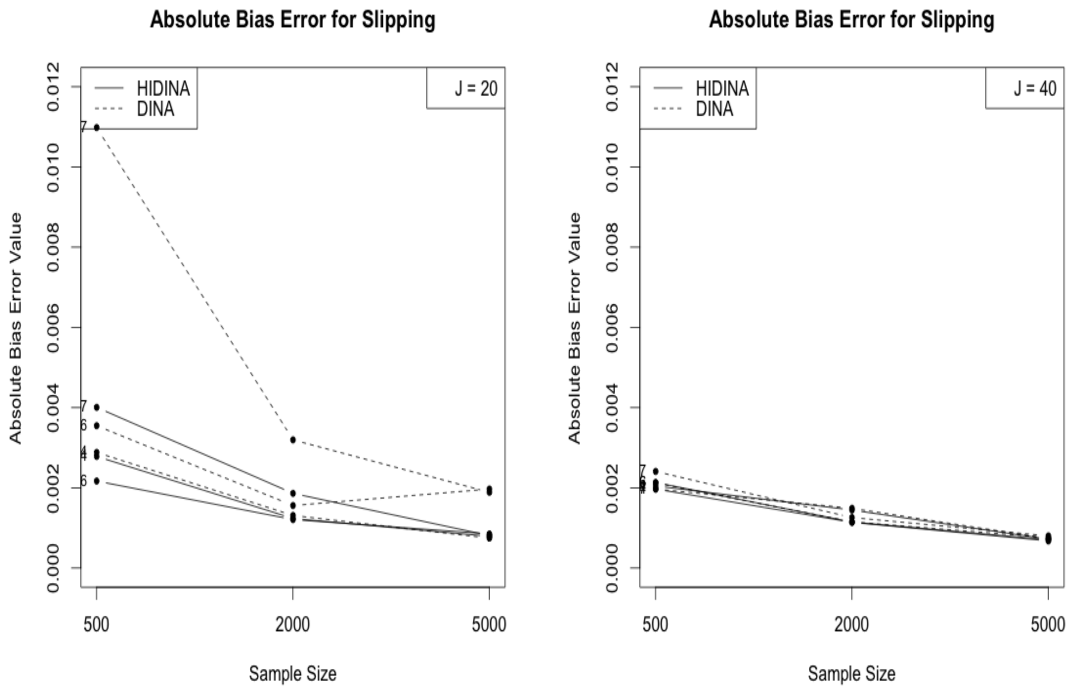

Figure 4.2: Absolute bias error for the slipping parameter

eters than does the DINA model. For the DINA-HC model, the bias will increase as the sample size decreases, thus sample size matters when it comes to the bias; also, the larger the sample size, the smaller the absolute bias error. As the number of attributes increases from 4 to 7, the absolute bias error will decrease. Also, the number of items has an effect on the absolute bias error: it will decrease as the number of items increase from J=20 to J=40. For the DINA model, the absolute bias error is always larger than it is for the DINA-HC model. The absolute bias error will increase as the sample size increases. The absolute bias error increases as the number of attributes increases and number of items decreases, so they both have effects on the value of absolute bias error. The absolute bias error for the slipping parameter normally is smaller than the guessing parameter.

To summarize: for the parameter estimates, the DINA-HC should be preferred over the DINA model when the attributes are convergent for the following reasons:

the absolute bias value is smaller for both the guessing and the slipping parameter; the DINA-HC model has a smaller ABE than does the DINA model for both the guessing and the slipping parameters.

4.1.4 The Classification Accuracy

An ANOVA test is conducted to investigate the performance of four factors (number of items, number of attributes, sample size, and the estimating model) in the relation to classification accuracy under all conditions. The results are presented in Table 4.12.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
NumOfAtt	2	25.14	12.57	29344.6	<.001
TestLength	1	24.76	24.76	57804.5	<.001
SampleSize	2	0.24	0.12	281.7	<.001
EstModel	1	13.73	13.73	32037.7	<.001
NumOfAtt:TestLength	2	5.33	2.66	6216.6	<.001
NumOfAtt:SampleSize	4	0.75	0.19	436.6	<.001
NumOfAtt:EstModel	2	2.03	1.02	2374.9	<.001
TestLength:SampleSize	2	0.22	0.11	255.8	<.001
TestLength:EstModel	1	0.79	0.79	1842.0	<.001
SampleSize:EstModel	2	0.01	0.01	14.6	<.001
Residuals	3580	20.36	0.000		

Table 4.12: ANOVA Table of Classification Accuracy

In terms of the main effects, the results show that there are significant effects of four factors on the classification accuracy rate at a 0.05 level. The main effect of the number of attributes on classification accuracy is statistically significant, such that as the number of attributes increase, the classification accuracy decreases: ($F(2, 3580)=29344.6$, $p<0.001$). The main effect of test length on classification accuracy is statistically significant, such that as the test length increase from 20 to 40 the classification accuracy increases: ($F(1, 3580)=57804.5$, $p<0.001$). The main effect of sample size on classification accuracy is statistically significant such that as the sample size increases, the classification accuracy will also increase: $(F(2, 3580)=281.7, p<0.001)$. The main effect of the estimating model on the classification accuracy is statistically significant, such that as the DINA switches to DINA-HC, the classification accuracy increases: $(F(1, 3580)=32037.7, p<0.001)$.

In terms of the interaction effects, there was a significant interaction between the number of attributes, and the test length: $(F(2, 3580)=6216.6, p<0.001)$. There was a significant interaction between the number of attributes and sample size: $(F(4, 3580)=436.6, p<0.001)$. There was a significant interaction between the number of attributes and the estimating model: $(F(2, 3580)=2374.9, p<0.001)$. There was a significant interaction between the test length and sample size: $(F(2, 3580)=255.8, p<0.001)$. There was a significant interaction between the test length and the estimating model: $(F(1, 3580)=1842, p<0.001)$. There was a significant interaction between the estimating model and the sample size: $(F(2, 3580)=14.6, p<0.001)$.

To further investigate differences among the conditions, a Tukey post hoc test was conducted. The results show that when the attribute is 4, there is no interaction between the sample sizes 500 and 2000 with an adjusted $p=0.602$. There is no interaction between the sample sizes 2000 and 5000 with an adjusted $p=0.108$ neither. When the items $J=40$, there is no interaction between the sample sizes 500 and 5000 with an adjusted $p=0.982$. When the estimating model is DINA-HC, there is no interaction between the sample sizes 2000 and 5000 with an adjusted $p=0.966$. All other comparisons show that there are statistically significant differences, with an adjusted $p<0.001$.

The attribute profile pattern classification accuracy is calculated and computed as the average value over 100 replications for all the conditions as may be seen in Table 4.13 and 4.14. The R plot is presented in the Figure 4.3. In terms of the attribute

		Classification Accuracy		
		K=4	K=6	K=7
N=500	DINA-HC	0.892900	0.76116	0.56528
	DINA	0.80760	0.53830	0.40608
N=2000	DINA-HC	0.857980	0.80511	0.62081
	DINA	0.79254	0.54124	0.47072
N=5000	DINA-HC	0.874212	0.77638	0.66485
	DINA	0.80286	0.56218	0.51921

Table 4.13: Attribute profile pattern classification accuracy for J=20

		Classification Accuracy		
		K=4	K=6	K=7
N=500	DINA-HC	0.910100	0.92978	0.83076
	DINA	0.854380	0.80406	0.73416
N=2000	DINA-HC	0.942315	0.93050	0.82730
	DINA	0.883365	0.80111	0.72917
N=5000	DINA-HC	0.934862	0.88665	0.85283
	DINA	0.880818	0.75305	0.76004

Table 4.14: Attribute profile pattern classification accuracy for J=40

profile pattern classification accuracy for J=20 and J=40, the DINA-HC model has higher classification accuracy value than does the DINA model for K=4, K=6, and K=7. The classification accuracy value will decrease as the number of attributes increases from K=4 to K=7. It seems the classification accuracy is invariant of sample size. The largest classification accuracy value is 0.942315 when the sample size is 2000, the number of attributes is 4, the number of items is 40, and the estimating model is DINA-HC. The smallest classification accuracy value is 0.53830 when the sample size is 500, the number of attributes is 6, the number of items is 20, and the estimating model is DINA.

In sum, the DINA-HC model achieves a higher classification accuracy rate than does the DINA model when the attributes display a convergent hierarchy.

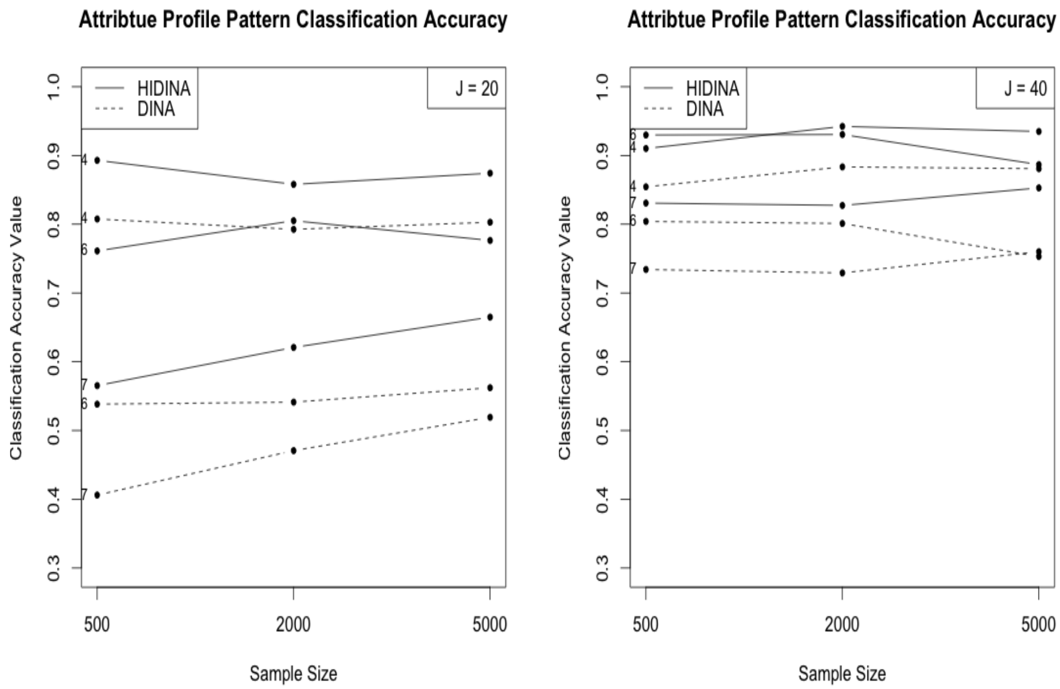

Figure 4.3: Attribute profile pattern classification accuracy

4.2 Real Data Results

This section presents the results gained from the ECPE data. The Examination for Certificate of Proficiency in English (ECPE) is a test whose attributes display a linear hierarchy. It was used to examine both the DINA-HL and the DINA model.

4.2.1 The ECPE Data Results

Both AIC and BIC were used as the model fit criterion. The log-likelihood value is -42853 for the DINA-HL approach and -42841 for the DINA model. The number of parameters is 59 for the DINA-HL approach and 63 for the DINA model. The AIC value for the DINA-HL approach is slightly larger than the value for the DINA model, and the BIC value for the DINA-HL approach is slightly smaller than the value for DINA model as the following table shows.

Model Fit	DINA-HL	DINA
AIC	85823	85809
BIC	86176	86186

Table 4.15: Model fit statistics

A likelihood-ratio test was conducted to determine whether the DINA-HL model is significantly differently from the DINA model. The DINA-HL (the null model) is a special case of the DINA (the alternative model). The number of skill class parameters for the DINA is 7 and the number of skill class parameters for the DINA-HL is 3. The alpha level was set at 0.05. The DINA-HL model is significantly different from the DINA model: χ^2 (2, N=2922)= 18.733, p<0.001. The DINA model should be preferred when the attributes have a linear hierarchy.

The item parameter estimates derived from the two models are shown in the following table. The guessing estimates from both models are rather large for almost all of the items except for items 12, 20, 22, 24, and 27. The problem sets are multiple choice, and the items in the multiple choice problem are linearly dependent, so the guessing parameter estimates are larger than the usual range. The slipping parameter estimates fall within the usual range.

Items	DINA(g)	DINA-HL(g)	DINA(s)	DINA-HL(s))
Item 1	0.70553	0.71194	0.078620	0.095823
Item 2	0.73776	0.74690	0.095239	0.107830
Item 3	0.43799	0.43873	0.265659	0.262771
Item 4	0.47876	0.47573	0.163020	0.163388
Item 5	0.76203	0.76114	0.040589	0.041234
Item 6	0.71600	0.71464	0.066903	0.067392
Item 7	0.54410	0.54621	0.085002	0.082712
Item 8	0.81608	0.82388	0.035982	0.046914
Item 9	0.53354	0.53392	0.200121	0.201889
Item 10	0.48586	0.49736	0.163093	0.160804
Item 11	0.55599	0.55858	0.098747	0.097322
Item 12	0.19437	0.19928	0.304976	0.304203
Item 13	0.63476	0.64349	0.121453	0.120693
Item 14	0.51864	0.52598	0.211604	0.208163
Item 15	0.74756	0.74580	0.040323	0.040560
Item 16	0.54898	0.55132	0.125487	0.124029
Item 17	0.81845	0.81366	0.054602	0.060795
Item 18	0.72820	0.72591	0.086387	0.086164
Item 19	0.47072	0.46922	0.150805	0.152159
Item 20	0.23859	0.24211	0.295337	0.293436
Item 21	0.62161	0.62348	0.096767	0.095338
Item 22	0.31860	0.31685	0.188670	0.190546
Item 23	0.66138	0.66790	0.067075	0.081358
Item 24	0.33987	0.35366	0.307981	0.330462
Item 25	0.51352	0.51743	0.271748	0.266840
Item 26	0.55391	0.55267	0.211364	0.212029
Item 27	0.26813	0.27911	0.368851	0.365456
Item 28	0.65729	0.65578	0.086418	0.087054

Table 4.16: The item parameter estimates

Chapter 5

Discussion

5.1 Summary of the Findings

This study utilized a convergent attribute hierarchy in the DINA model to analyze data having hierarchical attributes. The model fit, item fit, bias, absolute bias, absolute bias error, and attribute profile pattern classification accuracy of two different models (DINA-HC and DINA) were examined and compared across 100 replications under 36 conditions based upon the simulated data. The results indicate that the DINA-HC model has smaller AIC and BIC values for all 36 conditions than the DINA model, where smaller values indicate a better model fit. In terms of the bias for the guessing parameter, there is a significant effect of sample size on the bias. In terms of the absolute bias there are significant effects of sample size, test length, number of attributes, and the estimating models. There also are significant effects of all four factors on the bias and absolute bias of the slipping parameter. As the sample size or test length increases, the absolute bias will decrease. As the number of attributes increases, the absolute bias will increase as well. When the estimating model is DINA-HC, the absolute bias will be smaller than when it is the DINA. The results from the attribute profile pattern classification accuracy show that there are significant effects of all four factors on the classification accuracy. As the number of attributes and the sample size increase, the classification accuracy rate also will increase. When the estimating model is DINA-HC, the classification accuracy rate is significantly higher than it is with the DINA model.

Based on the results gained from the ECPE data, there is a statistically significant difference between the DINA-HL and DINA model. Most of the guessing parameter estimates are higher than expected. It seems the results are sensitive to the number of attributes and the types of attribute hierarches present.

In sum, the DINA having convergent attribute hierarchy (DINA-HC) is a better model choice, has less biased parameter estimates and does higher classification accuracy than the DINA model, and thus it should be used when the hierarchical attributes are present. When the attributes are non-hierarchical, the DINA model has a better model fit than does the DINA-HL.

5.2 Implications of the Findings

In practice, there will be situations when the attributes are hierarchical. Specifically, in the Mathematics, Science, and English areas, the mastery of higher level attributes sometimes requires the mastery of certain lower level attributes. The DINA-HC model can be applied to situations where cognitive skills are hierarchical. More importantly, it offers more diagnostic power and can potentially provide more detailed information on the mastery level of students. By designing items based on a particular hierarchy, the diagnostic information of the items can help instructors and thus to locate the learning stage of the students and thereby design more efficient curricula based on that stage. Another practical outcome of this finding is that the DINA-HC model can improve computing efficiency by reducing the size of the skill space.

This study has examined the performance of DINA-HC and the DINA model under different conditions. It represents an important addition to the study carried out by Su(2013) for using the convergent attribute hierarchy as its example. This study also is an addition to the DINA mode (Junker & Sijtsma, 2001). Although some previous studies (Templin et al., 2010, Leighton et al., 2004b) involved attribute hierarchies, none of them examined the performance of the DINA-HC model under different conditions. More detailed analysis of the tests could provide valuable information for instructors and policy makers on the topic of the learning progressions of students, and help them to understand not just the existence of certain attribute hierarchies within the Mathematics and Science domain, but also other concepts. It is believed that the method and analysis used in this study can be used to interpret many other subjects' data. In the future, if a test can be developed based on a specific hierarchical structure, there will be more information on students' learning paths and stages. Instructors can then design specific curricula to target their students' particular needs.

5.3 Limitations of the Findings

The Q-matrix of the ECPE data might be misspecified, because generally it is subjective to construct a Q-matrix. The attribute hierarchies also can be constructed in multiple ways, because different students approach the problems by different paths, thus there is no such thing as an absolute correct Q-matrix or attribute hierarchy. Although the focus of this study has not been on the construction of Q-matrix, it is possible that its parameter estimates have been affected by the Q-matrix. In addition there are only three attributes for the real data, so the number of attribute profiles has been reduced to 4 from the original 8, which is not a big difference. Unlike the simulation results, the DINA performs better than the DINA with a linear hierarchy (DINA-HL) for the real data. If there were more attributes in the data, the results might provide better diagnostic information.

Another limitation is the number of conditions in the simulation study. It is ideal to include as many conditions as possible, so that more information regarding other factors will be extracted. Different hierarchical structures probably would have different estimation results, so the results of this study can not be generalized to all possible hierarchical structures. The study uses only the convergent attribute hierarchy. Similar results also are found in Su (2013) for the linear and unstructured attribute hierarchies. The coverage of attribute hierarchies may not be enough to provide a comprehensive understanding of the behaviors of different hierarchies under different situations. The advantage of the DINA-HC may not be utilized when the number of attributes is small.

5.4 Future Research Suggestions

As mentioned in the previous section, this study uses only the convergent attribute hierarchy in simulation analysis. More reliable results would be provided if all four types of attribute hierarchies were included, examined, and compared. The convergent attribute hierarchy has many different types of structures. Depending on certain items, a specific hierarchical structure can be developed to analyze the data. Other models such as the GDM, NIDA, and NIDO having particular attribute hierarches can also be investigated in the future. As the popularity of online assessment increases, the amount of data also will increase. It will be more informative if more attributes are defined in a test. This study has used the EM-algorithm to estimate the guessing and slipping parameters; the procedure is based on de la Torre (2009) in which the parameters are optimized in a maximum likelihood fashion. The Markov chain Monte Carlo (MCMC), as an alternative method, can also be used in the future. The default values for the guessing and slipping parameters have been set at 0.2, but it would be interesting to see what results were obtained if other default values were set and compared with this study.

Bibliography

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.

Battista, M. T. (2004). Applying cognition-based assessment to elementary school students' development of understanding of area and volume measurement. *Mathematical Thinking and Learning*, 6(2):185–204.

Buck, G. & Tatsuoka, 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.

Chai, T. & Draxler, R. R. (2014). Root mean square error (rmse) or mean absolute error (MAE)?- Arguments against avoiding rmse in the literature. *Geoscientific Model Development*, 7(3):1247–1250.

de la Torre, J. (2009a). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33(3):163–183.

de la Torre, J. (2009b). Dina model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34(1):115–130.

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

DeCarlo, L. T. (2010). On the analysis of fraction subtraction data: The dina model, classification, latent class sizes, and the q-matrix. *Applied Psychological Measurement*.

DiBello, L. V., Roussos, L. A., & Stout, W. (2006). 31a review of cognitively diagnostic assessment and a summary of psychometric models. *Handbook of Statistics*, 26:979–1030.

French, G. A. & Dodd, B. G. (1998). Parameter recovery for the rating scale model using parscale. *Journal of Outcome Measurement*, 3(2):176–199.

Gierl, M. J. (2007). Making diagnostic inferences about cognitive attributes using the rule-space model and attribute hierarchy method. *Journal of Educational Measurement*, 44(4):325–340.

Gierl, M. J., Cui, Y., & Zhou, J. (2009). Reliability and attribute-based scoring in cognitive diagnostic assessment. *Journal of Educational Measurement*, 46(3):293–313.

Hartz, S. (2002). A bayesian framework for the unified model for assessing cognitive abilities: Blending theory with practicality. *University of Illinois at Urbana-Champaign: Unpublished doctoral dissertation*.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74(2):191–210.

Hudson, H. & Rottmann, R. M. (1981). Correlation between performance in physics and prior mathematics knowledge. *Journal of Research in Science Teaching*, 18(4):291–294.

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation*, 15(3):n3.

Junker, B. W. & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25(3):258–272.

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

Kuhn, D. (2001a). How do people know? *Psychological Science*, 12(1):1–8.

Kuhn, D. (2001b). Why development does (and does not) occur: Evidence from the domain of inductive reasoning. *Mechanisms of Cognitive Development: Behavioral and Neural Perspectives*, pages 221–249.

Leighton, J., Gierl, M., & Hunka, S. (2004a). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement*, 41(5):205–236.

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

Liu, Y., Douglas, J. A., & Henson, R. A. (2009). Testing person fit in cognitive diagnosis. *Applied Psychological Measurement*, 33(8):579–598.

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

Park, Y. S. & Lee, Y.-S. (2014). An extension of the dina model using covariates examining factors affecting response probability and latent classification. *Applied Psychological Measurement*, 38(5):376–390.

Robitzsch, A., Kiefer, T., George, A. C., & Ali, U. (2014). CDM: Cognitive Diagnosis Modeling. R package version 4.1. http://CRAN.R-project.org/package=CDM.

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

Rupp, A. A. & Mislevy, R. J. (2007). Cognitive foundations of structured item response models. *Cognitive Diagnostic Assessment for Education: Theories and Applications*, pages 205–240.

Rupp, A. A. & Templin, J. (2008a). The effects of q-matrix misspecification on parameter estimates and classification accuracy in the dina model. *Educational and Psychological Measurement*, 68(1):78–96.

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

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464.

Shepherd, D. (2005). *Negative masking and the units problem in audition*. PhD thesis, ResearchSpace@ Auckland.

Sternberg, R. J. & Ben-Zeev, T. (1996). *The Nature of Mathematical Thinking*. Routledge.

Su, Y. (2013). Cognitive diagnostic analysis using hierarchically structured skills. *University of Iowa: Unpublished doctoral dissertation.*

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4):345–354.

Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. *Diagnostic Monitoring of Skill and Knowledge Acquisition*, pages 453–488.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. *Cognitively Diagnostic Assessment*, pages 327–359.

Team, R. C. (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012.

Templin, J. (2006). CDM: Cognitive Diagnosis Modeling with Mplus [computer software].

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

Templin, J. et al. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. Guilford Press.

Templin, J., Henson, R., & Douglas, J. (2007). General theory and estimation of cognitive diagnosis models: Using mplus to derive model estimates. *Unpublished manuscript*.

Templin, J. L. & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11(3):287.

von Davier, M. (2005). A general diagnostic model applied to language testing data. *ETS Research Report Series*, 2005(2):i–35.

von Davier, M. (2014). The dina model as a constrained general diagnostic model: Two variants of a model equivalency. *British Journal of Mathematical and Statistical Psychology*, 67(1):49–71.

Vosniadou, S. & Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, 24(4):535–585.

Appendix A

Appendices

A.1 Q-matrix from Simulation Study

Item	skill 1	skill 2	skill 3	skill 4
Item 1	1	0	0	0
Item 2	0	1	0	0
Item 3	0	0	1	0
Item 4	0	0	0	1
Item 5	1	1	0	0
Item 6	1	0	1	0
Item 7	1	0	0	1
Item 8	0	1	1	0
Item 9	0	1	0	1
Item 10	0	0	1	1
Item 11	1	1	1	0
Item 12	1	0	1	1
Item 13	1	1	0	1
Item 14	0	1	1	1
Item 15	1	1	1	1
Item 16	1	0	0	0
Item 17	0	1	0	0
Item 18	0	0	1	0
Item 19	0	0	0	1
Item 20	1	1	0	0

Table A.1: The Q-matrix with 20 items and 4 skills

Item	skill 1	skill 2	skill 3	skill 4	skill 5	skill 6
Item 1	1	0	0	0	0	0
Item 2	0	1	0	0	0	0
Item 3	0	0	1	0	0	0
Item 4	0	0	0	1	0	0
Item 5	0	0	0	0	1	0
Item 6	0	0	0	0	0	1
Item 7	1	1	0	0	0	0
Item 8	1	0	1	0	0	0
Item 9	1	0	0	1	0	0
Item 10	1	0	0	0	1	0
Item 11	1	0	0	0	0	1
Item 12	0	0	1	1	0	0
Item 13	0	0	1	0	1	0
Item 14	0	0	1	0	0	1
Item 15	0	0	0	1	1	0
Item 16	0	0	0	1	0	1
Item 17	1	1	1	0	0	0
Item 18	1	1	0	1	0	0
Item 19	1	1	0	0	1	0
Item 20	0	0	1	1	0	1

Table A.2: The Q-matrix with 20 items and 6 skills

Item	skill 1	skill 2	skill 3	skill 4	skill 5	skill 6	skill 7
Item 1	1	0	0	0	0	0	0
Item 2	0	1	0	0	0	0	0
Item 3	0	0	1	0	0	0	0
Item 4	0	0	0	1	0	0	0
Item 5	0	0	0	0	1	0	0
Item 6	0	0	0	0	0	1	0
Item 7	0	0	0	0	0	0	1
Item 8	1	1	0	0	0	0	0
Item 9	1	0	1	0	0	0	0
Item 10	1	0	0	1	0	0	0
Item 11	1	0	0	0	1	0	0
Item 12	1	0	0	0	0	1	0
Item 13	1	0	0	0	0	0	1
Item 14	0	0	1	1	0	0	0
Item 15	0	0	0	1	1	0	0
Item 16	0	0	0	1	0	1	0
Item 17	1	1	0	1	0	0	0
Item 18	1	1	0	0	1	0	0
Item 19	0	0	1	1	0	1	0
Item 20	0	0	1	0	0	1	1

Table A.3: The Q-matrix with 20 items and 7 skills

Item	skill 1	skill 2	skill 3	skill 4
Item 1	1	0	0	0
Item 2	0	1	0	0
Item 3	0	0	1	0
Item 4	0	0	0	1
Item 5	1	1	0	0
Item 6	1	0	1	0
Item 7	1	0	0	1
Item 8	0	1	1	0
Item 9	0	1	0	1
Item 10	0	0	1	1
Item 11	1	1	1	0
Item 12	1	0	1	1
Item 13	1	1	0	1
Item 14	0	1	1	1
Item 15	1	1	1	1
Item 16	1	0	0	0
Item 17	0	1	0	0
Item 18	0	0	1	0
Item 19	0	0	0	1
Item 20	1	1	0	0
Item 21	1	0	1	0
Item 22	1	0	0	1
Item 23	0	1	1	0
Item 24	0	1	0	1
Item 25	0	0	1	1
Item 26	1	1	1	0
Item 27	1	0	1	1
Item 28	1	1	0	1
Item 29	0	1	1	1
Item 30	1	1	1	1
Item 31	1	0	0	0
Item 32	0	1	0	0
Item 33	0	0	1	0
Item 34	0	0	0	1
Item 35	1	1	0	0
Item 36	1	0	1	0
Item 37	1	0	0	1
Item 38	0	1	1	0
Item 39	0	1	0	1
Item 40	0	0	1	1

Table A.4: The Q-matrix with 40 items and 4 skills

Item	skill 1	skill 2	skill 3	skill 4	skill 5	skill 6
Item 1	1	0	0	0	0	0
Item 2	0	1	0	0	0	0
Item 3	0	0	1	0	0	0
Item 4	0	0	0	1	0	0
Item 5	0	0	0	0	1	0
Item 6	0	0	0	0	0	1
Item 7	1	1	0	0	0	0
Item 8	1	0	1	0	0	0
Item 9	1	0	0	1	0	0
Item 10	1	0	0	0	1	0
Item 11	1	0	0	0	0	1
Item 12	0	0	1	1	0	0
Item 13	0	0	1	0	1	0
Item 14	0	0	1	0	0	1
Item 15	0	0	0	1	1	0
Item 16	0	0	0	1	0	1
Item 17	1	1	1	0	0	0
Item 18	1	1	0	1	0	0
Item 19	1	1	0	0	1	0
Item 20	0	0	1	1	0	1
Item 21	1	0	0	0	0	0
Item 22	0	1	0	0	0	0
Item 23	0	0	1	0	0	0
Item 24	0	0	0	1	0	0
Item 25	0	0	0	0	1	0
Item 26	0	0	0	0	0	1
Item 27	1	1	0	0	0	0
Item 28	1	0	1	0	0	0
Item 29	1	0	0	1	0	0
Item 30	1	0	0	0	1	0
Item 31	1	0	0	0	0	1
Item 32	0	0	1	1	0	0
Item 33	0	0	1	0	1	0
Item 34	0	0	1	0	0	1
Item 35	0	0	0	1	1	0
Item 36	0	0	0	1	0	1
Item 37	1	1	1	0	0	0
Item 38	1	1	0	1	0	0
Item 39	1	1	0	0	1	0
Item 40	0	0	1	1	0	1

Table A.5: The Q-matrix with 40 items and 6 skills

Item	skill 1	skill 2	skill 3	skill 4	skill 5	skill 6	skill 7
Item 1	1	0	0	0	0	0	0
Item 2	0	1	0	0	0	0	0
Item 3	0	0	1	0	0	0	0
Item 4	0	0	0	1	0	0	0
Item 5	0	0	0	0	1	0	0
Item 6	0	0	0	0	0	1	0
Item 7	0	0	0	0	0	0	1
Item 8	1	1	0	0	0	0	0
Item 9	1	0	1	0	0	0	0
Item 10	1	0	0	1	0	0	0
Item 11	1	0	0	0	1	0	0
Item 12	1	0	0	0	0	1	0
Item 13	1	0	0	0	0	0	1
Item 14	0	0	1	1	0	0	0
Item 15	0	0	0	1	1	0	0
Item 16	0	0	0	1	0	1	0
Item 17	1	1	0	1	0	0	0
Item 18	1	1	0	0	1	0	0
Item 19	0	0	1	1	0	1	0
Item 20	0	0	1	0	0	1	1
Item 21	1	0	0	0	0	0	0
Item 22	0	1	0	0	0	0	0
Item 23	0	0	1	0	0	0	0
Item 24	0	0	0	1	0	0	0
Item 25	0	0	0	0	1	0	0
Item 26	0	0	0	0	0	1	0
Item 27	0	0	0	0	0	0	1
Item 28	1	1	0	0	0	0	0
Item 29	1	0	1	0	0	0	0
Item 30	1	0	0	1	0	0	0
Item 31	1	0	0	0	1	0	0
Item 32	1	0	0	0	0	1	0
Item 33	1	0	0	0	0	0	1
Item 34	0	0	1	1	0	0	0
Item 35	0	0	0	1	1	0	0
Item 36	0	0	0	1	0	1	0
Item 37	1	1	0	1	0	0	0
Item 38	1	1	0	0	1	0	0
Item 39	0	0	1	1	0	1	0
Item 40	0	0	1	0	0	1	1

Table A.6: The Q-matrix with 40 items and 7 skills

Appendix B

Appendices

B.1 Convergent attribute profile pattern for the 4, 6 and 7 attributes

Attribute Profile Pattern	A1	A2	A3	A4
1	0	0	0	0
2	1	0	0	0
3	1	1	0	0
4	1	0	1	0
5	1	1	1	0
6	1	1	0	1
7	1	0	1	1
8	1	1	1	1

Table B.1: Possible attribute profiles for 4 attributes

Attribute Profile Pattern	A1	A2	A3	A4	A5	A6
1	0	0	0	0	0	0
2	1	0	0	0	0	0
3	1	1	0	0	0	0
4	1	1	1	0	0	0
5	1	1	1	1	0	0
6	1	1	1	0	1	0
7	1	1	1	1	1	0
8	1	1	1	1	0	1
9	1	1	1	0	1	1
10	1	1	1	1	1	1

Table B.2: Possible attribute profiles for 6 attributes

Attribute Profile Pattern	A1	A2	A3	A4	A5	A6	A7
1	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0
3	1	1	0	0	0	0	0
4	1	0	1	0	0	0	0
5	1	1	1	0	0	0	0
6	1	1	0	1	0	0	0
7	1	0	1	1	0	0	0
8	1	1	1	1	0	0	0
9	1	1	0	1	1	0	0
10	1	0	1	1	1	0	0
11	1	1	1	1	1	0	0
12	1	1	0	1	0	1	0
13	1	0	1	1	0	1	0
14	1	1	1	1	0	1	0
15	1	1	0	1	1	1	0
16	1	0	1	1	1	1	0
17	1	1	1	1	1	1	0
18	1	1	0	1	1	0	1
19	1	0	1	1	1	0	1
20	1	1	1	1	1	0	1
21	1	1	0	1	0	1	1
22	1	0	1	1	0	1	1
23	1	1	1	1	0	1	1
24	1	1	0	1	1	1	1
25	1	0	1	1	1	1	1
26	1	1	1	1	1	1	1

Table B.3: Possible attribute profiles for 7 attributes