



# The Classification of Students Based on Mathematics Achievement Test in TIMSS 2015 Turkey Sample by Cognitive Diagnosis Models\*

Tuba GÜNDÜZ<sup>1</sup>, Mehtap ÇAKAN<sup>2</sup>

<sup>1</sup> Gazi University, Ankara, tubagunduz@gazi.edu.tr  
<http://orcid.org/0000-0002-0921-9290>

<sup>2</sup> Gazi University, Ankara, mehtapcakan@gazi.edu.tr  
<http://orcid.org/0000-0001-6602-6180>

Received : 22.05.2020

Accepted : 19.10.2020

Doi: 10.17522/balikesirnef.741401

---

*Abstract* – The aim of this study is to examine the classification of students in accordance with different attributes on the basis of mathematics test data in TIMSS 2015 Turkey sample. For this purpose, DINA and DINO models were used from Cognitive Diagnosis Models. Four different Q-Matrices were determined as the levels of the content domain (numbers, data display, geometric shapes and measures), the levels of the cognitive domain (knowing, applying, reasoning), the levels of both the content and the cognitive domain, and the levels obtained through Principal Components Analysis as the attributes dimension in the Q-Matrix in which matching of attribute and item is carried out. The accuracy and consistency of the classifications acquired with these four different Q-Matrices were examined as an evaluation criterion for which classification was performed better. The study group of the research consists of 463 fourth-grade students who have taken the first booklet within TIMSS 2015 Turkey sample. The data used in the study were obtained from 23 items selected from 25 items in the TIMSS 2015 4th grade mathematics booklet-1. According to the findings of the research, the Q-Matrix in which content domain levels are used as the attribute dimension, is the matrix that classified the students best. In conclusion, the classification was made by using this matrix according to the DINA Model that conducts the best classification within the scope of the research and its results were discussed.

*Key words:* TIMSS 2015, mathematics achievement, content domain, cognitive domain, Principal Component Analysis (PCA), Cognitive Diagnosis Models (CDM)

-----  
\*This paper was presented at the 6th International Congress on Measurement and Evaluation in Education and Psychology, as oral presentation.

## Introduction

Mathematics education, one of the most crucial building blocks of basic education, takes on a function beyond teaching operations and providing calculation skills and provides important support in terms of establishing connection with events, reasoning, and problem solving (Umay, 2003). As understood from this point of view, mathematics education has at least two dimensions like many other disciplines. Therefore, it is inevitable that these dimensions are taken into consideration while measuring mathematics achievement. Determining to what extent the objects of mathematics lesson have been achieved and whether students have improved in terms of mathematical knowledge and skills depends on the assessment and evaluation of student achievement with correct methods (Yıldız and Uyanık, 2004).

Accurate assessment of mathematical knowledge and skills supports students to learn mathematics, presents information on the developments in the field of mathematics education at national level, and provides the evaluation of the mathematics program (IEA-International Association for the Evaluation of Educational Assessment, 2011). As in many disciplines, assessment and evaluation applications at national level in the field of mathematics education offer a crucial source of information about education in a given country. However, international applications are also required to better evaluate the developments in education. Participation in international evaluation applications that include many countries is necessary in order to determine the current state of the education system in our country, to understand whether the students are trained in accordance with the needs of the information society, to improve the existing education system, and to compare with the education systems of other countries TIMSS (Trends in International Mathematics and Science Study), PISA (Program for International Student Assessment), and PIRLS (Progress in International Reading Literacy Study) can be given as examples of such applications.

The general object of TIMSS, which is the longest and most widely used comparative education study in mathematics and science, is to assess the versatile knowledge and skills of students (Mullis, Martin, Ruddock, O'Sullivan and Preuschoff, 2009). Information is collected about the performance of students in science and mathematics, the education systems of the countries, the education programs, the characteristics of students, teachers and schools in line with this purpose.

The international mathematics competence levels are defined by the IEA as advanced, upper level, intermediate level, and low level based on the mathematics achievement test scores of the students participating in the TIMSS study at the 4th grade level. General information on these levels of competence is given in Table 1 (Mullis, Martin, Foy and Arora, 2012).

**Table1.** TIMSS International Benchmarks of Mathematics Achievement for 4th grade.

<b>Advanced (625):</b>	Students can apply their understanding and knowledge in a variety of relatively complex situations and explain their reasoning. They can solve a variety of multi-step word problems involving whole numbers, including proportions. Students at this level show an increasing understanding of fractions and decimals. Students can apply geometric knowledge of a range of two- and three-dimensional shapes in a variety of situations. They can draw a conclusion from data in a table and justify their conclusion.
<b>High (550-625):</b>	Students can apply their knowledge and understanding to solve problems. Students can solve word problems involving operations with whole numbers. They can use division in a variety of problem situations. They can use their understanding of place value to solve problems. Students can extend patterns to find a later specified term. Students demonstrate understanding of line symmetry and geometric properties. Students can interpret and use data in tables and graphs to solve problems. They can use information in pictographs and tally charts to complete bar graphs.
<b>Intermediate (475-550):</b>	Students can apply basic mathematical knowledge in straightforward situations. Students at this level demonstrate an understanding of whole numbers and some understanding of fractions. Students can visualize three-dimensional shapes from two-dimensional representations. They can interpret bar graphs, pictographs, and tables to solve simple problems.
<b>Low (400-475):</b>	Students have some basic mathematical knowledge. Students can add and subtract whole numbers. They have some recognition of parallel and perpendicular lines, familiar geometric shapes, and coordinate maps. They can read and complete simple bar graphs and tables.

It is understood from Table 1 that the levels are determined by increasing the knowledge and skills from each field in each level increase in relation to the determined levels of competence. For instance, while the competence written in the field of geometry at the low level is "They have some recognition of parallel and perpendicular lines, familiar geometric shapes, and coordinate maps", at the intermediate level, it is "Students can visualize three-dimensional shapes from two-dimensional representations". While competence at the high level is " Students demonstrate understanding of line symmetry and geometric properties ", at the advanced level, it is "Students can apply geometric knowledge of a range of two- and three-dimensional shapes in a variety of situations". In other words, it assumes that each student has similar levels of competence in each field in determining these levels. On the other hand, student profiles for those who are competent in one field but less competent in another field are not included among

these levels. In other words, a student who is competent in algebra but less competent in geometry and shapes, and two students who are competent in geometry and shapes but less competent in algebra and take 500 are considered at the same level.

Assessment and evaluation, as elements of the education system, are required to reveal students' inadequacies and their reasons (Baykul, 1992). In order to increase achievement, firstly, individual inadequacies of students are required to be determined and precautions should be taken in this direction. According to the Learning for Mastery Model of Bloom (1968), which had a great influence, the main task in education is to find strategies that take individual differences into account in a way that promotes the individual's mastery learning (development). The type of evaluation that takes into account individual differences is the formative evaluation.

The total score obtained from the formative tests used with the purpose of formative evaluation has almost no meaning for this type of evaluation. Each student should be given information showing the patterns of his/her responses to the items and the mistakes he/she made. Because, this evaluation is carried out to provide information to the teacher and the student about the effectiveness of teaching in this way. Thus, it is possible to adjust the teaching according to the learning speed and learning strength of the students (Tekin, 2007, p.26).

Psychometric approaches such as Classical Test Theory (CTT) and Item Response Theory (IRT) generally focus on scaling or ranking individuals in accordance with some latent characteristics (Rupp, Templin, and Henson 2010). However, they cannot make a multivariate classification about individuals based on the presence or absence of more than one skill or attribute on an item basis. Modern assessment methods enable this and focus on cognitive skills diagnosis that can provide feedback on the strengths and weaknesses of specific learning goals (Rupp and Templin, 2008). This allows more informative feedback on student skill sets and enables more effective response strategies to be designed (Rupp et al., 2010).

### *Cognitive Diagnostic Models*

Through the evaluations made under Cognitive Diagnosis Models (CDMs), which is one of the modern assessment theories, an individual's detailed information about the knowledge/skills of the students in a learning domain can be obtained, and the efficiency of feedback element of the education system can be provided. For example, a student needs to

know the addition, multiplication, division, and operation priority to answer  $(3+2 \times 6) \div 3$  correctly. In this item, it is aimed to assess four characteristics: addition, multiplication, division, and operation priority. Concluding that a student who answers "10" to this question can perform addition, multiplication and division operations, but is inadequate in operation priority can provide diagnostic information in profiling the student in the context of the relevant attributes. Since CDMs have a purpose to identify students' strengths and weaknesses in a particular content domain or cognitive domain, they may include determining the best application strategy for the student through classroom applications (Rupp et al., 2010). It is understood that the student who made a mistake in this example is sufficient in the four operations skills, but needs to improve operation priority and the necessary precautions should be taken in this direction. The relevance of a test with Cognitive Diagnostic Models is to estimate student profiles rather than providing a general ability estimate. Separate attribute profiles designed for individual participants are presented in the reporting of CDM.

CDM is an estimate of the probability of a direct participant meeting the criteria for a given diagnosis. (Leighton and Gierl, 2007). In other words, CDM provides statistical classification of participants according to one or more diagnostic criteria. Based on the example given above, it is assumed that the number of attributes is three (Addition and Subtraction, Multiplication and Division, Operation Priority). For these three characteristics determined here, a total of eight (If  $k$  is the number of characteristics, the number of latent classes that can occur:  $2^k$ ) classes will be formed. These are (000)(001)(010)(100)(110)(011)(101)(111). The student in the example above can be considered to be in the (110) class. Classifications for CDM are based on observed responses collected by diagnostic evaluation.

Classification in CDMs can be carried out according to whether the latent predictor variable and observed response variables have two or multiple categories. In addition to this classification, there are two types of models in CDMs: compensatory and non-compensatory models. The main difference between compensatory and non-compensatory models is in determining how latent predictive variables are combined between different attributes in providing the observed response (Rupp and Templin, 2008). While the deficiency of one attribute can be fulfilled by the remaining attributes in compensatory models, each attribute is required for the correct answer in the non-compensatory model. In other words, the missing attribute in compensatory models can be fulfilled by another attribute, whereas the missing attribute cannot be fulfilled by any other attribute in non-compensating models. Among the

non-compensatory and compensatory models, the most basic models frequently used in the literature are DINA (Deterministic inputs, noisy 'and' gate) (Junker and Sijtsma, 2001) and DINO (Deterministic inputs, noisy 'or' gate) (Templin and Henson, 2006) models.

The common characteristics of all models of CDM is that the Q-matrix must be determined. Q-matrix is also called "item characteristic matrix" and/or "characteristic matrix" in the literature (Ma, 2014; Tatsuoka, 1984). In the Q-matrix, the item and the attribute or attributes required to answer the items correctly are associated. This matrix is two-dimensional, with items in a row and attributes in a separate column. For the attributes required to solve an item to the relevant section, "1" is written and "0" is written for unnecessary attributes. Table 2 contains an example Q-matrix.

**Table 2.** An Example Q-Matrix

	Attribute 1	Attribute 2	Attribute 3
Item 1	0	1	0
Item 2	1	0	1
Item 3	0	1	1
Item 4	1	1	1

In the Q-matrix prepared for four items and three attributes in Table 2, Item 1 is associated with only Attribute 2 and Item 4 is associated with all three attributes.

Although the Q-matrix consisting of 0-1 elements appears to have a simple structure, it has a crucial place for CDM since an error in determining the Q-matrix can completely and adversely affect the analysis. Different approaches can be used in the development of the Q-matrix. The most commonly used approach is the development of the Q-matrix in accordance with expert decision. For this reason, structuring the Q-matrix can be seen as the process with the most effort in CDMs. The process requires that items and attributes continue in an interactive manner, cyclically until the desired result is reached.

In determining the Q-matrix, expert opinions in the relevant discipline are typically used. Q-matrix is one of the most crucial concepts of CDM and determining the Q-matrix correctly and completely can affect the whole classification (Tatsuoka, 1984). Başokçu (2015) states that there are cases in which different characteristics not defined by the Q-matrix may be sufficient for an item to be answered correctly. This case will negatively affect the accuracy and consistency of the classification to be made.

In this study, it has been set out with the estimation that the Q-matrix can be better defined by using exploratory information (factor analytical approaches) of the test items in addition to determining the levels in different domains according to expert decisions in relation to the items in the Q-matrix in accordance with CDM. The consistency and reliability of the classification made in this way has been intended to be compared with the Q matrix results, which have been formed only by considering expert opinion. By using expert decisions (from a priori item information), Q-matrices containing the attributes of the cognitive domain, content domain and both domains were used. Based on the data, Principal Components Analysis, one of the exploratory factor analytical approaches suitable for dichotomous scoring, was employed. Principal Components Analysis, generally used for the purpose of eliminating the dependency structure between variables and/or dimension reduction, is used as an analysis on its own, as well as as a data preparation technique for other analyses (Tatlıdil, 1996).

The purpose of this study is to comparatively examine the accuracy and consistency of the classification made by determining the level of content domain, the levels of the cognitive domain belonging to the Q-matrix attribute dimension, the levels of both domains and the levels obtained through Principal Components Analysis under the DINA and DINO Models. In line with these findings, another purpose of the study is to examine the model that makes the classification relatively the best and the classification of students regarding the Q-matrix. It is considered that this study will contribute to the theory in determining the Q-matrix, which has a key place in CDM Evaluations, and may be important in suggesting a related method.

Answers to the following research questions have been sought in line with the purpose of the study:

1. Based on the mathematics achievement test in TIMSS 2015 Turkey sample, which of the four different Q-matrices formed by using DINA and DINO Models from Cognitive Diagnostic Models is used to make a better classification in terms of classification consistency and accuracy?
2. How are students classified in the model and Q-matrix that provides the best classification based on the mathematics achievement test?

## **Method**

### *Research Model*

The purpose of the study is to compare the classification of students taking mathematics tests in TIMSS 2015 Turkey sample in terms of different attributes in accordance with DINA  
**NEF-EFMED Cilt 14, Sayı 2, Aralık 2020/ NFE-EJMSE Vol. 14, No. 2, December 2020**

and DINO Models and to examine how students are classified according to the classification found relatively the best from the classifications. In line with this purpose, the study is a descriptive study as it is aimed to examine the classification accuracy and consistency of four different Q-matrices.

### *Working Group*

The study group of the research consists of a total of 463 fourth grade students, 232 of whom are girls (50.1%) and 231 of whom are boys (49.9%) who have taken the first booklet within the scope of TIMSS 2015 Turkey sample. Before making the analysis, the data were analysed and the missing data analysis was made. As a result of the missing data analysis, it was understood that the missing data did not have a certain pattern and a "Median of Nearby Points" was performed.

### *Data Collection Tool*

Research data, TIMSS 2015 4. It has been obtained from the 25 items in the Class 1 mathematics booklet, and it has been obtained from 23 items, nine of which are short-answered and 14 of which are multiple-choice items.

The assessment scope of the TIMSS study was organized in two dimensions: the cognitive domain and the content dimension at the 4th level of the mathematics achievement test. The content dimension takes the scope to be assessed, that is, mathematical subjects into account, while the cognitive domain dimension takes the thinking processes into consideration. At the 4th grade level, the content dimension consists of numbers, geometric shapes and measuring, and data display. Knowing, applying and reasoning constitute the cognitive domain dimension.

### *Organizing and Analysis the Data*

While organizing the Q-matrices used in the study, the item information in the international database of the IEA (2015) was used for the item attribute associations of the predetermined three Q-matrices (<https://timssandpirls.bc.edu/timss2015/international->



database/). Q-matrices as "Numbers, Data Display, Geometric Shapes and Measuring" under the content domain, "Knowing, Applying, Reasoning" under the cognitive domain, Q-matrices in which the attributes in both domains used together are listed in Table 3, Table 4 and Table 5, respectively.

**Table 3.** Q-Matrix Determined According to Content Domain

Items	$\alpha_S$	$\alpha_G$	$\alpha_V$	Items	$\alpha_S$	$\alpha_G$	$\alpha_V$	Items	$\alpha_S$	$\alpha_G$	$\alpha_V$
M041004	1	0	0	M041254	0	1	0	M061050	1	0	0
M041023	1	0	0	M041153	0	1	0	M061167	1	0	0
M041034	1	0	0	M041132	0	1	0	M061206	0	1	0
M041087	1	0	0	M041174	0	0	1	M061265A	0	1	0
M041124	1	0	0	M041191	0	0	1	M061265B	0	1	0
M041302A	0	1	0	M061272	1	0	0	M061185	0	1	0
M041302B	0	1	0	M061029	1	0	0	M061239	0	0	1
M041302C	0	1	0	M061031	1	0	0				

$\alpha_S$ : Numbers,  $\alpha_V$ : Data Display,  $\alpha_G$ : Geometric Shapes and Measures

**Table 4.** Q-Matrix Determined According to Cognitive Domain

Items	$\alpha_B$	$\alpha_U$	$\alpha_A$	Items	$\alpha_B$	$\alpha_U$	$\alpha_A$	Items	$\alpha_B$	$\alpha_U$	$\alpha_A$
M041004	1	0	0	M041254	1	0	0	M061050	0	1	0
M041023	1	0	0	M041153	0	1	0	M061167	1	0	0
M041034	1	0	0	M041132	1	0	0	M061206	0	1	0
M041087	1	0	0	M041174	0	1	0	M061265A	0	1	0
M041124	0	1	0	M041191	1	0	0	M061265B	0	1	0
M041302A	1	0	0	M061272	0	0	1	M061185	0	1	0
M041302B	0	0	1	M061029	0	0	1	M061239	0	1	0
M041302C	0	0	1	M061031	0	1	0				

$\alpha_B$ : Knowing,  $\alpha_U$ : Applying,  $\alpha_A$ : Reasoning

**Table 5.** Q-Matrix Determined According to Content and Cognitive Domain

Items	$\alpha_S$	$\alpha_G$	$\alpha_V$	$\alpha_B$	$\alpha_U$	$\alpha_A$	Items	$\alpha_S$	$\alpha_G$	$\alpha_V$	$\alpha_B$	$\alpha_U$	$\alpha_A$
M041004	1	0	0	1	0	0	M041191	0	0	1	1	0	0
M041023	1	0	0	1	0	0	M061272	1	0	0	0	0	1
M041034	1	0	0	1	0	0	M061029	1	0	0	0	0	1
M041087	1	0	0	1	0	0	M061031	1	0	0	0	1	0
M041124	1	0	0	0	1	0	M061050	1	0	0	0	1	0
M041302A	0	1	0	1	0	0	M061167	1	0	0	1	0	0
M041302B	0	1	0	0	0	1	M061206	0	1	0	0	1	0
M041302C	0	1	0	0	0	1	M061265A	0	1	0	0	1	0
M041254	0	1	0	1	0	0	M061265B	0	1	0	0	1	0
M041153	0	1	0	0	1	0	M061185	0	1	0	0	1	0
M041132	0	1	0	1	0	0	M061239	0	0	1	0	1	0
M041174	0	0	1	0	1	0							

$\alpha_S$ : Numbers,  $\alpha_G$ : Geometric Shapes and Measures,  $\alpha_V$ : Data Display,,  
 $\alpha_B$ : Knowing,  $\alpha_U$ : Applying,  $\alpha_A$ : Reasoning

As seen in Table 3 and Table 4, each item is associated with only one attribute. As seen in Table 5, each item is associated with two attributes in the Q-matrix formed by the content domain and the cognitive domain.

FACTOR 10.08.3 program, which enables analysis based on tetrachoric correlation matrix, was used to create a Q-matrix with Principal Components Analysis. The suitability of the data within the scope of the study for the analysis of principal components was determined by the Kaiser-Mayer-Olkin (KMO) coefficient and the Barlett Sphericity Test. Since the KMO coefficient of the data in PCA (Principal Components Analysis) was found to be 0,91 and greater than 0,60 and the Barlett Sphericity test was considered significant ( $p < 0,01$ ), it was decided that the test used was suitable for PCA.

For the four components in the analysis result, those with loads as 0,30 and greater than 0.30 a in the Q-matrix were determined as "1" and smaller ones as "0". Table 6 contains the Q-matrix determined according to PCA.

**Table 6.** Q-Matrix Determined According to Principal Components Analysis

Items	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	Items	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	Items	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$
M041004	0	0	0	1	M041254	1	0	0	0	M061050	0	1	0	0
M041023	0	0	0	1	M041153	1	0	1	0	M061167	0	0	0	1
M041034	0	0	0	1	M041132	0	1	1	0	M061206	1	0	0	0
M041087	0	1	0	0	M041174	0	0	1	1	M061265A	0	0	0	1
M041124	0	0	0	1	M041191	0	0	1	1	M061265B	0	1	0	0
M041302A	1	0	1	0	M061272	0	0	0	1	M061185	0	0	1	1
M041302B	0	0	1	0	M061029	1	0	0	1	M061239	0	0	0	1
M041302C	1	0	1	0	M061031	0	1	0	1	M041302C	1	0	1	0

$\alpha_1$ : First Component,  $\alpha_2$ : Second Component,  $\alpha_3$ : Third Component,  $\alpha_4$ : Fourth Component

When Table 6 is examined, nearly 60% of the items are associated with one attribute, while the remaining ones are associated with two attributes. The analysis of the different Q-matrices and the students' response patterns according to the DINA and DINO models was carried out with the CDM package (George, Robitzsch, Kiefer, Groß, and Ünlü, 2016) in the R 3.0 program (R Core Team, 2015).

### Findings and Comments

In this section, findings related to the research questions are presented accordingly.

1. *"Based on the mathematics achievement test in TIMSS 2015 Turkey sample, which of the four different Q-matrices formed by using DINA and DINO Models from Cognitive Diagnostic Models is used to make a better classification in terms of classification consistency and accuracy?"*

The accuracy and consistency of the classification obtained as a result of the analysis of the Q-matrix according to the previously known characteristics of the items (cognitive domain, content domain and the use of both domains) and the items according to the information obtained after the application (determining the characteristics using PCA) are presented in Table 7.

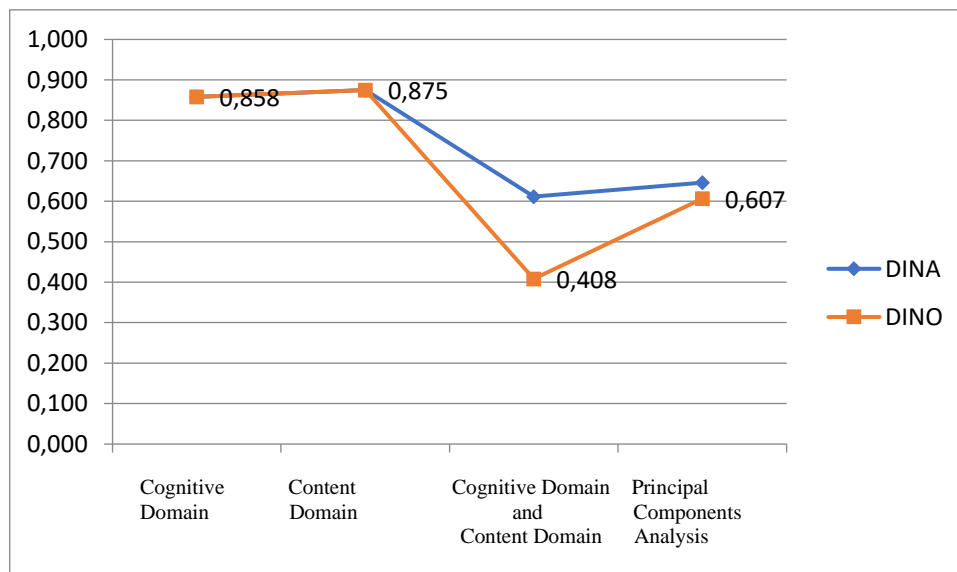
**Table 7.** Classification Accuracy and Consistency Information Obtained as a Result of Determination of Q-Matrix on the basis of Four Different Conditions

Attribute Dimension of Q-matrix	Classification Accuracy		Classification Consistency	
	DINA	DINO	DINA	DINO
Cognitive Domain	0.916	0.916	0.858	0.858
Content Domain	0.929	0.929	0.875	0.875
Cognitive and Content Domains	0.582	0.443	0.612	0.408
Domain revealed by PCA	0.699	0.712	0.646	0.607

When Table 7 is examined, it is understood that the highest classification accuracy and consistency are obtained from matrices created based on content domain and cognitive domain. Another remarkable result is that the DINA and DINO models in both cases form the same classification accuracy and consistency. The reason for this is that in both Q-matrices, only one column in each row is determined as 1. In other words, each item assesses only one domain or skill. This finding is a finding which supports that DINA and DINO Models are mathematically

equal models (Frey, 2018). When it is evaluated relatively and the Q-matrix is determined according to the content domain, it is classified under DINA and DINO Models with a higher rate of classification accuracy and consistency than all other Q-matrix determination methods. This is followed by the classifications made by determining the Q-matrix according to the cognitive domain, determining the Q-matrix using the PCA, and determining the matrix obtained by taking the cognitive and content domain together into the attribute set, respectively. Considering the number of attributes in the determined Q-matrix, it can be said that less consistent and accurate classification is made with Q-matrices with more attribute.

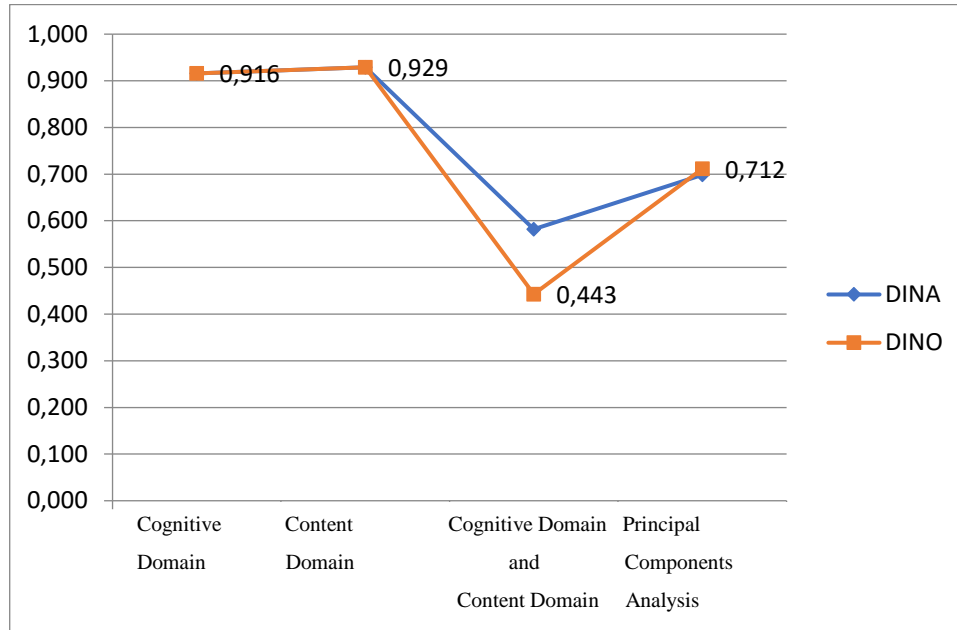
Within the scope of the study, the classification consistencies obtained from the classifications made with Q-matrices determined under four different conditions taking the DINA and DINO models into account are presented in Figure 1, and classification accuracies are presented in Figure 2 comparatively.



**Figure 1.** Classification Consistency under DINA and DINO Models with Differently Determined Q-matrices

When Figure 1 is examined, it is understood that the DINA Model makes a more consistent classification than the DINO Model. Considering that both cognitive and content domain characteristics are used together in the Q-matrix, where cognitive and content domains are used together, it can be concluded that the DINO Model gives less consistent results due to its compensatory structure (having at least one item-related characteristics in order to answer an item correctly). On the other hand, as a result of the analysis made with the Q-matrix created

by using Principal Components Analysis, the classification consistency gives similar values in both models.



**Figure 2.** Classification Accuracy under DINA and DINO Models with Differently Determined Q-matrices

When Figure 2 is examined, it is understood that the DINA Model makes a more accurate classification than the DINO Model. Considering that both the cognitive and content domain characteristics required in the solution of an item in the Q-matrix, where cognitive and content domains are used together as in the classification consistency, it can be concluded that the DINO Model gives results with less classification accuracy due to its compensatory structure. On the other hand, as a result of the analysis made with the Q-matrix created by using Principal Components Analysis, the classification accuracy gives similar values again in both models.

*2. How are students classified in the model and Q-matrix that provides the best classification based on the mathematics achievement test?*

As emphasized in the first research question, it is understood that the Q-matrix is structured based on the content domain with more accuracy and consistency compared to other cases. M2, RMSEA and SRMSR, which are among the absolute fit indices related to this classification made under the DINA Model, were examined and it was observed that the model data fit was achieved (M2=335,50, df= 223, p-value= 0; 0.033 with 90 % CI: [0.0255 , 0.0401]; SRMSR = 0.0617 ). Table 8 gives a summary of the item statistics obtained under the DINA Model of the Q-matrix structured by content domain.

**Table 8.** Summary of item characteristics for DINA model with Q-matrix of Content Domain

Item Characteristics	Minimum	Maximum	Average	Standard Deviation
p-value (item difficulty)	0.14	0.86	0.48	0.18
g parameter	0.05	0.70	0.26	0.18
s parameter	0.03	0.79	0.34	0.22
Item distinctiveness parameter $\omega_{1j} = 1 - g_j - s_j$	0.08	0.68	0.40	0.16
Item convenience parameter $\omega_{2j} = (g_j + (1 - s_j))/2$	0.14	0.83	0.46	0.19

The p value, which gives the item difficulty, describes the percentage of students who correctly solved each item. While 14% of the students answered the most difficult item correctly, 86% of the students answered the easiest item correctly. On average, 48% of the students answered the items correctly. As can be understood from this, the test items have a little less than medium difficulty for this group. As is known, there is a relationship between item difficulty and item convenience parameter, and Table 6 supports this (George and Robitzsch, 2015). The g parameter values range from 0.05 to 0.70 ( $ss = 0.18$ ) and has an average value of 0.26. The s parameter changes between 0.03 and 0.79, and its average value is 0.34. The low level of both parameters is another indicator that the items have medium difficulty (Zhang, 2006). Item distinctiveness parameters vary between 0.14 and 0.83 and their average is 0.46. Although there are items with low distinctiveness, it can be argued that the overall test is minimally distinctive based on the average distinctiveness parameter of the items.

Table 9 demonstrates the rates of students' having relevant attributes (skill distributions) and their standard errors.

**Table 9.** Skill Distributions and Standard Errors for DINA Model with Q-Matrix of Content Domain

	$\alpha_S$	$\alpha_G$	$\alpha_V$
P( $\alpha_k$ )	0.63	0.52	0.61
Standard Error	0.03	0.04	0.03

$\alpha_S$ : Numbers,  $\alpha_G$ : Geometric Shapes and Measures,  $\alpha_V$ : Data Display

It is found that 63% of the students have a comprehensive knowledge in the "numbers" subject field, 52% of them in "geometric shapes and measuring" subject field and 61% of them

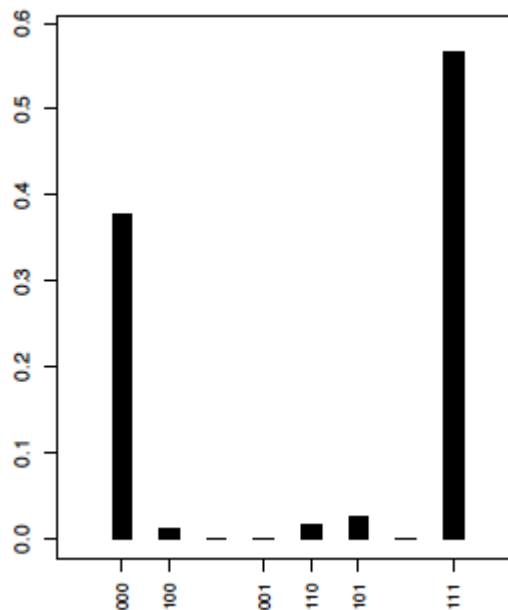
in the "data display" subject field. This also means that 37% of the students do not have a comprehensive knowledge in the "numbers" subject field; 48% of them in the "geometric shapes and measuring" subject field and 39% of them "data display" subject field.

In Table 10, each skill class probabilities (probability of having a combination of attributes that students have or do not have) and standard errors are given.

**Table 10.** Skill Class Probabilities and Standard Errors for DINA Model with Q-Matrix of Content Domain

	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$
	[0.0.0]	[1. 0.0]	[0.1.0]	[0. 0.1]	[1.1.0]	[1. 0.1]	[0. 1.1]	[1. 1.1]
$P(\alpha_i)$	0.35	0.04	0.00	0.02	0.00	0.07	0.00	0.52
Standard Error	0.03	0.01	0.01	0.01	0.01	0.02	0.01	0.04

When Table 10 is examined, it is understood that there are probability values of each class since  $2^3$  content domains are considered. Standard errors related to class probabilities in the classification of content domain vary between 0.01 and 0.04. The graph of the skill class probabilities formed by the individuals within the scope of the relevant attributes is given in Figure 1.



**Figure 3.** Skill Class Probabilities for DINA Model with Q-Matrix of Content Domain

Table 10 and Figure 3 demonstrate the distribution probabilities of different classes and enable the analysis of attribute combinations. It is understood from the distribution of attribute classes that most students either have all skills ( $P([1,1,1]) = 0.52$ ) or none ( $P([0,0,0]) = 0.35$ ). In other words, while approximately half of the students have a comprehensive knowledge in all three content domains, it is understood that 35% of the students do not have the comprehensive knowledge in all three content domains. In addition to this, students who have a comprehensive knowledge in the content domain of numbers and those who have a comprehensive knowledge of the content domain of data display constitute the vast majority with a rate like  $P([1,0,1]) + P([1,1,1]) = 0.07 + 0.52 = 0.59$ . There are no students ( $P([0,1,0]) + P([0,1,1]) = 0.00 + 0.00 = 0.00$ ) who have a comprehensive knowledge in the field of geometric shapes and measuring but do not have a comprehensive knowledge in the field of numbers. There are no students ( $P([0,1,0]) + P([0,1,1]) = 0.00 + 0.00 = 0.00$ ) who have a comprehensive knowledge in the field of geometric shapes and measuring but do not have a comprehensive knowledge in the field of numbers. In other words, this result demonstrates that for this sample, having the fields of the geometric shape and measuring requires comprehensive knowledge in the other two fields. It is understood that the class of a student, who has a comprehensive knowledge of numbers and data display content domains but does not have a comprehensive knowledge in geometry field, constitutes a 7% . Students who have a comprehensive knowledge in the field of numbers constituted the vast majority with the rate of  $P([1,0,0]) + P([1,0,1]) + P([1,1,0]) + P([1,1,1]) = 0.04 + 0.07 + 0.00 + 0.52 = 0.63$ .

### Conclusion

One of the aims of this study is to compare the accuracy and consistency of the classification made by determining the attribute dimension of the Q-matrix based on the data obtained at the end of the application, as well as considering the a priori item knowledge or field expert knowledge of the students who took the first booklet based on the mathematics achievement test in TIMSS 2015 Turkey sample. Another aim of the study is to examine the classification of students regarding the Q-matrix and the model that makes the classification relatively best in line with these findings.

According to the results of the first research question of the study, the Q-matrix, which is formed by using the attributes of the content domain, makes classifications with the best consistency and accuracy in both classifications based on DINA and DINO Models. On the other hand, the DINA Model generally provided higher classification accuracy and consistency



compared to the DINO Model. In addition to this, in the classifications made with the Q-matrix determined by using the factor analytical methods using the PCA, the classification was made with less accuracy and consistency than the classification made with the Q-matrix, which was determined based on the item information with a lower number of attributes. On the other hand, the classification was more accurate and consistent than the classification made with the Q-matrix, which was determined based on the item information with a higher number of attributes. It has been stated by many researchers that as the number of attributes in the Q-matrix increases, prediction errors increase (Bradshaw and Templin, 2014) and that as the number of attributes increases, less accurate classification can be made (Henson and Douglas 2005; Kunina-Habenicht, Rupp and Wilhelm, 2012). Nevertheless, using an exploratory factor analytical method by combining the determined Q-matrix with the item knowledge based on expert decisions can be a powerful tool to strengthen the theory in the development of the Q-matrix. Liu, Douglas, and Henson (2009) state that although factor analytical approaches are not generally used in determining the Q-matrix, these methods can provide a reasonable solution when the Q-matrix is not overly complex. Nonetheless, the item and attribute relationship of the Q-matrix, which is determined in the PCA within the scope of the study, is not based on theory. With another study, the Q-matrix can be revised again by applying expert decisions. The classification of individuals may be insufficient because the items in the assessment tool, from which the data used in this study were obtained, do not go through the process of developing tests suitable for Cognitive Diagnosis Models. This limitation can be eliminated with future studies.

According to the results of the study related to the other research question, it has been demonstrated how the students within the scope of the research are classified according to their content domain attributes. While this classification demonstrates the subject fields in which the students have the comprehensive knowledge, it also gives information about the subject fields that need to be developed. For instance, it has been observed that some of the students have comprehensive knowledge in the field of numbers, but do not have in geometry and shapes. It is suggested to students in this class to plan their education processes related to geometry and shapes rather than numbers. On the other hand, it is understood that geometry and shapes fields are the fields where the student have the least comprehensive knowledge compared to the other two fields. Studies can be conducted to reveal the factors that cause deficiencies in this field. Giving suggestions based on results in this direction reveals the advantage of using CDM to mathematics educators and researchers, unlike other common psychometric models. In addition

to this, it is understood that the classification of the students according to the classification made with CDM rather than the international mathematics competence level in the TIMSS study is more informative. It has been observed that the majority of the students in the study group have all or none of the skills and that this situation is parallel to the findings of some researchers using real data. (George and Robitzsch, 2015; Yi, 2017; Arıcan, 2019; Ardiç, 2020; Koyuncu, 2020). As the result of this research, the information obtained based on CDM is powerful and can be used to reveal information on the efficiency and effectiveness of teaching (Leighton and Gierl, 2007). In line with these results, it is recommended to combine the findings obtained from the classifications made on different samples of the same stage and to make the necessary improvement studies in our education system within a certain plan. For instance, regarding the effective provision of courses organized by the Ministry of National Education General Directorate of Assessment, Evaluation and Examination Services to support and educate willing students and trainees studying in official and private formal education institutions and non-formal education institutions, the information provided by CDM can be used in organizing the courses by knowing the strengths and weaknesses of the student groups.

## References

- Ardıç, E., Ö. (2020). *Bilişsel tanı ve çok boyutlu madde tepki modellerinin sınıflama doğruluğu ve parametrelerinin karşılaştırılması*. (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Arıcan, M., (2019). A diagnostic assessment to middle school students' proportional reasoning. *Turkish Journal of Education*, 8(4), 237-257. DOI: 10.19128/turje.522839
- Başokçu, T. O. (2015). The Cognitive Diagnostic Models for Estimating Students' Ability and Their Applications. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14 (1), 1-32. Retrieved from <http://dergipark.gov.tr/aibuefd/issue/1502/18183>
- Baykul, Y. (1992). Eğitim sisteminde değerlendirme. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 7(7).
- Bloom, B. S. (1968). Learning for Mastery. *Instruction and Curriculum*. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation comment*, 1(2), n2.
- Büyüköztürk, Ş., Çakan, M., Tan, Ş., ve Atar, H. Y. (2014). *TIMSS 2011 Ulusal Matematik ve Fen Raporu 4. Sınıflar*. Ankara: Milli Eğitim Bakanlığı Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü.

- EARGED (2003). *TIMSS 1999 üçüncü uluslararası matematik ve fen bilgisi çalışması*. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- Frey, B. B. (Ed.). (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications.
- George, A. C., & Robitzsch, A. (2015) CognitiveDiagnosisModels in R: A Didactic. *The Quantitative Methods for Psychology*, 11(3), 189-205.
- George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R Package CDM for Cognitive Diagnosis Models. *Journal of Statistical Software*, 74(2), 1–24.
- Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, 29(4), 262-277.
- International Association for the Evaluation of Educational Achievement – IEA. (2011). *Trends in International Mathematics and Science Study 2011*. <http://www.iea.nl/timss-2011> adresinden 17 Haziran 2018 tarihinde erişilmiştir.
- International Association for the Evaluation of Educational Achievement – IEA. (2015). *TIMSS 2015 Assessment Frameworks*. Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Karasar, N. (2005). *Bilimsel araştırma yöntemi*. Ankara: Nobel.
- Koyuncu, M., S. (2020). *Bilişsel tanı modellerinde yapısal eşitlik modeli ile Q-matris doğruluğunun belirlenmesi*. (Yayınlanmamış yüksek lisans tezi). Gazi Üniversitesi, Ankara.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models. *Journal of Educational Measurement*, 49(1), 59-81.
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Liu, Y., Douglas, J. A., & Henson, R. A. (2009). Testing person fit in cognitive diagnosis. *Applied Psychological Measurement*, 33(8), 579.
- Ma, L. (2014). Validation of the item-attribute matrix in TIMSS–Mathematics using multiple regression and the LSDM (Order No. 3616238). Available from ProQuest Dissertations
- NEF-EFMED Cilt 14, Sayı 2, Aralık 2020/ NFE-EJMSE Vol. 14, No. 2, December 2020**

- & Theses Global. (1525999295). Retrieved from <https://search.proquest.com/docview/1525999295?accountid=11054>
- Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O’Sullivan, C.Y. ve Preuschoff, C. (2009). *TIMSS 2011 assessment framework*. MA: TIMSS & PIRLS International Study Center, Boston College.
- R Core Team. (2015). *R: a language and environment for statistical computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org> sitesinden erişildi.
- Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, 6(4), 219-262.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: The Guilford Press.
- Tekin, H. (2007). *Eğitimde Ölçme ve Değerlendirme*. (Gözden Geçirilmiş 18. Baskı) Yargı Yayınevi.
- Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79(2), 317-339.
- Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287–305.
- Umay, A. (2003). Matematiksel muhakeme yeteneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 24(24).
- Tatlıldil, H. (1996). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Cem Web Ofset Ltd. Şti, Ankara.
- Yıldız, İ., & Uyanık, N. (2004). Günümüz matematik öğretimi ve yakın çevre etkileri. *Kastamonu Eğitim Dergisi*, 12(2), 437-442.
- Yi, Y. S. (2017). In search of optimal cognitive diagnostic model (s) for ESL grammar test data. *Applied Measurement in Education*, 30(2), 82-101.
- Zhang, W. (2006). *Detecting Differential Item Functioning Using the DINA Model*. Yayımlanmamış doktora tezi. The University of North Carolina at Greensboro.





# TIMSS 2015 Türkiye Örnekleminde Matematik Başarı Testine Dayalı Olarak Öğrencilerin Bilişsel Tanı Modelleri ile Sınıflandırılması\*

Tuba GÜNDÜZ<sup>1</sup>, Mehtap ÇAKAN<sup>2</sup>

<sup>1</sup> Gazi Üniversitesi, Ankara, tubagunduz@gazi.edu.tr  
<http://orcid.org/0000-0002-0921-9290>

<sup>2</sup> Gazi Üniversitesi, Ankara, mehtapcakan@gazi.edu.tr  
<http://orcid.org/0000-0001-6602-6180>

Gönderme Tarihi: 22.05.2020

Kabul Tarihi: 19.10.2020

Doi: 10.17522/balikesirnef.741401

*Özet* – Bu çalışmanın amacı TIMSS 2015 Türkiye örnekleminde matematik testi verilerine dayalı olarak öğrencilerin farklı niteliklere göre sınıflandırılmasını incelemektedir. Bu amaç için Bilişsel Tanı Modellerinden DINA ve DINO Modellerinden yararlanılmıştır. Nitelik ile madde eşleştirmesinin yapıldığı Q-Matrisinde nitelik boyutu olarak içerik alanına ait düzeyler (sayılar, veri gösterimi, geometrik şekiller ve ölçümler), bilişsel alana ait düzeyler (bilme, uygulama, akıl yürütme), hem içerik hem de bilişsel alana ait düzeyler ve Temel Bileşenler Analizi yoluyla elde edilen düzeyler olmak üzere dört farklı Q- Matrisi belirlenmiştir. Hangi sınıflamanın daha iyi yapıldığına ilişkin değerlendirme kriteri olarak bu dört farklı Q-Matrisiyle elde edilen sınıflamaların doğruluğu ve tutarlılığı incelenmiştir. Araştırmanın çalışma grubunu TIMSS 2015 Türkiye örnekleminde içinde birinci kitapçığı alan 463 dördüncü sınıf öğrencisi oluşturmaktadır. Araştırmada kullanılan veriler TIMSS 2015 4. Sınıf 1 numaralı matematik kitapçığındaki 25 sorudan seçilen 23 maddeden elde edilmiştir. Araştırmanın bulgularına göre nitelik boyutu olarak İçerik Alanı düzeylerinin kullanıldığı Q-Matrisi öğrencileri en iyi sınıflayan matris olmuştur. Son olarak bu matris kullanılarak, araştırma kapsamında en iyi sınıflama yapan DINA Modele göre sınıflandırma yapılmış ve sonuçlar tartışılmıştır.

*Anahtar kelimeler:* TIMSS 2015, Matematik başarıları, İçerik Alanı, Bilişsel Alan, Temel Bileşenler Analizi (TBA), Bilişsel Tanı Modelleri (BTM)

\*Bu çalışma 6. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde (VI. C-MEEP) sunulan sözlü bildirden yararlanılarak yapılmıştır.

## Giriş

Temel eğitimin en önemli yapıtaşlarından birisi olan matematik eğitimi sayıları, işlemleri öğretmekten, hesaplama becerilerini kazandırmaktan öte bir işlev üslenmekte ve olaylar arasında bağ kurma, akıl yürütme, problem çözme gibi önemli destekler sağlamaktadır (Umay, 2003). Buradan anlaşıldığı üzere matematik eğitiminin, diğer bir çok disiplinde olduğu gibi, en az iki boyutu bulunmaktadır. Dolayısıyla matematik başarısının ölçülmesinde bu boyutların dikkate alınması kaçınılmazdır. Matematik dersine ilişkin amaçlara ne kadar ulaşıldığının ve aynı zamanda öğrencilerin matematiksel bilgi ve beceriler bakımından gelişim gösterip göstermediklerinin belirlenmesi öğrenci başarısının doğru yöntemlerle ölçülmesine ve değerlendirilmesine de bağlıdır (Yıldız ve Uyanık, 2004).

Matematiksel bilgi ve becerilerin doğru biçimde ölçülmesi, öğrencilerin matematik öğrenmesini destekler, ulusal düzeyde matematik eğitimi alanındaki gelişmeler hakkında bilgi sağlar ve bunlarla birlikte matematik programının değerlendirilmesine imkan sağlar (IEA-International Association for the Evaluation of Educational Assessment, 2011). Bir çok disiplin alanında olduğu gibi matematik eğitimi alanında da ulusal düzeyde yapılan ölçme ve değerlendirme uygulamaları, o ülkedeki eğitim hakkında önemli bir bilgi kaynağıdır. Ancak eğitim alanındaki gelişmeleri daha iyi değerlendirmede uluslararası düzeyde yapılan uygulamalara da ihtiyaç duyulmaktadır. Ülkemizde var olan eğitim sisteminin mevcut durumunu tespit etmek, öğrencilerin bilgi toplumunun ihtiyaçlarına uygun yetişip yetişmediklerini anlamak, var olan eğitim sistemini geliştirmek ve diğer ülkelerin eğitim sistemleriyle karşılaştırma yapmak için çok sayıda ülkenin dahil olduğu uluslararası değerlendirme uygulamalarına katılmıştır. Bunlar bilindiği üzere; TIMSS (Trends in International Mathematics and Science Study - Uluslararası Matematik ve Fen Eğilimleri Çalışması), PISA (Programme for International Student Assessment - Uluslararası Öğrenci Değerlendirme Programı), PIRLS (Progress in International Reading Literacy Study - Uluslararası Okuma Becerilerinde Gelişim Projesi) uygulamalarıdır.

Matematik ve fen alanında en uzun süredir ve en yaygın olarak kullanılan karşılaştırmalı eğitim çalışması olan TIMSS'in genel amacı; öğrencilerin çok yönlü bilgi ve becerilerini ölçmektir (Mullis, Martin, Ruddock, O'Sullivan ve Preuschoff, 2009). Bu amaç doğrultusunda da öğrencilerin fen ve matematik alanlarındaki performansları, ülkelerin eğitim sistemleri, öğretim programları, öğrenci, öğretmen ve okulların özellikleri hakkında bilgi toplanmaktadır.

TIMSS çalışmasında 4.sınıf düzeyinde uygulamaya katılan öğrencilerin matematik başarı testi puanlarına dayalı olarak uluslararası matematik yeterlik düzeyleri IEA tarafından ileri, üst

düzey, orta düzey, alt düzey olarak tanımlanmıştır. Bu yeterlik düzeylerine ilişkin genel bilgiler Tablo 1’de yer almaktadır (Büyüköztürk, Çakan, Tan ve Atar, 2014).

**Tablo1.** TIMSS 4. Sınıf Uluslararası Matematik Yeterlik Düzeylerinin Tanımı

İleri Düzey (625):	Öğrenciler, anladıklarını ve bilgilerini çeşitli kompleks durumlara uygulayabilir ve nedenlerini açıklayabilirler. Öğrenciler, orantılar dahil tam sayıları içeren çok aşamalı sözlü ifade edilen çeşitli soruları çözebilirler. Bu seviyedeki öğrenciler yüzdeler ve ondalıklar hakkında daha yüksek bir kavrama sergilerler. Bir dizi iki veya üç boyutlu şeklin değişik durumlarına ilişkin geometrik bilgileri uygulayabilirler. Tablo verilerinden bir sonuç çıkarabilir ve bunu gerekçelendirebilirler.
Üst Düzey (550-625):	Öğrenciler bilgilerini ve anladıklarını problem çözmek için uygulayabilirler. Tam sayılarla işlem yapmayı içeren soruları çözebilirler. Çeşitli problem durumlarında bölme işlemini kullanabilirler. Bilinmeyenlere değer vererek problemleri çözerler. Daha sonra tanımlanacak bir terimi bulmak için örnekleri geliştirirler. Öğrenciler çizgi simetrisi ve geometrik şekillerden anladıklarını gösterebilirler. Öğrenciler tablo ve grafiklerdeki verileri yorumlar ve kullanır. Şekil ve çetele grafiği hakkındaki bilgilerini bar grafiklerini tamamlamada kullanırlar.
Orta Düzey (475-550):	Öğrenciler matematiğe ilişkin temel bilgileri basit durumlara uygulayabilirler. Öğrenciler, kesirler ve tamsayılar ile ilgili anladıklarını gösterebilirler. İki boyutlu gösterimden üç boyutlu gösterimi canlandırabilirler. Bar grafiği, şekil ve tabloları basit problemleri çözmek için yorumlayabilirler..
Alt Düzey (400-475):	Öğrenciler matematiğe ilişkin başlangıç düzeyindeki bilgileri bilir. Öğrenciler tamsayılarda toplama ve çıkarma işlemlerini yapabilirler. Yatay ve dikey çizgiler, basit geometrik şekiller, koordinat bilgisi farkındalığına sahiptir. Basit bar grafiği ve tabloyu okuyabilir ve tamamlayabilirler.

Tablo 1 incelendiğinde, belirlenen yeterlik düzeylerine ilişkin olarak her bir düzey artışında her bir alandan bilgi ve becerinin artırılarak düzeylerin belirlendiği görülmektedir. Örneğin en alt düzeyde geometri alanına ilişkin yazılan yeterlik “Yatay ve dikey çizgiler, basit geometrik şekiller, koordinat bilgisi farkındalığına sahiptir” iken orta düzeyde “İki boyutlu gösterimden üç boyutlu gösterimi canlandırabilirler” yeterliği eklenmiştir. Üst düzeyde “çizgi simetrisi geometrik şekillerden anladıklarını gösterebilirler” iken ileri düzeyde “Bir dizi iki veya üç boyutlu şeklin değişik durumlarına ilişkin geometrik bilgileri uygulayabilirler” yeterliği eklenmiştir. Yani bu düzeylerin belirlenmesinde her öğrencinin her bir alanda benzer seviyelerde yeterlik düzeylerine sahip olduğunu varsaymaktadır. Ancak bir alanda yetkin olup başka bir alanda daha az yetkin olan öğrenci profillerine bu düzeyler arasında yer verilmemiştir. Diğer bir deyişle cebir alanında yetkin olan ama geometri ve şekiller alanında daha az yetkin olan bir öğrenci ile geometri ve şekiller alanında yetkin olan ama cebir alanında daha az yetkin olan ve 500 alan iki öğrenci aynı düzeyde kabul edilmektedir.

Eğitim sisteminin bir ögesi olan ölçme ve değerlendirme, öğrencilerin yetersizlikleri ve bunların kaynaklarını ortaya çıkarıcı olmalıdır (Baykul, 1992). Başarıyı artırmak için öncelikli



olarak öğrencilerin bireysel olarak eksikliklerinin tespit edilmesi ve buna yönelik olarak önlemlerin alınması gerekir. Yine büyük bir etki alanı yaratmış olan Bloom'un (1968) Tam Öğrenme Modeline (Learning for Mastery) göre de; eğitimdeki temel görev, bireyin tam öğrenmesini (gelişimini) teşvik edecek şekilde bireysel farklılıkları dikkate alacak stratejiler bulmaktır. Bireysel farklılıkları dikkate alan değerlendirme türü biçimlendirmeye, yetiştirmeye dönük değerlendirmedir.

Biçimlendirme-yetiştirmeye yönelik değerlendirme maksadıyla kullanılan formatif testlerden elde edilen toplam puan bu tür değerlendirme için hemen hemen hiçbir anlam taşımaz. Her bir öğrenciye onun maddelere verdiği cevapların örüntülerini ve yaptığı hataları gösteren bilgi verilmelidir. Çünkü bir anlamda bu değerlendirme öğretmen ve öğrenciye öğretimin etkililiği hakkında bilgi sağlamak için yapılır. Böylece öğretimi öğrencilerin öğrenme hızı ve gücüne göre ayarlama olanağı elde edilir (Tekin, 2007, s. 26).

Klasik Test Kuramı (KTK), Madde Tepki Kuramı (MTK) gibi psikometrik yaklaşımlar, genellikle bireyleri bazı örtük özellik boyunca ölçeklendirmek veya sıralamak üzerine yoğunlaşmaktadır (Rupp, Templin ve Henson 2010). Ancak birden fazla becerinin veya niteliğin madde bazında varlığına veya yokluğuna dayanarak bireyler hakkında çok değişkenli sınıflama yapamazlar. Modern ölçme yöntemleri buna izin vererek belirli öğrenme hedeflerinin güçlü ve zayıf yönleri hakkında geri bildirim sağlayabilecek bilişsel beceri tanılarına odaklanır (Rupp ve Templin, 2008). Bu durum öğrenci becerileri setleri hakkında daha bilgilendirici geribildirim sağlar ve daha etkin müdahale stratejilerinin tasarlanmasına olanak tanır (Rupp ve diğerleri, 2010).

### *Bilişsel Tanı Modelleri*

Modern ölçme kuramlarından Bilişsel Tanı Modelleri (BTM) altında yapılan değerlendirmeler ile, öğrencilerin bir öğrenme alanı içindeki nitelik diye adlandırılan bilgi/becerileri hakkında bireysel olarak ayrıntılı bilgiler elde edilerek eğitim sisteminin dönüt ögesinin etkili olması sağlanabilir. Örneğin bir öğrencinin  $(3+2 \times 6) \div 3$  işlemine doğru cevap vermesi için toplama, çarpma, bölme ve işlem önceliğini bilmesi gereklidir. Bu madde ile toplama, çarpma, bölme ve işlem önceliği olarak dört özelliğin ölçülmesi hedeflenmektedir. Bu soruya "10" cevabını veren bir öğrencinin toplama, çarpma, bölme işlemlerini yapabildiği ancak işlem önceliği konusunda eksikliğini olduğu çıkarımını yapmak, öğrencinin ilgili nitelikler bağlamında profilinin çıkarılmasında tanılayıcı bilgiler sağlayabilir. BTM'ler belirli bir içerik alanında veya bilişsel alanda öğrencilerin güçlü ve zayıf yanlarını tanımlamayı

amaçladığından sınıf içi uygulamalarla öğrenci için en iyi alıştırmaya stratejisine belirlemeyi içerebilir (Rupp ve diğerleri, 2010). Bu örnekteki hatayı yapan öğrencinin örnekteki dört işlem becerileri konusunda yeterli olduğu ancak işlem önceliği konusunda geliştirilmesi gerektiği ve buna göre gerekli önlemlerin alınması gerektiği görülmektedir. Bilişsel Tanı Modelleri ile bir testin ilgisi, öğrencilerin genel bir yetenek kestirimini sağlamak yerine öğrenci profillerini kestirmektir. BTM'nin raporlanmasında bireysel katılımcılar için tasarlanmış ayrı ayrı nitelik profilleri verilmektedir.

BTM doğrudan bir katılımcının verilen bir tanılamaya için kriterleri sağlama olasılığının kestirimidir (Leighton ve Gierl, 2007). Yani BTM istatistiksel olarak bir ya da daha çok tanılamaya kriterine göre katılımcıların sınıflandırmasını sağlar. Yukarıda verilen örnek baz alınırsa nitelik sayısının üç olduğunu düşünelim (Toplama ve Çıkarma, Çarpma ve Bölme, İşlem Önceliği). Burada belirlenen bu 3 özellik için toplam 8 (k özellik sayısı ise oluşabilecek örtük sınıf sayısı:  $2^k$ ) sınıf oluşacaktır. Bunlar (000)(001)(010)(100)(110)(011)(101)(111) dir. Yukarıdaki örnekteki öğrencinin (110) sınıfında olduğu düşünülebilir. BTM için sınıflandırmalar tanılayıcı değerlendirme ile toplanan gözlenen tepkilere dayalıdır.

BTM'lerde sınıflama gizil yordayıcı değişken ve gözlenen tepki değişkenlerinin iki kategorili ya da çok kategorili olmalarına göre yapılabilir. Bu sınıflamanın yanında BTM'lerde telafi edici (compensatory) ve telafi edici olmayan (noncompensatory) model olarak iki tür model bulunmaktadır. Telafi edici ve telafi edici olmayan modeller arasındaki temel fark, örtük yordayıcı değişkenlerin, gözlenen cevabı vermede farklı nitelikler arasında nasıl bir araya getirildiğinin belirlenmesiyle ilgilidir (Rupp ve Templin, 2008). Telafi edici modellerde bir nitelikteki eksiklik diğer kalan nitelikler tarafından giderilebilirken, telafi edici olmayan modelde, her bir nitelik, doğru cevabın verilmesi için gereklidir. Yani telafi edici modellerde eksik olan nitelik, sahip olunan diğer bir nitelik tarafından tamamlanabilirken, telafi edici olmayan modellerde, eksik olan nitelik başka herhangi bir nitelik tarafından tamamlanamamaktadır. Telafi edici olmayan ve telafi edici modeller arasında alan yazında sıklıkla kullanılan en temel modeller sırasıyla DINA (Deterministic inputs, noisy 'and' gate) (Junker ve Sijtsma, 2001) ve DINO (Deterministic inputs, noisy 'or' gate) (Templin ve Henson, 2006) modelleridir.

BTM'nin tüm modellerinin ortak özelliği Q-matrisinin belirlenmesi gerekliliğidir. Q-matrisi alan yazında "madde özellik matrisi", "özellik matrisi" olarak da adlandırılmaktadır (Ma, 2014; Tatsuoka, 1984). Q-matrisinde madde ile maddeleri doğru cevaplamada gerekli olan nitelik ya da nitelikler ilişkilendirilir. Bu matris maddeler satırda, nitelikler ise ayrı sütunda

olacak şekilde iki boyutludur. İlgili hücreye bir maddenin çözümünde gerekli olan nitelikler için “1”; gerekli olmayan nitelikler için ise “0” yazılır. Tablo 2’de örnek bir Q-matrisi yer almaktadır.

**Tablo 2.** Örnek Bir Q-Matrisi

	Nitelik 1	Nitelik 2	Nitelik 3
Madde 1	0	1	0
Madde 2	1	0	1
Madde 3	0	1	1
Madde 4	1	1	1

Tablo 2’de bulunan dört madde ve üç nitelik için hazırlanmış Q-matrisinde, Madde 1 yalnızca Nitelik 2 ile ilişkilendirilmişken; Madde 4 her üç nitelik ile ilişkilendirilmiştir.

0-1 elemanlarından oluşan Q-matrisi basit bir yapıda görünse de BTM için önemli bir yere sahiptir; öyle ki Q-matrisinin belirlenmesinde meydana gelen bir hata, analizi tümüyle olumsuz etkileyebilir. Q-matrisinin geliştirilmesinde farklı yaklaşımlar kullanılabilir. Bunlardan en sık kullanılanı uzman yargılarına göre Q-matrisini geliştirilmesidir. Bu sebepten ötürü, Q-matrisinin yapılandırılması BTM’lerde en fazla çabanın gösterildiği süreç olarak görülebilir. Süreç, istenilen sonuca erişilinceye kadar madde ve niteliklerin birbiriyle etkileşimli bir şekilde döngüsel olarak devam etmesini gerektirir.

Q-matrisinin belirlenmesinde, tipik olarak, ilgili disiplin alanındaki uzman kanılarından faydalanılır. Q-matrisi BTM’nin en önemli kavramların biridir ve Q-matrisinin doğru ve tam belirlenmesi tüm sınıflamayı etkileyebilmektedir (Tatsuoka,1984). Başokçu (2015) Q-matrisi tarafından tanımlanmayan farklı özelliklerin de bir maddenin doğru cevaplanması için yeterli olabileceği durumların olduğunu belirtmiştir. Bu durum yapılacak sınıflamanın doğruluğunu ve tutarlığını negatif yönde etkileyecektir.

Bu çalışmada BTM altında Q-matrisindeki niteliklerin maddeler ile ilişkilendirilmesinde önsel olarak uzman yargılarına göre farklı alanlardaki düzeylerin belirlenmesi dışında ek olarak, test maddelerine ait açıklayıcı bilgilerin kullanılması (faktör analitik yaklaşımlar) ile Q-matrisinin daha iyi tanımlanabileceği ön görüşüyle yola çıkılmıştır. Bu şekilde yapılan sınıflama tutarlığı ve güvenilirliği yalnızca uzman görüşü dikkate alınarak oluşturulan Q matrisi sonuçlarıyla karşılaştırılmak istenmiştir. Uzman yargılarından (önsel madde bilgilerinden) faydalanarak bilişsel alan ile içerik alanının ve her iki alanın niteliklerinin bulunduğu Q-matrisleri kullanılmıştır. Verilere dayanarak ise ikili puanlamaya uygun olacak açıklayıcı

faktör analitik yaklaşımlardan Temel Bileşenler Analizi kullanılmıştır. Genel olarak değişkenler arasındaki bağımlılık yapısının yok edilmesi ve/veya boyut indirgeme amacıyla kullanılan Temel Bileşenler Analizi başlı başına bir analiz olduğu gibi, başka analizler için veri hazırlama tekniği olarak da kullanılmaktadır (Tatlıdil, 1996).

Bu çalışmanın amacı Bilişsel Tanı Modellerinden DINA ve DINO Modelleri altında Q-matrisinin nitelik boyutunun içerik alanına ait düzeyleri, bilişsel alana ait düzeyleri, her iki alana ait düzeyleri ve Temel Bileşenler Analizi yoluyla elde edilen düzeylerin belirlenmesiyle yapılan sınıflamanın doğruluğu ve tutarlılığını karşılaştırmalı olarak incelemektir. Bu bulguların ışığı altında göreceli olarak en iyi sınıflama yapan model ve Q-matrisine ilişkin öğrenci sınıflamasının incelenmesi araştırmanın diğer bir amacını oluşturmaktadır. Bu çalışmanın BTM Değerlendirmelerinde kilit bir yere sahip olan Q-matrisinin belirlenmesinde teoriye katkı sağlayacağı, buna ilişkin yöntem önermede önemli olabileceği düşünülmektedir.

Araştırmanın amacı doğrultusunda aşağıdaki araştırma sorularına yanıt aranmıştır:

3. TIMSS 2015 Türkiye örnekleminde matematik başarı testine dayalı olarak Bilişsel Tanı Modellerinden DINA ve DINO Modelleri kullanılarak ve belirtilen 4 farklı oluşturulan Q-matrislerinden hangisi kullanılarak sınıflama tutarlığı ve doğruluğu açısından daha iyi sınıflama yapılıır?
4. Öğrenciler, matematik başarı testine dayalı olarak en iyi sınıflamayı sağlayan model ve Q-matrisinde nasıl sınıflandırılmıştır?

## Yöntem

### *Araştırmanın Modeli*

Araştırmanın amacı TIMSS 2015 Türkiye örnekleminde matematik testine alan öğrencilerin farklı niteliklere göre sınıflandırılmasını DINA ve DINO Modelleri altında karşılaştırmak ve yapılan sınıflamalardan göreceli olarak en iyi bulunan sınıflamaya göre öğrencilerin nasıl sınıflandırıldığını incelemektir. Bu amaç doğrultusunda dört farklı belirlenmiş Q-matrisinin sınıflama doğruluğu ve tutarlılığının incelenmesi amaçlandığından araştırma betimsel düzeyde bir çalışmadır.

### *Çalışma Grubu*

Araştırmanın çalışma grubunu TIMSS 2015 Türkiye örnekleminde birinci kitapçığı alan 232'si kız (%50,1) ve 231'i erkek olan (%49,9) toplam 463 dördüncü sınıf öğrencisi oluşturmaktadır. Analize geçilmeden önce veriler incelenmiş, kayıp veri analizi yapılmıştır. Kayıp veri analizi sonucunda kayıp verilerin belli bir örüntüye sahip olmadığı görülmüş ve “Yakın Noktalar Medyan Ataması (Median of Nearby Points)” yapılmıştır.

#### *Veri Toplama Aracı*

Araştırma verileri, TIMSS 2015 4. Sınıf 1 numaralı matematik kitapçığındaki 25 maddeden ikili olarak puanlanmış dokuzu kısa cevaplı on dördü çoktan seçmeli olan 23 maddeden elde edilmiştir.

TIMSS çalışmasının ölçme kapsamı Matematik başarı testi 4. düzeyinde bilişsel alan ve içerik boyutu olmak üzere iki boyutta düzenlenmiştir. İçerik boyutu ölçülecek kapsamı yani matematiksel konuları, bilişsel alan boyutu ise düşünme süreçlerini dikkate almaktadır. 4. sınıf düzeyinde içerik boyutunu sayılar, geometrik şekil ve ölçümler ve veri gösterimi oluşturmaktadır. Bilişsel alan boyutunu ise bilme, uygulama ve akıl yürütme oluşturmaktadır.

#### *Verilerin Düzenlenmesi ve Analizi*

Çalışmada kullanılan Q-matrislerinin düzenlenmesinde önsel olarak belirlenen üç Q-matrisinin madde nitelik ilişkilendirmeleri için IEA'nin (2015) erişime açık olan uluslararası veri tabanındaki madde bilgilerinden yararlanılmıştır. (<https://timssandpirls.bc.edu/timss2015/international-database/>). İçerik alanına göre “Sayılar, Veri Gösterimi, Geometrik Şekil ve Ölçümler”, bilişsel alana göre “Bilme, Uygulama, Akıl Yürütme” ve her iki alandaki niteliklerin birlikte kullanıldığı Q-matrisleri sırasıyla Tablo3, Tablo 4 ve Tablo 5'te yer almıştır.

**Tablo 3.** İçerik Alanına Göre Belirlenmiş Q-Matrisi

Maddeler	$\alpha_S$	$\alpha_G$	$\alpha_V$	Maddeler	$\alpha_S$	$\alpha_G$	$\alpha_V$	Maddeler	$\alpha_S$	$\alpha_G$	$\alpha_V$
M041004	1	0	0	M041254	0	1	0	M061050	1	0	0
M041023	1	0	0	M041153	0	1	0	M061167	1	0	0
M041034	1	0	0	M041132	0	1	0	M061206	0	1	0
M041087	1	0	0	M041174	0	0	1	M061265A	0	1	0
M041124	1	0	0	M041191	0	0	1	M061265B	0	1	0
M041302A	0	1	0	M061272	1	0	0	M061185	0	1	0
M041302B	0	1	0	M061029	1	0	0	M061239	0	0	1
M041302C	0	1	0	M061031	1	0	0				

$\alpha_S$ : Sayılar,  $\alpha_V$ : Veri Gösterimi,  $\alpha_G$ : Geometrik Şekil ve Ölçümler

**Tablo 4.** Bilişsel Alana Göre Belirlenmiş Q-Matrisi

Maddeler	$\alpha_B$	$\alpha_U$	$\alpha_A$	Maddeler	$\alpha_B$	$\alpha_U$	$\alpha_A$	Maddeler	$\alpha_B$	$\alpha_U$	$\alpha_A$
M041004	1	0	0	M041254	1	0	0	M061050	0	1	0
M041023	1	0	0	M041153	0	1	0	M061167	1	0	0
M041034	1	0	0	M041132	1	0	0	M061206	0	1	0
M041087	1	0	0	M041174	0	1	0	M061265A	0	1	0
M041124	0	1	0	M041191	1	0	0	M061265B	0	1	0
M041302A	1	0	0	M061272	0	0	1	M061185	0	1	0
M041302B	0	0	1	M061029	0	0	1	M061239	0	1	0
M041302C	0	0	1	M061031	0	1	0				

$\alpha_B$ : Bilme,  $\alpha_U$ : Uygulama,  $\alpha_A$ : Akıl Yürütme

**Tablo 5.** İçerik ve Bilişsel Alana Göre Belirlenmiş Q-Matrisi

Maddeler	$\alpha_S$	$\alpha_G$	$\alpha_V$	$\alpha_B$	$\alpha_U$	$\alpha_A$	Maddeler	$\alpha_S$	$\alpha_G$	$\alpha_V$	$\alpha_B$	$\alpha_U$	$\alpha_A$
M041004	1	0	0	1	0	0	M041191	0	0	1	1	0	0
M041023	1	0	0	1	0	0	M061272	1	0	0	0	0	1
M041034	1	0	0	1	0	0	M061029	1	0	0	0	0	1
M041087	1	0	0	1	0	0	M061031	1	0	0	0	1	0
M041124	1	0	0	0	1	0	M061050	1	0	0	0	1	0
M041302A	0	1	0	1	0	0	M061167	1	0	0	1	0	0
M041302B	0	1	0	0	0	1	M061206	0	1	0	0	1	0
M041302C	0	1	0	0	0	1	M061265A	0	1	0	0	1	0
M041254	0	1	0	1	0	0	M061265B	0	1	0	0	1	0
M041153	0	1	0	0	1	0	M061185	0	1	0	0	1	0
M041132	0	1	0	1	0	0	M061239	0	0	1	0	1	0
M041174	0	0	1	0	1	0							

$\alpha_S$ : Sayılar,  $\alpha_G$ : Geometrik Şekil ve Ölçümler,  $\alpha_V$ : Veri Gösterimi,  
 $\alpha_B$ : Bilme,  $\alpha_U$ : Uygulama,  $\alpha_A$ : Akıl Yürütme

Tablo 3 ve Tablo 4’te görüldüğü gibi her bir madde yalnızca bir nitelik ile ilişkilendirilmiştir. İçerik alanı ve bilişsel alan ile oluşturulan Q-matriste ise Tablo 5’te görüldüğü gibi her bir madde iki nitelik ile ilişkilendirilmiştir.

Temel Bileşenler Analizi ile Q-matrisi oluşturmak için tetrakorik korelasyon matrisine dayalı analize izin veren FACTOR 10.08.3 programından yararlanılmıştır. Çalışma kapsamındaki verilerin temel bileşenler analizine uygunluğu Kaiser-Mayer-Olkin (KMO) katsayısı ve Barlett Küresellik Testi ile belirlenmiştir. TBA’da verilere ait KMO katsayısı 0,91 bulunup 0,60 ‘tan büyük olduğundan ve Barlett Küresellik testi manidar olduğundan ( $p < 0,01$ ) kullanılan testin TBA’ya uygunluğuna karar verilmiştir.

Analiz sonucundaki dört bileşen için Q-matriste yükleri 0.30 ve 0.30’dan daha büyük olanlar “1” ; daha küçük olanlar “0” olarak belirlenmiştir. Tablo 6’da TBA’ya göre belirlenmiş Q-matrisi yer almaktadır.

**Tablo 6.** Temel Bileşenler Analizine Göre Belirlenmiş Q-Matrisi

Maddeler	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	Maddeler	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	Maddeler	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$
M041004	0	0	0	1	M041254	1	0	0	0	M061050	0	1	0	0
M041023	0	0	0	1	M041153	1	0	1	0	M061167	0	0	0	1
M041034	0	0	0	1	M041132	0	1	1	0	M061206	1	0	0	0
M041087	0	1	0	0	M041174	0	0	1	1	M061265A	0	0	0	1
M041124	0	0	0	1	M041191	0	0	1	1	M061265B	0	1	0	0
M041302A	1	0	1	0	M061272	0	0	0	1	M061185	0	0	1	1
M041302B	0	0	1	0	M061029	1	0	0	1	M061239	0	0	0	1
M041302C	1	0	1	0	M061031	0	1	0	1	M041302C	1	0	1	0

$\alpha_1$ : Birinci Bileşen,  $\alpha_2$ : İkinci Bileşen,  $\alpha_3$ : Üçüncü Bileşen,  $\alpha_4$ : Dördüncü Bileşen

Tablo 6 incelendiğinde maddelerin %60 a yakını tek nitelik ile ilişkilendirilirken, kalanı iki nitelik ile ilişkilendirilmiştir. Elde edilen farklı Q-matrisleri ile öğrencilerin cevap örüntülerinin DINA ve DINO modellerine göre analizi R 3.0 programında (R Core Team, 2015) CDM paketi (George, Robitzsch, Kiefer, Groß, ve Ünlü, 2016) ile gerçekleştirilmiştir.

### Bulgular ve Yorumlar

Bu bölümde araştırma sorularına ilişkin bulgular sırasıyla sunulmuştur.

3. “TIMSS 2015 Türkiye örnekleminde matematik başarı testine dayalı olarak Bilişsel Tanı Modellerinden DINA ve DINO Modelleri kullanılarak ve belirtilen 4 farklı oluşturulan Q-matrislerinden hangisi kullanılarak sınıflama tutarlığı ve doğruluğu açısından daha iyi sınıflama yapılır?”

Q-matrisinin maddelerin önceden bilinen özelliklerine göre (bilişsel alan, içerik alanı ve her ikisinin kullanılması) ve maddelerin uygulamadan sonra elde edilen bilgilerine göre (TBA'dan faydalanılarak niteliklerin belirlenmesi) ele alınmasıyla yapılan analiz sonucu elde edilen sınıflama doğruluğu ve tutarlığı Tablo 7'de sunulmuştur.

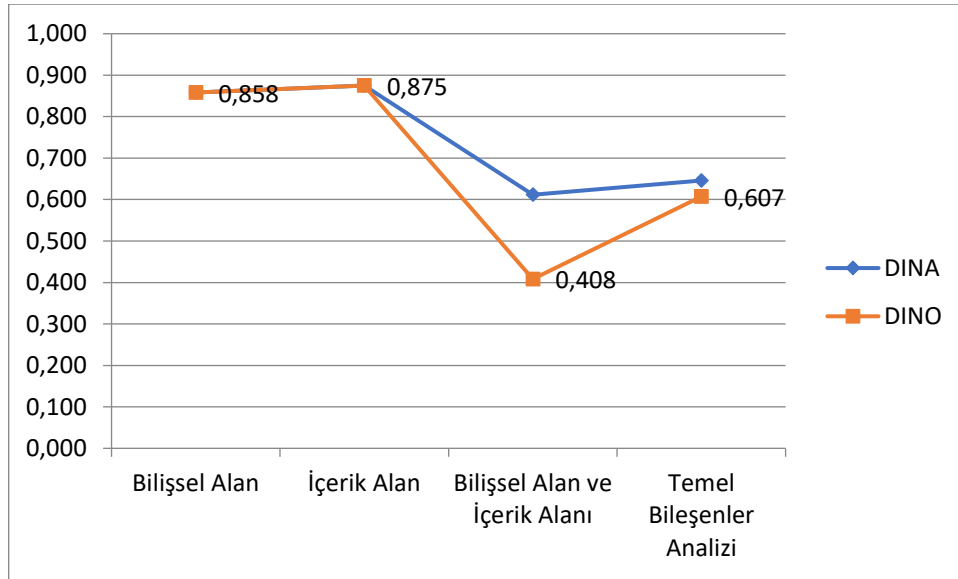
**Tablo 7.** Q-Matrisinin Dört Farklı Koşula Dayalı Belirlenmesi Sonucu Elde Edilen Sınıflama Doğruluğu Ve Tutarlığı Bilgileri

Q-matrisinin Nitelik Boyutu	Sınıflama Doğruluğu		Sınıflama Tutarlılığı	
	DINA	DINO	DINA	DINO
Bilişsel Alan	0,916	0,916	0,858	0,858
İçerik Alanı	0,929	0,929	0,875	0,875
Bilişsel ve İçerik Alanı	0,582	0,443	0,612	0,408
TBA'nın Ortaya Çıkardığı Alan	0,699	0,712	0,646	0,607

Tablo 7 incelendiğinde en yüksek sınıflama doğruluğu ve tutarlığının içerik alanı ve bilişsel alana dayalı oluşturulan matrislerden elde edildiği görülmektedir. Diğer dikkat çekici bir sonuç ise, her iki durumda da DINA ve DINO modellerinin aynı sınıflama doğruluğu ve tutarlığı üretmesidir. Bunun sebebi olarak belirlenen her iki Q-matriste de her bir satırda yalnızca tek bir sütunun 1 olarak belirlenmesi gösterilebilir. Yani, her bir madde yalnızca tek bir alan ya da beceriyi ölçmektedir. Bu bulgu DINA ve DINO Modellerinin matematiksel olarak eşit modeller olduğunu destekleyen bir bulgudur (Frey, 2018). Göreceli olarak değerlendirildiğinde ve Q-matrisi içerik alanına göre belirlendiğinde, DINA ve DINO Modelleri altında diğer tüm Q-matris belirleme yöntemlerine göre daha yüksek oranda sınıflama doğruluğu ve tutarlığına sahip sınıflama yapmıştır. Bunu sırasıyla bilişsel alana göre Q-matrisi belirleme, TBA'dan faydalanarak Q-matrisi belirleme ve bilişsel ve içerik alanının birlikte nitelik kümesine alınmasıyla elde edilen Q-matrisi belirleme ile yapılan sınıflamalar izlemektedir. Belirlenen Q-matristeki nitelik sayıları dikkate alındığında daha fazla niteliğe sahip Q-matrisleri ile daha az tutarlı ve doğru sınıflama yapıldığı söylenebilir.

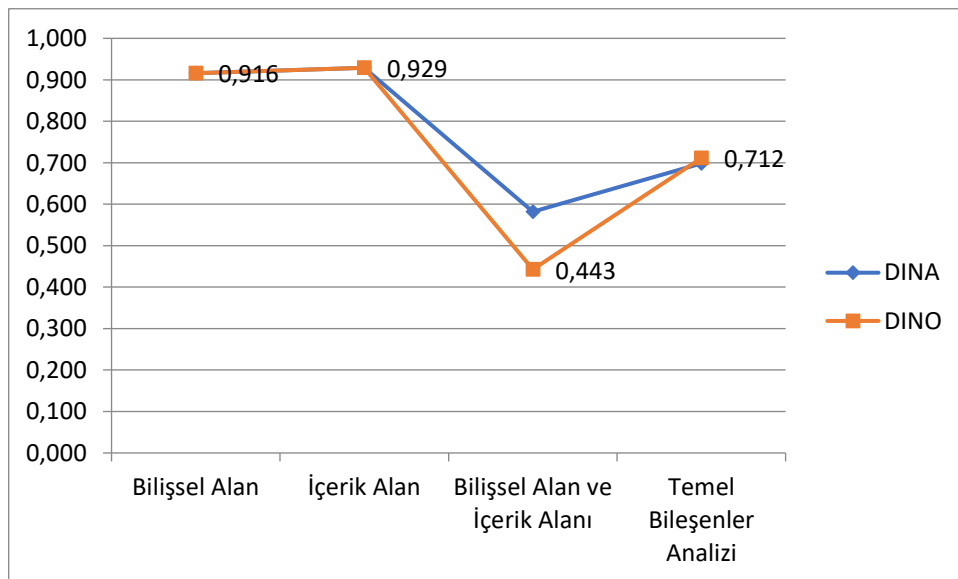
Çalışma kapsamında DINA ve DINO Modelleri dikkate alınarak dört farklı koşulda belirlenmiş Q-matrisleri ile yapılan sınıflamalarda elde edilen sınıflama tutarlıkları Şekil 1 ve sınıflama doğrulukları Şekil 2'de karşılaştırmalı olarak sunulmuştur.





**Şekil 1.** Farklı Belirlenmiş Q-matrisleri ile DINA ve DINO Modelleri altında yapılan Sınıflama Tutarlığı

Şekil 1 incelendiğinde DINA Modelinin DINO Modele göre daha tutarlı sınıflama yaptığı görülmektedir. Bilişsel ve içerik alanının birlikte kullanıldığı Q-matriste bir maddenin çözümünde gerekli olan hem bilişsel ve hem de içerik alanına ait özelliklerin birlikte kullanıldığı göz önüne alınırsa, DINO Modelin telafi edici (bir maddeyi doğru yanıtlayabilmek için madde ile ilişkili en az bir özelliğe sahip olunması) yapısından kaynaklı daha az tutarlı sonuçlar verdiği bulgusuna varılabilir. Ancak Temel Bileşenler Analizi'nden yararlanılarak oluşturulan Q-matrisi ile yapılan analiz sonucunda ise sınıflama tutarlılıkları her iki modelde de yakın değerler vermiştir.



**Şekil 2.** Farklı Belirlenmiş Q-matrisleri ile DINA ve DINO Modelleri altında yapılan Sınıflama Doğruluğu

Şekil 2 incelendiğinde DINA Modelinin DINO Modele göre daha doğru sınıflama yaptığı görülmektedir. Sınıflama tutarlığında olduğu gibi bilişsel ve içerik alanının birlikte kullanıldığı Q-matrisinde bir maddenin çözümünde gerekli olan hem bilişsel ve hem de içerik alanına ait özelliklerin kullanılması gerektiği göz önüne alınırsa DINO Modelin telafi edici yapısından kaynaklı daha az sınıflama doğruluğuna sahip sonuçlar verdiği bulgusuna varılabilir. Ancak Temel Bileşenler Analizi'nden yararlanılarak oluşturulan Q-matrisi ile yapılan analiz sonucunda sınıflama doğrulukları her iki modelde de yine yakın değerler vermiştir.

#### 4. Öğrenciler, matematik başarı testine dayalı olarak en iyi sınıflamayı sağlayan model ve Q-matrisinde nasıl sınıflandırılmıştır?

İlk araştırma sorusunda vurgulandığı üzere, Q-matrisinin içerik alanına dayalı yapılandırılması sonucunda diğer durumlara kıyasla daha fazla doğrulukta ve tutarlıkta sınıflama yaptığı görülmüştür. DINA Model altında yapılan bu sınıflamaya ilişkin mutlak uyum indekslerinden M2, RMSEA ve SRMSR incelenmiş ve model veri uyumunun sağlandığı görülmüştür (M2=335,50, df= 223, p-value= 0; RMSEA=0,033 %90 güven aralığı ile: [0,0255-0,0401]; SRMSR=0,0617). Tablo 8'de, içerik alanına göre yapılandırılmış Q-matrisinin DINA Modeli altında elde edilen madde istatistiklerinin özeti verilmektedir.

**Tablo 8.** İçerik Alanına ait Q-matrisi ile DINA modeli için madde özelliklerinin özeti

Madde Özellikleri	Minimum	Maksimum	Ortalama	Standart Sapma
p değeri (madde güçlüğü)	0,14	0,86	0,48	0,18
g parametresi	0,05	0,70	0,26	0,18
s parametresi	0,03	0,79	0,34	0,22
Madde ayırıcılık parametresi =1-g <sub>j</sub> -s <sub>j</sub>	0,08	0,68	0,40	0,16
Madde kolaylık parametresi( g <sub>j</sub> +(1-s <sub>j</sub> ))/2	0,14	0,83	0,46	0,19

Madde güçlüğü veren p değeri, her bir maddeyi doğru çözen öğrencilerin yüzdesini tanımlamaktadır. Öğrencilerin %14'ü en zor maddeyi doğru cevaplarırken , öğrencilerin %86'sı en kolay maddeyi doğru cevaplamıştır. Ortalama olarak, öğrencilerin %48'i maddeleri doğru cevaplamıştır. Buradan anlaşılacağı üzere test maddeleri bu grup için orta güçlüğü biraz altında güçlüğe sahiptir. Bilindiği üzere madde güçlüğü ile madde kolaylık parametresi arasında bir ilişki vardır ve Tablo 6'da bunu desteklemektedir (George ve Robitzsch, 2015). g

parametresi değerleri 0,05 ile 0,70 arasında değişmekte ( $ss = 0,18$ ) ve 0,26 ortalama değere sahiptir.  $s$  parametresi ise 0,03 ile 0,79 arasında değişmekte ve ortalamaları değeri ve 0,34'dir. Her iki parametrenin düşük olması maddelerin orta güçlükte olduklarının diğer bir göstergesidir (Zhang, 2006). Madde ayırıcılık parametreleri 0,14 ve 0,83 arasında değişmekte ve ortalamaları 0,46'dır. Ayırt ediciliği düşük maddeler olsa da maddelerin ortalama ayırıcılık parametresine dayanarak testin genelinin asgari düzeyde ayırt edici olduğu savunulabilir.

Tablo 9'da öğrencilerin ilgili niteliklere sahip olma oranları (beceri dağılımları) ve standart hataları yer almaktadır.

**Tablo 9.** İçerik Alanına Ait Q-Matrisi İle DINA Modeli İçin Beceri Dağılımları Ve Standart Hataları

	$\alpha_S$	$\alpha_G$	$\alpha_V$
$P(\alpha_k)$	0,63	0,52	0,61
Standart Hata	0,03	0,04	0,03

$\alpha_S$ : Sayılar,  $\alpha_G$ : Geometrik Şekil ve Ölçümler,  $\alpha_V$ : Veri Gösterimi

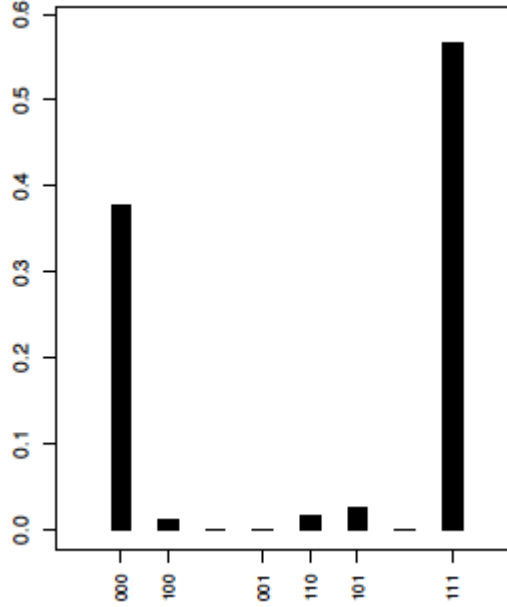
Öğrencilerin %63'ünün "sayılar" konu alanına; %52'sinin "geometrik şekil ve ölçümler" konu alanına; %61'inin de "veri gösterimi" konu alanına hakim olduğu görülmektedir. Bu aynı zamanda Öğrencilerin %37'sinin "sayılar" konu alanına; %48'sinin "geometrik şekil ve ölçümler" konu alanına; %39'unun da "veri gösterimi" konu alanına hakim olmadığını göstermektedir.

Tablo 10'da oluşan her bir beceri sınıf olasılıkları (öğrencilerin sahip olduğu veya olmadığı nitelik kombinasyonlarında bulunma olasılıkları) ile standart hataları verilmiştir.

**Tablo 10.** İçerik Alanına Ait Q-Matrisi İle DINA Modeli İçin Beceri Sınıf Olasılıkları Ve Standart Hataları

	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$	$\alpha_6$	$\alpha_7$	$\alpha_8$
	[0,0,0]	[1, 0,0]	[0,1,0]	[0, 0,1]	[1,1,0]	[1, 0,1]	[0, 1,1]	[1, 1,1]
$P(\alpha_i)$	0,35	0,04	0,00	0,02	0,00	0,07	0,00	0,52
Standart Hata	0,03	0,01	0,01	0,01	0,01	0,02	0,01	0,04

Tablo 10 incelendiğinde üç içerik alanı ele alındığından  $2^3$  adet sınıfa ait olasılık değeri olduğu görülmektedir. İçerik alanına ait yapılan sınıflamadaki sınıf olasılıklarına ilişkin standart hatalar 0,01 ve 0,04 arasında değişmektedir. Bireylerin ilgili nitelikler bağlamında oluşturduğu beceri sınıf olasılıklarına ilişkin grafik ise Şekil 1’de verilmektedir.



**Şekil 3.** İçerik Alanına Ait Q-Matrisi İle DINA Modeli İçin Beceri Sınıf Olasılıkları

Tablo 10 ve Şekil 3, farklı sınıfların dağılım olasılıklarını göstermektedir ve nitelik kombinasyonlarının analizine izin vermektedir. Nitelik sınıfı dağılımından çoğu öğrencinin ya tüm becerilere sahip olduğu ( $P([1,1,1]) = 0,52$ ) ya da hiç birine sahip olmadığı ( $P([0,0,0]) = 0,35$ ) görülmektedir. Diğer bir deyişle öğrencilerin yaklaşık yarısının her üç içerik alanına da hâkim iken öğrencilerin %35’inin her üç içerik alanına da hâkim olmadığı görülmektedir. Ayrıca, sayılar içerik alanına hakim olan öğrenciler ile veri gösterimi alanına hakim olan öğrenciler;  $P([1,0,1]) + P([1,1,1]) = 0,07 + 0,52 = 0,59$  gibi bir oranla büyük çoğunluğu oluşturmaktadır. Geometrik şekil ve ölçümler alanına hakim ancak sayılar alanına hakim olmayan ( $P[0,1,0] + P[0,1,1] = 0,00 + 0,00 = 0,00$ ) öğrenci bulunmamaktadır. Benzer şekilde geometrik şekil ve ölçümler alanına hakim ancak veri gösterimi alanına sahip olmayan ( $P[0,1,0] + P[1,1,0] = 0,00 + 0,00 = 0,00$ ) öğrenci de bulunmamaktadır. Yani bu sonuç, bu örneklem için, geometrik şekil ve ölçümler alanına sahip olabilmenin diğer iki alana hakim olmayı gerektirdiğini göstermektedir. Sayılar ve veri gösterimi içerik alanlarına hakim ancak geometri alanına sahip olmayan öğrenci sınıfının %7’lik bir dilim oluşturduğu görülmektedir. Sayılar

alanına hakim olan öğrenciler  $P([1,0,0]) + P([1,0,1]) + P([1,1,0]) + P([1,1,1]) = 0,04+0,07+0,00+0,52 = 0,63$ 'lük oranla büyük bir çoğunluğu oluşturmuştur.

## Sonuç

Bu çalışmanın bir amacı TIMSS 2015 Türkiye örnekleminde matematik başarı testine dayalı olarak birinci kitapçığı alan öğrencileri BTM altında DINA ve DINO Modelleri ile, Q-matrisinin nitelik boyutunun önsel madde bilgisini veya alan uzmanlık bilgisini dikkate almanın yanında, uygulama sonunda elde edilen verilere dayanarak belirlenmesiyle yapılan sınıflamanın doğruluğu ve tutarlılığını karşılaştırmaktır. Çalışmanın diğer bir amacı da bu bulguların ışığı altında göreceli olarak en iyi sınıflama yapan model ve Q-matrisine ilişkin öğrencilerin sınıflanmasını incelemektir.

Çalışmanın ilk araştırma sorusuna ilişkin sonuçlarına göre içerik alanına ait niteliklerden yararlanılarak oluşturulmuş Q-matrisi, DINA ve DINO Modellerine dayalı olarak yapılan her iki sınıflamada da en iyi tutarlıkta ve doğrulukta sınıflamalar yapmıştır. Bununla birlikte DINA Model, DINO Model'e kıyasla genel olarak daha yüksek sınıflama doğruluğu ve tutarlığı sağlamıştır. Ayrıca faktör analitik yöntemlerden TBA'dan faydalanarak belirlenen Q-matris ile yapılan sınıflamalarda nitelik sayısı daha az olan madde bilgisine dayalı olarak belirlenmiş Q-matris ile yapılan sınıflamaya göre daha az doğrulukta ve tutarlıkta sınıflama yapmıştır. Ancak nitelik sayısı daha fazla olan madde bilgisine dayalı olarak belirlenmiş Q-matris ile yapılan sınıflamaya göre daha fazla doğrulukta ve tutarlıkta sınıflama yapmıştır. Q-matriste yer alan nitelik sayısı arttıkça kestirim hatalarının arttığı (Bradshaw ve Templin, 2014) ve nitelik sayısı arttıkça daha az doğru sınıflamanın yapılabileceği bir çok araştırmacı tarafından belirtilmiştir (Henson ve Douglas 2005; Kunina-Habenicht, Rupp ve Wilhelm, 2012). Bununla birlikte açıklayıcı bir faktör analitik yöntemden faydalanarak belirlenmiş Q-matrisi ile uzman yargılarına dayanan madde bilgisinin birleştirilerek kullanılması Q-matrisin geliştirilmesinde teoriyi güçlendirmek için güçlü bir araç olabilir. Liu, Douglas ve Henson (2009), faktör analitik yaklaşımlarının Q-matrisin belirlenmesinde genel olarak kullanılmamasına rağmen, Q-matrisinin aşırı karmaşık olmadığı durumlarda bu yöntemlerin makul bir çözüm sunabileceğini belirtmiştir. Ancak çalışma kapsamında TBA'ya belirlenmiş olan Q-matrisin madde ve nitelik ilişkisi teoriye dayandırılmamıştır. Başka bir çalışma ile uzman yargılarına başvurularak Q-matrisi tekrar revize edilebilir. Bu çalışmada kullanılan verilerin elde edildiği ölçme aracında yer alan maddeler bilişsel tanı modellerine uygun test geliştirme sürecinden geçmedikleri için

bireylerin sınıflandırılması yetersiz olabilir. İleride yapılacak çalışmalar ile bu sınırlılık ortadan kaldırılabılır.

Çalışmanın diğer araştırma sorusuna ilişkin sonuçlarına göre araştırma kapsamındaki öğrencilerin içerik alanı niteliklerine göre nasıl sınıflandırdığı gösterilmiştir. Bu sınıflama öğrencilerin hakim olduğu konu alanlarını gösterirken, geliştirilmesi gereken konu alanlarının da ortaya çıkarması hususunda bilgi vermektedir. Örneğin öğrencilerin bir kısmının sayılar alanına hakim iken geometri ve şekiller alanına hakim olmadıkları görülmüştür. Bu sınıfta bulunan öğrencilere sayılar konu alanından ziyade geometri ve şekiller alanına ilişkin eğitim süreçlerinin planlanması önerilmektedir. Bununla birlikte geometri ve şekiller alanının diğer iki alana kıyasla en az hakim olunan alan olduğu görülmektedir. Bu alandaki eksikliklere neden olan etmenlerin ortaya çıkarılması için araştırmalar yapılabilir. Sonuçlar ve sonuçlara dayalı önerilerin bu yönde verilmesi, bu çalışma aracılığıyla matematik eğitimcilerine ve araştırmacılara, diğer yaygın psikometrik modellerin aksine BTM kullanmanın avantajını ortaya koymaktadır. Ayrıca öğrencilerin TIMSS çalışmasındaki uluslararası matematik yeterlik düzeylerine göre yapılan sınıflandırılması yerine BTM ile yapılan sınıflamaya göre sınıflandırılmasının daha fazla ayrıntılı bilgi verici olduğu görülmektedir. Çalışma grubundaki öğrencilerin çoğunluğunun tüm becerilere sahip olduğu ya da hiç birine sahip olmadığı nitelik sınıflarında olduğu görülmüş ve ortaya çıkan bu durum gerçek veri kullanılan bazı araştırmacıların bulgularıyla paralellik gösterdiği görülmüştür (George ve Robitzsch, 2015; Yi, 2017; Arıcan, 2019; Ardıç, 2020; Koyuncu, 2020). Bu araştırmanın sonucunda olduğu gibi, BTM'ye dayalı elde edilen bilgiler; öğretimin verimliliğine ve etkililiğine yönelik bilgileri ortaya çıkarmada güçlü olup kullanılabilir (Leighton ve Gierl, 2007). Bu sonuçlar ışığında aynı evrenin farklı örneklemi üzerinde yapılan sınıflamalardan elde edilen bulguların birleştirilmesi ve eğitim sistemimizde gerekli iyileştirme çalışmalarının belirli bir plan içerisinde yapılması önerilmektedir. Örneğin ülkemizde Milli Eğitim Bakanlığı Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü tarafından resmî ve özel örgün eğitim kurumları ile yaygın eğitim kurumlarında öğrenim gören istekli öğrenci ve kursiyerleri destekleme ve yetiştirme amacıyla düzenlenen kursların daha etkili verilmesine ilişkin olarak, öğrenci gruplarının güçlü ve zayıf yönlerinin bilinerek kursların düzenlenmesinde BTM'lerin sağladığı bilgilerden yararlanılabilir.

## Kaynakça

- Ardıç, E., Ö. (2020). *Bilişsel tanı ve çok boyutlu madde tepki modellerinin sınıflama doğruluğu ve parametrelerinin karşılaştırılması*. (Yayınlanmamış yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Arıcan, M., (2019). A diagnostic assessment to middle school students' proportional reasoning. *Turkish Journal of Education*, 8(4), 237-257. DOI: 10.19128/turje.522839
- Başokçu, T. O. (2015). The Cognitive Diagnostic Models for Estimating Students' Ability and Their Applications. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 14 (1), 1-32. Retrieved from <http://dergipark.gov.tr/aibuefd/issue/1502/18183>
- Baykul, Y. (1992). Eğitim sisteminde değerlendirme. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 7(7).
- Bloom, B. S. (1968). Learning for Mastery. *Instruction and Curriculum*. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation comment*, 1(2), n2.
- Büyüköztürk, Ş., Çakan, M., Tan, Ş., ve Atar, H. Y. (2014). *TIMSS 2011 Ulusal Matematik ve Fen Raporu 4. Sınıflar*. Ankara: Milli Eğitim Bakanlığı Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü.
- EARGED (2003). *TIMSS 1999 üçüncü uluslararası matematik ve fen bilgisi çalışması*. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Dairesi Başkanlığı.
- Frey, B. B. (Ed.). (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications.
- George, A. C., & Robitzsch, A. (2015) CognitiveDiagnosisModels in R: A Didactic. *The Quantitative Methods for Psychology*, 11(3), 189-205.
- George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R Package CDM for Cognitive Diagnosis Models. *Journal of Statistical Software*, 74(2), 1–24.
- Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement*, 29(4), 262-277.
- International Association for the Evaluation of Educational Achievement – IEA. (2011). *Trends in International Mathematics and Science Study 2011*. <http://www.iea.nl/timss-2011> adresinden 17 Haziran 2018 tarihinde erişilmiştir.
- International Association for the Evaluation of Educational Achievement – IEA. (2015). *TIMSS 2015 Assessment Frameworks*. Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Karasar, N. (2005). *Bilimsel araştırma yöntemi*. Ankara: Nobel.
- Koyuncu, M., S. (2020). *Bilişsel tanı modellerinde yapısal eşitlik modeli ile Q-matris doğruluğunun belirlenmesi*. (Yayınlanmamış yüksek lisans tezi). Gazi Üniversitesi, Ankara.

- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2012). The impact of model misspecification on parameter estimation and item-fit assessment in log-linear diagnostic classification models. *Journal of Educational Measurement*, 49(1), 59-81.
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Liu, Y., Douglas, J. A., & Henson, R. A. (2009). Testing person fit in cognitive diagnosis. *Applied Psychological Measurement*, 33(8), 579.
- Ma, L. (2014). Validation of the item-attribute matrix in TIMSS–Mathematics using multiple regression and the LSDM (Order No. 3616238). Available from ProQuest Dissertations & Theses Global. (1525999295). Retrieved from <https://search.proquest.com/docview/1525999295?accountid=11054>
- Mullis, I.V.S., Martin, M.O., Ruddock, G.J., O’Sullivan, C.Y. ve Preuschoff, C. (2009). *TIMSS 2011 assessment framework*. MA: TIMSS & PIRLS International Study Center, Boston College.
- R Core Team. (2015). *R: a language and environment for statistical computing*. ISBN 3-900051-07-0. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org> sitesinden erişildi.
- Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, 6(4), 219-262.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: The Guilford Press.
- Tekin, H. (2007). *Eğitimde Ölçme ve Değerlendirme*. (Gözden Geçirilmiş 18. Baskı) Yargı Yayınevi.
- Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79(2), 317-339.
- Templin, J., & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, 11, 287–305.
- Umay, A. (2003). Matematiksel muhakeme yeteneği. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 24(24).
- Tatlıdil, H. (1996). *Uygulamalı Çok Değişkenli İstatistiksel Analiz*, Cem Web Ofset Ltd. Şti, Ankara.
- Yıldız, İ., & Uyanık, N. (2004). Günümüz matematik öğretimi ve yakın çevre etkileri. *Kastamonu Eğitim Dergisi*, 12(2), 437-442.
- Yi, Y. S. (2017). In search of optimal cognitive diagnostic model (s) for ESL grammar test data. *Applied Measurement in Education*, 30(2), 82-101.
- Zhang, W. (2006). Detecting Differential Item Functioning Using the DINA Model. Yayınlanmamış doktora tezi. The University of North Carolina at Greensboro.