Monitoring Student Achievement with Cognitive Diagnosis Model

Levent YAKAR* Nuri DOĞAN ** Şenol DOST***
Nazan SEZEN YÜKSEL****

Abstract

In this study, it is aimed to show how student achievement can be monitored by using the cognitive diagnosis models. For this purpose, responses of the 6th, 7th, and 8th grade Mathematics subtests of High School Placement Tests (HSPT) in 2009, 2010, and 2011, which provide longitudinal data, were used, respectively. There were 49933 examiners' responses in data sets. The attributes examined by these tests were determined by the Mathematics experts, and the Q matrix consisting of five attributes was developed. As a result of the analysis, it was seen that the largest latent class for all three years consisted of those non-master for any attribute. It was observed that the probability of attribute mastery increased in the 7th grade and decreased in the 8th grade. The high classification accuracy seen as a result of the analysis applied to HSPT, which is not intended for the cognitive diagnosis, shows that the results can be used for monitoring student achievement.

Key Words: Cognitive diagnosis, student achievement, g-dina, attribute mastery probability, longitudinal data.

INTRODUCTION

Education includes the efforts made to gain individuals the pre-determined and necessary behaviors related to the cognitive, affective, and psychomotor areas. Gaining targeted behaviors are not operationsthat happen at once, but require a process. It can be said that this situation is also reflected in measurement and evaluation. Although, in measurement and evaluation practices,it is very common to collect data on the extent to which the product reached at the end of the process meets the expected qualifications, contemporary educational approaches accept that products are not independent ofthe processes and interactions in the process (Kutlu, Doğan & Karakaya, 2010). Therefore, it is necessary to measure the processes and interactions in the training process as well as the products at the end of the training process.

It is observed that as the importance is given to revealing the development of individuals in the process, practices and researches aimed at this purpose increase. If it is accepted that measurement practices related to the process are generally for monitoring the development, it can be said that the studies for gathering information about the process are carried out through both national and international exams (Ministry of NationalEducation [MoNE], 2017;Organisation for Economic Co-operation and Development [OECD], 2019). For example, through international exams such as PISA, TIMSS, and PIRLS,national-level development is tried to be monitored by making use of international comparisons in areas such as mathematics, science and technology, and reading comprehension. Although international exams give indirect information about the educational process in general, they do not provide information about the status of individual students who are the subjects of the process and cannot provide detailed information about the effectiveness of educational programs. In this regard, it is observed that in many countries, information about the process is collected through exams held at

* Assist. Prof. Dr., Kahramanmaraş Sütçü İmam Uni, Faculty of Education, Kahramanmaraş-Turkey, l_yakar@hotmail.com, ORCID ID: 0000-0001-7856-6926
** Prof. Dr., Hacettepe Uni, Faculty of Education, Ankara-Turkey, nuridogan2004@gmail.com, ORCID ID: 0000-0001-6274-2016
*** Prof. Dr., Hacettepe Uni, Faculty of Education, Ankara-Turkey, dost@hacettepe.edu.tr, ORCID ID: 0000-0002-5762-8056
**** Assoc. Prof. Dr., Hacettepe Uni, Faculty of Education, Ankara-Turkey, nsezen.hc@gmail.com, ORCID ID: 0000-0002-0539-3785

_____

different stages of education (OECD, 2019). For example, in the past Detection Exam of Student Achievement and today Monitoring and Evaluating Academic Skills exams aim to reveal the developments in main coursesin Turkey (MoNE, 2017). In addition to these, although the past High School Placement Exam (HSPT; "Seviye Belirleme Sınavı"), which students took three times in secondary school, is not an application for monitoring, it can be said that it is a test that provides information in terms of student development due to its multiple implementations (MoNE, 2008).

It is seen that research models based on repeated measurements come to the fore in order to determine the development of individuals in the process. In order to reveal whether the development of individuals is sufficient or not in research models based on repetitive measurements, the measurement results made at least two different times are compared using various statistical techniques. However, since the results obtained from such applications are based on the comparison of the average of the measurement results obtained attwo different timesat least, it does not give information about the development of individual students as well as neglecting the acquisitions and subject dimension. These techniques are criticized in this respect (Lohman, 1999).

In addition to traditional statistical techniques, cognitive diagnosis models (CDM), which is an effective technique to reveal the fine-grained ability parameters of individuals, can demonstrate level developments in repeated measures. It is stated that CDM, which will be discussed in this study, has become widespread, especially with the beginning of the 2000s, and its main purpose is to give cognitive feedback to teachers and families about their students (Embretson, 1998).

CDM is based on latent class analyses, which are used to identify subgroups and determine which individuals belong to these subgroups using multivariate categorical data and interrelated situations (Cheng, 2010). In this way, it is possible to calculate the structure of certain knowledge or the development of a skill in the student by taking into account the strengths and weaknesses of the student at the cognitive level (Leighton &Gierl, 2007). According to de la Torre (2009), with a test developed using CDM, it can be determined which skills the students have, which are predetermined by experts, and therefore, what their shortcomings are. Taking advantage of this feature of CDM, it may be possible to see the development of students in terms of relevant skills if the same skills are measured at different times.

By using CDM, psychological structures with more than one interrelated cognitive attributes can be measured with a single test. In practice, it is accepted that each item in the test measures one or more cognitive attributes. In CDM analyses, the Q matrix is used to determine which item measures which cognitive attribute. In the Q matrix, each column represents a cognitive attribute, and each row represents an item. The Q matrix is created by field experts by coding as 1 if the cognitive attributes specified in the column are measured with the item specified in the row, and 0 if not (de la Torre & Minchen, 2014). By the Q matrix used in CDM, 2^k latent classes are formed for k cognitive attributes defined by experts. There will be eight latent classes for k = 3; the latent class (000), indicating an individual who is non-master for any attributes; latent classes (100), (010), (001) indicating individuals with master one of the attributes; latent classes (110), (101), (011) indicating individuals with master two of the attributes and (111) latent class indicating individuals with master all the attributes. In addition to showing what attributes individuals have and which they do not, the latent classes also give an idea of which questions they are expected to answer correctly. CDM makes it possible to identify individuals in terms of cognitive attributes.

There are many CDM available; Deterministic inputs noisy and-gate (DINA; Junker and Sijtsma, 2001), Deterministic inputs noisy or-gate (DINO; Templin & Henson, 2006), re-parameterized unified model (R-RUM; Hartz, 2002), general diagnostic models (GDM; von Davier, 2008), generalized DINA Model (G-DINA; de la Torre, 2011), etc., that take different assumptions and parameters into account. Besides the various test and item parameters, the mastery probability of cognitive attributes in the Q matrix is calculated to determine which of the latent classes individuals will be included in, in these models. If the probability values calculated for a cognitive attribute are 0.5 and above, mastery of attribute is shown with "1"; if it is less than 0.5, it is indicated with "0". This process aims to make it easier to reveal the latent cognitive structures that individuals have.

_____

ISSN: 1309 – 6575 _Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi_
Journal of Measurement and Evaluation in Education and Psychology   304

Considering the example given above, it can be said that individuals in the "101" latent class have the first and third cognitive attributes and their probability of mastery of these attributes is 0.5 or above. On the other hand, it can be said that these individuals do not have the second cognitive attribute and their probability of mastery of the second cognitive attribute is less than 0.5. Therefore, while latent classes are obtained by rounding the probability value to 0 or 1, the differences between the probabilities of individuals are neglected. For example, an individual who has mastery probability of the first, second, and third cognitive attributes, respectively 0.55, 0.10 and 0.60; and an individual who has probability mastery 0.90, 0.45, and 0.95 are in the same latent class, which coded with "101". The fact that the transformation of mastery of attribute probability into binary category causes loss of information can be seen as the negative side of this transformation process.

It is one of the most important features of CDMs that they reveal the attributes they have in smaller parts instead of the holistic approach when diagnosing individuals. In this way, CDMs enable individuals to be diagnosed from different angles. The latent classes and attribute mastery probability outputs that are created with the help of the Q matrix input representing fine-grained small measurement units in CDMs provide detailed information for individuals. The fact that monitoring the cognitive characteristics of students in fine-grained skill with CDMs can provide more specific and relevant information compared to the general monitoring of students' cognitive level reveals that CDMs will be more useful in monitoring students' progress.

Considering that the main purpose of CDMs is to provide feedback to education stakeholders (Embretson, 1998), a detailed and fine-grained picture of the current situation can be taken through CDMs. Formative assessment, in which feedback is at the forefront, cannot be used adequately due to the high class size, the need for time and effort (Bennett, 2011). In this case, it is important to include high-stakes exams, which are not normally intended for formative assessment, in the feedback mechanism. In addition, the longitudinal feedback to be given for the same parts with the same method will be of great importance in terms of revealing the change and making the education even better.

Interest in CDMs is increasing both in the world and in Turkey. It can be said that the field of study of CDMs is mostly focused on simulation since the subject area is new with increasing interest. In some of these studies (Huang, 2017; Kaya &Leite, 2017; Wang, Yang, Culpepper & Douglas, 2018; Zhan, Jiao, Liao & Li, 2019), models are presented for the use of longitudinal data in CDMs. However, these studies are insufficient to show how the change in a large population is revealed by CDMs. The actual data in these studies consist of smaller datasets suitable for model use only. This study, on the other hand, is important by separating it from other studies in terms of targeting a wide audience.

In this study, it was aimed to apply cognitive diagnosis models to HSPT, which are repeated measures, and to monitor the development of students through their attributes. For this purpose, answers to the following sub-problems were sought;

1) What is the prevalence rate of the latent class patterns of students by years?

2) What is the rate of change in the students' mastery of each attribute by years?

3) What is the rate of change in the number of attributes mastered by students over the years?

4) What is the level of reliability and validity of the findings obtained?

METHOD

In this study, which aims to monitor the achievement of students with CDM, survey method, one of the quantitative research methods, was used.

Sample

The population of the study consists of approximately 1 million middle school students who started secondary school in 2009 and joined HSPT in 3 years. The answers of 131068 of these students in the

SBS every three years were given to the researchers by the MoNE. The data of 49933 students, who had complete data in allthree years, formed the sample of the research.

Data Collection Instruments

High Schools Placement Exams (HSPT) was organized by the Ministry of National Education (MoNE), Turkey. The data were obtained from the General Directorate of Measurement, Evaluation, and Examination Services of the MoNE upon the request of the researchers. HSPT was a central system high-stake exam which was held after the course period in every year in June, organized by the MoNE, and measures the level of achievement of the students related to the learning outcomes determined for the 6th, 7th, and 8th grades. The exam consisted of Turkish language, mathematics, science, social sciences, and English courses items within the scope of the middle school. The exam was prepared as multiple-choice tests based on the learning outcomes and is sufficient to measure the student's interpretation, analysis, critical thinking, predicting, and problem-solving skills, etc. (MoNE, 2010).

HSPT was an exam held once a year between 2008 and 2013 at the end of the spring term, and its scores are used to place students in high school. Approximately 1 million students had entered HSPT for each grade level each year. HSPT differs from the other old/new exams in terms of being held in 6th, 7th, and 8th grades among the exams held for transition to high schools. With this feature, HSPT is an important resource to examine the development of students over the years. Although the exam is not practiced today, HSPT was deemed suitable for this study because it has been measured more than twice, the answers of many students across the country have been obtained, and the study is on a theoretical and practical basis.

In the study, in which student progress was examined through math test items, 16, 18, and 20 math items were asked to students in the 6th, 7th and 8th grades, respectively.

Descriptive statistics regarding the test scores of the data used in the study are calculated and given in Table 1.

Table 1. Descriptive Statistics for Tests

Grade	Item	Mean	Std. Dev.	Mean Item Difficulty	d
6th	16	5.40	2.93	0.338	-.82
7th	18	7.62	4.44	0.423	-.4
8th	20	6.65	5.10	0.333	-.97

Table 1 shows that it was observed that the highest number of correct answers was in the 7th grade, while the lowest was in the 6th grade. It was seen that 8th grade students had correct responses on average 1.25 more questions than 6th grade. However, when the mean item difficulty, which indicates rates of correct responses, are examined, it is seen that the 6th and 8th grades are very close to each other due to the increasing number of questions over the years. Considering the relative variation coefficients showing the ratio of the standard deviation to the mean, it can be said that the groups become more heterogeneous from the sixth to the eighth grade. When the difficulty level of the tests is evaluated, the tests applied for the 6th and 8th grades have a similar difficulty, and the tests applied to the seventh grade are relatively easy. In the analysis for the item response theory, it was seen that all three data sets were two-dimensional. When the averages of the item difficulty parameters (*d*) obtained as a result of the analysis are examined, it can be said that the items in the 8th grade are easier. Although the 7th grade items are a little more difficult than the 8th grade, it is concluded that they are easy. It was seen that the 6th grade items were more difficult than the other grades but still close to the easy level.

_____

ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology   306

Procedure

Defining Attributes

In the primary mathematics teaching program, problem-solving ability is one of the basic skills that are stipulated to be provided to the students. Within the scope of the program, the problems are discussed under two headings as routine and non-routine. In general, problems are considered as questions, of which solutions are not foreknown and obvious, and in such questions, it is claimed that the students will reach a solution by making reasoning through their current knowledge(Sezen Yüksel, Sağlam Kaya, Urhan, & Şefik, 2019).In brief, problems that can be solved by using the information directly are described as "routine" problems, whereas problems that can be solved by interpreting existing information and by operations that are more complex are described as "non-routine problems".

Within the scope of this study, it was tried to determine the attributes of HSPT math items. For this purpose, the 6th, 7th, and 8th grade items were discussed primarily within the context of the problem types and then the mathematical skills that they require. While determining these skills, first of all, Math Taxonomy (Smith, Wood, Coupland, Stephenson, Crawford, &Ball, 1996) and "Mathematical content and process skills" (Tatsuoka, Corter, &Tatsuoka, 2004) in the literature were examined, and the operations required by the HSPT mathematics questions were grouped by the field expert researchers of the study. Operations (such as the application of a well-known algorithm, visual perception) that could not be classified into any of the existing skills were determined by field experts, gathered under common categories, and renamed. Five attributes have been created by making these skills more specific in accordance with the processes and subjects required by the items. The names and characteristics of these attributes are as follows:

Table 2. Defined Attributes' Code, Name and Definitions

Attribute Code	Attribute Name	Definition
Attribute 1	Operational Applications	Application of Basic Features of Numbers
Attribute 2	Mathematization Applications	Mathematization of a Word Problem
Attribute 3	Concept Calculations	Computational Application of Concept
Attribute 4	Concept's Advanced Applications	Application of the concept, in a different context in relation to other concepts, in a problem
Attribute 5	Geometric Manipulation	Application for Manipulation of Geometric Shapes

Attribute 1 covers the processes of "Routine operations by recalling a definition or a term, application of the formula, recalling the rules knowledge, classification knowledge, applying an algorithm, length measurement, numbers (fractions, decimal numbers, and percentages) and determining the number representations and making number conversions", which includes application of the basic features of the numbers.

Attribute 2 covers the mathematization of a word problem. In other words, it is the process of problem-solving through the use of mathematical representations of verbal expressions containing mathematical structures and taken from daily life.
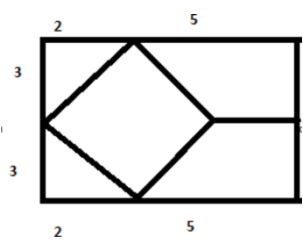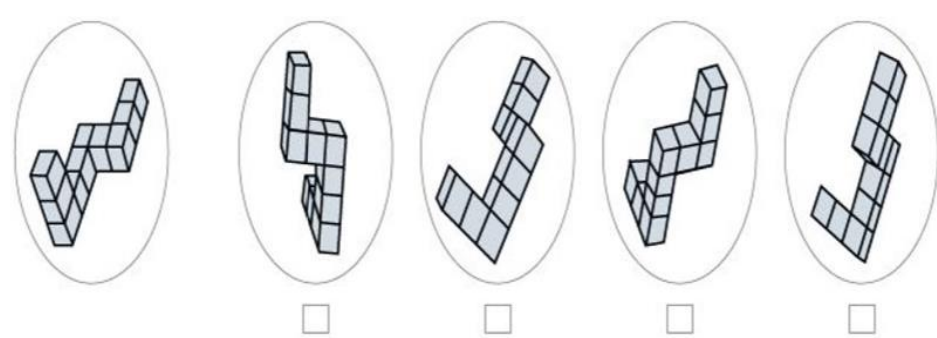
Attribute 3 includes the processes regarding the computational applications of concepts. This stage is the application of the processes required by the mathematical concept given in the problem expression.

Attribute 4 is operations of applying concepts in a different context and associated with other concepts.

Attribute 5 covers applications for the manipulation of geometric shapes. These applications involve the use of different forms of geometric shapes provided in the problem.

It would be beneficial to examine Table 3 to make the information on the attributes more understandable:

_____

ISSN: 1309 – 6575 *EğitimdevePsikolojideÖlçmeveDeğerlendirmeDergisi*
Journal of Measurement and Evaluation in Education and Psychology   307

Table 3. Example Items Which Examined Attributes

Attribute No	Example Items
Attribute 1	What is the result of the operation $(-2)^{-3} \cdot 4^2$?
Attribute 2	"Ayşe has 440 pieces 1 TL coins in her penny bank. Ayşe spent all her money to buy 5 dolls. In that case, what is the price of a doll?"
Attribute 3	How many unit squares is the area of a circumscribed circle of the square with a side length of 4 cm?
Attribute 4	2 triangular, 2 trapezoidal, and 1 equilateral rectangular regions are drawn in the rectangular region of the figure. When Ela throws a stone, what is the probability of the stone striking the triangular regions given in the figure?
Attribute 5	

Which one of the figures given in the above <u>cannot be obtained</u> by rotating the leftmost shape?

Five field experts were consulted for the mathematical attributes determined by the researchers. The field experts consisted of two academicians with specialisation in mathematics education and three mathematics teachers who were working in schools affiliated to the Ministry of National Education at the time of the study. Initially, their opinions about the names and contents of the attributes were elicited. The definitions and content of some attributes were modified based on these opinions. For example, due to the fact that the skills of the application of the basic features of the numbers given in the content of the Attribute 1 were perceived as four operations at first glance, an error was identified as considering that this attribute was included in all problems. In order to eliminate this error, it was decided to use more specific concepts in the definition of Attribute 1. Therefore, Attribute 1 was expressed as applications related to the basic characteristics defined on the number sets. Another correction suggestion encountered at this stage was related to the items of geometric shapes. Geometry has its own specific framework, and it is possible to solve some questions by known algorithms as such in mathematical questions. In accordance with the feedback taken from the experts, Attribute 5 was renamed as "Geometric Manipulation" because it did not address all geometry questions because the skills required by the solution of some geometry questions were the applications of the known algorithm. This led to the tagging of geometry items with other attributes, although the word geometry was not used in the attribute. Subsequently, the revised mathematical attributes were re-shared with the field experts in concern. In consequence, a consensus was reached on the attributes in accordance with the opinions received, and the attributes and their explanations were finalized accordingly.

_____

ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

308

Creating of Q-matrix

In general, the problems in mathematics differ from each other in the context of the mathematical skills required by their content and solution. This was taken into account when tagging the items according to the attributes established within the scope of the study. The lack of a hierarchical structure among mathematical skills leads to the lack of a hierarchy between the mathematical attributes prepared according to these skills. These facts played a significant role in the formation of the Q-matrix. For instance, a problem tagged as Attribute 4 may not contain other attributes. On the other hand, an item can be tagged with more than one attribute. For example, we may consider the problem of "Each one of the T-shirts purchased for TL 4,50 is printed on TL 1,25. When these t-shirts are sold to TL 9,50, which of the following is the algebraic expression of the profit earned from x unit?". This problem is tagged with Attributes of 1-2-3 because it includes basic operations with decimal numbers, mathematization of a word problem, and computational application of the concept of "profit". The four operations used on any problem are not always required to refer toAttribute 1, though.

In the process of obtaining the Q-matrix, the researchers firstly formed matrices individually. Then, they came together to compare the matrices. In this process, when the items were tagged with different attributes, the researchers finalized the matrix by reaching a consensus byrevising the mathematical skills included in the questions.

For the Q-matrix formed in the last case, the opinions of two academicians from the field experts who took part in the beginning of the process were obtained. The Q-matrix and items were submitted to the field experts together with the explanations of the attributes. The suggestions taken from both field experts were evaluated together. In order to give an example of the correction suggestions, item 19 of 8th graders' HSPT can be examined. In this item, a ramp image and the height of this ramp were given as 1 meter, and the slope was 10%, and if the slope was 8%, it was asked what point the ramp would start from the visual point. The researchers tagged this question with attributes 1, 4, and 5. The feedback received from the field expert was that this question did not include a geometric shape; therefore, it would not be related to Attribute 5. The researchers emphasized that the ramp image contained in this question covered a right triangle and that a solution could be reached through its manipulation. The field experts reached a consensus on this issue. A similar method was followed for other suggestions, and the Q-matrix was finalized by consensus with the field experts.

Some modification suggestions based on the results of the data-model fit of Q matrices, AIC (Akaike, 1974), BIC (Schwarzer, 1976), and the software package developed for Q matrix validation (Ma ve de la Torre, 2019) were conveyed to experts. The relative fit indices before and after the last recommendation are presented in Table 4.

Table 4. Relative Fit Indices Before and After the Last Recommendation

	AIC		BIC	
	Previous	Last	Previous	Last
6	923572.8	923036.4	924322.4	923741.9
7	1013470	1012564	1014114	1013225
8	1018767	1014285	1019499	1014999

As seen in Table 4, AIC and BIC relative model data fit indices at all three grade levels indicate a better fit for the Q matrices formed after the accepted recommendations. In line with the analyses and suggestions, the Q matrices were given their final form in Table 5.

According to Table 5, in the last case, Attribute 1 was examined in ten items in the 6th grade, four items in the 7th grade, and five items in the 8th grade. Attribute 2 was examined in six items in the 6th grade, three items in the 7th grade, and four items in the 8th grade. Attribute 3 was examined in seven items in the 6th grade, seven items in the 7th grade, and eight items in the 8th grade. Attribute 4 was examined in three items in the 6th grade, three items in the 7th grade, and four items in the 8th grade. Attribute 5 was examined in two items in the 6th grade, five items in the 7th grade, and six items in the 8th grade. In six of the 6th grade items, one attribute was examined, in nine of them two, and in one of them, four

_____

ISSN: 1309 – 6575 *EğitimdevePsikolojideÖlçmeveDeğerlendirmeDergisi*
Journal of Measurement and Evaluation in Education and Psychology

309

were examined. In 14 of the 7th grade items, one attribute was examined, and in four of them, two attributes were examined. In 13 of the 8th grade items, only one attribute and inseven of them, two attributes were examined.

Table 5. Q Matrix

6th Grade						7th Grade						8th Grade					
Item	A1	A2	A3	A4	A5	Item	A1	A2	A3	A4	A5	Item	A1	A2	A3	A4	A5
1	0	1	0	0	0	1	1	0	0	0	0	1	1	0	0	0	0
2	1	0	1	0	0	2	1	0	0	0	0	2	1	1	0	0	0
3	1	1	0	0	0	3	0	0	0	1	1	3	0	0	1	0	0
4	1	1	0	0	0	4	0	0	1	0	0	4	0	0	1	0	0
5	1	0	1	0	0	5	1	1	0	0	0	5	1	0	0	1	0
6	0	0	1	0	0	6	0	0	0	0	1	6	0	0	1	0	0
7	0	0	1	0	0	7	0	0	1	0	0	7	1	1	0	0	0
8	1	0	1	0	0	8	0	0	0	1	0	8	0	0	1	0	1
9	0	0	0	0	1	9	0	1	0	0	0	9	0	0	0	0	1
10	1	1	1	1	0	10	0	0	1	1	0	10	0	0	0	0	1
11	0	0	0	1	0	11	0	0	0	0	1	11	0	1	0	0	0
12	1	1	0	0	0	12	0	0	1	0	0	12	0	0	1	0	1
13	0	0	0	0	1	13	0	0	1	0	0	13	0	0	1	0	1
14	1	1	0	0	0	14	0	0	1	0	0	14	0	0	1	0	0
15	1	0	1	0	0	15	0	0	0	0	1	15	0	0	0	1	0
16	1	0	0	1	0	16	0	0	1	0	0	16	0	1	0	0	0
						17	0	1	0	0	0	17	0	0	0	1	0
						18	1	0	0	0	1	18	0	0	1	0	0
												19	1	0	0	1	0
												20	0	0	0	0	1

Data Analysis

In the selection of the model to be used in order to determine the cognitive classes of the students, the criterion of the model having the best fit at the item level was taken into consideration. For this purpose, the data sets and the Q matrices they were related to were subjected to model comparison analysis with the GDINA (Ma and de la Torre, 2018) package in the R software program. It was tested with Wald test that shows which of the G-DINA in the package or the restricted forms of G-DINA, DINA, DINO, ACDM, LLM (Maris, 1999), R-RUM (DiBello, Stout, & Liu Roussos, 2007) fit better. If there was no significant difference at the $p=.05$ level between the fit indices of G-DINA and its restricted forms, the restricted model with the simplest structure was chosen; otherwise, G-DINA was chosen as the model to be used for the relevant item. As a result of the analysis, in the 6th grade, LLM for the 3rd item, DINA for the 4th item, and R-RUM for the 10th and 12th items were determined as the most appropriate model. And in 8th grade, LLM for the 7th item and the R-RUM model for the 8th, 12th, and 19th items were determined as the most appropriate model. The GDINA model was determined as the most appropriate model for all the items in the 7th grade and for the other items in the 6th and 8th grades.

Analyses were performed using the R software program using the GDINA (Ma and de la Torre, 2018) package. Expected a Postieori (EAP) method was used to obtain individual parameters. For the first research question, the probability of mastering each attribute by years and the prevalence rates of the latent classes to which they were assigned as a result of the analysis were given. For the second sub-problem, the rates of change according to the years of mastery of each attribute are given. In the third sub-problem, the changes in the number of attributes of the students according to the years were reported. In the last sub-problem, the correct classification rates were examined for the reliability of the analysis results (Ciu, Gierl, Chang, 2012). For this, the accuracy of latent classifications was determined by Iaconangelo (2017), and the accuracy of classification by attribute was determined by Wang et al. (2015) with the help of indexes in the same package. In the examination of the validity of the analysis results, the proof of convergent validity was used. The correlation between the correct response rate and the probability of mastery of the attribute was examined as proof of convergent validity (Li, et al., 2020).

_____

RESULTS

Findings are given in order under sub-headings according to the sub-problem titles.

Students' Attributes by Years

The average of the students' attribute mastery probability for each year (grade level) was calculated and given in Table 6.

Table 6. Means of Attribute Mastery Probability

	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5	Mean
6th Grade	0.22	0.17	0.13	0.12	0.16	0.16
7th Grade	0.23	0.31	0.22	0.17	0.37	0.26
8th Grade	0.17	0.27	0.16	0.34	0.25	0.24

According to Table 6, it is seen that all attribute mastery probability increased with the transition from 6th to 7th grade. The probabilities of all the 8th grade attributes except attribute 1 were also found higher than the 6th grade levels. When the 8th grade probabilities were compared with those of the 7th grade, the values revealed closer to each other, but it is seen that the probabilities in the 7th grade were higher for all the qualities except attribute 4.

In the sixth grade, the most common attribute was attribute 1, followed by the 2nd and 5th attributes with similar rates. The least attribute mastery probability in the sixth grade was observed in attributes 4 and 3. In addition, the probability value of attribute 4 in the sixth grade was seen to have the lowest value among all the attribute probabilities covering three years. The highest attribute mastery probability in the seventh grade was in attribute 5, followed by attribute 2. In the 7th grade, attribute 4 had the lowest probability. The highest attribute mastery probability in the 8th grade was in attribute 4. In the 8th grade, it is seen that attributes 1 and 3 had the lowest probability average. The attribute mastery probability and the correct response rate for each class are given in Figure 1.
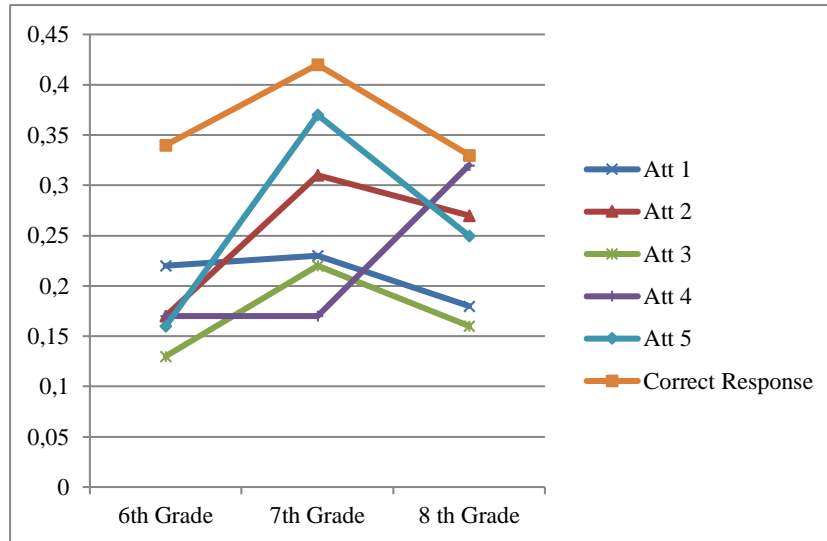

Figure 1. Correct Response Rate and Attribute Mastery Probability Means across Years

Figure 1 shows the variation of attribute mastery probability and correct response rate and their relation by years, clearly. It can be concluded that while the correct response rate and the mastery probability of attributes 2, 3, 4, and 5 increase visibly, the mastery probability of attribute 1 did not change much at the time of the transition from the 6th grade to the 7th grade. When transitioning from the 7th grade to the

_____

ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*   311
Journal of Measurement and Evaluation in Education and Psychology

8th grade, a decrease was observed in the rate of correct response and all attribute mastery probability except attribute 4. Among the attributes, attribute 4 had the lowest probability average in the 6th and 7th grades and the highest probability average in the 8th grade.

The average probability of having attributes may not provide sufficient information about the attribute patterns of students. For this, the latent attribute classes were examined. Table 7 shows the rate of students in latent attribute classes for 3 years.

Table 7. Prevalence of Latent Classes

Latent Class	6th Grade	7th Grade	8th Grade	Latent Class	6th Grade	7th Grade	8th Grade
00000	**0.75**	**0.62**	**0.67**	11100	0.00	0.00	0.00
10000	**0.07**	0.00	0.00	10110	0.00	0.00	0.00
01000	**0.02**	0.00	0.00	11010	0.00	0.00	0.00
00100	0.00	0.00	0.00	11001	0.00	0.01	0.00
00010	0.01	0.01	**0.04**	10110	0.00	0.00	0.00
00001	0.00	**0.07**	0.00	10101	0.00	0.00	0.00
11000	0.01	0.00	0.00	10011	0.00	0.00	0.00
10100	0.00	0.00	0.00	01101	**0.02**	0.01	0.00
10010	0.00	0.00	0.00	01011	0.00	0.00	**0.06**
10001	0.02	0.00	0.00	00111	0.00	0.00	0.00
01100	0.02	0.00	0.00	11110	0.00	0.00	0.00
01010	0.00	0.00	**0.04**	11101	0.01	**0.08**	0.00
01001	0.01	**0.06**	0.00	11011	0.00	0.00	0.01
00110	0.00	0.00	0.00	10111	0.00	0.00	0.00
00101	0.00	0.00	0.00	01111	0.00	0.00	0.01
00011	0.00	0.00	0.01	11111	**0.07**	**0.13**	**0.15**

For five attributes, 32 ($2^n = 2^5$) latent classes can be created. The highlighted characters in Table 7, which includes the rates of students' presence in the latent classes, indicated the most common five latent classes for each grade. When the table is examined, it is seen that the rate of students in the "00000" latent class, in other words, who had non-mastery for all attributes, was very high and close to each other for all three years. It was observed that approximately ¾ in 6th grade, in 7th and 8th grades 2/3 of the students were in the "00000" latent class. At the 7th and 8th grades, the second largest latent class is "11111", with rates of 13% and 15%, respectively. This latent class consists of students who mastered all attributes. At the 6th grade level, those with all the attributes constituted the 6th largest group. Considering the ratios, it is seen that the number of students who mastered all the attributes was far behind the group sizes of those without any attributes. It is seen that the ratio was 0.00 in many latent classes. Many of these latent classes appeared to have no students due to the rounding process. However, it was observed that there were no students in some latent classes before the rounding process. It can be said that the students were not homogeneously distributed in the latent classes.

Rates of Change in Students' Mastery of Each Attribute

The latent class sizes contain a general result about the latent class in which students are included according to the measurement made in the relevant year. It can be examined in Table 8 which attributes of the students changed in the 7th grade compared to the 6th grade.

Table 8. Proportion of Students Whose Attribute Mastery Changes in 7th Grade According to 6th Grade based on Attribute

Change	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5
Gainer	0.11	0.18	0.13	0.08	0.25
Loser	0.05	0.02	0.02	0.02	0.01
Total Change	0.17	0.20	0.14	0.10	0.26
Unchanging	0.83	0.80	0.86	0.90	0.73

_____

ISSN: 1309 – 6575 _Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi_
Journal of Measurement and Evaluation in Education and Psychology

312

When the attribute mastery status of the students as a result of the 7th grade measurements is compared with the results of the 6th grade from Table 8, it has been observed that the mastery status of approximately 4/5 of the students on the basis of the attribute did not change. The biggest change in the 7th grade was seen in attribute 5, in which 25% of the students gained the attribute and 1% lost. The smallest change was seen in attribute 4, where 8% of the students gained the attribute and 2% lost. When the changes are examined, it is seen that more students gained in all attributes. It can be examined in Table 9, which attributed mastery status of the students changed in the 8th grade when compared to the 7th grade.

Table 9. Proportion of Students Whose Attribute Mastery Changes in 8th Grade according to 7th Grade Based on Attribute

Change	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5
Gainer	0.02	0.05	0.01	0.20	0.03
Loser	0.08	0.08	0.08	0.01	0.15
Total Change	0.10	0.13	0.09	0.21	0.18
Unchanging	0.90	0.87	0.91	0.79	0.82

When the attribute mastery status of the students as a result of the 8th grade measurement was compared with the results of the 7th grade Table 9, it was seen that the mastery status of more than 4/5 of the students did not change on the basis of attributes. The biggest change was observed in the 8th grade in which 19% of the students gained the attribute and 1% lost. The smallest change was seen in attribute 3, in which 1% of the students gained the attribute and 8% lost. When the changes are examined, it is seen that more students lost in all attributes except attribute 4.

The Rate of Change in the Number of Attribute Mastered by Students

In order to see the reflection of the changes given in Tables 7 and 8 to the number of attributes mastered, the changes on student basis should be monitored. The rates of students gaining or losing the attribute in the 7th grade according to their 6th grade results are given in Table 10.

Table 10. Attribute Mastery Change Rates in 7th Grade According to 6th Grade on Student Basis

	No Gain	Gain 1	Gain 2	Gain 3	Gain 4	Gain 5
No Lost	0.62	0.08	0.08	0.05	0.05	0.03
Lost 1	0.06	0.01	0	0	0	
Lost 2	0.01	0.01	0	0		
Lost 3	0	0	0			
Lost 4	0	0				
Lost 5	0					

Values in Table 10 showed that 62% of the students remained in the same latent class in the 7th grade when compared with the 6th grade. It is seen that 29% of the students gained attribute/attributes without losing the attributes they have, while 7% lost one or two attributes without gaining attributes. It can be said that the change in the attributes of students was more in the direction of gaining. The rate of attribute changing from 7th to 8th grades on student basis is given in Table 11.

Table 11. Attribute Mastery Change Rates in 8th Grade According to 7th Grade on Student Basis

	No Gain	Gain 1	Gain 2	Gain 3	Gain 4	Gain 5
No Lost	0.66	0.07	0.03	0.02	0	0.01
Lost 1	0.06	0.02	0.03	0.01	0	
Lost 2	0.05	0.01	0	0		
Lost 3	0.01	0	0			
Lost 4	0.01	0				
Lost 5	0					

_____
ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

313

Values in Table 11 indicated that 2/3 of the students remained in the same latent class in the 8th grade when compared with the 7th grade. It is seen that 13% of the students gained new attributes/attributes without losing their attributes, while 13% lost their attributes/attributes without gaining attributes. It is seen that the change was in the direction of gaining or losing attributes of the students was more limited and balanced in the 8th grade. Table 12 shows the correlations between the correct response rate and the attribute mastery probability of the students across years.

Arguments of Reliability and Validity Regarding the Analysis Results

Table 12. Correlations between Correct Response Rate and Attributes Mastery Probability by Years

	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5
6-7	.63	.62	.65	.57	.57
7-8	.77	.77	.77	.61	.71
6. CRR-A	.78	.82	.82	.69	.84
7. CRR-A	.87	.89	.87	.77	.87
8. CRR-A	.87	.89	.86	.86	.89

Note: CRR-A Correlation between Correct Response Rate and Probability of Attribute Mastery

In the first two lines of Table 12, correlations between the attribute mastery probability of students calculated in consecutive years for each attribute were displayed, and in the last three lines correlations were found between the correct response rate and the attribute mastery probability of students in each year. The correlations found in the table were calculated with the Pearson coefficient, and all relationships were found to be significant at the $p<0.01$ level. When the first line is examined, it is seen that the 6th and 7th grades attribute mastery probability was moderately correlated. The lowest correlation coefficient found in the table was found to be between 0.47 belonging to the attribute 4 mastery probability in these years. It is seen that the correlation coefficients regarding the attribute mastery probability of 7th and 8th grades are higher than 6th -7th. The correlation coefficient calculated for attribute 4 was again lower than the other attributes. Other correlations contained high levels of correlation meanings.

When the correlations between the correct responserate and attribute mastery probability, which were carried out to examine the convergent validity of the analysis results, are examined, it is striking that the correlation coefficients were high.The correlation coefficients seen in the 6th grade were highly correlated. The values seen in the 7th grade were higher than the values seen in the 6th grade for all the attributes. In the 7th grade, it was observed that attribute 4 had lower than the other coefficients, again. When eighth-grade values are examined, higher correlation coefficients were observed the ones in previous years. It can be said that the relatively lower correlation coefficient observed for attribute 4 was not observed in the 8th grade values, and all correlation coefficients were close to each other. In Table 13, correct classification rates of students in terms of each attribute and latent classes in each class are given.

Table 13. Classification Accuracy

	Overall	Attribute 1	Attribute 2	Attribute 3	Attribute 4	Attribute 5
6th Grade	.73	.87	.92	.95	.94	.93
7th Grade	.81	.96	.94	.97	.95	.91
8th Grade	.88	.98	.96	.98	.94	.97

Table 13 shows that the lowest classification accuracy which can be taken as the reliability of findings, is in the 6th grade with 0.70; the classification accuracy increased to 0.81 in the 7th grade and 0.88 in the 8th grade. It was observed that the classification accuracy on the basis of attributes was higher than the values obtained for the whole latent class as expected. The average classification accuracy of the attributes was 0.91 for the 6th grade, 0.95 for the 7th grade, and 0.96 for the 8th grade. It is seen that the

correct classification rates on the basis of attributes increased over the years. The fact that the value of the 7th grade for attribute 4 was slightly lower than the value of the 8th grade is considered as an exception for this information. The high rates given provided important information about the accuracy of the classification resulting from the analysis.

DISCUSSION and CONCLUSION

In this study, it was aimed to monitor student achievement with cognitive diagnosis models (CDM). For thispurpose, 2009 6th grade, 2010 7th grade, and 2011 8th grade HSPT mathematics test data were used. Analyses were carried out with the help of Q matrix developed by the experts. When the students' achievements are examined with their raw score in the exam, it is seen that there was an increase (0.34; 0.42) in the transition from the 6th to 7th grade, and a decrease in the transition from the 7th to the 8th grade (0.42; 0.33). On the basis of the HSPT results of a year ago, the opposite changes were observed in the study conducted by Yakar (2011). When acting according to the classical test theory and monitoring student achievements, different results can be obtained due to the effect of the item difficulty.

When it is aimed to obtain qualified and in-depth information about students' achievements, the use of cognitive diagnostic models can be a source of detailed information. The 6th, 7th, and 8th grade HSPT mathematics items were examined during the Q-matrix creation stage by experts. They decided that items require attributes called "Operational Applications" (Attribute 1), "Mathematization Applications" (Attribute 2), "Concept Calculations" (Attribute 3), "Concept Advanced Applications" (Attribute 4), and "Geometric Manipulation" (Attribute 5). Each exam may require specific attributes. Considering the purpose and results of HSPT, it can be said that it is not designed for cognitive diagnosis. In order to benefit from cognitive diagnosis at the highest level, there are many studies that pre-design questions to reveal the existence of qualifications that students should have (Akbay, Terzi, Kaplan, Karaarslan, 2017; de la Torre, van der Ark, & Rossi, 2017; Sorrel et al., 2016; Templin & Henson, 2006; Tjoe& de la Torre, 2014). However, exams with different purposes (Chen & Chen, 2016; Liu, Huggins-Manley &Bulut, 2018; von Davier; 2008) can be used later for cognitive diagnosis by retrofitting. It can be said that while developing the Q matrix in retrofitting studies, examining the AIC and BIC model data fit indexes, making decisions with consensus by experts, and examining the Q-matrix validity with software, are the factors that make the use of the test for cognitive diagnosis functional and meaningful in this study.

As a result of the analysis made using the Q matrices created, the attribute mastery probabilities were generally low, an increase in the transition from the 6th to 7th grade (0.16-0.26) and a partial decrease in the transition from the 7th to 8th grade (0.26-0, 24) were seen. It can be said that the direction of the change (except for Attribute 4 in 8th grade) was similar to the change in the correct response rates of students over the years. It is thought that this situation may be related to the curriculum. Indeed, it is seen that the concepts at the 7th grade were designed as the application of the concepts addressed in the 6th grade, but there are concepts (irrational numbers, inequalities, etc.) that students encounter for the first time at the 8th grade. This opinion is supported by the study of Kablan, Baran, and Hazer (2013). In this study, it is stated that the behaviors targeted according to grade levels were at the comprehension level at the 6th and 8th grades and at the application level at the 7th grade.

In the trend of change in mastery probability of attribute 4, it was seen that the ratio increased slightly in the transition from the 6th to 7th grade, and there was a noteworthy increase (0.17-0.34) in contrast to the general change in the transition from the 7th to 8th grade. It can be thought that the items at the 7th grade were mostly the basic applications of the 6th grade concepts, and the 8th grade items were designed to cover previous learning.

CDMs basically classify individuals according to their attributes. Those with an attribute mastery probability of 0.5 and above were classified as attribute master and those below 0.5 were classified as non-master. When the attribute mastery probability obtained as a result of the analysis wastransformed into the latent class, as expected, the largest latent class was realized as the "00000" group in which the students non-master any attributes. More than 60% of the students took part in this latent class in three years. This situation may mean that the students did not acquire the behaviors targeted in the curriculum

or that the exam does not have the quality to measure these behaviors. The next largest latent class was seen as the "11111" latent group in which the students had all the attributes. However, according to the 6th, 7th, and 8th grades, only 6%, 13%, and 15% of the students were in this latent class, respectively. Although there was no linear hierarchy among the qualifications, it was anticipated that the qualification in the higher group would correspond to a more advanced structure. Accordingly, due to our education system, it is an expected result that as the grade level increased, the probability of having qualifications and even higher-level qualifications would increase in students who encountered different concepts and question types. However, the fact that a student appeared to have qualifications at a grade level should not mean that the relevant student would maintain the same qualifications or have more of that qualification as the grade level increased. Qualifications were not directly subject or curriculum based. The nature of the questions, in which the learning outcomes required by the subject or curriculum were tried to be determined, indicated the mathematical qualifications of the student. For this reason, the properties of the questions selected to measure the learning outcomes in classifying students were crucial. Within the scope of the exams examined in this study, it is noteworthy that the questions for Qualification 1 in the 6th Grade, Qualification 3 in the 7th Grade, and Qualification 3 in the 8th Grade were predominant. According to this situation, one of the most expected learning outcomes from 6th grade students was to complete operational practices, whereas one of the most expected learning outcomes from 7th and 8th grade students was to perform the operations for the computational applications of the concepts. Considering the developmental characteristics of the students, although it was appropriate to expect the applications of the basic qualities of numbers from the 6th grade students, questions that support mathematical thinking beyond the application of operations were expected at the next grade levels. However, the current results did not reflect this expectation. Uğurel, Moralı, and Kesgin (2012) also support this result by stating that HSPT includes information transfer in 6th grade, routine operations in 7th grade, and questions at both knowledge transfer and routine operations level in 8th grade. It can be said that other latent group sizes differed according to grade levels. Şen and Arıcan (2015) conducted a cognitive diagnosis analysis based on TIMSS 2011 8th grade mathematics responses of Turkish students, and they found 13% mastery for all attributes and 1% non-mastery for all attributes. When many variables such as the number of attributes defined for the test, the measurement frequency of the attributes, the examination of the attributes in the same item, and the model used for analysis are partially or completely different, the results to be obtained can vary significantly. Although the findings obtained were specific to the study, the fact that the majority of the students had no attributes was one of the prominent results of the study.

On the basis of attributes, it has been observed not a big change was observed in students' attributes in the 7th grade when compared to the 6th grade. It has been observed that approximately 80% of the students in each attribute did not change. It was observed that the change in the students' attributes in the 8th grade was less than the previous year. It was observed that the status of the attributes mastery did not change between 80-90%. Another prominent result was that the change in attribute 4 in 8th grade was in the opposite direction with the changes in other attributes. Accordingly, while attributes 1, 2, 3, and 5 moved together in terms of the direction of change according to the years, attribute 4 changed in the opposite direction of the others.

When attribute mastery status changes over the years were examined on the basis of students, a little more than half of the students who did not lose or gain any attribute in the 7th grade were compared with the 6th grade. While 38% of the students gained/lost their attributes, it has been observed that most of these students gained new attributes/attributes. When the attributes they mastered in the 8th grade were compared with the 7th grade, it is observed that the change was more limited when compared to the previous year. While no change was observed in 2/3 of the students, it was observed that the number of students who gained and lost their attributes wasclose to each other. When the changes by years on the basis of students and attributes are examined together, it is concluded that the change seen in the 7th grade was more and more positive than the one seen in the 8th grade.

It has been observed that the attribute mastery probabilities had high correlation values for consecutive years. The high correlation value confirmed the conclusion that the stability of the measurements and the changes in attribute mastery probabilities were limited across years. The fact that the correlation

_____
ISSN: 1309 – 6575 _Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi_
Journal of Measurement and Evaluation in Education and Psychology

316

values between the 7th and 8th grades were higher than the correlation values between the 6th and 7th grades shows that the differentiation in the change by years was also reflected in the correlation values.

The correlation coefficients between the number of correct answers and the probability of having the qualifications of the students can be seen as the convergent validity coefficient (Li et al., 2020). These values were found to be high. Thus, it can be said that the obtained results have an argument of validity. It is noteworthy that these correlation values, which could be observed for three different years, generally increased over the years. The fact that the lowest values for all correlation values belonged to attribute 4 continuously can be considered as a reflection of the direction of the change in this attribute's being in the opposite change direction when compared to the other attributes. The primary factor that may cause this situation seems to be that the number of questions related to this qualification was higher in the 8th grade compared to other grade levels. Another factor is the natural consequence of seeing an advanced qualification such as advanced applications of the concept in 8th grade students. Another factor is the natural result of 8th grade students' having an advanced qualification such as advanced applications of the concept. Along with the advanced grade level, the vast knowledge of the students enables them to perform more complex operations on mathematical concepts beyond operational applications.

For CDM, the Q matrix is considered to be the basic element that reflects the design of the assessment tool and determines the quality of the feedback obtained from the assessment tool (Rupp & Templin, 2008). In order to increase the robustness of the CDM results, experts and statistical validation opportunities were used in creating a Q matrix. The accuracy of the classification rates revealed at the end of the analysis was 70% for the 6th grade, 81% for the 7th grade, and 88% for the 8th grade. The accuracy rates on the basis of attributes were found between 87% and 97%. It can be said that the analysis produces more accurate results over theyears. In the studies conducted (de la Torre, Yong, & Deng, 2010; Madison & Bradshaw, 2015), no threshold value for classification accuracy is specified. However, it is seen that the classification accuracy revealed in the study has a higher level than similar studies based on real data (Cui, Gierl, & Chang, 2012; Li et al., 2020; Ma, Iaconangelo, & de la Torre, 2016). The high rates obtained reveal the accuracy of the analysis results and indicate that the comments made on the results can be trusted.

It is among the limitations of this study that HSPT did not have a diagnostic purpose and therefore did not have a predetermined Q matrix. Analysis of the data with a purpose or method other than its original purpose or analysis method is called retrofitting, and potential problems such as model-data fit fatigue may be encountered. Although it is desirable to prepare items based on the Q matrix, there are many studies performed through retrofitting (Chen & Chen, 2016; Liu, Huggins-Manley &Bulut, 2018; von Davier; 2008). The fact that the research data belongs to the previous years can be seen as a limitation. However, the same person in the succession of tests is limited, and HSPT was the only repeated measure high-stake exam for Turkey. There are suggested models in the analysis of longitudinal data with CDM (Huang, 2017; Kaya &Leite, 2017; Zhan et al., 2019). However, it was not possible to use it in this study since all of the suggested models are based on the common item.

It is one of the main advantages of CDMs that they provide detailed information about individuals. The fact that this benefit is also for monitoring student development makes CDMs more functional in evaluation. With the use of CDMs in large-scale exams, the knowledge that students get from the exams will not be limited to the correct numbers they make. By determining what level of deficiencies in which skill they have, the first step will be taken towards making up these deficiencies of students. Other stakeholders in education, such as the school, decision-makers, and parents, will also have the option to act on these deficiencies. When this feature is transferred to the exams held in series, student progress can be examined over the years over common attributes, as shown in the study. In this context, CDMs can be used for monitoring purposes in schools. Based on research results, suggestions for researchers are as follows;

- The research had a design in which the Q matrix was subsequently determined. In future studies, if the Q matrix is determined in advance and the items are created based on this, the classification accuracies can be examined.

_____

ISSN: 1309 – 6575 *EğitimdevePsikolojideÖlçmeveDeğerlendirmeDergisi*
Journal of Measurement and Evaluation in Education and Psychology

317

- Analysis was done repeatedly due to the lack of a suitable growth model. In particular, existing models can be developed to analyze data used in research at once.

- If such a model is developed, the research data can be analyzed again, and the attribute mastery probabilities can be examined.

- The learning outcomes aimed within the scope of the curriculum are subject-curricular-based and remain only within their own context, and it cannot be examined to what extent the students acquire the skills required by these learning outcomes. For this reason, it is not possible to observe the qualifications properly at all levels. It will be more meaningful to determine the skills expected from students in such leveling exams in advance and to create questions in the context of these skills in order to determine the qualifications of the student who will proceed to the next level.

REFERENCES

Akaike, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automated Control*, 19, 716–723. doi:10.1109/TAC.1974.1100705

Akbay, L., Terzi, R., Kaplan, M., &Karaaslan, K. G. (2017). Expert-based attribute identification and validation on fraction subtraction: A cognitively diagnostic assessment application. *Journal on Mathematics Education, 9*(1), 103-120.

Chen, H., ve Chen, J., (2016). Retrofitting Non-cognitive-diagnostic Reading Assessment Under the Generalized DINA Model Framework, *Language Assessment Quarterly,13*(3), 218-230, DOI: 10.1080/15434303.2016.1210610

Cheng, Y. (2010). Improving cognitive diagnostic computerized adaptive testing by balancing attribute coverage: the modified maximum global discrimination index method. *Educational and Psychological Measurement, 70* (6), 902-913.

Cui, Y., Gierl, M. J., & Chang, H. H. (2012). Estimating classification consistency and accuracy for cognitive diagnostic assessment. *Journal of Educational Measurement, 49*(1), 19-38.

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

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika, 76*(2), 179-199.

de la Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement, 47*(2), 227-249.

de la Torre, J., &Minchen N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *PsicologiaEducativa, 20*(2), 89-97.

de la Torre, J., van der Ark, L. A., & Rossi, G. (2017). Analysis of clinical data from a cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, 1-16.

DiBello, L. V., Stout, W. F., & Roussos, L. A. (2007). Cognitive diagnosis Part I. In C. R. Rao & S. Sinharay (Eds.), Handbook of statistics (Vol. 26): *Psychometrics* (pp. 979-1029). Amsterdam: Elsevier

Embretson, S. E. (1998). A cognitive design system approach to generating valid tests: Application to abstract reasoning. *Psychological Methods*, 3, 380-396

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

Huang, H. (2017), Multilevel cognitive diagnosis models for assessing changes in latent attributes. *Journal of Educational Measurement*, 54: 440-480. doi: 10.1111/jedm.12156

Iaconangelo, C.(2017). *Uses of classification error probabilities in the three-step approach to estimating cognitive diagnosis models.* (Unpublished doctoral dissertation). New Brunswick, NJ: Rutgers University.

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.

Kablan, Z., Baran, T. & Hazer, Ö. (2013). İlköğretim matematik 6-8 öğretim programında hedeflenen davranışların bilişsel süreçler açısından incelenmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi, 14* (1), 347- 366.

Kaya, Y., &Leite, W. L. (2017). Assessing change in latent skills across time with longitudinal cognitive diagnosis modeling: An evaluation of model performance. *Educational and Psychological Measurement, 77*(3), 369-388.

Kutlu, Ö., Doğan, C., & Karakaya, İ. (2010). *Öğrenci başarısının belirlenmesi: Performansa ve portfolyoya dayalı durum belirleme.* Pegem, Ankara

_____

ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology   318

Li, L., Zhou, X., Huang, J., Tu, D., Gao, X., Yang, Z., & Li, M. (2020). Assessing kindergarteners' mathematics problem solving: The development of a cognitive diagnostic test. *Studies in Educational Evaluation*, *66*, 100879.

Liu, R., Huggins-Manley, A. C., &Bulut, O. (2018). Retrofitting diagnostic classification models to responses from irt-based assessment forms. *Educational and Psychological Measurement, 78*(3), 357–383. https://doi.org/10.1177/0013164416685599

Leighton, J. P. veGierl M. J. (2007). Why cognitive diagnostic assessment, Leighton, J. P. Gierl M. J. (Eds). *Cognitive Diagnostic Assessment for Education*. Cambridge University Press, New York, USA.

Lohman, D. F. (1999). Minding our p's and q's: On finding relationships between learning and intelligence. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds*.), The future of learning and individual differences: Process, traits, and content* (55f72). Washington, DC: American Psychological Association.

Ma, W. & de la Torre, J. (2019). GDINA: The generalized DINA model framework. R package version 2.3. Retrived from https://CRAN.R-project.org/package=GDINA

Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, *40*(3), 200-217.

Madison, M. J., & Bradshaw, L. P. (2015). The Effects of Q-Matrix Design on Classification Accuracy in the Log-Linear Cognitive Diagnosis Model. *Educational and Psychological Measurement, 75*(3), 491–511. https://doi.org/10.1177/0013164414539162

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212.

Ministry of National Education (MoNE) (2018). *Matematik dersi öğretim programı (İlkokul ve Ortaokul 1, 2, 3, 4, 5, 6, 7 ve 8. Sınıflar)*. Ankara: Talim ve Terbiye Kurulu Başkanlığı.

Ministry of National Education (MoNE), (2008), *64 Soruda Ortaöğretime Geçiş Sistemi ve Seviye Belirleme Sınavı Örnek Sorular. Ankara: MEB Yayınları*

Ministry of National Education (MoNE), (2010), *Seviye belirleme sınavının değerlendirilmesi.* Turkish Ministry of National Education, Retrieved from https://www.meb.gov.tr/earged/earged/sbs_deger.pdf

Ministry of National Education (MoNE), (2017), *Akademik becerilerin izlenmesi ve değerlendirilmesi: 8. Sınıflar raporu.* Turkish Ministry of National Education, Retrieved from http://edirne.meb.gov.tr/meb_iys_dosyalar/2018_06/08104327_ABYDE_Turkiye.pdf.

Organisation for Economic Co-operation and Development (OECD), (2019), *PISA 2018: Insights and Interpretations.*, Retrieved from https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF..

Rupp, A. A., & Templin, J. (2008). The effects of q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement, 68*(1), 78–96. https://doi.org/10.1177/0013164407301545

Schwarzer, G. (1976). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464. doi:10.1214/aos/1176344136

Sezen Yüksel, N., Sağlam Kaya, Y., Urhan, S. & Şefik, Ö. (2019). *Matematik Eğitiminde Modelleme Etkinlikleri (Ed: Şenol Dost)*. Ankara: Pegem Akademi

Smith, G.H., Wood, L.N., Coupland, M., Stephenson, B., Crawford, K. &Ball, G. (1996). Constructing mathematical examinations to assess a range of knowledge and skills. *Int. J. Math. Educ. Sci. Technol., 27*(1), 65-77.

Sorrel, M., Olea, J., Abad, F., de la Torre, J., Aguado, D., &Lievens, F. (2016). Validity and reliability of situational judgement test scores: A new approach based on cognitive diagnosis models. *Organizational Research Methods. 19*(3), 506-532, doi: 10.1177/1094428116630065

Şen, S., &Arıcan, 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.

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.

Templin, J., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287-305.

Tjoe, H., & de la Torre, J. (2014). The identification and validation process of proportional reasoning attributes: An application of a cognitive diagnosis modeling framework. *Mathematics Education Research Journal,* 26, 237-255.

Uğurel, I., Moralı, H.S. &Kesgin, Ş. (2012). OKS, SBS ve TIMSS matematik sorularının 'math taksonomi' çerçevesinde karşılaştırmalı analizi. *Gaziantep Üniversitesi Sosyal Bilimler Dergisi, 11* (2), 423- 444.

von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, 61, 287–307. doi:10.1348/000711007X193957

_____

ISSN: 1309 − 6575 *EğitimdevePsikolojideÖlçmeveDeğerlendirmeDergisi*
Journal of Measurement and Evaluation in Education and Psychology

319

_____

Wang, S., Yang, Y., Culpepper, S. A., & Douglas, J. A. (2018). Tracking skill acquisition with cognitive diagnosis models: a higher-order, hidden markov model with covariates. *Journal of Educational and Behavioral Statistics*, *43*(1), 57-87.

Wang, W., Song, L., Chen, P., Meng, Y., & Ding, S. (2015). Attribute-level and pattern-level classification consistency and accuracy indices for cognitive diagnostic assessment. *Journal of Educational Measurement, 52* , 457-476

Yakar, L. (2011). *İlköğretim ikinci kademe öğrencilerinin SBS puanları ve akademik başarı puanları değişimlerinin izlenmesi ve SBS puanlarının kestirilmesi*. Unpublished Master dissertation. Abant İzzet Baysal Üniversitesi /Eğitim Bilimleri Enstitüsü, Bolu, Turkey.

Zhan, P., Jiao, H., Liao, D., & Li, F. (2019). A longitudinal higher-order diagnostic classification model. *Journal of Educational and Behavioral Statistics*, *44*(3), 251-281.

_____

ISSN: 1309 – 6575 *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*
Journal of Measurement and Evaluation in Education and Psychology

320