

What CDM Can Tell About What Students Have Learned: An Analysis of TIMSS Eighth Grade Mathematics

Kyong Mi Choi
University of Iowa, USA
Young-Sun Lee
Columbia University, USA
Yoon Soo Park
University of Illinois, USA

•Received 15 December 2013•Revised 2 March 2014 •Accepted 9 June 2014

International trended assessments have long attempted to provide instructional information to educational researchers and classroom teachers. Studies have shown that traditional methods of item analysis have not provided specific information that can be directly applicable to improve student performance. To this end, cognitive diagnosis models (CDMs) were developed to identify the presence or the absence of multiple fine-grained skills or attributes. This study reanalyzed the TIMSS 2003 8th grade mathematics assessment by applying a cognitive diagnosis modeling approach to provide useful information about test takers. Specifically, the deterministic, inputs, noisy, “and” gate (DINA; Junker & Sijtsma, 2001) model was used to provide more valuable diagnostic information about how well students perform on underlying skills and cognitive processes necessary to answer mathematics problems. Results showed that the interpretable benefits of using the DINA discrimination index provided additional information about comparative differences between the U.S. and Korea.

Keywords: cognitive diagnosis model, DINA, TIMSS, 8th grade mathematics

INTRODUCTION

Educational researchers have been seeking to provide useful information for teachers through assessments, however criticism has been made that assessments are hardly useful, especially TIMSS was criticized as being barely useful, hardly applicable in existing school programs, and not considerate of the cultural context (Holliday & Holliday, 2003; Wang, 2001). Ferrini-Mundy and Schmidt (2005) advised the necessity for mathematics educators, mathematicians, measurement experts, and educational statisticians to collaborate in research projects to recognize the potential values of additional conceptual discussions and secondary analyses that are directly applicable to the existing school systems.

Correspondence: Kyong Mi Choi,
Department of Teaching and Learning, College of Education, University of Iowa, N 291
Lindquist Center, 319-335-5118, USA.
E-mail: kyongmi-choi@uiowa.edu
doi: 10.12973/eurasia.2015.1421a

In the need of readily useful evidence to classroom instruction, Cognitive Diagnostic Models (CDMs) have been developed to provide additional information to researchers and educators on fine-grained cognitive skills or attributes that are required to solve a particular item, which can consequently be applied to various instructional practices. Previous studies define an “attribute” as posited knowledge and thinking skill (Tatsuoka, Corter, & Tatsuoka, 2004), a description of the procedures, skills, processes, strategies, and knowledge a student must possess to solve a test item (Dogan & Tatsuoka, 2008), or a generic term for a skill or knowledge state (de la Torre & Douglas, 2004). As previously mentioned, we define the term *attribute* as ‘a mathematical skill or content knowledge that is required to solve a test item.’

Conventional test theories such as the classical test theory (CTT) and unidimensional item response theory (IRT) have been providing valuable information such as single overall scores that indicate students’ relative positions on an ability scale; however, they do not provide diagnostic information on the mastery or non-mastery of skills. Furthermore, the notion of identifying if an examinee has mastered specific skills required to solve an item differs from the objective of traditional measurement models, which posits an examinee on a continuous scale to gauge relative achievement (i.e., a continuous latent score of the examinee’s ability). That is, an advantage of CDMs is its ability to identify the strength and weakness in a sets of fine-grained skills (or attributes) when inferring skill mastery profiles of examinees can be difficult when traditional methods such as CTT and IRT are used.

The purpose of this study is to introduce a new methodology (CDM) to assess and evaluate student achievement in an informative way that stakeholders – teachers and students – can readily apply the result of assessment to improve teaching and learning of mathematics. When teachers have better understanding what students learned or did not learn, it would be helpful to plan lesson for the future. Through this study, we conduct a secondary comparative analysis of the TIMSS 2003 mathematics items between the U.S. and South Korea (henceforth, Korea) using the CDM framework. The results from this study can offer a unique perspective on what international comparison study can offer other than average achievement scores and proportion corrects on items, domains, or sections of mathematics. Instead, we will compare what students attained fine-grained attributes, rather than simply “correctly answered” items because answering the items correctly does not always mean that the examinee attained all the knowledge required to solve the items. Similarly, answering incorrectly could mean that the examinee made a mistake such as marking a wrong bubble even though all required knowledge is attained.

State of the literature

- International comparison studies such as TIMSS have a potential value to provide useful information applicable in existing school programs through secondary analyses.
- Cognitive Diagnostic Models (CDMs) have been developed to provide additional information to researchers and educators on fine-grained cognitive skills or attributes that are required to solve a particular item, which can consequently be applied to various instructional practices.
- Korean students consistently perform well in various international assessments, especially in mathematics.

Contribution of this paper to the literature

- Despite of Korean students’ high performance in international assessments, their high success to find correct answers is likely to rely on guessing works. It is critical for educators to understand this concept, because measuring student mastery that relies only on their achievement scores could yield to take guessing as understanding.
- An area of improvement for Korean students is Probability and Statistics.
- There is no significant difference between American and Korean students in Probability and Statistics and an area of Measurement.

BACKGROUND

This section is devoted to review the psychometric model – CDM, more specifically, DINA model. Brief description of Korean mathematics education and TIMSS mathematics assessment are followed.

Cognitive diagnosis modeling (CDM)

CDMs have been developed to provide fine-grained information to researchers and educators on cognitive skills or attributes that are required to solve a particular item, which will allow applications to various instructional practices as well as resolve the limitations of item response theory (IRT) models (de la Torre, 2009). Since most conventional test theories such as the classical test theory (CTT) and unidimensional IRT provide single overall scores on an ability scale, they do not provide diagnostic information on the mastery of skills. Furthermore, the notion of classifying examinees as having mastered skills required to solve an item differs from the objective of traditional measurement models, which derives a continuous latent score of examinee ability.

A Q-matrix is a $J \times K$ matrix with zeroes and ones that consists with J items that require portions of K attributes. In the construction of Q-matrix (Embretson, 1984; Tatsuoka, 1985), required attributes need to be determined. Previous studies define an “attribute” as posited knowledge and thinking skill (Tatsuoka, Corter, & Tatsuoka, 2004), a description of the targeted procedures, skills, processes, strategies, and knowledge a student must possess to solve a test item (Dogan & Tatsuoka, 2008), or a generic term for a skill or knowledge state (de la Torre & Douglas, 2004). In this study, we adopt previous definitions and redefine the term *attribute* as ‘a mathematical skill or content knowledge that is required to solve a test item.’

Among several available CDMs (de la Torre & Douglas, 2004), this study employed the deterministic, inputs, noisy, “and” gate (DINA; Junker & Sijtsma, 2001) model. DINA model proves to be one of the most tractable and interpretable CDMs available (de la Torre, 2009) and was employed to present benefits of interpreting its item parameters (i.e., guessing and slip) and to demonstrate cognitive diagnostic information on students’ mastery/non-mastery of each attribute and its prevalence. Using the DINA model in an analysis of TIMSS data, Lee, Park, and Taylan (2011) report the results of their study can be used to formulate practical guidelines for teachers and educators in linking students’ cognitive processes to answer items correctly to professional development, curriculum reform, and teacher training.

The assumptions of the DINA model are related to IRT models in that local independence is required. In addition, the response patterns of examinees are also assumed to be independent of one another (Junker & Sijtsma, 2001). To assume monotonicity, which is another key assumption for IRT models, the probability that students do not slip has to be greater than the guessing probability (i.e., $1 - s > g$). A key assumption of CDMs is that the Q-matrix satisfies the appropriate relationship between the items and the attributes. In addition, CDMs, including the DINA model, are also related to multidimensional IRT models that measures multiple traits (Junker & Sijtsma, 2001); however, the fundamental difference between the DINA model and multidimensional IRT models is in classifying individuals as masters. The DINA model classifies individuals using a discrete latent variable (η_{ij}), unlike IRT which treats ability as a continuous latent variable. Further discussion of assumptions and statistical requirements of the DINA model are described in de la Torre (2009) and Lee et al. (2011).

Korean mathematics education

A few distinctive aspects of Korean mathematics education include its highly centralized and equalized characteristics. It includes textbooks either published or approved by the ministry of education (MOE) (Choi & Park, 2013; Pang, 2008), which equalizes quality of textbooks in terms of the content presentation as well as sizes of textbooks. From the development of national curriculum to the final revision of the textbooks and teacher guidebooks to be published, the approval committee evaluates and examines carefully to ensure its alignment with the national mathematics curriculum framework. The emphasis is made on internal connection of related mathematical topics within and across grades (Choi, & Park, 2013; Pang, 2004). In addition to concise textbooks, each publishing company publishes a self-learning workbook that includes many problems of various types and difficulty levels.

In terms of characteristics of mathematics instruction, Korean teachers “orchestrate their teacher-centered lessons more systematically, coherently, completely, and progressively than their U.S. counterparts do” and students are deeply engaged and respond enthusiastically to lessons (Pang, 2001). Influences of such unique school mathematics practices on students and findings from this study could explain Korean students’ high performance from international comparison studies.

TIMSS results

TIMSS 2003 international report (Mullis et al., 2004) includes an extensive summary of students’ achievement and other surveys. One of the most noted parts of the report is about how students achieved. Table 1 presents the overall average scale scores for the U.S. and Korea, which are 504 and 589, respectively. For each domain, average scale scores range from 472 to 527 for American students and for Korean counterparts, the range was from 569 to 598 while international average scale scores were 467.

Another aspect, Table 2 shows average percent correct in each content domain for the U.S. and Korean students. Per each strand, 42 to 72 percents of American students answered items in each domain correctly while 63 to 80 percents of Korean students responded with correct answers to mathematics items across the five domains. On average, 51 percent of American students and 72 percent of Korean students answered items with correct response while international averages are

Table 1. 8th grade average scale score

Content Area	U.S.	Korea	International
Number	508 (3.4)	586 (2.1)	467 (0.5)
Algebra	510 (3.1)	597 (2.2)	467 (0.5)
Measurement	495 (3.2)	577 (2.0)	467 (0.5)
Geometry	472 (3.1)	598 (2.6)	467 (0.5)
Data	527 (3.2)	569 (2.0)	467 (0.5)
Average Scale Score	504 (3.3)	589 (2.2)	467 (0.5)

Note. Data drawn from TIMSS 2003 International Mathematics Report (Mullis et al., 2004)

Table 2. Average percent correct in mathematics content area

Mathematics Content Area	U.S.	Korea	International
Number	54 (0.9)	73 (0.6)	50 (0.2)
Algebra	50 (1.0)	71 (0.6)	45 (0.2)
Measurement	42 (1.0)	63 (0.7)	41 (0.2)
Geometry	45 (0.9)	75 (0.6)	50 (0.2)
Data	72 (0.8)	80 (0.4)	62 (0.2)
Total	51 (0.9)	72 (0.5)	48 (0.2)

Note. Adopted and modified from TIMSS 2003 International Mathematics Report (Mullis et al., 2004)

between 41 and 62 percent. These values that solely examine the holistic performance may lead researchers to conclude that Korean students have a greater understanding of the secondary school mathematics over their counterparts from the U.S.

Another type of information that TIMSS report provides for item-level is proportion correct per item, which indicates what percentage of students provided a correct response to each item. Table 3 shows the proportion correct of US, Korea and all participating countries on 43 selected items. For examples, 81.8% of American and 80.3% Korean students answered an item (M032609) correctly while 61.1% of all participating students did so. For another item (M012026), 36.3% of American and 83.5% of Korean students provided correct answers while 45.9% of all participating 8th graders responded correctly.

Table 3. Proportion correct per item

Item	U.S.	Korea	International
1	44.9	51.1	31.0
2	69.4	81.3	54.9
3	55.3	57.2	45.8
4	51.7	68.3	38.4
5	60.5	86.8	51.8
6	29.2	70.3	28.4
7	22.5	52.7	20.1
8	57.4	81.8	44.8
9	50.2	78.3	36.6
10	29.6	60.4	22.1
11	19.4	48.0	14.0
12	14.2	43.9	12.0
13	67.2	82.0	49.4
14	48.1	65.7	32.8
15	6.1	12.0	4.3
16	23.1	46.0	24.7
17	74.2	86.6	61.3
18	24.8	42.7	26.0
19	52.5	10.9	33.6
20	50.1	80.2	41.6
21	48.4	43.2	39.1
22	48.9	76.2	42.4
23	62.4	77.2	52.0
24	39.2	61.8	37.6
25	36.9	81.2	43.2
26	78.2	78.7	60.4
27	86.7	93.9	76.9
28	70.0	92.2	59.9
29	49.6	80.0	48.4
30	33.4	41.0	39.4
31	40.0	77.1	38.6
32	35.3	44.6	31.7
33	48.6	75.2	47.6
34	46.7	82.2	49.7
35	47.4	68.5	55.0
36	79.7	88.5	62.4
37	36.3	83.5	45.9
38	50.9	56.3	35.6
39	65.6	87.9	47.1
40	81.8	80.3	61.1
41	45.0	71.2	45.2
42	67.9	85.2	51.0
43	58.5	89.2	50.2

METHODS

Data

This study compared data of two countries: the U.S. and Korea. Data were taken from the TIMSS 2003 8th grade Mathematics test, which consists of 14 assessment blocks (M01 – M14), assembled to create balance across blocks and booklets with respect to content domain, cognitive domain, and item format (Neidorf & Garden, 2004). Each booklet, which includes 2 to 4 blocks of mathematics, consists of total six mathematics and science blocks with a balanced score distribution of content and cognitive domains that are closely tied to the targeted proportion specified in the framework (Martin, 2005, pp. 2-4). Each booklet is designed to reflect similar proportions of item format, which are extended responses, short answer items, and multiple choice items. On average, each block has 6 to 7 multiple choice items and 5 to 6 short response items assessing five content strands: Number and Operation (28%); Algebra (25%); Geometry (16%); Measurement (16%); and Data Analysis and Probability (15%).

It is worth noting that not all mathematics items used in TIMSS 2003 were released to the public, which means that some booklets may have items that are not available to public on the TIMSS 2003 website (<http://timss.bc.edu>). Fortunately, however, among 14 tested mathematics booklets, all items in the booklet 5 were released at the time of research, which is a good represented sample due to its design that reflects balanced distribution of content and cognitive domains and item formats. Items were dichotomously scored; omitted or unreached responses to items were scored as incorrect. A total of 1,179 (740 U.S. and 439 Korean) students were sampled as they represent the number of students that took booklet 5 among 8,912 U.S. and 5,309 Korean students that took TIMSS 2003.

A construction of Q-Matrix

Attributes used to construct the Q-matrix in this study were adopted from the National Council of Teachers of Mathematics (NCTM) *Principles and Standards for School Mathematics (PSSM)* published in 2000, which outlines crucial learning points in mathematics. Adopted learning points were modified to fit the required knowledge for items used in the study. Although the *PSSM* is not a national curriculum, it provides guidance for teachers and educational professionals to make

Table 4. Attributes adopted from NCTM Principles and Standards (2000) to analyze performance on mathematics items from the TIMSS 2003 for 8th grade

Strand	Attribute	Frequency
Number	1. Understand numbers, ways of representing numbers, relationships among numbers, and number systems	19
	2. Understand meanings of operations and how they relate to one another	9
	3. Compute fluently and make reasonable estimates	7
Algebra	4. Understand patterns, relations, and functions	9
	5. Represent and analyze mathematical situations and structures using algebraic symbols	8
	6. Use mathematical models to represent, understand, and analyze quantitative relationships	3
Geometry	7. Analyze characteristics and properties of two- and three-dimensional geometric shapes and develop mathematical arguments about geometric relationships	10
	8. Specify locations and describe spatial relationships using coordinate geometry and other representational systems / Apply transformations and use symmetry to analyze mathematical situations	5
	9. Use visualization, spatial reasoning, and geometric modeling to solve problems	4
Measurement	10. Understand measurable attributes of objects and the units, systems, and processes of measurement	9
	11. Apply appropriate techniques, tools, and formulas to determine measurements	5
Data	12. Understand and apply basic concepts of probability	3

decisions about the content and characteristics of school mathematics. A list of attributes is presented in Table 4 that range from 3 to 19 in the number of times it was specified in the Q-matrix.

The Q-matrix in Table 5 was specifically constructed for the TIMSS 8th grade mathematics test used in this study. Individual coders solved items independently in a step-by-step manner and matched each step with one learning objective listed in *PSSM*. Coders are modified according to the NCTM standards so that mathematics educators can deliver reinforced instruction based on findings from this study. Two mathematics educators with advanced degrees in mathematics education and teaching experience at the secondary school mathematics and a domain-expert

Table 5. The Q-Matrix for the 8th grade TIMSS mathematics test

Item	Attribute											
	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0	1	1	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0
3	1	0	1	0	0	0	0	0	0	1	0	0
4	0	1	1	0	0	0	0	0	0	1	0	0
5	0	0	0	0	0	0	1	0	0	0	1	0
6	0	0	0	0	0	0	1	0	0	1	1	0
7	0	0	0	0	0	0	1	0	0	0	1	0
8	0	1	0	1	0	0	0	0	0	0	0	0
9	0	0	0	1	0	0	1	0	0	0	0	0
10	0	0	0	1	1	1	0	0	0	0	0	0
11	0	0	0	1	1	1	0	0	0	0	0	0
12	1	0	0	0	1	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	1	1	0	0	0
14	0	0	0	0	0	0	0	1	1	0	0	0
15	0	0	0	0	0	0	0	1	0	0	0	0
16	0	0	0	1	1	0	0	0	0	0	0	0
17	1	0	0	0	0	0	0	0	0	1	0	0
18	1	1	1	0	0	0	0	0	0	1	0	0
19	0	0	0	1	0	0	0	0	0	0	0	1
20	1	0	0	0	0	0	1	0	0	0	0	0
21	1	1	0	0	0	0	0	0	0	0	0	0
22	0	0	1	1	1	0	0	0	0	0	0	0
23	1	0	0	0	0	0	0	0	0	1	0	0
24	0	0	0	0	0	0	0	0	1	1	0	0
25	0	0	0	0	0	0	1	0	0	0	0	0
26	1	0	0	0	0	0	0	0	0	0	0	1
27	1	0	0	0	0	0	0	0	0	0	0	0
28	1	1	0	0	0	0	0	0	0	0	0	0
29	1	1	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	1	1	0
31	0	0	1	0	0	0	1	0	0	0	1	0
32	1	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	1	1	0	0	0	0
34	0	1	0	0	1	0	0	0	0	0	0	0
35	0	0	0	0	0	0	1	0	0	1	0	0
36	0	0	0	0	1	0	0	1	1	0	0	0
37	0	0	0	0	0	0	1	0	0	0	0	0
38	0	0	0	0	1	1	0	0	0	0	0	0
39	1	0	0	0	0	0	0	0	0	0	0	0
40	1	1	1	0	0	0	0	0	0	0	0	0
41	1	0	0	0	0	0	0	0	0	0	0	0
42	1	1	0	1	0	0	0	0	0	0	0	0
43	1	0	0	0	0	0	0	0	0	0	0	1

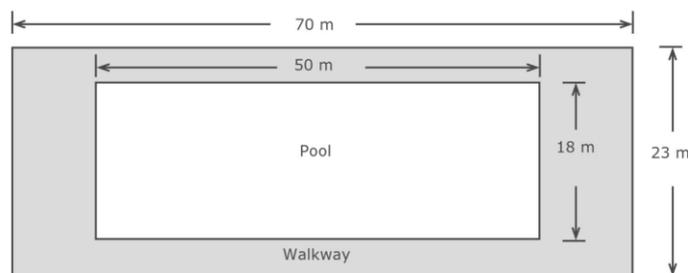
researcher identified the attributes required for each item. Each coder independently constructed the Q-matrix and subsequently regrouped to discuss discrepancies until a consensus was reached to develop the final Q-matrix (Table 5). This process ensured the validity and usefulness of the attributes and the Q-matrix.

Attributes in the Q-matrix are independently manifested during the process that an item is solved; in other words, each step required to solve a problem utilizes the mastery of a specific attribute that is mapped to the Q-matrix. For example, Figure 1 shows item 31 as presented in TIMSS, which asks student to obtain the area of the paved walkway that surrounds a pool. In order to solve this item correctly, a student needs to master three attributes, which were validated to be the most generic and dominant method by the three coders. First, a student must recognize that the area of the walkway is the difference between the area of the pool and the joint area of the two rectangles (walkway and the pool) - attribute 7. Secondly, the student should make fluent computations to find areas of the two rectangles - attribute 3. Finally, knowing and understanding that the product of two meters (m) becomes meters squared (m^2) - attribute 11 - completes the final answer.

Analysis

Given the Q-matrix, the data were analyzed using the Expectation Maximization (EM) algorithm, which provides an efficient iterative procedure to estimate the DINA model and its item parameters. The EM algorithm reduces the computational burden of parameters greatly and is described as accurately recovering attribute parameters than using the Markov chain Monte Carlo procedure (de la Torre, 2009). The computer program Ox (Doornik, 2002), available free of charge for academic research and teaching purposes, was implemented to run the EM algorithm for the DINA model and is available upon request.

A rectangular shaped swimming pool has a paved walkway around it as shown.



What is the area of the paved walkway?

- (A) 100 m^2
- (B) 161 m^2
- (C) 710 m^2
- (D) 1610 m^2

In order to solve this problem, a student needs to understand

1. The area of the paved walkway is the difference between the two areas: the area of the pool (the 50 m by 18 m rectangle) and the area of pool and the walkway (the 70 m by 23 m rectangle) – attribute 7
2. Making fluent computations to find areas of two rectangles – attribute 3

Understanding that the multiplication of two meters becomes m^2 – attribute 11

© Copyright 2005

International Association for the Evaluation of Educational Achievement (IEA)

Figure 1. Mapping of attributes to a solution of the item 31

Using the calculated δ values, high discrimination items and low discrimination items were identified. High discrimination items had δ values greater than 0.63, which was derived from the mean discrimination of both Korea and U.S., which ranged between 0.44 and 0.45 with a standard deviation range of 0.16 to 0.18. Thus, a δ value greater than 0.63 was determined to be the cutoff value for the high discrimination items, because it was one standard deviation above the mean discrimination. Similarly, items with δ values less than 0.27 were selected as the low discrimination items.

RESULT

Item discrimination

This subsection describes the trends of item discrimination that involves in guess and slip parameters. The guess and slip parameters are item-level parameters that provide what would be the percentage of students who answered a certain item correctly without possessing all required attributes for the item (guess) or the percentage of students who answered another item incorrectly although having all required attributes for the item (slip).

The DINA discrimination index provides information to identify significant differences in item discrimination. The advantage of using the DINA model to obtain the δ parameter is its beneficial interpretation. Recall that a high discrimination implies a greater clarity in an examinee's latent group membership to either $\eta = 1$ or $\eta = 0$, having all required skills to answer the item correctly or not. This is different from the CTT and IRT interpretation of the discrimination parameter, which are point-biserial correlation and the slope of the linear predictor for the logistic IRT model, respectively. Furthermore, it represents the probability that an examinee solves the item correct given that he or she does not guess and does not slip.

In addition, combinations of a country's discrimination index to its achievement on each item will provide valuable information on the accuracy of the attainment of specific attributes. There are four possible scenarios worth to consider: (1) high discrimination with high performance; (2) high discrimination with low performance; (3) low discrimination with high performance; and (4) low discrimination with low performance. The first and second scenarios can be interpreted with straight forward explanations. However, the third and fourth have more to explore because low discrimination could be high guess, high slip, or high guess and slip at the same time. An item may have low discrimination yet students may have performed poorly. Alternatively, an item could also have low discrimination with high student performance. These results can be explained naturally within the DINA framework, which assumes that a student guessed if they have not mastered an attribute yet was able to get the item correct; likewise, a student slipped if they mastered the attribute yet was unable to respond correctly to the item.

Among highly discriminating items, the Korean students scored well on items 4, 6, 7, 10, 11, 12, and 24. The most prevalent attributes from these items were attributes 5 and 10, which are skills that were instructed well and mastered by the Korean 8th graders. On the other hand, for the three items identified in the first category from the U.S. results, attribute 10 appeared more than once. Attribute 10 requires students to understand measurable attributes of objects and the units, systems, and possesses of measurement or to apply appropriate techniques, tools, and formulas to determine measurement and is the attribute that 8th graders in both

countries learned and performed highly without a significant difference between the two groups.

Among the low discrimination items with high performance for Korean students, attributes 1 and 8 appeared most frequently. From the U.S., attributes 1 and 2 from the same Number & Operations content domain were the most frequent. Based on the results from both countries, students *guessed* to answer items that required attribute 1 correctly even though the attribute was not necessarily mastered by the students.

Due to the large guessing parameter estimates from the Korean cohort, the differences between two countries' discrimination parameter patterns were not too different. Among the 43 items analyzed, the guessing parameters of 34 items present larger values in Korean data, underlying causes of which can be explained by its distinctive educational atmosphere.

Attribute prevalence

In this subsection, the prevalence of attribute mastery and attribute pattern probability are presented. This will demonstrate unique characteristics of the DINA model that provides cognitive diagnostic information that other traditional methods of item analysis such as the CTT and IRT models cannot.

An advantage of the CDM is the availability of cognitive diagnostic information, fine-grained attributes that students mastered/did not master that can be used to aid instruction for educators and mathematics researchers. Figure 2 shows the prevalence of attribute mastery level for the U.S. and Korea.

Similar to results from Table 6, Korea had a significantly greater prevalence of attribute mastery for all attributes, except attributes 10 and 12. Although this validates the results from previous figures and tables, this result is, again, not intuitive solely based on the overall performance of Korean students, which is dominant in all content areas over the U.S. However, to illustrate that the prevalence of attribute mastery for Measurement (attribute 10) and Data Analysis & Probability were not significantly different warrants further investigation for both countries.

Table 7 shows the top 5 highest latent classes based on their posterior probability. It should be noted here that there are 12 total attributes, resulting in $2^{12} = 4096$ possible latent classes for all possible combinations of attribute patterns. The posterior probability of Korean students (0.310) in the latent class with all possible attributes was almost double the posterior probability of U.S. students (0.155). The second largest latent class is the group of students that

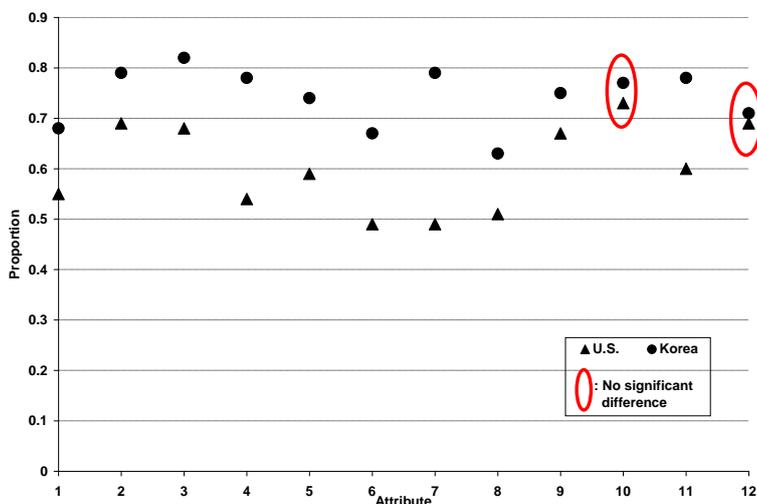


Figure 2. Attribute mastery level (Prevalence)

Table 6. A comparison of attributes attainment proportions

Strand	Attribute	z-score	p-value
Number & Operations (NO)	1	4.341	< .001***
	2	3.661	< .001***
	3	5.179	< .001***
Algebra (A)	4	8.186	< .001***
	5	5.145	< .001***
	6	5.955	< .001***
Geometry (G)	7	10.113	< .001***
	8	3.947	< .001***
	9	2.831	.005**
Measurement (M)	10	1.454	.146
	11	6.274	< .001***
Data Analysis & Probability (DP)	12	0.658	.511

Note. ** p-value < .01 (Two-tailed z-test for two proportions); *** p-value < .001 (Two-tailed z-test for two proportions)

Table 7. Attribute pattern probability

U.S.												Korea													
Attribute Pattern												Prob.	Attribute Pattern												Prob.
1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12		
1	1	1	1	1	1	1	1	1	1	1	1	0.155	1	1	1	1	1	1	1	1	1	1	1	0.310	
1	1	1	1	1	0	1	1	1	1	1	1	0.054	1	1	1	1	1	0	1	1	1	1	1	0.108	
1	1	1	1	1	0	1	1	1	1	0	1	0.032	1	1	1	1	1	1	1	1	1	1	0	0.046	
1	1	1	1	1	1	1	1	1	1	0	1	0.026	1	1	1	1	1	1	0	1	1	1	1	0.015	
1	1	1	1	1	1	0	1	1	1	1	1	0.019	1	1	1	0	1	1	1	1	0	1	1	0.013	

Note. Top 5 highest attribute pattern probabilities; Prob.: posterior probability given latent class.

mastered all attributes, but attribute 6, which uses mathematical models to represent, understand, and analyze quantitative relationships from the Algebra strand. The latent class size for Korea is again almost twice as greater than the U.S., which were 0.108 and 0.054, respectively. The next largest latent classes differ for both countries. For the U.S., the third largest latent class lacks attributes 6 and 11, with a latent class size of 0.032, where attribute 11 comes from the Measurement strand. This is followed by the latent class that lacks attribute 11 with size 0.026 and finally, the group that lacks attribute 8 from the Geometry strand with size 0.019. For Korea, the third largest latent class lacks attribute 12 from the Data Analysis & Probability strand with size 0.046. This follows by the group that lacks attribute 8 from the Geometry strand with size 0.015, then the group that lacks attribute 4 from the Algebra strand and attribute 9 from the Geometry strand with a size of 0.013.

DISCUSSION AND CONCLUSION

This study sought to introduce a new methodology to assess and evaluate student achievement to utilize the information to mathematics instruction. Due to the criticism about a lack of practical information to educators and applied researchers, identifying specific mathematical attributes that students either mastered or needed is desirable to develop and improve instruction and to reinforce areas of deficiency. The CDM approach was selected as it provided information about an examinee’s mastery or non-mastery of fine-grained attributes. Unlike other common psychometric models, through this study, it is possible to show mathematics educators and researchers the advantage of using a CDM via the DINA model as it provides the following: (1) more interpretable explanation of the discrimination index and (2) cognitive diagnostic information.

Results compared the discrimination parameter of DINA across the U.S. and Korea, which has several implications for mathematics educators. Although

differences in achievement scores between Korean and the U.S. 8th graders were significant, the analysis of discrimination indices revealed that out of the 29 items that showed significant differences between the two countries, the American cohort had 14 items with higher discrimination. This is unintuitive and worth to pay attention due to the strong performance of the Korean cohort.

A closer examination of the overall discrimination parameter by strand showed that there was no significant difference between the U.S. and Korean cohort for Algebra, Geometry, Measurement, and Data Analysis & Probability. Only the Number domain exhibited significant difference with American students presenting significant discrimination over Korean students. Here, the interpretational benefit of the discrimination index should also be reiterated. It proves to be a simple statistic to assess the magnitude of classification accuracy for latent group membership, because it is the difference in probabilities for the latent response groups.

As a result of the DINA model, diagnostic information, specific enough for teachers and education researchers can readily use, has been collected for students from two countries. Multiple analyses from various models address Korean students' weakness emerging from skills in the Data and Probability strand. Clearly, there is room for improvement. Researchers (Han, 1986) have focused on improving performance related to this skill as this result has been not been unprecedented. Mathematics educators must continue to look for other ways to refine this attribute as well as to develop strategies that can overcome student difficulty in mastering this skill. Also, Korean students need improvement in Algebra and Geometry, which seems to contradict their high achievement over all areas in mathematics studies. As such, further research must be followed to uncover why Korean students exhibit deficiencies in large portions of Algebra and Geometry.

There are two major implications drawn from this study. One is to apply CDM-based assessment results to instructional practices. Teachers have been experiencing difficulties when receiving assessment results: from how to interpret the results to where to implement the reports. With fine-grained information that is the basis of analysis, teachers could easily understand what the report tells about students' mastery or non-mastery of important learning objectives. While typical assessment reports present individual achievement on domain or a subject, which are rather broad to implement in teaching practices, reports from CDM could offer individual mastery of learning objectives that are smaller that could be used in classrooms. As expectations from *PSSM* are used for attributes in our study, teachers could work with measurement experts to design their own formative and summative assessments with attributes as small as lesson objectives that they want to ensure students to learn.

When conducting international assessment study, disseminated results are typically national average scores, ranking order of countries, etc., which can hardly offer to learn from other countries. Also, there is a danger of prematurely adopting certain educational practices of a high achieving nation that do not take adoptee's culture and practices into a consideration. When we can learn how exactly a group of teachers teach in relation of a certain topic, which their students are highly achieving in, it could be easier to learn from the group of teachers. Also, adopting other's practices in a smaller scale (attribute-level) in the beginning and expanding it as it progresses positively.

Limitations

When constructing a Q-matrix, only prominent strategy to solve each item was considered rather than taking every possible approach into consideration. This does not fully support the way we encourage students to learn and explore school mathematics, allowing multiple ways to represent mathematical ideas. For the

future studies, adopting a method to include multiple strategies to solve a mathematics problem when constructing a Q-matrix would be encouraged, which reflect the idea of allowing multiple mathematical approached to solve a problem.

This study examined examinees' skill mastery profiles using the DINA model. Yet, TIMSS data were originally designed and developed under an IRT approach. This may raise issues on problems of retrofitting data designed using a different model. However, as noted in Junker and Sijtsma (2001), the DINA model satisfies generalizations of standard IRT assumptions; it was also found that the DINA model was sensitive to attributes even when the data were designed to fit the Rasch model (an IRT model). More recently, studies have shown that parameters of the DINA model were highly correlated with IRT models and that retrofitting IRT data into a CDM framework uncovers important cognitive and content attributes when applied to the TIMSS data (Lee & de la Torre, 2008; Park & Lee, 2010; Dogan & Tatsuoka, 2008). Similarly, in Lee et al. (2011), TIMSS data were retrofitted using the DINA model, which revealed diagnostic information on students' attribute mastery. As this study revealed, the DINA model can be a meaningful tool to provide information for policymakers and instructors to aid mathematics education – even when data were developed for a different model.

Using the same rationale, implications of model fit can also be extended to this assumption; that is, items used in TIMSS have been calibrated using unidimensional IRT models. Based on model comparison using the information criteria, the DINA model showed better fit over 1PL, 2PL, and 3PL IRT models, which supports the use of CDMs to examine student mastery of fine-grained attributes. Understanding that attribute-based diagnostic information of students would be beneficial for instructors to prepare lessons, the contribution this study brings to mathematics education would be an important step forward in bridging test development to instruction.

Conclusion

This study examined the utility of the CDM approach to compare the performance of the TIMSS mathematics assessment between the US and Korean 8th grade samples. The main idea was to identify students' strengths and weakness of mathematical attributes that were used to construct the Q-matrix in the study. Through a careful investigation, the results can help mathematics educators and researchers identify possible causes of subpar achievement in specific content skills and make curriculum-based decisions and, in turn, enhance classroom practices. It was found that there is still a room for improvement although a country is known for their students' high performance from international assessments.

The results from this study provided fine-grained information on areas with room for improvement in both countries. As the CDM approach provides detailed information, future studies should examine the mastery of more-specific attributes. In such further studies, we may separate some attributes that include the broad scope of knowledge, skills, and processes. To investigate in greater detailed, specific information about each strand, it would be interesting to conduct studies that analyze an individual strand or a smaller number of strands.

One simple deception that educators easily forget to be cognizant is that high scores do not always translate to a student's mastery of the required skills to solve the item. Consequently, educators have made numerous attempts to adopt or trail high performing country's educational practices with no significant results. However, employing the CDM approach provides better understanding about a population's or an individual's true mastery of certain attributes that extends the simple notion of the number of correctly solved problems. It is critical for educational researchers to understand this concept, because measuring student

mastery that rely only on their achievement could yield to following inappropriate examples of other countries' teaching and learning practices.

Based on the demonstration of the interpretable discrimination index and the cognitive diagnostic information, the DINA model should be avidly implemented and applied to empirical research. This will enable mathematics instructors to directly address attributes in the classroom that students can benefit that is derived from a systematic model of cognitive diagnosis. As a next step, demonstrating the usefulness of DINA-based classroom assessment would warrant the feasibility to administer such assessment in their everyday instructions.

REFERENCES

- Choi, K. M., & Park, H-J. (2013). A comparative analysis of geometry education on curriculum standards, textbook structure, and textbook items between the U.S. and Korea. *Eurasia Journal of Mathematics, Science, and Technology Education*, 9(4), 379-391. 10.12973/eurasia.2013.947a
- de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333-353.
- de la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34, 115-130.
- Dogan, E., & Tatsuoka, K.K. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics*, 68(3), 263-272.
- Doornik, J.A. (2002). *Object-oriented matrix programming using Ox* (Version 3.1). [Computer software]. London: Timberlake Consultants Press.
- Embretson, S. (1984). A general latent trait model for response processes. *Psychometrika*, 49, 175-186.
- Ferrini-Mundy, J., & Schmidt, W.H. (2005). International comparative studies in mathematics education: Opportunities for collaboration and challenges for researchers. *Journal for Research in Mathematics Education*, 36, 164-175.
- Han, J. (1986). *The cognitive development of secondary school students in the republic of Korea*. Paper presented at the United States-Japan Seminar on Science Education sponsored by the National Science Foundation and the Japan Society for the Promotion of Science, Honolulu, HI. Retrieved from www.csa.com
- Holliday, W. G., & Holliday, B. W. (2003). Why using international comparative math and science achievement data from TIMSS is not helpful. *The Educational Forum*, 67, 250-258.
- Martin, M.O. (Ed.). (2005). *TIMSS 2003 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Junker, B.W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Lee, Y.-S., & de la Torre, J. (2008, March). *Cognitive diagnosticity of IRT-constructed assessment: An empirical investigation*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.
- Lee, Y.-S., Park, Y.S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the U.S. national sample using the TIMSS 2007. *International Journal of Testing*.
- Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004). *TIMSS 2003 international mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
- Neidorf, T.S., & Garden, R. (2004). Developing the TIMSS 2003 mathematics and science assessment and scoring guides. In Martin, M. O., Mullis, I. V. S., & Chrostowski, S. J. (Eds.). *TIMSS 2003 Technical Report*. Chestnut Hill, MA: Boston College.

- Pang, J.S. (2001). *Challenges of reform: Utility of sociomathematical norms*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Pang, J.S. (2004). *Development and characteristics of Korean elementary mathematics textbooks*. Paper presented at ICME-10 as part of KSICMI (Korean Sub-commission of ICMI session), Copenhagen, Denmark.
- Pang, J.S. (2008). Design and implementation of Korean mathematics textbooks. In Z. Usiskin & E. Willmore (Eds.), *Mathematics curriculum in pacific rim countries: China, Japan, Korea, and Singapore* (pp. 95-125). Charlotte, NC: Information Age.
- Park, Y.S., & Lee, Y.-S. (2010, July). *CTT, IRT, and G-DINA analysis of TIMSS 2007*. Paper presented at the 7th conference of the International Test Commission at the Chinese University of Hong Kong, Shatin, Hong Kong.
- Tatsuoka, K.K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, 12, 55-73.
- Tatsuoka, K.K., Corter, J., & Tatsuoka, C. (2004). Patterns of Diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American Educational Research Journal*, 41(4), 901-926.
- Wang, J. (2001). TIMSS primary and middle school data: Some technical concerns. *Educational Researcher*, 30, 17-21.

