1-1-2010

Cognitive Diagnostic Assessment Of Timss-2007 Mathematics Achievement Items For 8th Graders In Turkey

Turker Toker
University of Denver

COGNITIVE DIAGNOSTIC ASSESSMENT OF TIMSS-2007 MATHEMATICS ACHIEVEMENT ITEMS FOR 8TH GRADERS IN TURKEY

_____

A Thesis

Presented to

Morgridge College of Education

University of Denver

_____

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

_____

by

Turker Toker

August, 2010

Advisor: Kathy Green

Author: Turker Toker
Title: COGNITIVE DIAGNOSTIC ASSESSMENT OF TIMSS-2007 MATHEMATICS ACHIEVEMENT ITEMS FOR 8TH GRADERS IN TURKEY
Advisor: Kathy Green
Degree Date: August, 2010

ABSTRACT

This study investigated students' responses to released TIMSS mathematics items and considered what those responses might show about participants' mastery level of cognitive areas. The least squares distance method (LSDM) was used in this cognitive diagnostic analysis. There were 4,498 8th-grade students from seven geographical regions of Turkey who took TIMSS items. In this study, using the responses of Turkish students to the released math items of TIMSS-2007, an IRT analysis was also conducted to both compare results with international item parameters A Pearson correlation between these two sets of item logit positions produced a correlation of $r = .82$, and this correlation represents a high degree of relationship and so overall consistency in item logit position.

Results indicated that the majority of the items can be well explained by a set of 20 attributes (mean MAD = .075; $r = .68$). As a result of this study, educators in Turkey should note that students have weaker knowledge in geometry, graphics, charts, figures, rule application in algebra, and data management and stronger skills with. The results of the present study can help teachers and curriculum developers ensure that the utilized educational policies and methodologies would help students to improve their cognitive mastery levels in general.

ACKNOWLEDGEMENT

It has been a wonderful adventure pursuing my master's degree, which I could not have done without the encouragement, guidance, support, and wisdom of several people. I am heartily thankful to my advisor, Kathy Green, whose encouragement, supervision and support from the preliminary to the concluding level enabled me to develop an understanding of the subject. I also would like to thank my committee members Duan Zhang and Charles Reichardt whose comments were always helpful. Furthermore, I would like to thank mathematics experts Muhammet Talu and Araz Ismailov who helped me to develop the heart of this study.

Additionally, I would like to thank my wonderful wife, Pinar Toker, and beautiful daughter, Erem Leyl Toker, for their support. Their support and love lightened my way. I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

Lastly, I would like to dedicate this thesis to those who have helped me become who I am today and without them I could have never achieve such a success. For the long nights without sleep, for caring for me while I was sick, for always believing in me when others would not, for the unconditional love, care, and support, I will forever be thankful to two woman in my life, my mom, Emine Degiz, and my love, Pinar Toker.

(Turkish Translation)

Mastır eğitimim boyunca karsılaştığım pek çok zorlukta yanımda olan ve desteklerini benden hiç bir zaman esirgemeyen herkese çok teşekkür ederim. Özellikle eimin ve annemin yillar boyu bana göstermiş oldukları sevgi, saygı ve destekten dolayı

onlara sonsuz teşekkürlerimi sunarım. Her zaman yanımda hissettiğim karşılıksız sevginiz ve beni mutlu etmek için göstermiş olduğunuz fedakarlıklarınız bugüne kadar kazanmış olduğum ve bundan sonra da kazanmayı umut ettiğim başarılarımın temelini oluşturmaktadır.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER1

Introduction

Educational assessment has evolved from grading one's achievement level to being diagnostically useful at every step in education (Bolt, 2007). Assessment affects grades, placement, advancement, instruction, curriculum, and in some cases, funding. Assessment is an integral part of education, as it defines whether the goals of education are being met or not. Assessment has extensive influence on educators and leads them to address questions such as: "How well are we teaching?", "Are students learning what they need to learn?", "Are there other ways to teach better?", "How much funding do we need to teach better?". Assessment, thus, is critical to both evaluating the effects of educational programs and also to directing those programs.

Outcomes of education need to meet globally accepted criteria. Students must be able to think wisely and critically, to examine in detail, and to make inferences (Gierl, 2007). Changes in the skills base and the knowledge our students need require new learning goals; these new learning goals change the relationship between assessment and instruction and so teachers need to be informed regarding both. Teachers need to take an active role in making decisions about the purpose of assessment and the content that is being assessed (Wiggins, 1998). Kellough and Kellough (2007) stated that there are six purposes for assessment:

1. To help student learning
2. To identify student's academic strengths and weaknesses

3. To assess the effectiveness of a specific instructional strategy

4. To assess and improve the effectiveness of curriculum

5. To supply data that can assist in decision making process

6. To communicate with and involve parents.

Turkey has been affected by international changes in educational assessment as have other nations. Although the Turkish educational system is affected by its own cultural foundations, its geographic location and internal political changes make changes in the world external to Turkey more influential. In addition, the process of gaining membership in the European Union (EU) forces Turkey to take steps to conform to EU requirements regarding educational structures and assessment in education. A short description of the Turkish educational system is given in the next section.

Education in Turkey

The educational system in Turkey is structured as pre-school, primary, secondary, and higher education levels as seen in Figure 1.

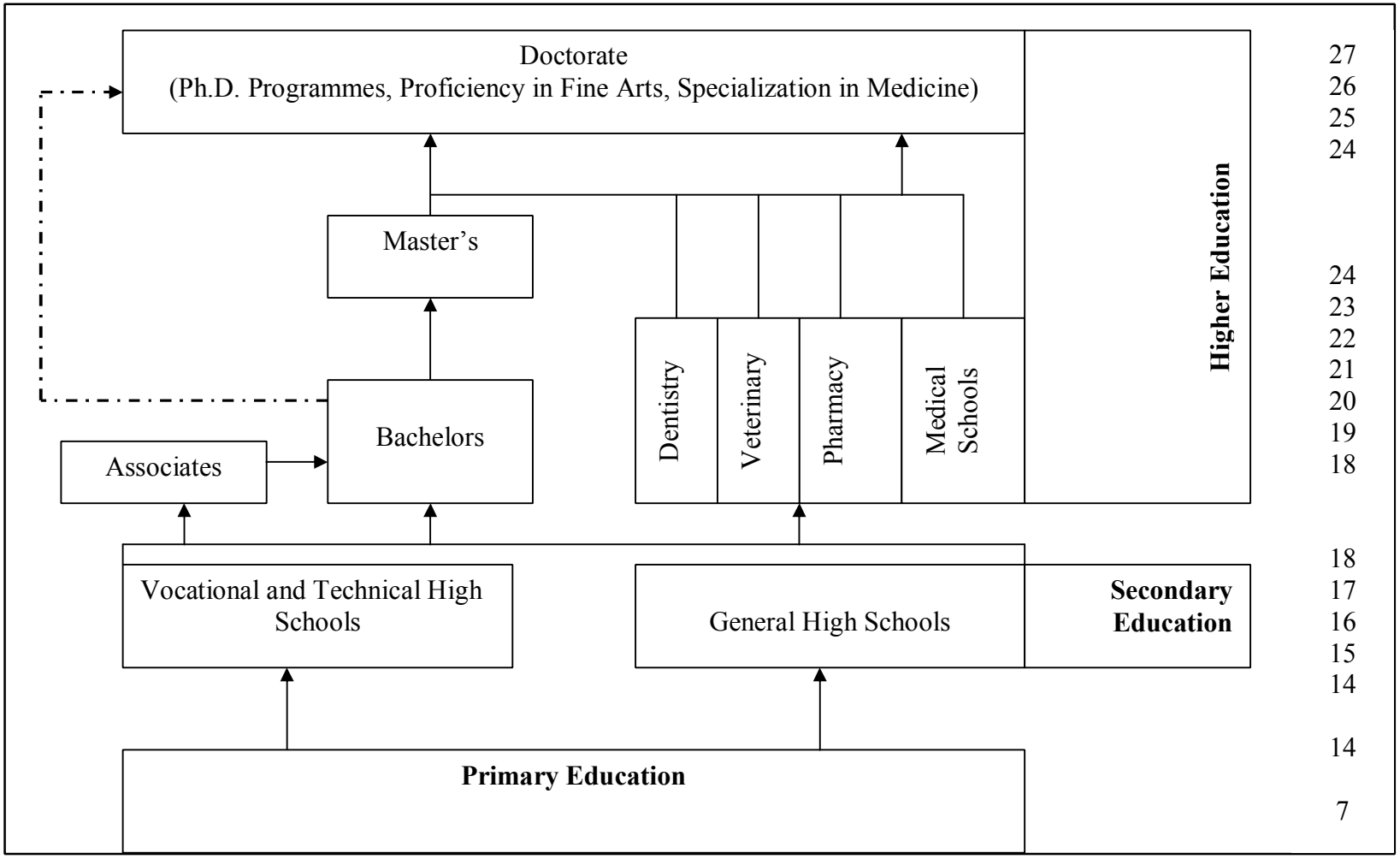


Figure 1. Structure of the Turkish Educational System (Age shown on the right) (MONE, 2001).

Turkey has a population of over 72 million, estimated to rise to about 82 million by 2015. Currently the pre-school education schooling rate is 16% (OOEGM, 2005), and an objective of the Ministry of National Education (MONE) was to increase the rate to 25% by 2010. There are around 13 million students at the formal primary and secondary education levels with more than 500,000 teachers (MONE, 2001). Formal education in Turkey may be preceded by a non-compulsory pre-school year.

Compulsory education was increased from 5 to 8 years in 1997. There are more than ten million students receiving 8-year primary education in schools with about 400,000 teachers at that level. The enrollment at this level is approximately 91.9% of the relevant age population (UNDP, 2004). Secondary school is not yet compulsory but was extended from 3 to 4 years in September 2005. After receiving 8 years of compulsory education, students may choose to go to a general high school, which prepares them for institutions of higher education, or a vocational/technical high school, which provides occupational education. There are around 1.5 million students in general high schools, and 820,000 students in vocational/technical high schools. The combined general and vocational/technical high school enrolment rate is 73.3% of the age group (UNDP, 2004). At the higher education level, there are 95 state universities and 51 private universities established by foundations. Entry into tertiary level education is extremely competitive. In 2008, some 1.5 million students sat a national examination for about 850,000 places in tertiary education institutions, including vocational ones and the Open University. The vocational high school graduates may apply to 2-year vocational schools without taking the university entrance exam.

Institutions of higher education, in total, serve some 1.9 million students. The enrollment rate at this level is about 23.8% (UNDP, 2004). The enrollment rate in Turkey, when primary, secondary, and tertiary education is combined, is 61%. The female overall enrollment is 54%, and the male enrollment is 65% (UNDP, 2004). The adult literacy rate, for people aged 15 and over, is 86.5%, the male and female rates being 94.4% and 78.5%, respectively (UNDP, 2004).

Education in Turkey is affected by impermanent political solutions. Rather than having long-term solutions, running political parties try to save the day. This affects all parts of education. Districts have no control over assessment with decisions made based on national examinations. Large scale assessments are a critical part of the educational system. Students take around six countrywide tests throughout their education which have major effects on their lives.

The results of national research conducted to determine students' achievement (MEB-OBBS, 2002; MEB-OBBS, 2005 and MEB-OBBS,2007) and PISA, TIMSS and PIRLS international exams show that Turkish students' learning outcomes were insufficient (MEB-PISA, 2005). Political parties made changes in the educational system from the 2005–2006 school year on, with new curricula developed progressively for grades 1-5 of the primary school and from the 2006–2007 school year on for other grades of primary schools. The new curriculum is based on a constructivist approach. Knowledge is to be constructed by students. This change in the curriculum also caused alterations in content, teaching methods, and instructional materials as well as in

assessment. The new curriculum gives the teachers roles and missions different than those they held before (Gelbal & Kelecioglu, 2009).

Recently curricula and associated teaching methods have changed, but large scale assessments are still traditional and high-stakes. Students have started to take performance-based exams during the year. Unfortunately, their performance scores have little or no effect on the large scale assessments that they take each year. This leads them to take school entrance exams more seriously than their process exams during the year. The research has shown that teachers have difficulty using assessment methods according to the new instructional programs. They think that they are less qualified in the area of measurement and evaluation than in other areas. Also they state that they need education about assessment methods (Gozutok, Akgun, & Karacaoglu 2005; Yapıcı & Demirdelen, 2007; Yaşar et al., 2005). A constructivist approach to learning is mandated while high-stakes examinations are in traditional form. Teachers feel compelled to prepare students for high-stakes examinations while also providing instruction for general success in education. Identification of the discrete skills necessary to examination success and to learning in general would benefit teachers in their quest to educate students.

In this study, data were taken from one of the most respected international exams, the Trends in International Mathematics and Science Study (TIMSS-2007), used to assess cognitive abilities of 8^{th} graders' mathematics achievement in Turkey. The purpose of this study was to validate cognitive attributes on the released TIMSS-2007 mathematics test items with respect to the cognitive attributes developed by Tatsuoka and her associates and so to extend use of the Least Square Distance Method (LSDM) to skills

and cognitive processes as well as knowledge attributes. The study was specific to attribute identification for Turkish students; thus, experts familiar with the educational system in Turkey served as experts.

The main goal of this study was to provide information about attributes of TIMSS 2007 8th grade mathematics results for Turkey to inform teachers in the Turkish educational system. Even though TIMSS 2007 has extensive content validity support based on MONE's information given to the test administrators (TIMSS-2007 Technical Report, 2008), results show that instruction methods are not adequate as students' scores fall short. This study investigated students' responses to the released mathematics items and then considered what those responses might show about participants' mastery level of cognitive areas. By taking the results to the item level, instructors can see whether the mistakes made by participating students in the sample are also mistakes made by their own students. So, teachers can focus on those areas in need of more work during their instruction to improve cognitive mastery levels of students.

TIMSS-2007

TIMSS was designed to assess trends in students' mathematics and science achievement. TIMSS-2007 is the fourth in a four-year-cycle of assessments (previously administered in 1995, 1999 and 2003). It was designed to align with mathematics and science curricula in the participating countries. TIMSS results assess the degree to which students have learned mathematics and concepts and skills likely to have been taught in school. TIMSS tests put an emphasis on questions and tasks that offer insight into the analytical, problem-solving, and inquiry skills and capabilities of students. In addition,

students, teachers, and school principals in each participating country are asked to complete questionnaires concerning the context for learning mathematics and science, so as to provide a resource for interpreting the achievement results and to track changes in instructional practices. TIMSS-2007 assesses the mathematics and science achievement of children in two target populations: fourth grade and eighth grade students (Shen, 2000).

At the beginning of every TIMSS cycle, a committee comprising curriculum experts in mathematics and science from participating countries constructs the framework for the coming assessment. This framework should be confirmed by all member countries as being representative of their country's curricula. However, there will always be some content that is not covered in the curriculum of every participating country. This is solved at the data analysis stage of the project by excluding the items to which a participant country objects. This item exclusion rarely has any positive or negative effect on a country's score.

Turkey has participated in two TIMMS studies (1999 and 2007). There were 38 countries that participated in TIMSS-1999. This was Turkey's first TIMSS experience. According to the 8th grade mathematics results, Turkey placed 31st in the study. The mean mathematics score was 500 with a standard deviation of 100. The average score for Turkey was 429. The overall international average was 487 (TIMMS-R Turkey Report, 2003). TIMSS-2007 was the second study in which Turkey participated. The mean mathematics score and standard deviation were the same as they were in TIMSS-1999. There were 48 countries at 8th grade level and Turkey placed 31st in the study with a score

of 432. Turkey's mathematics score was higher than that of 19 countries, and lower than that of 28 countries. For comparison, the United Stated placed 19th in 1999 and 9th in 2007; China placed 3rd in 1999 and 3rd in 2007.

The quality of the TIMSS exam is supported by experts in the area of educational measurement and evaluation. In most of the participating countries, the TIMSS studies have had political and educational effects. Opportunities are presented to compare results across countries. In other words, participants can see how badly or how well they do worldwide. All of the TIMSS studies have had scientific impact (Tatsuoka, Corter & Tatsuoka, 2004). It is important to participate in all four-year cycles to see the development of a country in educational basics; however, Turkey participated in only two of the exams at 8th grade and not 4th grade level, which shows the effect of political decisions on the Turkish educational system. Politicians appoint the people who are in the position of decision making. Newcomers do not continue with recent changes. They to change the system to correspond with what they think would be most successful.

The general framework for the TIMSS-2007 test of mathematics has two dimensions. The first is related to content and the second is related to cognition. There were three domains in mathematics at Year 4 and four at Year 8 in the content dimension. And, there were three cognitive domains in each curriculum area: *knowing, applying,* and *reasoning.* Both of the dimensions and their domains were the bases of the mathematics test. The content domains defined the specific subject matter covered by the assessment, and the cognitive domains defined the sets of skills expected of students as they engage with the content (Martinez, 2001).

Reliability and validity were concerns of TIMSS developers. The organization has procedures to ensure that the tests are reliable. Test construction is accompanied by specific instructions and assessment procedures. The test is developed with good quality items to provide reliable measurement. It is important to ensure that the results are not impacted by outside effects. Because reliability is not enough for good quality measurement, an effort also is directed to assessing the validity of the tests. The validity of test items includes unified agreement in mathematics and science for both 4th- and 8th-grade students. Agreement means that the items included in the tests measure those agreed-upon elements of mathematics and science (TIMSS-2007 Technical Report, 2008). For extensive information regarding the reliability and validity of the TIMSS-2007, see the TIMSS-2007 Technical Report. The TIMSS-2007 items were assembled into 14 blocks of mathematics items and 14 blocks of science items, and then the blocks were assembled into 14 booklets, each one including 2 blocks of mathematics items and 2 blocks of science items assembled according to a rotated design. Each student had 45 minutes of test-taking time for each part at the 8th-grade level (TIMSS-2007 Technical Report, 2008).

A cognitive diagnostic model was used for validation of TIMSS 8th grade mathematics items. To better understand the idea behind the model, a review is given below.

Cognitive Diagnostic Assessment (CDA)

One of the most important concerns in education systems is summative assessment, which evaluates the student after instruction. In order to have powerful,

effective, and meaningful summative assessment, evaluation should also be formative, which means it has to support teaching and learning processes with results (DiBello & Stout, 2007). An ideal assessment would not only be able to meet precise psychometric standards, but would also be able to provide specific feedback about how students learn and what attributes they need to achieve goals.

Diagnostic assessment is a form of examination of the cognitive processes necessary to successful task completion, because it supplies specific information about each attribute students need to master instead of a single score result (McGlohen, 2004). Traditional assessment determines what an individual has learned, but not what s/he has the capacity to learn (Embretson, 1983). A good diagnostic assessment provides good quality feedback. In other words, it shows how effective instruction and the curriculum or program is.

From a cognitive point of view "feedback is regarded as a source of information necessary for verification, elaboration, concept development, and meta-cognitive adaptation" (Narciss, 1999, p. 3). Guskey (2001) emphasizes that feedback should be diagnostic, prescriptive, and appropriate to the students' level of learning. Useful feedback comes from a constructive and objective appraisal of performance. It is given to improve a student's behavior or skills. It can be formative in nature for the purpose of modifying the learner's behavior or it may be a summative evaluation, in which a judgment is made about performance and for comparisons among learners (Bienstock et al., 2007). According to Altun (2001), feedback has two important functions in education;

a. It shows the gap between what a student learned and what s/he was supposed to learn.

b. It helps to fill the gap between these two dimensions.

In western education systems process evaluation is a large part of assessment. Ma, Çetin, and Green (2009) argue that, for better evaluation of learning, evaluators need to focus on the learning processes of individuals with stated performance outcomes. To evaluate learning processes, educational researchers have been studying the interplay of cognitive psychology and educational assessment for last 30 years. Educators, cognitive psychologists, and psychometricians have been engaged in diagnostic assessments of students' learning skills and which attributes they use during assessment processes (Birenbaum & Tatsuoka, 1993). Educational tests should characterize aspects of learning and return information to the system. They should be diagnostic (Baek, 1994).

The current Turkish examination system is based on the idea of selecting participants for opportunities in the educational system. According to the Higher Education Committee (YOK) and Student Selection and Placement Center (OSYM), students from cities on the west side of Turkey are more successful than those on the east side of Turkey. Large-scale exams select students based on the knowledge they gained in private teaching institutions in Turkey. Even though students need content knowledge to pass the exams, knowledge of test taking skills is seen as more important than content knowledge. The knowledge referred to here is test techniques rather than knowledge gained through learning the curricula. Students try to master test-taking techniques which has no effect on their cognitive development.

CDA methods have been used to structure instruction to help individuals succeed in educational opportunities, rather than selecting them for those opportunities (Stiggins, 1991). A brief explanation of the development of CDA is given below.

One of the first cognitive diagnostic models was Fischer's Linear Logistic Test Model (LLTM), created in 1972. The LLTM is an extension of the basic Rasch model to take into account the cognitive steps needed for correct item response. In the LLTM, the Rasch item difficulty is computed as a sum of discrete cognitive attribute-based difficulties. The point of the LLTM that makes it appropriate for diagnostic assessment is that the item difficulty is the composite of the influences of the basic cognitive levels, or "factors," critical for properly solving an item (Fischer, 1972).

The rule space methodology (RSM) was developed by Tatsuoka and her associates (1983), and comprises two parts. The first part involves determining the relation between the items of a test and the attributes that they are assessing. Each participant may or may not hold a mastery-level understanding of each attribute. The set of attributes which are mastered and not mastered by an individual participant are described in an attribute vector, which is also known as a "knowledge state."

The description of which items measure which attributes is shown in a Q-matrix. A Q-matrix is sometimes alluded to as an incidence matrix, but it is not used in this form in CDA. The Q-matrix is a [K x n] binary matrix, where K is the number of attributes to be measured and n is the number of items on the test. For a given element of the Q-matrix in the kth row and the ith column, a one indicates that item i does indeed measure

attribute *k* and a zero indicates it does not. Figure 2 provides an example of a 3x3 Q-matrix:

		i1	**i2**	**i3**
	k1	1	1	0
Q =	**k2**	0	0	0
	k3	1	0	1

Figure 2. Sample Q-matrix.

The first item in the above Q-matrix shows that students who mastered first and third attributes should have correct responses to item one. For a correct response to item two students should master the first attribute. A correct response to item three means they should have mastery of attribute three. Attributes which are not necessary for a correct response to the item are shown by zeros that meaning students do not need to master those attributes to correctly respond to that item. Expert consultation in the content area is a way to build the structure of the Q-matrix to determine if an item measures a specific attribute (see Appendix). Some other ways to build the structure of the Q-matrices can be borrowing from the test blueprint or subjectively evaluating each item to draw conclusions about which attributes are being measured (Tatsuoka, 1983).

DiBello, Stout, and Rousses (1995) established a new approach, the unified model, based on the rule space method. They tried to fix one of the fundamental problems of the rule space approach. In the unified model, the source of random error is

separated into different four types of systematic error. They looked at the possible sources of random errors and divided them into four different groups. The fusion model was founded on the unified model (Hartz, Roussos, & Stout, 2002) which, in other words, was based on Tatsuoka's rule space method (DiBello, Stout, & Roussos, 1995). The fusion model includes the advantages of the unified model while decreasing the number of parameters so that they can be more statistically recognizable. The unified model has $2Ki+3$ parameters for each item, while the fusion model only has $K+1$ (where K is the number of attributes).

Although interest in cognitive assessment has increased in past decades, traditional analysis methods still have a strong effect on mathematics education (Tatsuoka, Corter, & Tatsuoka, 2004). Given current issues related to differences in students' learning characteristics, effective sampling across the curricula, and recent emphases on measuring cognitive abilities, traditional mathematics tests debatably do not validly assess student abilities (Watt, 2005). According to Polya (1954), students should always be trying to understand problem-solving methods in mathematics. Polya's theories about instruction and mathematical thinking are still accepted today and supported by substantial data (Pressley, 1995). Referring to Polya's theories, CDA methods assess what mastery levels students should achieve during their mathematics instruction.

There are different cognitive assessment models in learning and teaching mathematics (Dogan & Tatsuoka, 2008; Duval, 2006; Küchemann, 1981; Tatsuoka et al.,

2004). Several studies have applied CDA to mathematics items. Their results are summarized in the following section.

Cognitive Diagnostic Assessment of Mathematic Items

Tatsuoka, Corter, and Tatsuoka (2004) examined TIMSS-R math items across 20 countries. Their results showed that high-achieving countries in the eighth-grade TIMSS-99 mathematics assessment showed their level of performance in different ways. Students in high achieving countries mostly had higher level thinking skills. Although higher level skills are very important for students to master in order to succeed at school or employment in science and technology fields, students in the industrialized countries that were not in the highest achieving cluster of the study tended to show poor scores in these higher level mathematical thinking skills.

In addition, Chen, Gorin, Thompson, and Tatsuoka (2006) conducted a cognitive diagnostic assessment of TIMSS-99. They used three analyses, including calculation of classification rates, multiple regression analyses, and comparisons of attribute mastery probabilities across four booklets. In general, a list of cognitive attributes and the incidence matrix used for the rule space model in the study predicted the performance of Taiwanese eighth graders on the TIMSS-1999 mathematics tests very well.

Dogan and Tatsuoka (2008) examined Turkish students' mathematics performance on TIMSS-R. Their study was conducted using the RSM. They used a Q-matrix that included a set of 23 attributes. Results showed that when compared to the American students, Turkish students were poor in mastering attributes such as P10

(quantitative reading), S4 (approximation/estimation), S6 (patterns and relationships), and S10 (solving open-ended problems).

Ma, Çetin, and Green (2009) studied a Turkish national assessment of eighth-grade students' performance in math administered in 2005. They used Tatsuoka and her associates' attributes to examine the validation of the test via the least squares distance method (Dimitrov, 2007). They found that many attributes predictive of the item difficulties from the Rasch model were beyond the students' abilities. The given time for students to complete the 25- item test was short. Furthermore, their results suggest that the students at lower ability levels might guess at some attributes. Overall, they also found item difficulties to be strongly predicted from the set of attributes used.

The results of the studies mentioned above are consistent, since they all used Tatsuoka's attributes. Mostly, their results are based on the subjective judgement of experts leading to the development of the Q-matrices. Generally speaking, studies that used the same data can yield different results because of use of different Q-matrices.

For the purpose of this study, a relatively new approach to cognitive diagnostic analysis called the Least Squares Distance Method (LSDM) was used to explore the validation of cognitive attributes on the released TIMSS-2007 mathematics test items with respect to the cognitive attributes developed by Tatsuoka and her associates. This approach has not been used previously with TIMSS items. The intent of this study was to extend use of the LSDM to skills and cognitive processes as well as knowledge attributes, specific to attributes as identified by Turkish education experts.

Least Squares Distance Method

The least squares distance method (Dimitrov, 2007) uses the item parameters of binary test items, gained in the framework of item response theory (IRT), to (a) validate cognitive attributes that underlie item responses and (b) assess the probability of correct processing of such attributes across levels of the scale. The LSDM is a conjunctive model in which a correct answer on a test item requires mastery of all cognitive attributes associated with that item. The cognitive attributes for all test items are outlined in a Q-Matrix, where a "1" shows an attribute is needed for an item and a "0" means an attribute is not needed. The basic assumption with LSDM is that, theoretically, the probability of correct item response is equal to the probability that all required attributes are correctly applied; which is,

$$P_{ij} = \prod_{k=1}^{K} \left[P\left(A_k = 1 \middle| \theta_i \right) \right]^{q_{jk}} \qquad (1)$$

where

$\boldsymbol{P_{ij}}$ is the probability of correct response on item j at ability level $\boldsymbol{\theta_i}$,

$\boldsymbol{P(A_k = 1 | \theta_i)}$ is the probability of correct answer on attribute A_k at ability level θ_i ,

and $\boldsymbol{q_{jk}}$ is a 0 or 1 element of the Q-matrix for item j and attribute $\boldsymbol{A_k}$.

Formula 2 is produced by taking the natural logarithm of both sides of Formula 1:

$$\ln P_{ij} = \sum_{k=1}^{K} q_{ik} \ln P\left(A_k = 1 \middle| \theta_i \right) \qquad (2)$$

Then, Formula 2 is simplified to:

$$\mathbf{L = QX} \qquad (3)$$

where

L is the vector with known elements $\mathbf{ln}\ \boldsymbol{P_{ij}}$,

Q is the Q-matrix,

and **X** is the vector with unknown elements $\mathbf{ln}\ \boldsymbol{P(A_k = 1|\ \theta_i)}$.

According to Dimitrov, generally speaking, Equation 3 does not have exact solutions since it is "overdetermined"—the number of equations [$i*j$] is greater than the number of unknowns [$k*i$]" (p. 372). To solve this problem, LSDM is used to minimize the Euclidean norm of the vector ||QX - L||. For a participant with ability level θ_i, the probability of a correct answer on attribute A_k is $P(A_k = 1|\ \theta_i) = \exp(X_k)$. Item probabilities are recovered from the attribute probabilities $P(A_k = 1|\ \theta_i)$ across ability levels. The graphical image of this probability across ability levels shows the probability curve for cognitive attribute A_k. The Rasch item characteristic curve (ICC) is represented by the recovered item probabilities (P_{rec}), or the recovery curve. The ICC recovery compared to LSDM provides information about how well the required attributes describe the item across ability levels. The mean absolute difference (MAD) between the LSDM curve and the ICC provides for validation of attributes for each item across ability levels.

MAD equal to 0.0 would indicate perfect ICC recovery. According to Dimitrov (2007), a classification for level of ICC recovery was developed: "(a) very good ($0.00 \leq$ MAD < 0.02), (b) good ($0.02 \leq$ MAD < 0.05), (c) somewhat good ($0.05 \leq$ MAD < 0.10), (d) somewhat poor ($0.10 \leq$ MAD < 0.15), (e) poor ($0.15 \leq$ MAD < 0.20), and (f) very poor (MAD ≥ 0.20)." (p. 373). These criteria were applied in this study.

Moreover, Dimitrov (2007) pointed out an interpretation of the LSDM results with respect to heuristic criteria for validation of cognitive attributes: "(1) The smaller the LSD…, the better the cognitive attributes hold together (jointly for all items) at this ability level; (2) The attribute probability curves (APCs) should exhibit logical and substantively meaningful behavior in terms of monotonicity, relative difficulty, and discrimination; (3) The better the ICC recovery for an item, the better the required attributes explain the item" (pp. 372-373).

CHAPTER 2

Method

Participants

Released items and Turkish students responses to TIMSS-2007 in mathematics administered in 2007 were used in this study. There were 4,498 8th-grade students from seven geographical regions of Turkey. There were 14 booklets which were randomly assigned to students (see Table 1). Booklets contained at least one released item. Since there are limited numbers of released items which were repeated in different booklets, all booklets were used in this study.

Table 1

Booklet Assignments to TIMSS-2007 Examinees in Turkey

Booklet Number	Number of Students
Booklet 1	314
Booklet 2	320
Booklet 3	314
Booklet 4	320
Booklet 5	320
Booklet 6	331
Booklet 7	327
Booklet 8	315
Booklet 9	324
Booklet 10	324
Booklet 11	325
Booklet 12	323
Booklet 13	325
Booklet 14	316

The study used data from 2,093 female students with a mean age of 13.96. There were 2,405 male students with a mean age of 14.09 (see Table 2).

Table 2

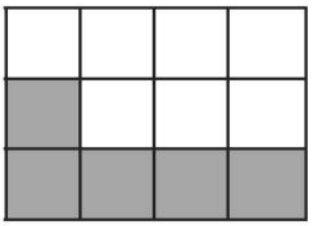
Gender and Age of TIMSS-2007 Examinees

Gender	Count	Mean Age
Girl	2093	13.96
Boy	2405	14.09

A Q-matrix was developed by two Turkish speaking mathematics teachers and the author of this paper. All of the experts have bachelor's degrees in teaching. All experts were male. Two of them work as mathematics teachers in the U.S. The ages of experts are 27, 35 and 30. The author of this paper has three years of teaching experience in Turkish schools. The other two experts have also teaching experience of three and five years.

Instrument

The TIMSS-2007 for 8th grade consisted of 179 questions which includes 96 multiple-choice questions in 14 different booklets. There were only 51 multiple-choice items released. No information was found about why TIMSS administrators released only those items. Two of the items were dropped since they did not provide any variation at all in student responses (i.e., all students answered correctly or all answered incorrectly). The final dataset was based on 49 items. For the Q-matrix, only released multiple-choice items were used. This test covered content domains of numbers, geometric shapes and measures, and data display. The cognitive domains included in the exam were knowing, reasoning, and applying. Two examples of the released items for this exam are displayed in Figures 3 and 4 below:



Which circle has approximately the same fraction of its area shaded as the rectangle above?

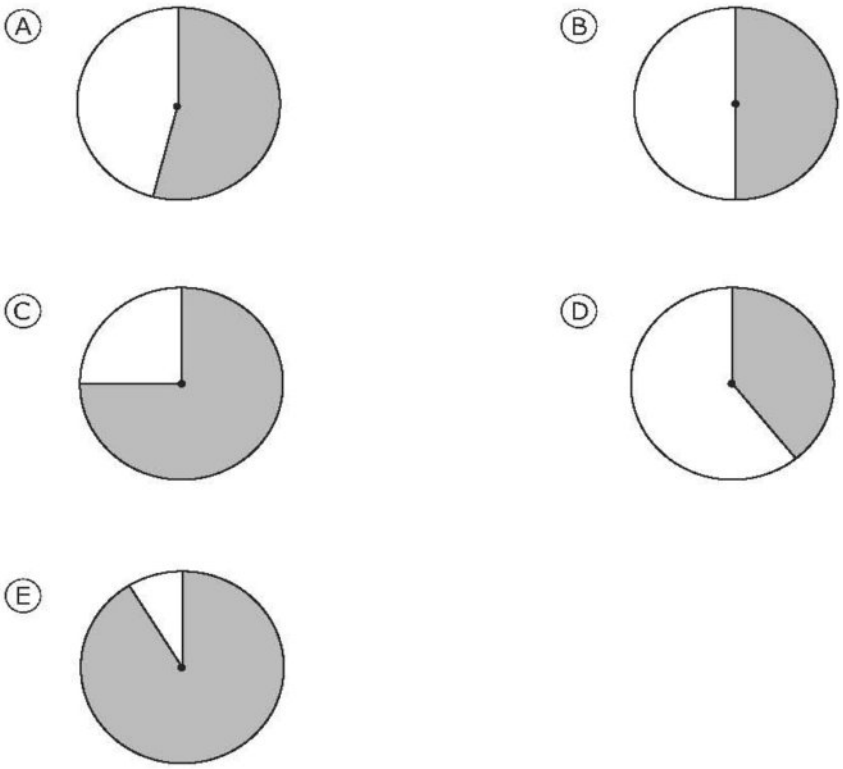


Figure 3. Sample item (TIMSS-2007 Technical Report, 2008)

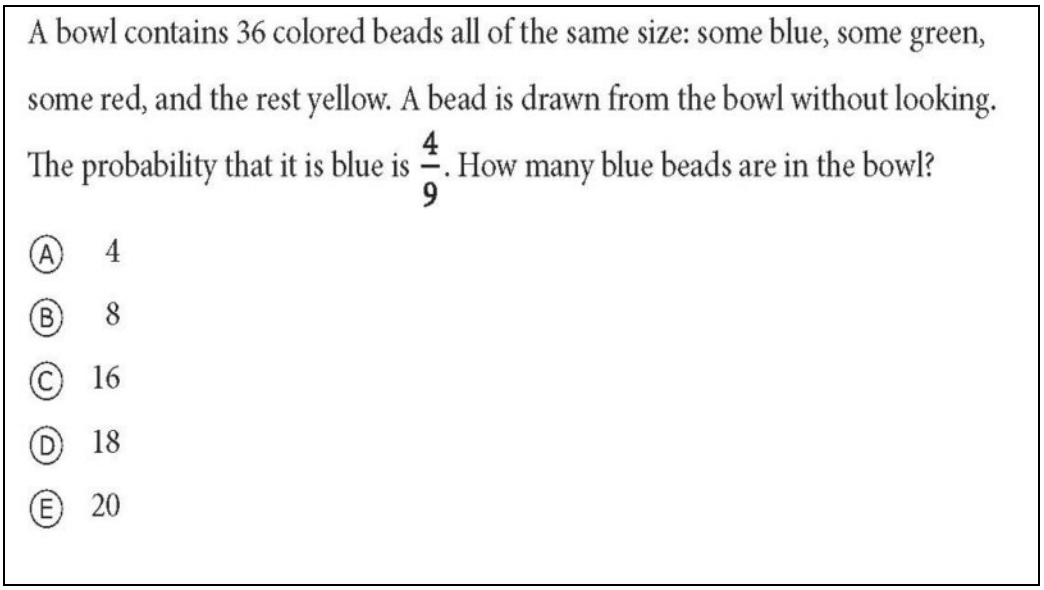


Figure 4. Sample item (TIMSS-2007 Technical Report, 2008)

Procedure

An institutional review board application was submitted prior to the study. Since the data are publicly available, the IRB committee decided that there was no need for a committee review on 06/10/2010.

TIMSS-2007 was administered to approximately 425,000 students from 59 countries around the world. Developing the TIMSS tests for 2007 was a cooperative project including experts from the participating countries. The TIMSS and PIRLS International Study Center started the process with an item-writing workshop for the National Research Coordinators from the participating countries and their colleagues.

Countries then submitted items that were reviewed by mathematicians. Participating countries pre-tested the items with samples of students. Additionally, all of the new items were reviewed by the TIMSS-2007 Science and Mathematics Item Review Committee of subject area experts (TIMSS-2007 Technical Report, 2008). After conducting the test in each participating country, data were released on the official website of TIMSS. Data used in this study were taken from the internet release of item statistics and item responses. Item parameters were also reported by the TIMSS organization.

A Q-matrix was developed by the experts. Released multiple choice items were e-mailed to the experts with an excel form of the Q-matrix to complete that included the cognitive attributes developed by Tatsuoka and her associates. The interrater reliability was .68. Most of the disagreement centered on the process attributes. After collecting independent judgments, a final meeting was organized to discuss the final version of the Q-matrix. At that meeting, agreement on the Q-matrix was negotiated.

Analysis

A Q-matrix of cognitive attributes (see Appendix) of each item was developed based on Tatsuoka and her associates' classification of cognitive attributes (K. Tatsuoka, Corter, & C. Tatsuoka, 2004). The Q-matrix includes 27 attributes divided into the three categories of content, process, and skill. Cognitive attributes which are represented by all items or not represented by all items were not used since they would provide no variation. The final version of the Q-matrix included 20 attributes. To ensure the consistency of the subjective judgment of attributes, the cognitive attribute matrix was developed based on

the independent identifications of two Turkish mathematics teachers and the author of this paper.

A correct answer on an item means that a student has mastered all attributes required, within a margin of error. In order to examine the validity of the Q-matrix, a linear regression analysis was conducted to see if the Q-matrix could explain item difficulty. This study also compared item parameters for Turkish students to the item parameters from the international data. The overall IRT parameters were taken from the official web site of TIMSS-2007 and parameters for Turkish students were obtained via use of Winsteps (Linacre, 2007) Item parameters were correlated and a scatterplot constructed to identify items that had distinctly different positions for the Turkish and international data.

The probability of correct response was calculated via the WINSTEPS program for each item for Turkish students (Linacre, 2007). Obtained results were used to perform the LSDM. The individual attribute probabilities and the average LSDs for the 49 items across ability levels were estimated using the MATLAB computer program (The MathWorks, Inc., 2005). Finally, the mean absolute difference between the ICC and the LSDM recovery curve for each item was calculated to validate the cognitive attributes.

CHAPTER 3

Results

Item Parameter Comparison

Using the responses of Turkish students to the released math items of TIMSS-2007, an IRT analysis was conducted to both compare results with international item parameters and to use for LSDM analysis. It was hypothesized that the item parameters for Turkish students would be related to the item parameters for the international data. To test this hypothesis, for the released 49 items, WINSTEPS analysis results for Turkish students and item parameters from the official website of TIMSS were used. TIMSS administrators did not report an item parameter for some of the items. Because of this, only released overall item parameters were used in the analysis. There were some items which were statistically significantly different at $p < .01$ in logit position such as items 1 (t =5.69), 25 (t =7.35), and 37 (t =8.45).

A Pearson correlation between these two sets of item logit positions produced a correlation of $r = .82$, and this correlation represents a high degree of relationship and so overall consistency in item logit position. Both overall item parameters and Turkish students' item parameters are given in Table 3.

Cognitive Attributes Matrix

Based on the author and two experts' independent identifications the cognitive attribute matrix was developed. The overall agreement level was 68%. After having a final meeting, three judges agreed on the final version of the Q-matrix. Seven attributes

(P5, P6, S4, S5, S8, and S10) which were in the original classification of cognitive attributes were deleted because no items required the attributes. One more attribute (S11) was excluded because it was present in all items. The last version of the cognitive attribute matrix had 6 content, 8 process, and 6 skill attributes, or 20 total attributes (see Table 3).

A correlation analysis was run to see the relationship between attributes. There were some statisticially significant relationships between some of the attributes. The relationship between attributes C1 (Basic concepts and operations in whole numbers and integers) and P9 (Management of data and procedures) was statisticially significant (r = .43, $p < .05$). Additionally, the conducted Pearson correlation test produced a correlation of r = ..62 between C1 (Basic concepts and operations in whole numbers and integers) and C4 (Basic concepts and operations in two-dimensional geometry), and this correlation was also significant (r = .62, $p < .05$). Moreover, for attributes P7 (Generating, visualizing, and reading figures and graphs) and S3 (Using figures, tables, charts, and graphs) the Pearson correlation was high and statistically significant (r = .89, $p < .05$). From this results, we can conclude that some attributes are comprising each other. We can also say in order to master one attribute, one needs to master another attribute.

A multiple regression analysis showed that most of the variance in item difficulties was accounted for by the identified cognitive attributes; however as can be seen from Table 3, none of the individual attributes was statistically significant. The

estimates of R^2 and adjusted R^2 were .65 and .42, respectively. Identified cognitive attributes account for approximately 65% of the variation in item difficulties.

Table 3

Item Difficulty and the 20 Initial Attributes for the 49 Items

Item	Item Difficulty Int.	Item Difficulty Turkey	Content Attributes						Cognitive Process Attributes								Skill Attributes					
			C1	C2	C3	C4	C5	C6	P1	P2	P3	P4	P7	P8	P9	P10	S1	S2	S3	S6	S7	S9
1	0.12	-0.45	1	1	0	0	0	1	0	1	1	0	0	0	1	1	0	0	1	1	1	1
2	0.95	0.42	1	1	1	0	0	0	1	1	1	1	1	1	0	1	0	1	1	0	0	1
3	-0.35	-1.08	1	1	1	0	0	0	1	1	1	0	1	1	0	0	1	1	1	1	0	1
4	0.55	-0.17	1	1	1	0	1	0	1	1	0	1	1	1	0	0	1	0	0	0	0	1
5	0.63	0.63	0	0	1	1	0	1	0	1	1	0	0	0	1	1	1	1	0	0	1	1
6	0.097	-0.4	1	1	1	0	0	0	1	1	1	1	1	1	0	1	0	1	1	0	0	0
7	-	-0.95	1	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0
8	-	-0.56	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0
9	-	0.99	0	1	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
10	-	0.70	1	1	1	0	0	0	0	1	0	0	1	1	0	0	0	0	1	0	0	1
11	-	-0.59	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	1	0	0	0	0
12	-	1.09	0	0	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1	0	1	0
13	-	0.14	0	0	1	1	0	1	1	1	1	0	0	0	1	1	0	1	0	0	1	1
14	-	-0.86	0	0	0	1	0	0	0	0	1	1	1	0	1	0	0	1	1	0	1	0
15	-	-1.89	1	0	1	0	1	0	0	1	1	0	1	0	1	0	0	1	0	1	1	0
16	-0.19	0.26	0	1	1	0	1	0	0	0	1	1	1	0	0	1	1	0	0	0	1	0
17	-0.50	-1.11	1	0	1	0	1	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1
18	-0.49	0.01	0	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	0	1	0	0
19	0.70	0.68	0	0	0	1	0	1	0	0	1	0	1	1	1	0	1	0	1	1	1	0
20	0.20	0.30	0	1	1	1	0	1	1	1	0	0	1	1	1	0	0	0	0	1	1	1
21	-0.68	-0.90	0	0	0	0	1	1	0	1	0	0	1	1	1	1	1	0	1	0	1	0
22	1.13	0.91	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	1
23	-0.93	-1.89	1	0	1	0	0	0	0	1	0	0	1	0	0	0	1	1	0	1	0	1
24	-0.01	-0.68	1	1	1	0	0	1	1	0	1	0	1	1	0	0	1	0	0	1	0	1
25	0.31	-0.48	1	0	1	0	0	0	0	1	1	0	0	0	0	1	0	1	0	1	0	0
26	-0.15	-0.63	0	0	1	1	0	1	0	1	1	0	0	0	1	1	1	0	0	0	1	1
27	-	-1.45	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0
28	-	0.20	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0

29	-	0.82	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0
30	-	0.28	0	0	1	0	0	1	0	0	1	0	1	0	1	0	1	0	0	1	0	0
31	-	0.39	1	0	1	0	0	0	0	1	0	1	0	1	0	0	0	1	0	0	0	0
32	-	1.53	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0
33	-	1.13	1	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0
34	-	0.11	0	0	0	1	0	0	0	0	1	0	1	0	1	0	1	0	1	0	1	0
35	-	0.64	0	0	1	1	0	1	1	1	1	0	0	0	1	1	0	1	0	0	1	1
36	-	0.03	1	0	1	0	1	0	0	0	1	0	1	0	0	1	1	1	1	0	0	1
37	1.05	-0.43	0	1	1	0	0	0	0	1	1	0	0	1	1	0	1	0	1	1	0	1
38	0.64	0.79	0	0	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0
39	0.03	0.37	0	0	1	1	0	1	0	1	1	0	1	0	1	0	1	1	1	0	1	0
40	0.64	0.64	0	1	1	0	1	0	0	0	1	1	1	0	0	1	1	0	1	0	1	0
41	0.89	0.25	0	1	1	0	0	0	0	1	1	1	0	0	0	1	1	1	0	1	0	0
42	1.23	1.09	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	1
43	-0.23	-1.17	0	0	1	0	0	0	1	1	1	1	0	0	1	1	1	0	0	1	1	1
44	0.05	-1.11	0	1	1	0	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0
45	0.53	-0.09	0	0	1	0	0	0	1	1	1	1	0	0	1	0	1	0	0	0	0	1
46	1.30	0.43	0	0	1	1	0	1	1	1	0	1	0	1	0	1	1	0	1	0	1	0
47	0.51	0.81	0	0	1	1	0	1	1	1	0	0	1	0	1	0	0	1	1	0	1	1
48	-0.15	0.03	0	0	1	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	1	0
49	0.91	1.22	1	1	1	0	0	0	0	1	0	1	1	0	0	0	1	1	1	1	0	1
	Regression Coefficients																						
	b		-.49	-.12	.17	.17	-.41	.59	-.34	.09	.01	.45	.07	-.12	-.91	-.44	-.27	-.26	.32	-.40	.17	.35
	p		.22	.77	.75	.75	.43	.20	.44	.87	.99	.23	.87	.75	.18	.24	.42	.46	.37	.28	.71	.42
	β		-.28	-.07	.07	.09	-.17	.32	-.18	.05	.01	.27	.04	-.07	-.54	-.26	-.17	-.16	.19	-.24	.11	.21

The LSDM was conducted across 17 ability levels in the interval from −4.0 to 4.0, with an increment of 0.5 on the logit scale. The 20 attributes were applied more accurately and consistently by higher ability examinees. The LSDM estimates of the probabilities of correct performance on each attribute across ability levels are presented in Table 4. The attribute probability curves (APCs) monotonically increase across the ability levels and provide information about the relative difficulty and discrimination of the 20 attributes. C4 (basic concepts and operations in two-dimensional geometry) was the most difficult attribute because its APC was consistently below the other APCs across all ability levels. Other attributes can be put in order (in increasing difficulty) as follows: C1, C5, P1, P3, P8, P10, S1, S9, C3, C2, S2, P9, C6, S7, P2, P7, S6, P4, S3, and C4.

Table 4

Estimates of Probability for Correct Performance on 20 Abilities across 17 Ability Levels

		Content Attributes						Cognitive Process Attributes								Skill Attributes					
Ability Level	LSD's	C1	C2	C3	C4	C5	C6	P1	P2	P3	P4	P7	P8	P9	P10	S1	S2	S3	S6	S7	S9
-4.00	0.136	1.000	1.000	1.000	0.168	1.000	0.804	1.000	0.421	1.000	0.265	0.393	1.000	0.869	1.000	1.000	0.907	0.198	0.391	0.596	1.000
-3.50	0.127	1.000	1.000	1.000	0.180	1.000	0.795	1.000	0.443	1.000	0.293	0.446	1.000	0.988	1.000	1.000	0.952	0.233	0.446	0.663	1.000
-3.00	0.118	1.000	0.977	1.000	0.199	1.000	0.800	1.000	0.479	1.000	0.321	0.505	1.000	1.000	1.000	1.000	0.976	0.278	0.564	0.775	1.000
-2.50	0.109	1.000	0.941	1.000	0.221	1.000	0.816	1.000	0.521	1.000	0.357	0.571	1.000	1.000	1.000	1.000	0.909	0.336	0.676	0.896	1.000
-2.00	0.100	1.000	0.924	1.000	0.253	1.000	0.850	1.000	0.564	1.000	0.404	0.634	1.000	1.000	1.000	1.000	0.996	0.409	0.789	1.000	1.000
-1.50	0.089	1.000	0.925	1.000	0.313	1.000	0.919	1.000	0.617	1.000	0.470	0.698	1.000	1.000	1.000	1.000	0.983	0.500	0.882	1.000	1.000
-1.00	0.078	1.000	0.948	1.000	0.392	1.000	0.945	1.000	0.673	1.000	0.533	0.751	1.000	1.000	1.000	1.000	0.966	0.602	0.943	1.000	1.000
-0.50	0.065	1.000	0.980	0.578	0.514	1.000	1.000	1.000	0.740	1.000	0.640	0.800	1.000	1.000	1.000	1.000	0.937	0.705	0.972	1.000	1.000
0.00	0.053	1.000	1.000	0.743	0.654	1.000	1.000	1.000	0.808	1.000	0.734	0.850	1.000	1.000	1.000	1.000	0.921	0.797	0.975	1.000	1.000
0.50	0.041	1.000	1.000	0.876	0.785	1.000	1.000	1.000	0.873	1.000	0.822	0.901	1.000	1.000	1.000	1.000	0.924	0.873	0.972	1.000	1.000
1.00	0.030	1.000	1.000	0.958	0.885	1.000	1.000	1.000	0.928	1.000	0.889	0.944	1.000	1.000	1.000	1.000	0.940	0.930	0.973	1.000	0.991
1.50	0.021	1.000	1.000	0.990	0.945	1.000	1.000	1.000	0.963	1.000	0.938	0.972	1.000	1.000	1.000	1.000	0.962	0.965	0.981	1.000	0.992
2.00	0.014	1.000	1.000	0.999	0.975	1.000	1.000	1.000	0.983	1.000	0.969	0.987	1.000	1.000	1.000	1.000	0.979	0.983	0.989	1.000	0.996
2.50	0.009	1.000	1.000	1.000	0.989	1.000	1.000	1.000	0.993	1.000	0.986	0.945	1.000	1.000	1.000	1.000	0.990	0.993	0.995	1.000	0.998
3.00	0.006	1.000	1.000	1.000	0.995	1.000	1.000	1.000	0.997	1.000	0.994	0.998	1.000	1.000	1.000	1.000	0.996	0.997	0.998	1.000	0.999
3.50	0.004	1.000	1.000	1.000	0.998	1.000	1.000	1.000	0.999	1.000	0.997	0.999	1.000	1.000	1.000	1.000	0.998	0.999	0.999	1.000	1.000
4.00	0.002	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.999	1.000	0.999	1.000	1.000	1.000	1.000	1.000	0.999	0.999	0.100	1.000	1.000

The APCs obtained with these probabilities are graphed in Figure 5. C4 was the most difficult attribute (the lowest curve), followed in decreasing difficulty by S3, P4, P2, P7, S6, S7, C6, P9, S2, and C2. All the other attributes have similar difficulty values. For example, the probability of correct performance on C4 at the ability level y=0 was .654 (see Table 4). That is, the likelihood of examinees with ability at the origin of the logit scale to correctly process geometrical operations in two dimensional geometry (C4) was 65.4 %. For the same examinees, the likelihood to correctly process algebra operations in elementary algebra (C3) was higher (74.3%).

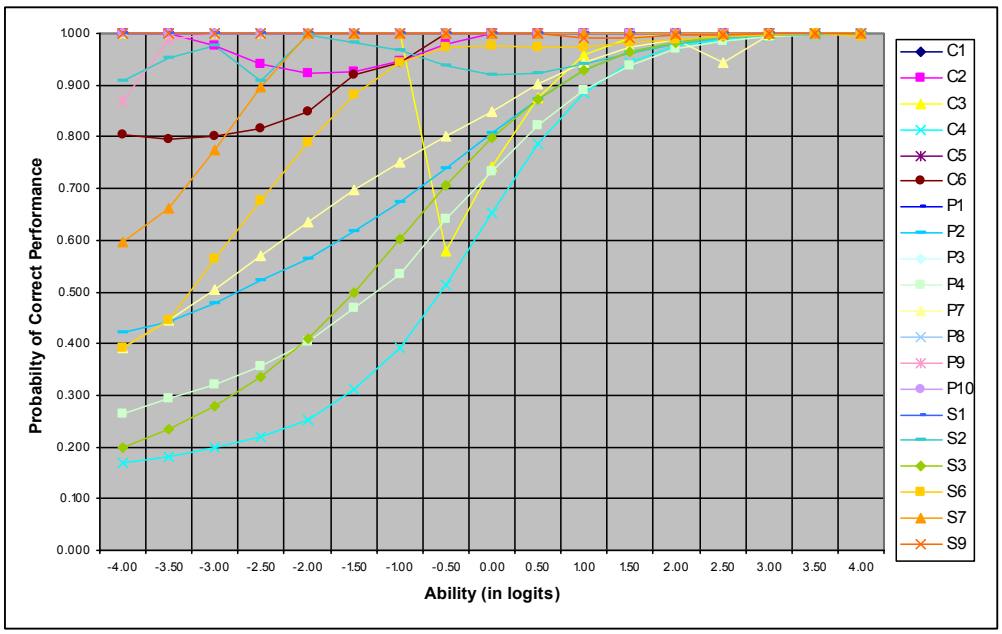


Figure 5. Attribute Probability Curves

For each item, the mean absolute differences for the ICC recovery across ability levels are given in Table 5. Graphically, the ICC recovery is presented (Figure 5) for four examples of items (47, 40, 19, and 17) with differing MAD levels. According to Dimitrov's criteria, item 47 is in the category of very good with a MAD of .017, item 40 is in the category of good with a MAD of 0.036, item 19 is in the category of somewhat

good with a MAD of .060 and item 17 is in the category of somewhat poor with a MAD of .14. With the conventional rule for degree of ICC recovery described earlier in the LSDM section of this paper, the examination of all 49 graphs for ICC recovery and their MAD values revealed that the ICC recovery was very good for four items (20, 36, 41, and 47), good for 16 items (1, 2, 5, 11, 13, 16, 24, 25, 28, 31, 34, 35, 39, 40, 45, and 46), somewhat good for 15 items (4, 7, 10, 18, 19, 21, 22, 29, 30, 33, 37, 38, 42, 48, and 49), somewhat poor for eight items (6, 8, 9, 12, 17, 26, and 27), poor for four items (3, 14, 43, and 44) and very poor for two items (15 and 23). Generally speaking, such diagnostic information on ICC recovery can be particularly useful in validating math sub-skills for students.

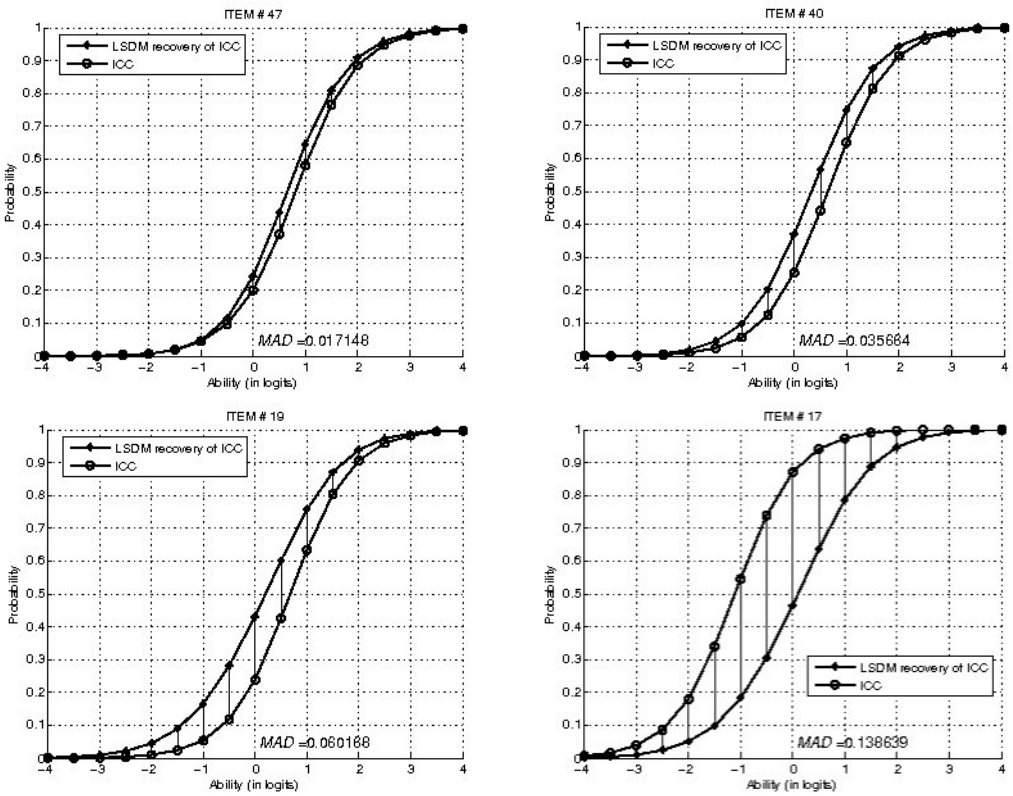


Figure 6. Item Characteristic Curve Recovery with the LSDM for Four Sample Items

Overall, the findings indicate that the 20 attributes relate to difficulties in mathematical skills of students (Mean MAD = .075). The mean MAD value suggests that overall item recovery is somewhat good based on the Dimitrov's criteria. For this reason, the APCs of the 20 attributes provide valuable information in terms of their difficulty (see Figure 5). But the results for ICC recovery suggest that there is room for improvement regarding the set of attributes and their links to items in the Q-matrix. Indeed, using Dimitrov's criteria compared to the MAD values of items in Table 5, 35 items have very good, good, or somewhat good ICC recovery while 14 items do not. According to these results, it might be said that the Q-matrix can be improved to get better results since 14 items were not well-recovered; for example, see the location of Items 47 and 17 in Figure 6 above (p. 33).

Table 5

Absolute Differences for Item Characteristic Curve Recovery with the Least Squares Distance Method

Item	Ability (logits)																	
	-4.0	-3.5	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	MAD
1	0.024	0.006	0.013	0.030	0.067	0.143	0.282	0.479	0.683	0.834	0.922	0.965	0.985	0.993	0.997	0.999	1.000	0.043
2	5.403	0.001	0.003	0.007	0.016	0.037	0.082	0.173	0.329	0.534	0.729	0.863	0.936	0.972	0.988	0.995	0.998	0.024
3	0.007	0.016	0.037	0.082	0.173	0.329	0.534	0.729	0.863	0.936	0.972	0.988	0.995	0.998	0.999	1.000	1.000	0.173
4	0.002	0.003	0.008	0.019	0.043	0.094	0.196	0.363	0.572	0.758	0.880	0.945	0.976	0.990	0.996	0.998	0.999	0.058
5	3.780	8.848	0.002	0.005	0.011	0.026	0.059	0.128	0.255	0.445	0.652	0.815	0.912	0.960	0.983	0.993	0.997	0.033
6	0.002	0.005	0.012	0.027	0.062	0.133	0.265	0.458	0.664	0.822	0.916	0.962	0.984	0.993	0.997	0.999	0.999	0.119
7	0.006	0.013	0.030	0.067	0.143	0.282	0.479	0.683	0.834	0.922	0.965	0.985	0.993	0.997	0.999	1.000	1.000	0.066
8	0.003	0.007	0.016	0.036	0.079	0.168	0.321	0.526	0.722	0.859	0.934	0.971	0.987	0.995	0.998	0.999	1.000	0.102
9	2.048	4.790	0.001	0.003	0.006	0.014	0.033	0.073	0.156	0.303	0.504	0.704	0.848	0.929	0.968	0.986	0.994	0.112
10	3.355	7.855	0.002	0.004	0.010	0.023	0.053	0.115	0.233	0.416	0.625	0.796	0.901	0.955	0.980	0.992	0.996	0.055
11	0.003	0.007	0.016	0.037	0.083	0.175	0.332	0.538	0.732	0.865	0.937	0.972	0.988	0.995	0.998	0.999	1.000	0.049
12	1.728	4.046	9.471	0.002	0.005	0.012	0.028	0.063	0.135	0.268	0.462	0.668	0.825	0.917	0.963	0.984	0.993	0.119
13	8.869	0.002	0.005	0.011	0.026	0.058	0.126	0.252	0.441	0.649	0.812	0.910	0.960	0.982	0.992	0.997	0.999	0.025
14	0.005	0.011	0.026	0.058	0.126	0.252	0.441	0.649	0.812	0.910	0.960	0.982	0.992	0.997	0.999	0.999	1.000	0.165
15	0.027	0.061	0.131	0.262	0.453	0.660	0.820	0.914	0.962	0.983	0.993	0.997	0.999	0.999	1.000	1.000	1.000	0.232
16	7.093	0.002	0.004	0.009	0.021	0.048	0.105	0.215	0.391	0.601	0.779	0.892	0.951	0.978	0.991	0.996	0.998	0.023
17	0.007	0.017	0.039	0.086	0.180	0.340	0.547	0.739	0.869	0.939	0.973	0.988	0.995	0.998	0.999	1.000	1.000	0.139
18	0.001	0.003	0.006	0.014	0.032	0.071	0.152	0.296	0.496	0.697	0.844	0.927	0.967	0.986	0.994	0.997	0.999	0.054
19	3.472	8.127	0.002	0.004	0.010	0.024	0.054	0.118	0.239	0.424	0.633	0.802	0.904	0.957	0.981	0.992	0.997	0.060
20	6.627	0.002	0.004	0.008	0.020	0.045	0.099	0.204	0.375	0.584	0.767	0.885	0.948	0.977	0.990	0.996	0.998	0.019
21	0.005	0.012	0.027	0.062	0.133	0.265	0.458	0.664	0.822	0.916	0.962	0.984	0.993	0.997	0.999	0.999	1.000	0.078
22	2.347	5.496	0.001	0.003	0.007	0.016	0.037	0.083	0.175	0.332	0.538	0.732	0.865	0.937	0.972	0.988	0.995	0.068
23	0.027	0.061	0.131	0.262	0.453	0.660	0.820	0.914	0.962	0.983	0.993	0.997	0.999	0.999	1.000	1.000	1.000	0.233
24	0.004	0.008	0.019	0.043	0.096	0.199	0.367	0.576	0.761	0.882	0.946	0.976	0.990	0.996	0.998	0.999	1.000	0.041
25	0.003	0.006	0.014	0.031	0.070	0.150	0.292	0.492	0.694	0.841	0.926	0.967	0.986	0.994	0.997	0.999	1.000	0.043
26	0.003	0.008	0.017	0.040	0.089	0.185	0.348	0.555	0.745	0.873	0.941	0.974	0.989	0.995	0.998	0.999	1.000	0.103
27	0.013	0.030	0.067	0.143	0.282	0.479	0.683	0.834	0.922	0.965	0.985	0.993	0.997	0.999	1.000	1.000	1.000	0.100
28	7.855	0.002	0.004	0.010	0.023	0.053	0.115	0.233	0.416	0.625	0.796	0.901	0.955	0.980	0.992	0.996	0.998	0.044
29	2.736	6.405	0.002	0.004	0.008	0.019	0.043	0.096	0.199	0.367	0.576	0.761	0.882	0.946	0.976	0.990	0.996	0.075
30	6.856	0.002	0.004	0.009	0.020	0.046	0.102	0.210	0.383	0.593	0.773	0.889	0.949	0.978	0.990	0.996	0.998	0.076
31	5.686	0.001	0.003	0.007	0.017	0.039	0.086	0.180	0.340	0.547	0.739	0.869	0.939	0.973	0.988	0.995	0.998	0.022
32	8.173	1.914	4.481	0.001	0.003	0.006	0.013	0.031	0.069	0.148	0.289	0.487	0.690	0.839	0.924	0.966	0.985	0.145
33	1.614	3.780	8.848	0.002	0.005	0.011	0.026	0.059	0.128	0.255	0.445	0.652	0.815	0.912	0.960	0.983	0.993	0.098
34	9.154	0.002	0.005	0.012	0.027	0.061	0.131	0.262	0.453	0.660	0.820	0.914	0.962	0.983	0.993	0.997	0.999	0.022
35	3.716	8.699	0.002	0.005	0.011	0.026	0.058	0.126	0.252	0.441	0.649	0.812	0.910	0.960	0.982	0.992	0.997	0.034
36	0.001	0.003	0.006	0.013	0.031	0.069	0.148	0.289	0.487	0.690	0.839	0.924	0.966	0.985	0.994	0.997	0.999	0.019
37	0.002	0.005	0.012	0.029	0.065	0.139	0.275	0.470	0.675	0.830	0.919	0.964	0.984	0.993	0.997	0.999	1.000	0.063
38	2.879	6.740	0.002	0.004	0.009	0.020	0.045	0.100	0.207	0.379	0.588	0.770	0.887	0.948	0.977	0.990	0.996	0.058

39	5.883	0.001	0.003	0.008	0.017	0.040	0.089	0.185	0.348	0.555	0.745	0.873	0.941	0.974	0.989	0.995	0.998	0.034
40	3.716	8.699	0.002	0.005	0.011	0.026	0.058	0.126	0.252	0.441	0.649	0.812	0.910	0.960	0.982	0.992	0.997	0.036
41	7.215	0.002	0.004	0.009	0.021	0.048	0.107	0.218	0.395	0.605	0.782	0.894	0.952	0.979	0.991	0.996	0.998	0.018
42	1.728	4.046	9.471	0.002	0.005	0.012	0.028	0.063	0.135	0.268	0.462	0.668	0.825	0.917	0.963	0.984	0.993	0.089
43	0.008	0.019	0.043	0.094	0.196	0.363	0.572	0.758	0.880	0.945	0.976	0.990	0.996	0.998	0.999	1.000	1.000	0.161
44	0.007	0.017	0.039	0.086	0.180	0.340	0.547	0.739	0.869	0.939	0.973	0.988	0.995	0.998	0.999	1.000	1.000	0.166
45	0.001	0.003	0.007	0.016	0.037	0.083	0.175	0.332	0.538	0.732	0.865	0.937	0.972	0.988	0.995	0.998	0.999	0.028
46	5.312	0.001	0.003	0.007	0.016	0.036	0.081	0.170	0.325	0.530	0.725	0.861	0.935	0.971	0.988	0.995	0.998	0.030
47	2.783	6.515	0.002	0.004	0.008	0.019	0.044	0.097	0.201	0.371	0.580	0.764	0.883	0.947	0.977	0.990	0.996	0.017
48	0.001	0.003	0.006	0.013	0.031	0.069	0.148	0.289	0.487	0.690	0.839	0.924	0.966	0.985	0.994	0.997	0.999	0.053
49	1.385	3.243	7.592	0.002	0.004	0.010	0.022	0.051	0.111	0.227	0.408	0.617	0.790	0.898	0.954	0.980	0.991	0.065

Discussion

Cognitive diagnostic assessment is generally said to be a useful way to examine validation of test items (Tatsuoka, Corter, & Tatsuoka, 2004). The approach used in this study tests validation of cognitive attributes required by each item across all ability levels. To do this, the method requires item position parameters from Rasch or IRT models and correctly identified attributes by experts in the area. The idea behind the model is to overcome the difficulty in obtaining data on student performance on an individual attribute. With information about student performance on an individual attribute in hand, instruction can be tailored to the individual attributes level and then to overall success on item solution. With the LSDM, the present study investigated the validation of cognitive attributes on TIMSS-2007 items for Turkish students. The attributes used in this paper were based on prior research by using the three types of attributes defined by Tatsuoka and her associates. First, the cognitive validation of the items was evaluated using the 27 initially-identified attributes. Once independent responses of experts were collected, a lack of variability was found resulting in deletion of seven of the attributes. In addition to this, two of the items were deleted due to a lack of variability. The degree of the LSDM recovery of ICC was then assessed for the 49 items. Additionally, item parameters for both international and Turkish students were analyzed in order to assess the correlation between those parameters and so the generalizability of results.

The results showed the validation of cognitive attributes for the test with respect to the 20 revised attributes. The monotonic decrease of LSDs across ability levels

demonstrated that it was unlikely that different cognitive strategies had been employed by different ability students, and that the higher ability students applied the attributes more accurately and consistently . The APCs generally indicated relative attribute difficulties and clear discriminations. The LSDM recovery of ICCs showed that on average, 71% of the items were recovered well by the attribute probabilities, revealing that the most of the 49 items could be substantially explained by the identified attributes. But the LSDM recovery also suggested that there is need for modifying attributes for some items such as Items 3, 6, 8, 9, 12, 14, 15, 17, 23, 26, 27, 43,and 44, because they were not explained very well by the attributes.

With respect to the three categories of attributes, the present research generated some useful results. It was found that identifying the content attributes was easier than identifying the cognitive process and skill attributes. The MAD values also showed that the content attributes accounted better for the items than the other two kinds of attributes and so, the LSDM recovery of ICCs was more accurate for content attributes. The findings exhibited that although all attributes together contributed to the correct response for an item, their overlap might lead to unclear results. Some of the attributes include the same mastery areas. For example P7 and S3 are overlapping on usage of figures and graphs. This shows that it might not be reasonable to combine different types of attributes in one analysis. The most difficult attribute for students to master was C4 (basic concepts and operations in two-dimensional geometry). This indicates that Turkish students' level of mastery is not sufficient in geometry. Teachers should consider concentrating on students geometry knowledge, in order to get better outcomes.

To assess the stability of results, it is suggested that in future study an LSDM analysis be run with a random Q-matrix or with Q-matrices of experts run separately. In addition to this, an aggregate Q-matrix can be iteratively revised for poor items to decrease the MAD and increase the multiple correlation between attributes and item difficulties. If an item could be written to uniquely address a specific attribute, these results would provide concurrent validation evidence and could be the best scenario to see how well students are doing on certain attributes. It is feasible that items might be constructed for content attributes and less so for process and skill attributes, which are conveyed via a content item. Thus process and skill attributes would be overlaid on content.

This study was well developed with a large sample of students and the findings represented most of the released items in the mathematics test for the eighth grade in Turkey. But some caveats should be considered. The first limitation is the selection of the attributes used in this research. With respect to the cognitive model, if the same attributes are shown as relevant for two different items, the item difficulties of the two items should be close. Therefore, large differences in item difficulties would signify the misspecification or inadequacy of the chosen attributes for the items. Because of this, items 2 (b = .42) and 3 (b = -1.08), items 23 (b = -1.89) and 28 (b = .20), and items 27 (b = -1.45) and 33 (b = 1.13) were flagged for potential problems in the specification of the attributes since similar attributes were identified but the logit difficulties were very different. Additionally, to clearly see the relationship between items and attributes, simple attributes are better than complex attributes.

Moreover, no guessing or omissions were taken into account in the LSDM (Dimitrov, 2007). In the current study, since the results from the IRT model showed that most of the items were beyond the students' abilities, it is reasonable that there were slips and guessing. Students at lower ability levels might guess at some items. Nevertheless, these problems with the LSDM were not taken into account in this study. Future studies might consider mistakes and guessing with the LSDM.

As a result of this study, educators in Turkey should note that students have weaker knowledge in geometry (C4), graphics, charts, figures (P7), rule application in algebra (P4), and data management (P9). This study showed that students' levels of mastery of cognitive attributes are important in order to get better results at the international level. Outcomes of education need to meet international criteria and the needs of a global economy. Since economic relationships are becoming more global in the world, it is more important than ever to succeed at the international level.

Industrialization of the country is very important to supply more jobs to unemployed members of the younger generation. During its development process, educational policies must shape Turkey's steps. This study shows that students have weak levels of mastery in certain areas, some of which are related to engineering and physical sciences. The results indicates that the administrators of educational systems in Turkey should consider groundbreaking changes in teaching geometry, because geometry may enable teaching of important mathematical thinking skills which are needed in physical science and engineering. These skills are of course very important in maintaining technological development in today's global industries (see Attributes C4,

P2, and S6). Furthermore, students are not good at reading figures, graphs and tables. Also, the results are not good enough in basic algebra for a country that is on its way to taking part in the European Union. Teachers might choose to focus on cognitive process attributes more because they were four of the seven most difficult attributes. Improving students' processing levels could produce better results at the international level.

For example, to improve mastery of attribute C4, teachers could consider student-centered learning theories. Instead of saying what a triangle is, they might give examples from students' lives. Students could participate in several activities during their classes, which is possible if classrooms are not overcrowded. The number of students in one class should not be more than 24 students (EU Criteria). In addition, although the average number of students per teacher for Turkish 8th grade classrooms is lower than in many European countries, the Turkish government should balance it throughout the country. There are some areas where there are around 50-55 students in a classroom. There are also some areas where there are only 10-12 students in a classroom. The government should build new schools in more crowded areas of the country to allow greater use of class activities. Districts must also supply class activity materials to schools. Although eight years of education in elementary schools is compulsory and free for all students, schools are struggling with economic hardships. The ministry of national education should revise its financial role in compulsory education system. Districts should supply money that schools can use for materials based on the number of students attending.

A country's development is highly connected to its curricula. Better quality outcomes from the educational system can only be supplied by better quality curricula.

The results of the present study could help teachers and curriculum developers ensure that the utilized educational policies and methodologies would help students to improve their cognitive mastery levels.

References

Altun, M. (2005). Ogretim hizmetinin ogeleri. *Acik Ogretim Fakultesi,* 59. Ankara: Anadolu Universitesi.

Baek, S. G. (1994). Implications of cognitive psychology for educational testing. *Educational Psychology Review*, 6(4), 373-389.

Bienstock, J., Katz, N., Cox, S., Hueppchen, N., Erickson, S., & Puscheck, E. (2007). To the point: Medical education reviews-providing feedback. *American Journal of Obstetric & Gynecology*, 508-513.

Birenbaum, M., & Tatsuoka, K. K. (1993). Applying an IRT-based cognitive diagnostic model to diagnose students' knowledge states in multiplication and division with exponents. *Applied Measurement in Education, 6,* 255-268.

Bolt, D. (2007). The present and future of IRT-based cognitive diagnostic models (ICDMs) and related methods. *Journal of Educational Measurement*, *44*(4), 377-383.

Chen, Y., Gorin, J., Thompson, M., & Tatsuoka, K. (2006, April). *Verification of cognitive attributes required to solve the TIMSS-1999 mathematics items for Taiwanese students.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, US.

DiBello, L. V., & Stout, W. (2007). Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, *44*, 285–291.

DiBello, L., Stout, W., & Rousses, L. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman, and R. L. Brennan (Eds.), *Cognitive Diagnostic Assessment* (pp. 327-361). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dimitrov, D. M. (2007). Least squares distance method of cognitive validation and analysis for binary items using their item response theory parameters. *Applied Psychological Measurement, 31*(5), 367-387.

Dogan, E., & Tatsuoka, 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.

Duval, R. (2006). A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics*, *61*, 103–131.

Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin, 93*, 179-197.

EU-European Union, 2006. Summaries of legislation. Available from /http://europa.eu.int/scadplus/leg/en/cha/c11063.htmS.Access date: 24 Apr 2010.

Fisher, G. (1973). Linear logistic test model as an instrument in educational research. *Psychologica, 37,* 359-374.

Gelbal, S., & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33*, 135-147.

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

Gözütok, D., Akgün, Ö. E., & Karacaoğlu, C. (2005). İlköğretim programlarının öğretmen yeterlilikleri açısından değerlendirilmesi. Yeni İlköğretim programlarını değerlendirme sempozyumu, 14-16 Kasım 2005, Kayseri. 17-39.

Guskey, T. (2001, April) *Bloom's Contributions to Curriculum, Instruction, and School Learning.* Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.

Hartz, S.. Roussos, L., & Stout, W. (2002) *Skills diagnosis: Theory and practice*. User Manual for Arpeggio software. ETS.

Kellough, R. D., & Kellough, N. G. (2007). *Secondary school teaching: A guide to methods and resources* (3rd ed.). Upper Saddle River, NJ: Pearson.

Kuchemann, D. (1981). Cognitive demands of secondary school mathematics items. *Educational Studies in Mathematics, 12*, 301-316.

Linacre, J. M. (2007). Winsteps (Version 3.64.2) [Computer Software]. Chicago: Winsteps.com.

Ma, L., Cetin, E., & Green, K., E. (2009, April). *Cognitive assessment in mathematics with the least squares distance method.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, US.

Martinez, J. (2001) Exploring, inventing, and discovering mathematics: A pedagogical response to TIMSS. *Mathematics Teaching in Middle School, 7(2),* 114-120.

The MathWorks, Inc. (2005). MATLAB (Version 7.1) [Computer software]. Natick, MA: The MathWorks, Inc.

McGlohen, M. (2004). *The application of cognitive diagnosis and computerized adaptive testing to a large-scale assessment.* Unpublished Doctoral Dissertation, University of Texas at Austin, TX.

MEB (2005). EARGED İlköğretim 1-5 Sınıf Pilot Uygulama Sonuçlarının Değerlendirilmesi.

MEB (2005). EARGED ÖBBS Projesi (Bilgisayar ve İngilizce Okuryazarlığı) 2004 Uygulama Raporları.

MEB (2005). EARGED PISA Projesi 2003 Uygulaması Ulusal Raporu.

MEB (2006). EARGED İlköğretim 6. Sınıf Pilot Uygulama Sonuçlarının Değerlendirilmesi.

MEB (2007). EARGED ÖBBS Projesi (Türkçe, Matematik, Fen Bilgisi ve Sosyal Bilgiler) 2005 Uygulama Raporları.

MEB-EARGED (2003). Üçüncü Uluslararası Fen ve Matematik Çalışması (TIMSS 1999) Ulusal Rapor [Third international mathematics and science study (TIMSS 1999) national report]. Ankara: Turkey.

MONE, 2001. Turkish education system and developments in education. Forty-sixth session of the international conference on education for all learning to live together: contents and learning strategies—problems and solutions. Available from /http://www.ibe.unesco.org/International/ICE/natrap/Turkey.pdfS. Access date: 1 May 2010.

Narciss, S. (1999, April). *Motivational effects of the informativeness of feedback.* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada.

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). *TIMSS 2007 Technical Report.* Chestnut Hill, MA: Boston College, TIMSS & PIRLS International Study Center, 2008; http://timss.bc.edu/TIMSS2007/techreport.html.

OOEGM—Okul Oncesi Egitim Genel Mudurlugu , 2005.Yapılan ve devam eden calısmalar. Available from /http://ooegm.meb.gov.tr/faaliyetler.htmlS. Access date: 20 Apr 2010.

Polya, G. (1954). *Mathematics and plausible reasononing I. Induction and analogy in mathematics II. Patterns of plausible inference.* Princeton, NJ: Princeton University Press.

Pressley, M. (1995). *Cognition, teaching and assessment.* New York: HarperCollins.

Shen,C (2000).The relationship between student's achievement and their self-perception of competence and rigour of mathematics and science: A cross-national analysis. *Assessment in Education, 7*,237-253.

Stiggins, R. J. (1991). Facing the challenges of a new era of educational assessment. *Applied Measurement in Education, 4,* 263-273.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement 20*(4), 901-926.

Tatsuoka, K. K., Corter, J. E., & 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.

UNDP, 2004. Country evaluation: Assessment of development results. Turkey. Available from /http://www.undp.org.trS.Access date: 17 Apr 2010.

Watt, H. (2005). Attitudes to the use of alternative assessment methods in mathematics: A study with secondary mathematics teachers in Sydney, Australia. *Educational Studies in Mathematics, 58*, 21-44.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance.* San Francisco: Jossey-Bass.

Yapıcı, M.& C. Demirdelen (2007). İlköğretim 4. sınıf programına ilişkin öğretmen görüşleri. İlköğretim Online, 6(2), 204-212.

Yaşar, Ş., M. Gültekin, B. Türkkan, N. Yıldız & P. Girmen (2005). Yeni ilköğretim programlarının uygulanmasına ilişkin sınıf öğretmenlerinin hazırbulunuşluk düzeylerinin ve eğitim gereksinimlerinin belirlenmesi: Eskişehir ili örneği. Yeni İlköğretim programlarını değerlendirme sempozyumu, 14-16 Kasım 2005, Kayseri. 50-63.

Appendix

The Content, Cognitive Process and Skill Attributes for the TIMSS-R (1999)

Content attributes

C1 Basic concepts and operations in whole numbers and integers
C2 Basic concepts and operations in fractions and decimals
C3 Basic concepts and operations in elementary algebra
C4 Basic concepts and operations in two-dimensional geometry
C5 Data, probability, and basic statistics
C6 Measuring or estimating: length, time, angle, temperature, etc.

Cognitive Process attributes

P1 Translate/formulate equations and expressions to solve a problem
P2 Computational applications of knowledge in arithmetic and geometry
P3 Judgmental applications of knowledge in arithmetic and geometry
P4 Applying rules in algebra
P5 Logical reasoning-includes case reasoning, deductive thinking skills, if-then, necessary and sufficient, generalization skills (Deleted)
P6 Problem search; analytic thinking, problem restructuring; inductive thinking (Deleted)
P7 Generating, visualizing, and reading figures and graphs
P8 Applying and evaluating mathematical correctness
P9 Management of data and procedures
P10 Quantitative and logical reading

Skill (item type) attributes

S1 Unit conversion
S2 Apply number properties and relationships; number sense/number line
S3 Using figures, tables, charts, and graphs
S4 Approximation/estimation (Deleted)
S5 Evaluate/verify/check options (Deleted)
S6 Patterns and relationships (inductive thinking skills)
S7 Using proportional reasoning
S8 Solving novel or unfamiliar problems (Deleted)
S9 Comparison of two/or more entities
S10 Open-ended items, in which an answer is not given (Deleted)
S11 Understanding verbally posed questions (Deleted)