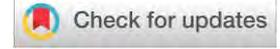# A diagnostic assessment to middle school students' proportional reasoning

Muhammet, Arıcan 🆔

Kırşehir Ahi Evran University, Faculty of Education, Kırşehir, Turkey, muhammetarican@gmail.com

Check for updates

ABSTRACT   This study investigated Turkish middle school students' proportional reasoning and provided a diagnostic assessment of their strengths and weaknesses on the ratio and proportion concepts. A proportional reasoning test with 22 multiple-choice items was developed from the context of the log-linear cognitive diagnosis model. The test was developed around four core cognitive skills (attributes) that required in solving middle school ratio and proportion problems. These skills included understanding ratios, directly, inversely, and nonproportional relationships. The test was applied to 282 seventh grade students, and the collected data were analyzed using the Mplus software. The analysis showed that approximately 62% of the students were able to recognize directly proportional relationships. Whereas, roughly 48% of them were able to recognize inversely proportional relationships. Moreover, while 25% of the students did not master any of the four cognitive skills, 39.1% mastered all four of these skills. In addition, many students had difficulty distinguishing proportional relationships from nonproportional relationships. Diagnostic feedbacks on the students' strengths and weaknesses were provided based on the findings.

Keywords: *Diagnostic assessment, Diagnostic classification models, Middle school students, Proportional reasoning, Ratios and proportions.*

Ortaokul öğrencilerinin orantısal akıl yürütmeleri üzerine tanısal bir değerlendirme

ÖZ   Bu çalışmada ortaokul öğrencilerinin orantısal akıl yürütmeleri araştırılmış ve oran ve orantı konuları için güçlü ve zayıf yönlerinin bilişsel bir tanısal değerlendirmesi sağlanmıştır. Yirmi iki çoktan seçmeli madde içeren bir orantısal akıl yürütme testi log-linear bilişsel tanı modeli perspektifinden faydalanılarak geliştirilmiştir. Test, ortaokul öğrencilerinin oran ve orantı problemlerini çözmeleri için gerekli olan dört temel bilişsel beceri etrafında tasarlanmıştır. Bu beceriler sırasıyla oran, doğru orantılı ilişki, ters orantılı ilişki ve orantısal olmayan ilişki kavramlarını anlamayı içermektedir. Test 282 yedinci sınıf öğrencisine uygulanmış ve toplanan veriler Mplus yazılımı kullanılarak analiz edilmiştir. Yapılan analizler neticesinde öğrencilerin en çok (yaklaşık 62%) doğru orantılı ilişkileri tanıma becerisine ve en az (yaklaşık 48%) ters orantılı ilişkileri tanıma becerisine sahip oldukları görülmüştür. Ayrıca, öğrencilerin 25%'inin dört temel becerinin hiçbirisine sahip olmadıkları, 39,1%'inin ise bütün becerilere sahip oldukları görülmüştür. Bunlara ek olarak, pek çok öğrencinin orantısal ilişkileri orantısal olmayanlardan ayırt etmede zorlandıkları görülmüştür. Elde edilen bulgular yorumlanarak öğrencilerin güçlü ve zayıf yönleri ile ilgili tanısal geri bildirimler verilmiştir.

Anahtar   *Bilişsel tanı modelleri, Ortaokul öğrencileri, Oran ve orantı, Orantısal akıl yürütme, Tanısal*
Sözcükler: *değerlendirme.*

INTRODUCTION

Ratios, proportions, and proportional reasoning are three important concepts in school mathematics. Although understanding these concepts constitutes a key area of school mathematics, teaching these concepts might be difficult for teachers (Lobato & Ellis, 2010). These concepts are also essential in understanding many situations in science and everyday life (Cramer & Post, 1993). Hence, students should understand and use ratios and proportions to represent relationships between quantities in the middle grades (National Council of Teachers of Mathematics [NCTM], 2000). Similarly, they should be able to determine proportional relationships and use these relationships to solve real-world problems (Common Core State Standards Initiative [CCSSI], 2010). New mathematics standards released by Turkish Ministry of Education [MEB] (2018) also expect middle school students to "determine quantities that are given in a ratio, determine proportional situations by examining real-life examples, and solve problems by understanding directly and inversely proportional relationships" (p. 12).

In mathematics education literature, the term *ratio* is defined as "multiplicative comparison of two quantities, or it is a joining of two quantities in a composed unit" (Lobato & Ellis, p. 12). In a ratio, the quantities compared may have the same units as in $\frac{15\ boys}{20\ girls}$ in which a multiplicative comparison of number of boys to girls in a classroom is presented. This multiplicative comparison can be stated as "The number of girls is four-thirds of the number of boys" or as "The number of boys is three-fourths of the number of girls." Moreover, in a ratio, the quantities compared may have different units, which is also referred as *rate*, as in $\frac{3\ kg\ apple}{5\ TL}$. The multiplicative comparison can be stated as either "three-fifths kg of apple costs 1 Turkish Lira" or as "Five-thirds Turkish Lira per 1 kg apple." Furthermore, the term *proportion* is defined as a "statement of the equality of two ratios (i.e., a/b = c/d)" (Fisher, 1988, p. 157). Finally, proportional reasoning is defined as "detecting, expressing, analyzing, explaining, and providing evidence in support of assertions about proportional relationships" (Lamon, 2007, p. 647). Hence, proportional reasoning requires noticing and representing proportional relationships between covarying quantities (CCSSI, 2010). Proportional reasoning, which is a specific form of multiplicative reasoning (Lesh, Post, & Behr, 1988), plays a significant role in the development of students' mathematics. Thus, it has been regarded as an important concept in students' elementary school mathematics and in advanced mathematics (Kilpatrick, Swafford, & Findell, 2001; NCTM, 2000). Large scale international studies such as TIMSS and PISA also treat proportional reasoning as a benchmark for students' mathematical competence.

Although proportional reasoning has a significant role in students' mathematical development, researchers (e.g., Ayan & Isiksal-Bostan, 2018; De Bock, Verschaffel, & Janssens, 1998; Misailadou & Williams, 2003; Modestou & Gagatsis, 2007) reported their difficulties with this concept. One of the most reported difficulties in developing students' proportional reasoning is that traditional instruction places more attention to the rule memorization and rote computations. Hence, students usually refer to the cross-multiplication and across-multiplication strategies when solving missing-value problems (Fisher, 1988; Arican, 2018; Arican, 2019). Although these strategies are effective in obtaining correct answers, students apply them without understanding multiplicative relationships presented (Lamon, 2007; Stemn, 2008). Another important difficulty is that students, even preservice and in-service teachers, have a tendency to determine nonproportional relationships as proportional and apply proportional strategies for solving nonproportional problems (e.g., Atabas & Oner, 2017; Izsák & Jacobson, 2017; Lim, 2009; Modestou & Gagatsis, 2007; Van Dooren, De Bock, Janssens, & Verschaffel, 2007). In addition, some researchers (e.g., Degrande, Van Hoof, Verschaffel, & Van Dooren, 2017; Van Dooren, De Bock, Hessels, Janssens, & Verschaffel, 2005) also observed the opposite of this incorrect tendency—that is assuming proportional relationships as nonproportional.

238

Two quantities can form either a directly proportional relationship or an inversely proportional relationship. In a directly proportional relationship, two quantities, *x* and *y*, remain in a constant ratio relationship (i.e., $x/y =$k). On the other hand, in an inversely proportional relationship, the product of values of two quantities is equal to a constant (i.e., $x*y =$ k). As explained in the previous paragraph, students usually refer to the rote computations and formulas when solving proportion problems without understanding constant ratio and constant product relationships between quantities. Moreover, they frequently attend to the simultaneous increases or decreases and rate of change when determining directly and inversely relationships which are not sufficient in distinguishing proportional relationships from nonproportional relationships (Arican, 2018; Arican, 2019). Therefore, it is essential for mathematics teachers to develop their students' meaningful understanding of ratios, directly, inversely, and nonproportional relationships. A meaningful understanding of ratios, directly, inversely, and nonproportional relationships necessitates students to recognize a ratio is a multiplicative comparison of two quantities, to detect directly and inversely proportional relationships between quantities, and to distinguish proportional relationships from nonproportional relationships. Two kinds of proportional relationship problems have been used in mathematics education literature: missing-value and comparison problems. In missing-value problems, students are presented with three of the four values of a proportion, and they are asked to determine the missing-value. Whereas, in comparison problems, students compare two ratios to decide whether they are equal, or if one is larger or smaller. In this study, both missing-value and comparison problems were used in exploring students' proportional reasoning.

Regarding the difficulties mentioned above, researchers have been conducting qualitative and quantitative studies to examine students' proportional reasoning. Although these studies provide useful information, they lack diagnostic information about students' difficulties and incorrect tendencies. Moreover, generalizing findings over a population in qualitative studies is quite difficult because they have been usually conducted over a small sample. Hence, the size of the research sample is regarded as one of the main challenges for the generalization of findings in qualitative studies. On the other hand, as stated by Ranjbaran and Alavi (2017), quantitative studies usually employ traditional techniques to obtain examinees' total, average, or individual scores to measure their mathematical knowledge or skills (e.g., Kuzu, 2017). While these scores provide valuable insights on examinees' overall performance, they do not offer diagnostic information on their strengths and weaknesses in terms of subject areas. In addition, the traditional techniques used in quantitative studies are not that useful in identifying multidimensional characteristics of a research topic (Bradshaw, Izsák, Templin, & Jacobson, 2014). Therefore, in recent years, researchers (e.g., Arican & Kuzu, 2019; Bradshaw et al., 2014; Choi, Lee, & Park, 2015; Dogan & Tatsuoka, 2008; Im & Park, 2010; Jurich & Bradshaw, 2014; Lee, Park, & Taylan, 2011; Sen & Arican, 2015; Terzi & Sen, 2019; Toker & Green, 2012) have been applying cognitive diagnostic models (CDMs) to provide diagnostic feedback on examinees' performance in mathematics.

In this study, inspired by the new developments in measuring students' mathematical knowledge, a multidimensional proportional reasoning test (PRT) was developed from the perspective of a general CDM, the log-linear cognitive diagnostic model (LCDM) (Henson, Templin, & Willse, 2009). The PRT intended to measure the students' mastery of four core cognitive skills, which are also referred to as *attributes*, students rely on when solving ratio and proportion problems. These four cognitive skills are determined as follows: Understanding the concept of a ratio and determining the value of a quantity in a given ratio; Recognizing directly proportional relationships and solving daily-life problems involving this type of relationships; Recognizing inversely proportional relationships and solving daily-life problems involving this type of relationships; and Recognizing nonproportional relationships and solving daily-life problems involving this type of relationships. Thus, the current study contributes to the mathematics education literature by investigating middle school students' proportional reasoning and providing diagnostic feedback for their particular strengths and weaknesses on the ratio and proportion concepts.

239

BACKGROUND

Cognitive Diagnosis Models

CDMs, which are also known as diagnostic classification models (DCMs), are a family of psychometric models that categorize examinees as either a master or nonmaster of an attribute, which is a categorical latent variable, according to their item responses. "CDMs predict the probability of an observable categorical response from unobservable (i.e., latent) categorical variables" (Ravand & Robitzsch, 2015, p. 2). The term *attribute* is used to define cognitive skills that are required in solving a specific item. The aim of CDMs is to offer diagnostic feedback with regard to these carefully defined attributes (Bradshaw et al., 2014). One of the strengths of CDMs over the Classical Theories is that they provide highly reliable examinee estimates with a small number of items (Templin & Bradshaw, 2013) and small sample sizes (Bradshaw & Cohen, 2010). Therefore, in this study, a CDM analysis was conducted to understand middle school students' understanding of ratios, directly, inversely, and nonproportional relationships and diagnostic feedback were provided for their particular strengths and weaknesses.

There are three types of CDMs: compensatory, noncompensatory, and general. In the compensatory models (i.e., deterministic input, noisy-or-gate model [DINO], Templin & Henson, 2006; and compensatory reparameterized unified model [C-RUM], Hartz, 2002), mastering one or some of the attributes, which are required to obtain a correct answer, can compensate for nonmastery of other attributes. Hence, mastery of at least one of the required attributes is enough to solve an item correctly. On the other hand, in the noncompensatory models (i.e., deterministic input, noisy-and-gate model [DINA], Junker & Sijtsma, 2001; and noncompensatory reparameterized unified model [NC-RUM], DiBello, Stout, & Roussos, 1995, Hartz, 2002), the lack of mastery of one of the required attributes cannot be completely compensated by the mastery of other attributes. Therefore, possession of all required attributes is needed to obtain a correct answer. However, the general models (i.e., general diagnostic model [GDM], von Davier, 2005; log-linear cognitive diagnostic model [LCDM], Henson, Templin, & Willse, 2009; and generalized DINA [G-DINA], de la Torre, 2011) can behave either as a compensatory model or noncompensatory model (Ravand & Robitzsch, 2015).

Using a generalized linear model, the LCDM can estimate attribute effects on each item response in a compensatory or noncompensatory manner and maps those item responses onto latent attributes (Bradshaw et al., 2014). Hence, it provides more flexibility to the researchers by helping them to detect patterns of attribute mastery (Bradshaw et al., 2014). Therefore, in this current study, the PRT was developed from the perspective of the LCDM. The LCDM estimates the probability of an examinee's (*e*) correct response for an item (*i*), which measures two attributes (α_{e1} and α_{e2}), by applying the following equation:

$$ln\left(\frac{P(X_{ei} = 1|\alpha_e)}{P(X_{ei} = 0|\alpha_e)}\right) = \lambda_{i,0} + \lambda_{i,1(1)}(\alpha_{e1}) + \lambda_{i,1(2)}(\alpha_{e2}) + \lambda_{i,2(1*2)}(\alpha_{e1}\alpha_{e2}) \qquad (1)$$

In Equation 1, $ln\left(\frac{P(X_{ei} = 1|\alpha_e)}{P(X_{ei} = 0|\alpha_e)}\right)$ models the log-odds of a correct response which is conditional on an examinee's attribute pattern. The log-odds is called a *logit* and represented as $ln\left(\frac{P(X_{ei} = 1|\alpha_e)}{P(X_{ei} = 0|\alpha_e)}\right) = $ Logit$(X_{ei} = 1|\alpha_e)$ in which $P(X_{ei} = 1|\alpha_e) = \frac{\text{Exp}(\text{Logit}(X_{ei} = 1|\alpha_e)}{1+\text{Exp}(\text{Logit}(X_{ei} = 1|\alpha_e)}$. In Equation 1, the parameter $\lambda_{i,0}$ is the intercept of the LCDM and represents the predicted log-odds of a correct response for individuals who have not mastered Attribute 1 or Attribute 2 (Bradshaw et al., 2014). Parameter $\lambda_{i,1(1)}$ is the simple main effect that represents the predicted log-odds of a correct response for individuals who have mastered Attribute 1, but not Attribute 2. Similarly, parameter $\lambda_{i,1(2)}$ is the simple main effect that represents the predicted log-odds of a correct response for individuals who have mastered Attribute 2,

240

but not Attribute 1. Finally, parameter $\lambda_{i,2(1*2)}$ is the interaction effect that represents the predicted log-odds of a correct response for individuals who have mastered both Attribute 1 and Attribute 2. The following example was provided for making sense of Equation 1:

Example: Let students are given the following question: 2 * 4 + 6 = ? To solve this question, they need to understand multiplication (α_1) and addition (α_2). The LCDM analysis provides estimations of the intercept, main, and interaction effect parameters. Assume the following parameter estimations are provided by the LCDM: $\lambda_{i,0} = -2, \lambda_{i,1(1)} = 3, \lambda_{i,1(2)} = 2$, and $\lambda_{i,2(1*2)} = 1$. Hence, replacing these parameters in Equation 1, we obtain $\text{Logit}(X_{ei} = 1|\alpha_e) = -2 + 3(\alpha_{e1}) + 2(\alpha_{e2}) + 1(\alpha_{e1}\alpha_{e2})$. In CDMs, there are 2^n attribute mastery profiles in which *n* responds to the number of attributes. Hence, we have $2^2 = 4$ attribute profiles in this example. Table 1 presents LCDM predicted logits and probabilities of mastery for each of these four attribute profiles. According to Table 1, 12% of the nonmasters of two attributes were able to solve this item. Whereas, 73% of the students who had mastered attribute 1 (multiplication), and 50% of the students who had mastered attribute 2 (addition) solved this item correctly. Finally, 98% of the students who had mastered both multiplication and addition solved this item correctly.

Table 1.
LCDM Predicted Logits and Probabilities

α_1	α_2	The LCDM Logit function	Logit	Probability
0	0	$\lambda_{i,0} + \lambda_{i,1(1)} * (0) + \lambda_{i,1(2)} * (0) + +\lambda_{i,2(1*2)} * (0) * (0)$	-2	$\frac{\text{Exp}(-2)}{1+\text{Exp}(-2)} = 0.12$
1	0	$\lambda_{i,0} + \lambda_{i,1(1)} * (1) + \lambda_{i,1(2)} * (0) + +\lambda_{i,2(1*2)} * (1) * (0)$	-2+3=1	$\frac{\text{Exp}(0)}{1+\text{Exp}(0)} = 0.73$
0	1	$\lambda_{i,0} + \lambda_{i,1(1)} * (0) + \lambda_{i,1(2)} * (1) + +\lambda_{i,2(1*2)} * (0) * (1)$	-2+2=0	$\frac{\text{Exp}(1)}{1+E\text{xp}(1)} = 0.5$
1	1	$\lambda_{i,0} + \lambda_{i,1(1)} * (1) + \lambda_{i,1(2)} * (1) + +\lambda_{i,2(1*2)} * (1) * (1)$	-2+3+2+1=4	$\frac{\text{Exp}(4)}{1+\text{Exp}(4)} = 0.98$

METHODOLOGY

This quantitative study was conducted following three corresponded steps: test development, acquisition of Q-matrix, and administration of the test and Q-matrix validation. In the following pages, these three steps were discussed in details.

Step 1: Test Development

In the first step, the PRT was developed. Hence, core skills required in solving middle school ratio and proportion problems were determined. In doing so, I examined the Turkish middle school mathematics curricula and referred to the CCSS and NCTM standards. Therefore, I determined a list of critical skills and asked three mathematics educators and a mathematics teacher to provide feedback on these skills. After receiving expert feedback, four core skills that served as attributes were determined (Table 2).

Table 2.
Attributes Required in Solving Turkish Middle School Ratio and Proportion Problems

Attributes			
Attribute 1. Understanding the concept of a ratio and determining the value of a quantity in a given ratio.	Attribute 2. Recognizing directly proportional relationships and solving daily-life problems involving this type of relationships.	Attribute 3. Recognizing inversely proportional relationships and solving daily-life problems involving this type of relationships.	Attribute 4. Recognizing nonproportional relationships and solving daily-life problems involving this type of relationships.

241

The test items were developed by a group of six people (myself and five preservice teachers [PSTs]) around the four attributes by following the problem examples provided in the Turkish middle school mathematics curricula, Turkish national tests, and mathematics education literature. The five PSTs were student teaching under my guidance at the middle school in which the study was conducted. To develop test items, I designed various problems around the four attributes and asked each one of these five PSTs to write at least one problem for each attribute. After collecting items that I and the PSTs wrote, I checked those items for their contents and included items in the test if they were measuring the intended attributes. As a result of these efforts, we had 25 items in our test. Next, we asked a middle school mathematics teacher to control these items for their suitability to the middle school mathematics program. Based on his feedback, we decided to delete three items. Thus, the final PRT included 22 multiple-choice items (see Appendix 1). In the PRT, Item 5 was adapted from Lim (2009); Item 14 was adapted from Cramer, Post, and Currier (1993); and Item 22 was adapted from Beckmann (2011).

Step 2: Acquisition of the Q-Matrix

In the second step, using the four attributes in Table 2, the test items were independently coded for their measured attribute by four mathematics educators, including myself, and a middle school mathematics teacher with more than 10 years teaching experience. Three mathematics educators and the mathematics teacher were instructed about the coding process. The independent codes received from these experts enabled me to assess whether the test items were measuring the intended attributes. This step of the study was very crucial because correct alignment of items with attributes helps researchers obtain high CDM classification accuracy (Rupp & Templin, 2008). Hence, I obtained the Q-matrix (Tatsuoka, 1985), which expresses the item–attribute alignment, using the independent codes received from these experts. In the Q-matrix (Table 3), code "1" represents an item measuring corresponding attribute(s); whereas, code "0" represents an item not measuring corresponding attribute(s). In the Q-matrix, code "1" was applied if at least three coders agreed that an item measured the corresponding attribute, else a code "0" was applied for that attribute. Therefore, in the PRT, two items measured three attributes; 14 items measured two attributes; and six items measured a single attribute.

Table 3.
The Q-matrix

Attributes	Items																						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Total
Attribute 1	1	1	1	1	0	1	1	0	0	0	1	0	1	0	1	1	1	1	0	1	0	0	13
Attribute 2	1	1	1	1	0	1	1	0	0	0	0	0	1	0	0	1	0	1	1	1	1	1	13
Attribute 3	0	0	0	0	0	0	0	1	1	1	1	0	0	0	1	0	1	0	1	0	1	1	9
Attribute 4	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	1	0	6

Note: Code "1" represents items measuring the corresponding attribute, and Code "0" represents items not measuring the corresponding attribute.

Step 3: Administration of the Test and Q-Matrix Validation

In the last step, conducted during the Spring 2018 semester, the PRT was administered to 282 seventh grade students enrolled at a middle school in a province located in the central Turkey. The school was among one of the high achieving schools in this province. There were eight seventh grade classrooms in this school, and the PRT was administered to the volunteered students in these eight classrooms. In Turkey, ratio, proportion, and proportional relationships are taught in sixth and seventh grades. Hence, the students are expected to have the required knowledge on these topics before their participation in the study. Before the administration of the PRT, the necessary permissions were obtained from the Ministry of Education, school administration, and students' parents. Therefore, the students participated in this study on a voluntary basis. They were given 40 minutes to complete this test, and five PSTs assisted me in collecting the students' responses to test items. Next, the students' responses to the test items were coded dichotomously (i.e., correct responses were coded as "1," and incorrect responses were coded as "0") in which missing responses were coded as "9". Finally, the LCDM analysis was conducted using the Q-matrix and dichotomously coded responses.

242

When developing a test from the perspective of Classical Test Theories, an item analysis needs to be performed, which was accomplished by examining item difficulty and discrimination indices, in order to detect problematic items. However, because CDMs are designed for diagnostic purposes, they use different measures for defining an item as good or bad. According to CDMs, a reliable test is one that correctly estimates examinees' profiles (Templin, 2008). CDMs examine test quality by determining item-attribute discrimination indices (e.g., Henson & Douglas, 2005; Henson, Roussos, Douglas, & He, 2008) that highlight how well an item estimates the required attribute or attributes. Having well discriminating items results in obtaining high attribute classification reliability that indicates overall quality of the test items. Item-attribute discrimination indices of the PRT items were determined by using an executable CDM file, which is specifically designed to determine item-attribute discrimination indices (Table 4). If the item-attribute discrimination index is 0, masters and nonmasters of attribute(s) have the same probability of answering the item correctly. In addition, discrimination index with a value of 1 indicates that the correct answer rate is higher for masters of the attribute(s). In addition, a negative index indicates that the correct answer rate is higher for nonmasters of the attribute(s). Although there is not a clear cut-off score for determining poor item-attribute discrimination indices, de la Torre (2008) stated a discrimination index .31 as low. Hence, the item-attribute discrimination indices in Table 4 show that the PRT items discriminated well between masters and nonmasters of an attribute. Therefore, as required, while the masters of attribute(s) tended to answer the items correctly, nonmasters tended to answer these items incorrectly. Thus, the analysis of the data was continued using 22 items and the Q-matrix described in Table 3.

Table 4.
Item-Attribute Discrimination Indices

Item	Attribute 1	Attribute 2	Attribute 3	Attribute 4
1	.46	.49		
2	.85	.69		
3	.88	.85		
4	.51	.37		
5				.78
6	.71	.70		
7	.42	.32		
8			.77	
9			.81	
10			.89	
11	.85		.73	
12				.84
13	.59	.52		
14				.84
15	.85		.79	
16	.78	.77		
17	.83		.69	
18	.81	.77		
19		.59	.71	.97
20	.71	.80		
21		.77	.78	.91
22		.76	.73	

Item difficulty index, which ranges between 0 and 1, expresses the proportion of students that answer an item correctly. Although reporting this index is usually a desired method in Classical Test Theories, in this study, the difficulty indices together with the students' responses were presented to better grasp the students' strengths and weaknesses in ratio and proportion concepts. According to Table 4, the item difficulty index ranged between .25 and .93, with a mean of .60. There were seven medium difficulty items (index values between .40 and .60), four difficult items (index values less than .40), and 11 easy items (index values more than .60). Hence, there was a good balance among the items in terms of their level of difficulty. Table 5 shows that Items 5 and 14 were perceived as the most difficult items, which were answered by only 26% and 25% of the students, respectively. On the other hand, Item 7 was the easiest item having been answered correctly by 93% of the students. No high or low level outliers were

243

deleted at this level because the purpose of the PRT was to determine the students' strengths and weaknesses and did not aim to measure their academic achievement.

Table 5.
The Students' Responses and Item Difficulty Indices

Items	A	B	C	D	NA	Dif.	Items	A	B	C	D	NA	Dif.
1	39	2	237 *	2	2	.84	12	54	16	187 *	22	3	.66
2	31	24	20	203 *	4	.72	13	230 *	7	32	10	3	.82
3	48	23	64	128	19	.45	14	69	186 *	9	13	5	.25
4	7	245 *	2	26	2	.87	15	32	50	29	163 *	8	.58
5	184 *	6	73	8	11	.26	16	21	15	207 *	21	18	.73
6	23	172 *	26	56	5	.61	17	110 *	72	20	61	19	.39
7	262 *	5	8	7	0	.93	18	177 *	33	14	44	14	.63
8	77	30	172 *	2	1	.61	19	49	33	88	105 *	7	.37
9	25	138 *	4	112	3	.49	20	18	23	17	215 *	9	.76
10	20	66	19	127 *	50	.45	21	41	156 *	50	22	13	.55
11	170 *	9	17	84	2	.60	22	29	158 *	29	21	45	.56

Note: * indicates the correct answer, NA: No answer, and Dif.: Difficulty index.

The Data Analysis and Model Fit

After testing the overall quality of the PRT items, the next stage was to obtain examinee estimates. Estimates from the data were calculated using Mplus 6.12 statistical software (Muthen & Muthen, 2011). The LCDM codes for Mplus were generated by R software (R Core Team, 2017) using the "Mplus Input Generator," which was written by Dr. Olga Kunina-Habenicht. The first step of the data analysis was to determine the best-fit LCDM model. To determine this best-fit model, one-way, two-way, and three-way log-linear structural model parameterizations were compared (Table 6). In Table 6, Model A represents the one-way structural model, Model B represents the two-way structural model, and Model C represents the three-way structural model. While Model A only included intercepts (i.e., $\lambda_{i,0}$ in Equation 1) and main effects (i.e., $\lambda_{i,1(1)}$ and $\lambda_{i,1(2)}$), Model B included intercepts, main effects, and two-way interaction effects (i.e., $\lambda_{i,2(1*2)}$). Moreover, Model C, which was the full model, included intercepts, main effects, and both two- and three-way interaction effects. These three models were compared with each other conducting a chi-square difference test using the log-likelihood values with the Maximum Likelihood Robust (MLR) estimator (Satorra & Bentler, 2010). In Table 6, a significant chi-square difference ($p < .05$) recommends that the larger model with more freely estimated parameters would better fit the data than the smaller model with less freely estimated parameters (Werner & Schermelleh-Engel, 2010). On the other hand, a nonsignificant chi-square difference indicates that both models fit the data equally well. Thus, the data estimation should be carried out using the model with smaller information model fit indices (i.e., AIC, BIC, and SSA BIC).

Table 6.
Model Modifications and Fit Indices for One-Way, Two-Way, and Three-Way Models and the Final Model

Model	Description	AIC	BIC	SSA BIC	LL	NPR	Chd	*df*	*p*
A	One-way structural model	6080.07	6360.50	6116.34	-2963,04	77	-	-	-
B	Two-way structural model	6052.63	6405.89	6098.31	-2929.32	97	50.26	1	.000
C	Three-way structural model	6056.29	6416.84	6102.91	-2929.15	99	.15	1	.696
Y	Best-fit model	6037.716	6332.71	6075.86	-2937.86	81	-	-	-

Note: AIC: Akaike's information criteria; BIC: Bayesian information criteria; SSA BIC: sample size adjusted Bayesian information criteria; LL: Log-likelihood; NPR: number of estimated parameters; Chd: chi-square difference; df: degrees of freedom.

Table 6 showed that Model B, two-way structural model, was superior over Model A, one-way structural model, ($p = .00$). Next, comparing Model B with Model C, three-way structural model, a nonsignificant chi-square difference ($p = .696 > .05$) was obtained. Hence, Model B was a better fit for the data than Model C because it yielded smaller model fit indices. The next step was to remove nonsignificant two-way interaction effects in Model B since they did not contribute to the estimation of attribute mastery

244

profiles. In Model B, there were a total of 20 two-way interaction effects, and the LCDM analysis showed that seven of them were nonsignificant. Hence, these two-way interaction effects were removed one by one, starting from the interaction effect with the highest *p*-value. After removing the interaction effect with the highest *p*-value, the model was run again and the model fit indices were checked in order to determine if removing this interaction effect improved the model fit. If this new model yielded smaller AIC, BIC, and SSA BIC indices, then estimations were preceded with this new model. The last step of obtaining the best-fit model was to remove nonsignificant main effects. In Model B there were 40 main effects, and nine of these main effects were nonsignificant. Therefore, by removing these nonsignificant main effects, Model Y was obtained, which was the best-fitting model. Model Y estimated 81 free parameters, and model fit indices are presented in Table 6.

RESULTS

The LCDM places test-takers into some latent classes based on their mastery of attributes. The total number of possible attribute profiles is calculated by the formula 2^A in which A represents the number of attributes. Hence, there were $2^4 = 16$ distinct attribute profiles since four attributes were estimated. Attribute profiles, number of students belonging to each profile, and proportions of these counts are presented in Table 7. The LCDM does not provide exact numbers for students belonging each profile but presents an estimation of these numbers. When reporting number of students in each profile, the LCDM provides decimal numbers, which were rounded up to closest whole number. Table 7 shows that approximately 111 students (39.1%) mastered all four attributes. On the other hand, roughly 71 students (25%) did not master any attributes. Thus, most students either mastered all four of the attributes or none of these attributes. Moreover, about 43 students (15.3%) mastered only Attribute 2, "Recognizing directly proportional relationships and solving daily-life problems involving this type of relationships," and none of the students belonged to the Latent Classes 2 (0001), 3 (0010), 4 (0011), 6 (0101), 7 (0110), 8 (0111), and 13 (1100). This finding showed that none of the students mastered Attribute 3, "Recognizing inversely proportional relationships and solving daily-life problems involving this type of relationships" and Attribute 4, "Recognizing nonproportional relationships and solving daily-life problems involving this type of relationships" alone, Attribute 1, "Understanding the concept of a ratio and determining the value of a quantity in a given ratio" and Attribute 2 together, Attributes 2 and 3 together, Attributes 2 and 4 together, Attributes 3 and 4 together, or Attributes 2, 3, and 4 together.

Table 7.
Estimated Number of Students Belonging to each Attribute Profile and their Proportions

Class	Attribute Profile	Count	Proportion
1	0000	71	0.250
2	0001	0	0
3	0010	0	0
4	0011	0	0
5	0100	43	0.153
6	0101	0	0
7	0110	0	0
8	0111	0	0
9	1000	5	0.016
10	1001	21	0.076
11	1010	2	0.008
12	1011	8	0.029
13	1100	0	0
14	1101	7	0.024
15	1110	14	0.049
16	1111	111	0.392

245

Note: Attribute profiles show the students' mastery of Attributes 1 through 4. For instance, attribute profile 0101 indicates the students' mastery of Attributes 2 and 4 and nonmastery of Attributes 1 and 3.

Proportions of attribute mastery are calculated by summing up corresponding proportions in Table 7. For instance, mastery of Attribute 1 can be calculated by the sum of proportions beginning from Latent Class 9 and ending up with Latent Class 16. The proportions for the mastery of remaining attributes are also calculated summing up the corresponding proportions in Table 7. Table 8 shows that roughly 59.5% of the students mastered Attribute 1, "Understanding the concept of a ratio and determining the value of a quantity in a given ratio"; 61.9% of the students mastered Attribute 2, "Recognizing directly proportional relationships and solving daily-life problems involving this type of relationships"; 47.9% of the students mastered Attribute 3, "Recognizing inversely proportional relationships and solving daily-life problems involving this type of relationships"; and 52.2% of the students mastered Attribute 4, "Recognizing nonproportional relationships and solving daily-life problems involving this type of relationships." The proportions presented in Table 8 suggest that, among the four attributes, the students were better at mastering Attribute 2. On the other hand, they obtained the lowest proportion of mastery from Attribute 3. To assess how accurately proportions of attribute mastery in Table 8 were estimated, the CDM measure of reliability indices (Templin & Bradshaw, 2013) were calculated. According to Templin and Bradshaw (2013), the CDM measure of reliability relies upon "the correlation of mastery statuses between two hypothetical independent administrations of the same test" (p. 259). Hence, this procedure was equivalent to determining test–retest reliability. According to Table 8, Attributes 1 and 3 were estimated at 99% certainty and Attribute 2 and Attribute 4 were estimated at 97% and 96% certainty, respectively. Although there is no systematic investigation on the effect of sample size on CDM classification accuracy, some researchers (e.g., Tatsuoka, 1983; Lei & Li, 2016) stated that CDM classification accuracy increases with the sample size (Ravand & Robitzsch, 2018). On the other hand, CDMs can provide highly reliable examinee estimates with small sample sizes (Bradshaw & Cohen, 2010). As a rule of thumb, it is suggested that the number of items and sample size should exceed the number of latent classes (e.g., Huebner, 2010; Ravand & Robitzsch, 2018). In this study, there were 16 latent classes which was less than number of items (i.e., 22) and sample size (i.e., 282). Moreover, reliability indices in Table 8 indicate that the LCDM provided highly reliable attribute mastery estimations that suggested overall validity of the PRT.

Table 8.
The Students' Attribute Mastery Proportions and Reliability Indices

	A1	A2	A3	A4
Mastery	.595	.619	.479	.522
Reliability	.99	.97	.99	.96

Note: The letter *A* represents the term *Attribute*.

To understand the contribution of attributes on the respondents' success, the LCDM provides item parameter estimates. These item parameter estimates are converted to proportions that show the respondents' success rates. Item parameter estimates, standard errors, and estimated probabilities are provided in Appendix 2. In Appendix 2, estimated proportions for intercepts were calculated using the $\frac{Exp(\lambda_{i,0})}{1+Exp(\lambda_{i,0})}$ formula in which the letter *i* represents the item number. Estimated proportions for main and interaction effects were also calculated by adapting the same formula. For instance, $\frac{Exp(\lambda_{i,0}+\lambda_{i,1(1)})}{1+Exp(\lambda_{i,0}+\lambda_{i,1(1)})}$ calculates the estimated proportions for Attribute 1 (i.e., $\lambda_{i,1(1)}$ is item parameter estimate for Attribute 1 and item *i*). On average, items had an intercept of -0.55 that suggested that roughly 36.6% of the students who had not mastered any of the attributes (i.e., nonmasters) answered the items correctly evidently by guessing. Appendix 2 shows that Items 9 and 19 were the most difficult items for the nonmasters, which were answered by 18.8% and 19.5% of them, respectively. On the other hand, Items 4 and 7 were the easiest items for the nonmasters, which were answered by 73.2% and 73.1% of them, respectively. Moreover, Appendix 2 shows that the attributes were partly compensatory. For example, the proportion of a correct response to Item 16 increases from .407 to .723 to .739 to 1 when comparing examinees who mastered neither attribute, only Attribute 1, only Attribute 2, and both attributes,

246

respectively. Thus, the students who had mastered both Attribute 1 and Attribute 2 had more chance of answering Item 16 than the ones who did not master any attributes or than the ones who mastered a single attribute.

One of the advantages of CDMs is that they can be used in providing diagnostic feedback on individual performance. Mplus calculates respondents' posterior probabilities using an expected a posteriori (EAP) estimation for each latent class. Each respondent's estimations of attribute mastery can be calculated by the sum of posterior probabilities in corresponding latent classes. For instance, estimation of Attribute 1 is calculated by summing up posterior probabilities beginning from Latent Class 9 and ending up with Latent Class 16. Table 9 presents four selected students' response patterns and estimated proportions of attribute mastery. Table 9 showed that while Student 195 had a 99.8% chance of mastering Attribute 1, Student 147 had only 28.9% chance of mastering this attribute. Although there is not an exact cutoff score when deciding respondents' mastery status, proportions less than .40 suggested nonmastery, proportions over .60 suggested mastery, and proportions in between .40 and .60 are regarded as questionable (Rupp, Templin, & Henson, 2010). Hence, in Table 9, Student 53 could not master any attributes; Student 147 mastered only Attribute 2; Student 195 mastered Attributes 1, 2, and 4; and Student 261 mastered all four attributes. Therefore, in Table 9, as in this example, classroom teachers and educators may use these individual scores to understand each student's mastery of selected skills. In terms of questionable proportions of attribute mastery, CDMs use a model-internal classification criterion. Hence, classifications of respondents into the latent classes are "the direct result of the application of a psychometric model" (Rupp et al., 2010, p. 86). Thus, CDMs classify each respondent in one of the latent classes using posterior probabilities calculated for each attribute profile.

Table 9.
Four Selected Students' Item Response Patterns and Estimated Proportions of Attribute Mastery

ID	Response Pattern	P(A1)	P(A2)	P(A3)	P(A4)
53	10*100100*0110**0*0*0*	.289	.102	.002	.279
147	1001*11000101001000101	.054	.905	.000	.018
195	111111100*11111101110*	.998	.990	.075	.966
261	1111101101011011001101	.918	.866	.649	.812

Note: The words *A* and *C* represent attribute and latent class, respectively, and the asterisks indicate the missing responses.

DISCUSSION and CONCLUSIONS

The purpose of the current study was to investigate Turkish seventh grade middle school students' understanding of ratios and proportions and to provide diagnostic feedback on their strengths and weaknesses on these concepts. By carefully examining national and international standards and large-scale assessments, four core cognitive skills that were deemed to be necessary for solving middle school ratio and proportion problems were identified. In the next step, a multidimensional test, the PRT, was designed around these four attributes. Diagnostically reliable interpretations of the students' understanding of ratios and proportions were provide using the LCDM in analyzing their responses to the PRT items. In this quantitative study, the development of the PRT and the findings that were subsequently obtained from the application of this test were explained.

In terms of the students' mastery of the four attributes, the LCDM results suggested that there were not extreme differences among the students' mastery levels of the attributes. However, the students were better at "Recognizing directly proportional relationships and solving problems involving this type of relationships" (Attribute 2) and "Understanding the concept of a ratio and determining the value of a quantity in a given ratio" (Attribute 1) in comparison to "Recognizing inversely proportional relationships and solving problems involving this type of relationships" (Attribute 3) and "Recognizing

247

nonproportional relationships and solving problems involving this type of relationships" (Attribute 4). Although the students obtained the highest mastery proportions for Attribute 2 (i.e., .619), roughly 38% of them could not master this attribute. Similarly, more students could not master the remaining attributes. Hence, this finding suggested that many students had difficulty understanding ratio concept, recognizing directly, inversely, and nonproportional relationships and solving daily-life problems involving these three types of relationships.

Among the four attributes, the lowest proportion of mastery was for Attribute 3 (i.e., .479). According to this result, roughly 52% of the students had difficulty recognizing inversely proportional relationships and solving problems involving this type of relationships. The analysis showed that in Items 8, 9, and 11, which measured this attribute, respectively 77, 112, and 84 students solved these items as if they were direct proportion problems (Table 5). In accordance with this result, researchers such as Degrande et al. (2017) and Van Dooren et al. (2005) reported students' tendency to assume inversely proportional relationships to be proportional and use of proportional strategies for solving inverse proportion problems. In addition, many students had difficulty mastering Attribute 4, approximately 48% of them could not master this attribute. This finding suggested constraints in their understanding of nonproportional relationships and differentiating this type of relationships from directly and inversely proportional relationships. The literature on ratios and proportions has many examples (e.g., Atabas & Oner, 2017; Modestou & Gagatsis, 2007; Van Dooren et al., 2007) expressing students' challenges in differentiating proportional relationships from nonproportional relationships and vice versa.

Table 7 showed that 25% of the students could not master any attributes and 39.2% of them mastered all four attributes. Hence, roughly 64% of the students either mastered all four attributes or did not master any attributes. This finding suggested a knowledge gap between the selected students. In my discussion with the mathematics teachers of the eight classrooms, they pointed out this gap by stating that they had high achieving and low achieving students in each class. Thus, teachers may use information obtained studies like this one in detecting and narrowing this knowledge gap between high achieving and low achieving students.

In the item level, Items 4 and 7 appeared as the easiest items having been answered correctly by 87% and 93% of the students, respectively (Table 5). Similarly, Appendix 2 presented that 73.2% and 73.1% of the nonmasters were also able to answer these two items, respectively. However, regarding with these two items, the students who had mastered Attribute 1 or Attribute 2 or both of them had higher correct answer rates than the nonmasters. High correct answer rates by the nonsmasters could be explained by both items included a simple direct proportion, which is one of the most popular and repeated concepts in mathematics textbooks. On the other hand, Table 5 pointed out that Items 5 and 14 were the most difficult items for the students having been answered correctly by 26% and 25% of the students, respectively. Both items included nonroutine real-life contexts with nonproportional relationships. Hence, the students' poor performance on these two items showed their insufficient experiences with this type of problems. Thus, this result suggests that classroom teachers should expand their teaching by using nonroutine problems, which may also expand students' proportional reasoning by forcing them to think out of the box, and real-life contexts with nonproportional relationships.

Table 5 also showed the students' confusion over the directly and inversely proportional graphs. In Item 19, 63% of the students had difficulty identifying graphs that were not depicting a directly proportional relationship. By the same token, in Item 21, 45% of the students had difficulty identifying the inversely proportional graph. In both items, many students assumed nonproportional relationships presented in options A as directly and inversely proportional, respectively. Similar incorrect tendencies reported by Arican (2018), Arican (2019), De Bock et al. (1998), Modestou and Gagatsis (2007), and Van Dooren et al., (2007) who observed students' over attention to linearity when inferring directly and inversely proportional relationships.

Studies on students' proportional reasoning usually report difficulties that they have with ratio and proportion concepts. However, they lack diagnostic information about these difficulties. Hence, this

study differs from those studies by providing students' mastery levels of four core cognitive skills which they rely on when solving ratio and proportion problems. Furthermore, rather than providing an overall score of students, DCMs present attribute mastery profiles of individuals that help educators detecting each individual's strengths and weaknesses in terms of subject areas. Therefore, as in Table 9, teachers and educators can use CDMs in determining students' mastery of cognitive skills and use this information on detecting their strengths and weaknesses and provide diagnostic feedback if necessary.

Implications for Teaching and Suggestions

Classical Test Theories offer only limited information about students' strengths and weaknesses because they provide a single overall score for each student. Hence, in recent years, researchers have been paying more attention to CDMs in order to provide diagnostic feedback on students' performance. Briefly, in the current study, the LCDM analysis suggested Turkish seventh grade middle school students' overall weaknesses in all four attributes. Therefore, teachers should work to expand students' knowledge of ratios, proportions, and proportional relationships in order that they can develop meaningful understanding of these concepts. The results obtained from the LCDM analyses may provide new insights into students' understanding of these concepts.

In Turkey, for many years, the national tests that are conducted at the end of eighth and twelfth grades only included multiple-choice items. Moreover, parents and school administrators have paid attention to students' number of correct answers, and teachers' success was measured by these numbers. Hence, teachers directly or indirectly directed to teach rules and rote computations rather than developing their students' meaningful understanding of mathematics. Therefore, Turkish students' weaknesses on the PRT items might be a reflection of the traditional mathematics instruction that has been used in classrooms for many years. Thus, considering findings obtained from research such as the current study and other factors (i.e., students' educational background, socio-economic status, access to educational technologies, etc.) that influence students' academic achievements can effect policymakers' decision-making on curricular choices and can therefore help policymakers in developing effective educational systems.

This study included 282 students, which was a relatively small sample, from an over achieving middle school. Hence, future studies should examine students' proportional reasoning using a larger sample and including schools with varying educational achievements. Moreover, there is a need for future studies to apply qualitative methods (i.e., interviews, classroom observations, etc.) together with CDM analysis to better understand students' strengths and weaknesses on the concepts of ratios, proportions, and proportional relationships.

REFERENCES

Arican, M. (2018). Preservice middle and high school mathematics teachers' strategies when solving proportion problems. *International Journal of Science and Mathematics Education, 16*(2), 315–335. DOI: 10.1007/s10763-016-9775-1

249

Arican, M. (2019). Preservice mathematics teachers' understanding of and abilities to differentiate proportional relationships from nonproportional relationships. *International Journal of Science and Mathematics Education, 17*(7), 1423–1443. DOI: 10.1007/s10763-018-9931-x

Arican, M., & Kuzu, O. (2019). Diagnosing preservice teachers' understanding of statistics and probability: Developing a test for cognitive assessment. *International Journal of Science and Mathematics Education,* 1–20. DOI: 10.1007/s10763-019-09985-0

Atabas, S., & Oner, D. (2017). An examination of Turkish middle school students' proportional reasoning. *Boğaziçi University Journal of Education, 33*(1), 63–85.

Ayan, R., & Isiksal-Bostan, M. (2018). Middle school students' proportional reasoning in real life contexts in the domain of geometry and measurement. *International Journal of Mathematical Education in Science and Technology*, 1–17. DOI: 10.1080/0020739X.2018.1468042

Beckmann, S. (2011). *Mathematics for elementary teachers* (3rd. ed.). Boston, MA: Pearson.

Bradshaw, L., & Cohen, A. (2010). Accuracy of multidimensional item response model parameters estimated under small sample sizes. In A. Izsák (Chair), Using cognitive attributes to develop mathematics assessments, opportunities, and challenges. Symposium conducted at the annual American Educational Research Association conference in Denver, CO.

Bradshaw, L., Izsak, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understandings of rational numbers: Building a multidimensional test within the diagnostic classification framework. *Educational Measurement: Issues and Practice, 33*(1), 2–14. DOI: 10.1111/emip.12020

Choi, K. M., Lee, Y. S., & Park, Y. S. (2015). What CDM can tell about what students have learned: An analysis of TIMSS eighth grade mathematics. *Eurasia Journal of Mathematics, Science & Technology Education, 11*(6), 1563–1577. DOI: 10.12973/eurasia.2015.1421a

Common Core State Standards Initiative. (2010). *The common core state standards for mathematics.* Washington, D.C.: Author.

Cramer, K., & Post, T. (1993). Making connections: A case for proportionality. *Arithmetic Teacher, 60*(6), 342–346.

Cramer, K., Post, T., & Currier, S. (1993). Learning and teaching ratio and proportion: Research implications. In D. Owens (Ed.), *Research ideas for the classroom: Middle grades mathematics* (pp. 159–178).   New York, NY: Macmillan.

De Bock, D., Verschaffel, L., & Janssens, D. (1998). The predominance of the linear model in secondary school students' solutions of word problems involving length and area of similar plane figures. *Educational Studies in Mathematics, 35*(1), 65–83. DOI: 10.1023/A:1003151011999

Degrande, T., Van Hoof, J., Verschaffel, L., & Van Dooren, W. (2017). Open word problems: Taking the additive or the multiplicative road? *ZDM, 50*(1), 1–12. DOI: 10.1007/s11858-017-0900-6

de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of Educational Measurement, 45*(4), 343–362. DOI: 10.1111/j.1745-3984.2008.00069.x

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179–199. DOI: 10.1007/s11336-011-9207-7

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood–based classification techniques. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361–390). Hillsdale, NJ: Lawrence Erlbaum.

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. DOI: 10.1007/s10649-007-9099-8

Fisher, L. C. (1988). Strategies used by secondary mathematics teachers to solve proportion problems. *Journal for Research in Mathematics Education, 19*(2), 157–168.

Hartz, S. (2002). *A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practice* (Unpublished doctoral dissertation). University of Illinois at Urbana–Champaign.

Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnostics. *Applied Psychological Measurement, 29*(4), 262–277. DOI: 10.1177/0146621604272623

Henson, R., Roussos, L., Douglas, J., & He, X. (2008). Cognitive diagnostic attribute–level discrimination indices. *Applied Psychological Measurement, 32*(4), 275–288. DOI: 10.1177/0146621607302478

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log–linear models with latent variables. *Psychometrika, 74*(2), 191–210. DOI: 10.1007/s11336-008-9089-5

Huebner, A. (2010). An overview of recent developments in cognitive diagnostic computer adaptive assessments. *Practical Assessment, Research & Evaluation, 15*(3), 1–7.

Im, S., & Park, H. J. (2010). A comparison of US and Korean students' mathematics skills using a cognitive diagnostic testing method: Linkage to instruction. *Educational Research and Evaluation, 16*(3), 287–301. DOI: 10.1080/13803611.2010.523294

250

Izsák, A., & Jacobson, E. (2017). Preservice teachers' reasoning about relationships that are and are not proportional: A knowledge-in-pieces account. *Journal for Research in Mathematics Education, 48*(3), 300–339. DOI: 10.5951/jresematheduc.48.3.0300

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258–272. DOI: 10.1177/01466210122032064

Jurich, D. P., & Bradshaw, L. P. (2014). An illustration of diagnostic classification modeling in student learning outcomes assessment. *International Journal of Testing, 14*(1), 49–72. DOI: 10.1080/15305058.2013.835728

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Kuzu, O. (2017). Matematik ve fen bilgisi öğretmen adaylarının integral konusundaki kazanımlarının incelenmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi, 18*(3), 948–970. DOI: 10.29299/kefad.2017.18.3.049

Lamon, S. (2007). Rational numbers and proportional reasoning: Toward a theoretical framework for research. In F. K. Lester, Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol 1, pp. 629–667). Charlotte, NC: Information Age Publishing.

Lee, Y. S., Park, Y. S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the US national sample using the TIMSS 2007. *International Journal of Testing, 11*(2), 144–177. DOI: 10.1080/15305058.2010.534571

Lei, P. W., & Li, H. (2016). Fit indices' performance in choosing cognitive diagnostic models and Q-matrices. Paper presented at the annual meeting of the National Council on Measurement in Education (NCME), Philadelphia, PA.

Lesh, R., Post, T., & Behr, M. (1988). Proportional reasoning. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 93–118). Reston, VA: National Council of Teachers of Mathematics.

Lim, K. (2009). Burning the candle at just one end: Using nonproportional examples helps students determine when proportional strategies apply. *Mathematics Teaching in the Middle School, 14*(8), 492–500.

Milli Eğitim Bakanlığı (2018). *Matematik dersi öğretim programı* (1, 2, 3, 4, 5, 6, 7 ve 8. Sınıflar) [*Mathematics curriculum* (1, 2, 3, 4, 5, 6, 7, and 8. Grades]. Ankara: Talim ve Terbiye Kurulu Başkanlığı.

Misailadou, C., & Williams, J. (2003). Measuring children's proportional reasoning, the "tendency" for an additive strategy and the effect of models. In N. A. Pateman, B. J. Dougherty, & J. T. Zilliox (Eds.), *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 293–300). Honolulu, HI: University of Hawaii.

Modestou, M., & Gagatsis, A. (2007). Students' improper proportional reasoning: A result of the epistemological obstacle of "linearity". *Educational Psychology, 27*(1), 75–92. DOI: 10.1080/01443410601061462

Muthen, L. K., & Muthen, B. O. (2011). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthen & Muthen.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Ranjbaran, F., & Alavi, S. M. (2017). Developing a reading comprehension test for cognitive diagnostic assessment: A RUM analysis. *Studies in Educational Evaluation, 55*, 167–179. DOI: 10.1016/j.stueduc.2017.10.007

Ravand, H., & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment, Research & Evaluation, 20*(11), 1–12.

Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology, 38*(10), 1255–1277. DOI: 10.1080/01443410.2018.1489524

R Core Team. (2017). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R–project.org/

Rupp, A., & Templin, J. (2008). Effects of Q–matrix misspecification on parameter estimates and misclassification rates in the DINA model. *Educational and Psychological Measurement, 68*(1), 78–98. DOI: 10.1177/0013164407301545

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

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi–square test statistic. *Psychometrika, 75*(2), 243–248. DOI: 10.1007/s11336-009-9135-y

Sen, S., & Arican, M. (2015). A diagnostic comparison of Turkish and Korean students' mathematics performances on the TIMSS 2011 assessment. *Journal of Measurement and Evaluation in Education and Psychology, 6*(2), 238–253. DOI: 10.21031/epod.65266

Stemn, B. S. (2008). Building middle school students' understanding of proportional reasoning through mathematical investigation. *Education 3–13, 36*(4), 383–392. DOI: 10.1080/03004270801959734

251

Tatsuoka, K. (1985). A probabilistic model for diagnosing misconceptions by the pattern classification approach. *Journal of Educational Statistics, 10*(1), 55–73. DOI: 10.3102/10769986010001055

Templin, J. (2008). *Test construction item discrimination*. Lecture presented at the Diagnostic Modelling Seminar at the University of Georgia, Athens. Retrieved from https://jonathantemplin.com/files/dcm/ersh9800f08/ersh9800f08_lecture11.pdf

Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification, 30*(2), 251–275. DOI: 10.1007/s00357-013-9129-4

Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods, 11*(3), 287–305. DOI: 10.1037/1082-989X.11.3.287

Terzi, R., & Sen, S. (2019). A nondiagnostic assessment for diagnostic purposes: Q-matrix validation and item-based model fit evaluation for the TIMSS 2011 assessment. *SAGE Open*, 1–11. DOI: 10.1177/2158244019832684

Toker, T., & Green, K. (2012). *An application of cognitive diagnostic assessment on TIMMS–2007 8th grade mathematics items*. Paper presented at the annual meeting of the American Educational Research Association, Vancouver, British Columbia, Canada.

Van Dooren, W., De Bock, D., Hessels, A., Janssens, D., & Verschaffel, L. (2005). Not everything is proportional: Effects of age and problem type on propensities for overgeneralization. *Cognition and Instruction, 23*(1), 57–86. DOI: 10.1207/s1532690xci2301_3

Van Dooren, W., De Bock, D., Janssens, D., & Verschaffel, L. (2007). Pupils' overreliance on linearity: A scholastic effect? *British Journal of Educational Psychology, 77*(2), 307–321. DOI: 10.1348/000709906X115967

von Davier, M. (2005). *A general diagnostic model applied to language testing data.* ETS Research Report. Princeton, NJ: Educational Testing Service.

Werner, C., & Schermelleh-Engel, K. (2010). Deciding between competing models: Chi–square difference tests. In *Introduction to Structural Equation Modeling with LISREL* (pp. 1–3). Frankfurt, Germany: Goethe University..

APPENDIX 1.

Proportional Reasoning Test

1) Two numbers are in a ratio of $\frac{3}{5}$. If small number is 45, what is the value of large number?

   A) 27   **B)** 55   **C)** 75   **D)** 85

2) There are some number of red and blue balls in a container. The ratio of the number of red balls to the number of blue balls is $\frac{3}{4}$. If there are a total of 42 balls in this container, how many of them are blue?

   A) 6   **B)** 12   **C)** 18   **D)** 24

3) The measures of outer angles of a triangle are proportional to the numbers 9, 12, and 15. What is the measure of the greatest internal angle of this triangle?

   A) 45   **B)** 60   **C)** 75   **D)** 90

4) A teacher gives her students 2 candies for every 5 questions they answer correctly. If a student received a total of 14 candies, how many correct answers is this student had?

   A) 28   **B)** 35   **C)** 45   **D)** 70

5) Two different but identical candles, A and B, are burning at the same constant rate but they are lit at different times. We know that when candle A has burned 24 mm, candle B has burned 18 mm. When 36 mm of candle B has burned, how many millimeters will candle A have burned?

   A) 48   **B)** 44   **C)** 42   **D)** 40

6) Lemonade A is made by mixing 4 cups of water with 2 tablespoons of sugar, Lemonade B is made by mixing 7 cups of water with 5 tablespoons of sugar, and Lemonade C is made by mixing 3 cups of water with 1 tablespoon of sugar. Which one of the following information is true about the flavors of these lemonades?

   A) The least taste of sugar is taken from Lemonade A   **B)** The most taste of sugar is taken from Lemonade B.   **C)** The same taste of sugar is taken from Lemonade A and Lemonade C.   **D)** The most taste of sugar is taken from Lemonade C.

7) An athlete, who runs at a constant speed, consumes 150 calories in 2 hours. If this athlete keeps running at this same constant speed, how many calories does he consume after 3 hours?

   A) 225   **B)** 200   **C)** 175   **D)** 120

8) Three painters working at the same pace can paint a building in 12 days. In how many days 4 painters can paint the same building?

   A) 16   **B)** 8   **C)** 9   **D)** 10

9) A car travels the distance between two cities in 6 hours driving at a constant speed of 60 kilometers per hour. If this car increases its speed to 90 kilometers per hour, how many hours can it take between the two cities?

   A) 3   **B)** 4   **C)** 6   **D)** 9

10) Twelve workers working at the same pace can a finish a certain job in 15 days working 7 hours each day. How many days are needed for 14 workers to finish the same job working 5 hours each day?

   A) 10   **B)** 12   **C)** 16   **D)** 18

11) x and y are two inversely proportional quantities. We know that when x is 12, y is 4. When x is 6, what is the value of y?

   A) 8   **B)** 6   **C)** 4   **D)** 2

12) Mehmet bought a 90 cm long plant from the market. This plant grows 30 cm each year. After three years, the height of the plant was measured as 180 cm. According to this information, how many centimeters should be the height of the plant after 6 years?

   A) 360   **B)** 300   **C)** 270   **D)** 240

13) a and b two numbers that vary in a directly proportional relationship. We know that when a is 5, b is 15. So, when b is 24, what is a?

   A) 8   **B)** 12   **C)** 14   **D)** 16

14) Ayse and Esra, who are running at the same speed, started running around an 800 meters circular track at different times. It is known that when Ayse ran 300 meters, Esra ran 200 meters. According to this information, when Esra ran 400 meters, how many meters can Ayse ran?

   A) 500   **B)** 600   **C)** 700   **D)** 800

15) Three painters working at the same pace can paint a building in 8 days. Which one of the following tables depicts the relationship between the number of painters and painting time correctly?

   A)

Number of Painters	1	2	3	4
Painting Time	2	4	8	16

   B)

Number of Painters	1	2	3	4
Painting Time	24	16	8	4

   C)

Number of Painters	1	2	3	4
Painting Time	16	12	8	6

   D)

Number of Painters	1	2	3	4
Painting Time	24	12	8	6

253

16) The table in below shows the relationship between the number of chickens on a farm and the average number of eggs they produce per week. According to this table, how many chickens are required to have 240 eggs per week?

Number of Chickens	Number of Eggs per Week
4	12
6	18
9	27
12	36

A) 120 **B)** 100 **C)** 80 **D)** 60

17) (**m+3**) and (**n+1**) are two numbers that vary in an inversely proportional relationship. We know that when *m* is 1, *n* is 2. When *m* is 3, what is **n**?

A) 1 **B)** 2 **C)** 3 **D)** 4

18) The graph below shows that 30 kg sugar is present in an 80 kg mixture. How many kg sugar are there in a 120 kg mixture?
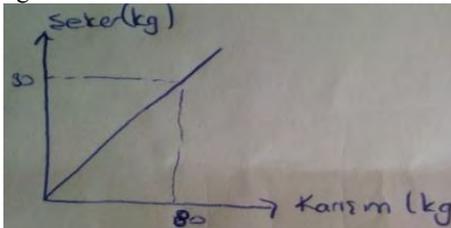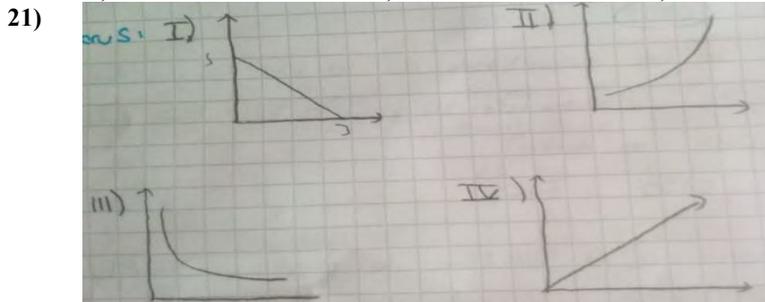

A) 45 **B)** 40 **C)** 35 **D)** 20
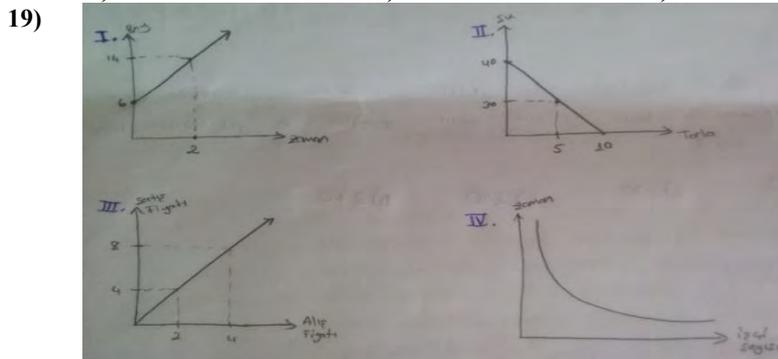
19)

Which one(s) of the above graphs do not depict a directly proportional relationship?

A) Only IV **B)** Only II **C)** II and IV **D)** I, II, and IV

20) At a certain time of a day, the length of the shadow of a 60 meters long apartment is measured as 12 meters. At the same certain time, what is the length of the shadow of a 5 meters long tree?
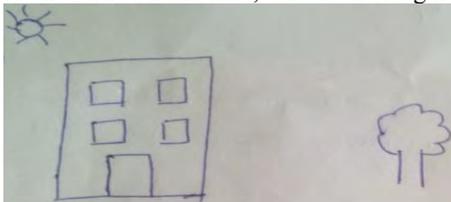
A) 4 **B)** 3 **C)** 2 **D)** 1

21)

Which one(s) of the above graphs depict an inversely proportional relationship?

A) Only I **B)** Only III **C)** I and III **D)** I, II, and IV

22) In a bakery, 3 people working at the same constant rate can frost 12 cakes in 4 hours. In this bakery, how many hours are needed by 4 people to frost 8 cakes?

A) 1 **B)** 2 **C)** 3 **D)** 4

254

APPENDIX 2.

Item Parameter Estimates, Standard Errors, and Estimated Proportions

i	λi,0	Pr.	λi,1(1)	Pr.	λi,1(2)	Pr.	λi,1(3)	Pr.	λi,1(4)	Pr.	λi,2(1,2)	Pr.	λi,2(2,3)	Pr.	λi,2(2,4)	Pr.	λi,2(3,4)	Pr.
1	.09 (.27)	.521	27.26 (1.04)	1.00	1.29 (.53)	.798					-23.75 (1.53)	.992						
2	-.04 (.26)	.491	2.62 (.99)	.930	-1.67 (.74)	.153					3.30 (1.38)	.985						
3	-1.36 (.23)	.204									2.69 (.33)	.791						
4	1.01 (.22)	.732	2.48 (.60)	.97														
5	-.29 (.21)	.427							-1.57 (.42)	.14								
6	-.65 (.26)	.343					1.98 (.36)	.790										
7	1.00 (.29)	.731	2.56 (.87)	.97	26.5 (.85)	1.00												
8	-.60 (.19)	.355							2.71 (.37)	.89								
9	-1.46 (.24)	.188							2.97 (.38)	.82								
10	-1.11 (.23)	.249							2.96 (.39)	.86								
11	-.88 (.21)	.293	2.41 (.31)	.82														
12	-.58 (.21)	.359							3.59 (.96)	.95								
13	.41 (.20)	.600	1.72 (.77)	.89							26.39 (1.23)	1.00						
14	-.68 (.20)	.337							-.91 (.37)	.17								
15	-.96 (.22)	.277	1.26 (.59)	.58					1.50 (.67)	.63								
16	-.38 (.28)	.407	1.34 (.56)	.72	1.42 (.54)	.739					25.56 (1.22)	1.00						
17	-.91 (.22)	.286	-2.71 (1.90)	.03					4.10 (1.89)	.96								
18	-.54 (.19)	.368									4.02 (.67)	.97						
19	-1.42 (.32)	.195			1.19 (.51)	.443	-22.4 (.84)	.00	-19.8 (.84)	.00			22.36 (1.27)	.00	2.39 (1.30)	.00		
20	-1.29 (.39)	.217	27.95 (1.08)	1.00	3.64 (.93)	.914					-26.14 (1.55)	.985						
21	-.89 (.27)	.290			2.04 (.56)	.759	-23.3 (.94)	.00					44.13 (1.15)	.00	-44.65 (.96)	.76	23.80 (1.08)	.00
22	-.59 (.27)	.358			2.59 (1.1)	.881	2.56 (1.35)	.88					-3.10 (1.81)	.99				

Note: The terms *i* and *Pr* stands for items and proportions, respectively, and numbers between parentheses are standard errors.

255

TÜRKÇE GENİŞLETİLMİŞ ÖZET

Orantısal akıl yürütme çarpımsal akıl yürütmenin özel bir biçimi olup (Lesh, vd., 1988), öğrencilerin matematiksel gelişiminde önemli bir rol oynamaktadır. İlkokul aritmetiğinde ve daha üst düzey matematikte önemli bir kavram olan orantısal akıl yürütme (Kilpatrick, vd., 2001; NCTM, 2000), "orantısal ilişkileri tespit etme, temsil etme, analiz etme, açıklama ve kanıt sunma" (Lamon, 2007, s. 647) olarak tanımlamıştır. Bu nedenle, orantısal akıl yürütme, iki nicelik arasındaki orantısal ilişkileri tespit etmeyi ve ifade etmeyi gerektirir (CCSSI, 2010). Orantısal akıl yürütme öğrencilerin matematiksel gelişiminde önemli bir rol oynasa da, araştırmacılar (ör., Ayan & Isiksal-Bostan, 2018; De Bock, vd., 1998; Misailadou & Williams, 2003; Modestou & Gagatsis, 2007) öğrencilerin oran, orantı ve orantısal ilişkiler konularında zorlandıklarını belirtmişlerdir. Geleneksel matematik öğretiminin kural ve ezbere dayalı hesaplamalara daha fazla önem vermesi öğrencilerin orantısal akıl yürütmelerinin gelişimine engel olarak görülen en önemli yanlışlardan birisidir. Bu nedenle, öğrenciler, hatta öğretmen adayları ve öğretmenler, eksik-değer problemlerini çözerken daha çok çapraz-çarpım yöntemini tercih etmektedirler (Fisher, 1988). Her ne kadar bu yöntem öğrencilerin doğru cevabı elde etmelerine imkân verse de, öğrenciler bu yöntemi problemde verilen nicelikler arasındaki çarpımsal ilişkileri anlamadan mekanik olarak uygulamaktadırlar (Lamon, 2007). Öğrencilerin nicelikler arasındaki çarpımsal ilişkileri anlamadan verilen problemleri çözmeleri, orantılı olan ve olmayan ilişkileri birbirleriyle karıştırmalarına neden olabilmektedir. Bundan dolayı, yapılan pek çok çalışmada (ör., Atabas & Oner, 2017; Izsák & Jacobson, 2017; Modestou & Gagatsis, 2007; Van Dooren, vd., 2007) öğrencilerin orantısal olmayan ilişkileri orantısal olarak düşünme eğilimine sahip oldukları ve bu tip ilişkiler içeren soruları orantısal yöntemler kullanarak çözmeye çalıştıkları rapor edilmiştir. Bunun yanı sıra, bazı araştırmacılar (ör., Degrande, vd., 2017; Van Dooren, vd., 2005) öğrencilerin doğru ve ters orantılı ilişkileri orantısal olmayan olarak düşünme eğilimini de gözlemlemişlerdir.

Öğrencilerin orantısal akıl yürütmelerini incelemek adına araştırmacılar nitel veya nicel çalışmalar yürütmektedirler. Bu çalışmalar öğrencilerin orantısal akıl yürütmeleri hakkında faydalı bilgiler sağlasa da, öğrencilerin güçlü ve zayıf yönleri hakkında tanısal bilgiden yoksundurlar. Bundan dolayı, son yıllarda bazı araştırmacılar (ör., Bradshaw, vd., 2014; Lee, vd., 2011) öğrencilerin, öğretmen adaylarının ve öğretmenlerin matematik performansları hakkında tanılayıcı geribildirim sağlamak için bilişsel tanı modellerine başvurmaktadırlar. Bu sebeple, bu çalışmada öğrencilerin matematik bilgilerini ölçmedeki yeni gelişmeler takip edilerek, 22 çoktan-seçmeli sorudan oluşan bir orantısal akıl yürütme testi genel bilişsel tanı modellerinden log-lineer cognitive diagnostic model (LCDM) (Henson, vd., 2009) perspektifinden geliştirilmiştir. Bu test yardımıyla öğrencilerin oran ve orantı problemlerini çözerken kullandıkları bilişsel becerilerdeki ustalık derecelerini belirleme amaçlanmıştır. LCDM analizlerinden elde edilen sonuçlar kullanılarak yedinci sınıf öğrencilerinin orantısal akıl yürütmeleri incelenmiş olup, güçlü ve zayıf yönleri için tanısal geri bildirimler sağlanmıştır.

Bu çalışma birbirini izleyen üç adımı takip ederek gerçekleştirilmiştir: test geliştirme, Q-matris'in elde edilmesi ve testin uygulanması ve Q-matris'in doğrulanması. İlk olarak, orantısal akıl yürütme testini geliştirmek için gerekli olan temel beceriler belirlenmiştir. Bunun için Milli Eğitim Bakanlığı ortaokul matematik müfredatı, uluslararası yayınlar ve öğretim standartları incelenmiştir. Bunun sonucunda şu dört temel beceri belirlenmiştir: Oran kavramını anlama ve verilen bir orandaki bir niceliğin değerini belirleme (Beceri 1); Doğru orantılı ilişkileri tanıma ve bu tür ilişkileri içeren günlük yaşam problemlerini çözme (Beceri 2); Ters orantılı ilişkileri tanıma ve bu tür ilişkileri içeren günlük yaşam problemlerini çözme (Beceri 3); Orantısal olmayan ilişkileri tanıma ve bu tür ilişkileri içeren günlük yaşam problemlerini çözme (Beceri 4). Daha sonra, bu dört temel beceri etrafında 22 çoktan-seçmeli sorudan oluşan bir test geliştirilmiştir. İkinci adımda, her bir test maddesi, dört matematik eğitimcisi ve bir ortaokul matematik öğretmeni tarafından ölçtükleri beceriler için bağımsız olarak kodlanarak Q-

256

matris elde edilmiştir. Son adımda ise geliştirilen test bir ortaokulda öğrenim gören 282 yedinci sınıf öğrencisine uygulanmıştır.

Toplanan öğrenci yanıtları doğru yanıtlar "1," yanlış yanıtlar "0" ve eksik cevaplar ise "9" olarak kodlanmış ve elde edilen Q-matris ile birlikte Mplus programı kullanılarak analiz edilmişlerdir. Elde edilen sonuçlar öğrencilerin her dört beceriye sahip olma derecelerini gözler önüne sermiştir. Öğrencilerin %59,5'nin birinci beceriye; %61,9'nun ikinci beceriye; %47,9'nun üçüncü beceriye ve %52,2'nin dördüncü beceriye sahip oldukları görülmüştür. Bu sonuç, öğrencilerin ters orantılı ve orantısal olmayan ilişkileri belirleme de ve bu tür ilişkiler içeren problemleri çözme de zorlandıklarını göstermiştir. Öğrencilerin becerilere sahip olma olasılıkları sırasıyla 0,99, 0,97, 0,99 ve 0,96 güvenirlikle ölçülmüştür. Ayrıca, analizler neticesinde öğrencilerin %25'nin hiçbir beceriye sahip olmadığı ve %39,1'nin ise her dört beceriye de sahip olduğu belirlenmiştir. Bu sonuç, çalışmaya katılan öğrenciler arasında becerilere sahip olma bakımından büyük bir fark olduğunu ortaya koymuştur. Bunlara ek olarak, hiçbir beceriye sahip olmayan öğrencilerin ortalamada %36,6'nın testteki maddeleri tahmin yöntemini kullanarak doğru cevaplayabildikleri belirlenmiştir.

Yapılan çalışmalar son yıllarda araştırmacıların öğrencilerin matematik performanslarıyla ilgili tanısal geri bildirim sağlamak için bilişsel tanı modellerine daha fazla başvurduklarını göstermektedir. Bu nedenle, mevcut çalışma bulgularının eğitimcilere müfredat seçimleri ve daha etkili öğretim metotları geliştirmede yardımcı olabileceği düşünülmektedir.

257