BUEFAD | 2021, Volume 10, Issue 1, 200-217
Bartın University Journal of Faculty of Education
dergipark.org.tr/buefad
DOI: 10.14686/buefad.801589

Adaptation Study of Mathematical Ability Test (TOMAGS) to Turkish

Duygu Özdemir [a*], Mine Işıksal Bostan [b*]

a* Asst. Prof. Dr., İstanbul Aydın University, Faculty of Education, (https://orcid.org/0000-0001-5841-3404) *duyguozdemir5@aydin.edu.tr
b* Prof. Dr., Middle East Technical University, Faculty of Education, (https://orcid.org/0000-0001-7619-1390)

Research Article | Received: **29.9.2020** | Revised: **12.1.2021** | Accepted: **24.1.2021**

Abstract

Today, it is an important issue to diagnose and provide educational support to mathematically gifted students who are seen as the potential to develop societies. Literature review and existed practices reveal that there is no common and clear way to diagnose mathematically gifted students. In this study, it was aimed to explain the adaptation studies of Test of Mathematical Abilities for Gifted Students (TOMAGS) into Turkish Language. TOMAGS was a norm referenced, standardized test and in this study, steps for adaptation of an achievement test were followed. In this regard, first of all, language and cultural adaptations were conducted and then, psychometric analysis was carried out based on the results obtained from implementation of the test with the sample consisting of 563 students whom aged ranged between 9 and 12 in different cities of Turkey. Results show validity and reliability evidences from the implementation and it was concluded that the test can also be used in Turkish Language for identifying mathematical abilities of gifted students.

Keywords: Mathematics, gifted students, test adaptation, diagnose

Üstün Yetenekli Çocuklar için Matematiksel Yetenek Testi'nin (TOMAGS) Türkçe'ye Uyarlama Çalışması

Öz

Günümüzde, toplumları ileriye taşıyabilecek potansiyeller olarak görülen matematikte üstün yetenekli çocukların tanılanmaları ve ihtiyaç duydukları eğitimsel desteğin sağlanabilmesi oldukça önemli bir konu olarak görülmeye başlanmıştır. Ancak var olan uygulamalar ve erişilebilen alan yazın taramaları bu çocukları tanılamak için çok yaygın ve net bir yöntem bulunmadığını ortaya koymaktadır. Bu sebeple, bu çalışmada matematikte üstün yetenekli çocukları belirlemek amacıyla geliştirilmiş olan TOMAGS (Test of Mathematical Abilities for Gifted Students)'ın Türkçe uyarlama çalışmaları süreci ve bulguları hakkında bilgi vermek amaçlanmıştır. Standardize edilmiş, norm referanslı bir test olan TOMAGS'ın Türkçe uyarlama sürecinde başarı testinin uyarlama sürecine ilişkin önerilen adımlar izlenmiştir. Bu süreçte önce dil ve kültür uyarlamaları yapılmış, ardından testin 9-12 yaş aralığındaki 563 kişiden oluşan örneklemde uygulanmasından elde edilen veriler ile psikometrik analizleri yapılmıştır. Güvenilirlik ve geçerlik kanıtları sunulan uygulama sonuçlarında ise testin Türkçe dilinde de üstün yetenekli çocukların matematiksel yeteneğin seviyelerini belirlemede kullanılabileceği sonucuna varılmıştır.

Anahtar kelimeler: Matematik, üstün yetenekli çocuklar, test uyarlama, tanılama

To cite this article in APA Style:
Özdemir, D. & Işıksal, M. (2021). Adaptation study of mathematical ability test (TOMAGS) to Turkish. *Bartın University Journal of Faculty of Education*, *10*(1), 200-217. https://doi.org/10.1016/buefad.801589

1 | Introduction

Gifted children in mathematics who can see the world mathematically (Krutetski, 1976) have a special potential that can contribute to the development and future of societies (Davaslıgil, 2004; Hannah, James, Montelle, & Nokes, 2011). In order to reach the desired level economically and socially, countries have initiated various studies to diagnose their gifted children in mathematics (Fıçıcı & Siegle, 2008). In order to reveal and use their existing potential, identification of gifted students is an important step. In this context, when the studies in the literature are examined, they (Basister & Norimune, 2018; Fıçıcı & Siegle, 2008; Gadanidis, Hughes, & Cordy, 2011; Glavche, Anevska, & Malcheski, 2019; Johnson, 2000; Singer, Sheffield, Freiman, & Brandl, 2016) focus on the characteristics of gifted students in mathematics, their cognitive, affective and social needs, need for differentiated or enriched education in classrooms, and teachers' knowledge, attitudes and beliefs towards these students.

How to identify mathematically gifted students is seen as one of the important questions to answer for providing appropriate educational opportunities to them (Greenes, 1981). In this context, in line with the theory of Vygotsky's (1978) Zone of Proximal Development (ZPD), theoretical framework of this study, this study focuses on the recognition of gifted students in mathematics in order to respond to their differentiated needs in classroom. Although many approaches have been developed to identify gifted students in mathematics, there is no common and clear definition of giftedness in mathematics (Özdemir 2016; Pitta-Pantazi, Christou, Kontoyianni, & Kattou, 2011). Krutetski (1976) stated that there is a mathematical thinking style in mathematically gifted children and therefore they can see the world from a mathematical perspective and many researchers have developed arguments on it. Thus, it has been mentioned that gifted students in mathematics have some important and common features even if a clear definition cannot be made. For example; having a relational understanding between numbers and symbols, associating and interpreting them in real life, solving mathematical concepts and problems in different ways in an unusual speed and accuracy, are some of these features (Fıçıcı & Siegle, 2008; Sriraman, Haavold & Lee, 2013). In addition, being able to make different and creative interpretations about mathematical concepts and abilities of solving complex problems and related interest compared to their peers are the characteristics of gifted students in mathematics (Ashley, 1973; Greenes, 1981; House, 1987). In addition, mathematical creativity is seen by many researchers (Leikin, 2009; Pitta-Pantazi et al., 2011; Sheffield, 1994; Sriraman, 2005; Sriraman et al., 2013) as an important factor in determining mathematical giftedness.

In general framework, giftedness in mathematics has been a remarkable subject since the 1900s. In addition to the common characteristics of gifted students in mathematics, various studies are carried out to determine this ability. Since the questions in the tests organized according to the students' own age and class levels are insufficient in determining gifted students, testing above the level has been one of the most used methods. As one of the first studies carried out in this context (Stanley, 1991), a criterion to get 500-800 points from the Scholastic Abilities Test - Mathematics (SAT-M)) was set for children who will attend special mathematics-related programs such as Study of Mathematically Precocious Youth (SMPY) at Johns Hopkins University. Later on, this limit score was changed in order to force the students more and to choose higher level children, and the students who could score between 700-800 as a result of the evaluation were selected for the program. Similarly, students were selected with this kind of above-level test exams in the study conducted by Lupkowski-Shoplik and Kuhnel (1995) at Carnegie Mellon University. However, since measuring the students' mathematical ability with methods appropriate for their developmental levels are more meaningful, it was thought that these above-level tests can be problematic (Ryser & Johnsen, 1998). Thus, some additional methods like open-ended questions or student interviews have been seen as one of the effective methods that can be used to evaluate giftedness in mathematics (Sheffield, 1994). Nonetheless, in spite of the advantages of student interviews, due to its difficulties in standardizing and application-evaluation process, it has not found much use. In this context, in determining the superior ability in mathematics; the use of standardized tests including open-ended questions, was considered one of the most appropriate methods.

Test of Mathematical Abilities for Gifted Students (TOMAGS) is a measurement tool with validity and reliability values that can meet the needs of researchers who need standardized measurement tools to determine giftedness in mathematics (Ryser & Johnsen, 1998). The test consists of items that will require students to use their mathematical thinking skills, mathematical reasoning and problem-solving skills, and the test has been developed taking the principles of the curriculum and characteristics of gifted children into account (Sriraman, 2008). Mathematical problem solving, mathematical communication/language and mathematical reasoning skills, which are the three basic principles that exist in the National Council of Teachers of Mathematics (NCTM) as well as the mathematics education programs of the Ministry of National Education (MoNE) in Turkey, are the principles and basic skills that have been taken into consideration in the formation of TOMAGS (MoNE, 2013).

When the national literature is examined, in addition to the test adaptation process of general intelligence, the adaptation studies of the scales developed for determining the sub-dimensions of the intelligence can be seen. For instance, the Schutte Emotional Intelligence Scale (Austin, Saklofske, Huang, and McKenney, 2004) by Tatar, Tok and Saltukoğlu (2011), BarOn Emotional Intelligence Test (BarOn & Parker, 2000) by Karabulut (2012); Perception of Gifted Education Scale (Jeong, 2010) by Tortop and Sarar (2018); Multiple Intelligence Scale (McClellan and Conti, 2008) by Babacan & Dilci (2012); Scale of Cultural Intelligence (Ang et al. 2007) by İlhan and Çetin (2014) were achieved reliable and valid results during their adaptation process into Turkish. Regarding giftedness in mathematics, which is one of the important sub-dimensions of giftedness, no similar study has been found that aims to introduce test studies that can determine giftedness in mathematics with the help of standardized scales. Additionally, as it is stated before, gifted students in mathematics are really important values for countries because these children have the potential to lead the developments in the country. Therefore, in terms of Turkey context, to be able to support the gifted students in classrooms, the first step for their progress is to recognize, identify and reveal their giftedness . For this reason, as can be seen in the relevant international and national literature, it is necessary to identify gifted students in mathematics, thus it is needed to develop measurement tools that allow us to identify them. In this connection, in the present study, it was aimed to analyze and present a measurement tool that allows the identification of gifted students in mathematics in Turkey. Hence, the validity and reliability evidences with the analysis conducted for the adaptation study of TOMAGS were presented to make contribution to the related national literature. That is, based on the need in mathematics education field, in this study, adaptation studies for the TOMAGS (intermediate level) developed for children between the ages 9-12, and the methods and its findings will be conveyed.

2 | METHOD

The adaptation process of TOMAGS, a standardized, norm-referenced test developed to identify mathematically gifted children, will be described in this study, (Callahan, 2006; Ryser & Johnsen, 1998). There are two different parts, as elementary and intermediate level, in the test. The first level of the test was developed for children aged 6 to 9, while the intermediate level was developed for children between the ages of 9-12. TOMAGS-intermediate test, which will be adapted in the scope of the study, was developed as a measurement tool that can measure the limits of gifted students, consisting of 47 open-ended questions with the required difficulty level in problem solving format. In addition, this test measures students' ability to transfer their mathematical knowledge to new and different situations or to produce new solution strategies for existing problem situations (Ryser & Johnsen, 1998).

In the study carried out by Ryser and Johnsen (1998), within the scope of the reliability studies of the original TOMAGS test, the analysis values for 3 types of error conditions were obtained as shown in Table 1 and since these values were greater than .80, it was stated that the test results can be used safely, that is, the TOMAGS test results can be used safely.

Table 1. Reliability studies of TOMAGS (Ryser & Johnsen, 1998, p. 28).

	Test Errors				
	Sample		Time Sampling	Scorer	Average
	Regular Students	Gifted Studets			
TOMAGS - Intermediate	.88	.86	.94	.99	.93

In addition, content, criterion and construct validity studies were carried out regarding the validity studies of the TOMAGS original test. During the content validity period, the content was determined by considering NCTM standards and literature reviews. Besides, the content validity of the measurement tool was strengthened by pilot study and the classical item analysis studies which enabled the reconstruction and keeping the most suitable items in the content. In addition, TOMAGS original study was examined in terms of criterion validity by making correlation study with two separate tests. That is, the Cognitive Abilities Test (CogAT) (Thorndike & Hagen, 1986) with 55 students and Iowa Tests of Basic Skills: ITBS, applied to 38 students who were diagnosed as gifted in mathematics (Hambleton, Hieronymous, & Hoover, 1987) showed that TOMAGS provides statistically significant criterion validity on the total math scores (see Table 2).

For the evidence of construct validity of the original test, group differentiation, factor analysis, item bias and item validity studies were conducted. In this context, a statistically significant difference was found on the assumption that the TOMAGS test could distinguish gifted students in mathematics and other students. While factor analysis studies concluded that the test was parallel to NCTM standards, it was also concluded that there was a scarcely any significant bias between the groups in the sample. Thus, the item validity studies reflected that the test could be used safely (see Ryser & Johnsen, 1998).

Table 2. Correlation Between TOMAGS Intermediate and Selected Tests (Ryser & Johnsen, 1998, p. 35).

Criterion Measures	TOMAGS-Intermediate
CogAT Quantitative Battery	.67
ITBS Mathematics Total	.44

Test adaptation process of TOMAGS to Turkish language

In this section, it is aimed to present the psychometric properties of this ability test adapted to Turkish as the main purpose of the study and to give information about the road map followed in the adaptation study in general. In the adaptation process of the success tests, there are steps that contain a great deal of similarity to the scale adaptation process. Hambleton (2002) stated that there are different stages that make the adaptation process of a success test successful. In the context of these steps, the steps followed during this adaptation study are summarized and explained below.

Step 1. Expert opinion on the structure to be measured is similar in both the original language and the target language.

In this study, the suitability of the adaptation process was accepted because the psychological structure of the test has similar behavioral indicators in both cultures.

Step 2. Make sure that the adaptation is the best option

As an alternative to adaptation of a test, development process of this instrument is an option. However, if there is a measurement tool that provides up-to-date, sufficient psychometric evidence in a different language, adaptation will be meaningful if other conditions are met. In addition to its psychometric strength, adaptation of the measurement tool is also useful for comparing the feature to be measured between the target language and the original language. In the adaptation phase of this test, it has been found that it is more convenient and effective to adapt a reliable and valid test that has been developed before by considering the appropriate criteria regarding giftedness.

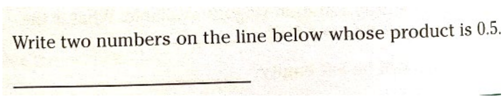
Step 3. Working with translators who know both languages well

In terms of the adaptation process, because it is vital to maintain the structural equivalence of the measuring tool, the translators who carry out the translation and adaptation process should know both the original and target language well and, if possible, should be knowledgeable in terms of the measured psychological characteristics. In the study, we have worked with translators who know both target languages and support has been received from relevant experts in terms of field knowledge.

Step 4. Translation and adaptation of the measurement tool

After the first three steps, the content of the measurement tool needs to be translated and adapted to ensure equivalence in terms of two languages. In the adaptation process of the study, the content was first translated from the original language to the target language (forward-translation,) then the translated content was translated back to the original language (back-translation) (Figure 1). In this way, the original content and the translated content were compared in terms of equivalence, and the forward and reverse translation stages were completed in this way.

Original Form of the Question:


Write two numbers on the line below whose product is 0.5.

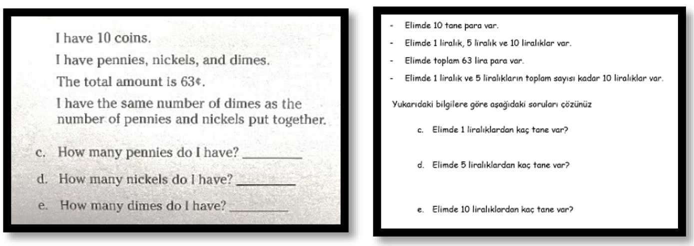
Forward Translation: Aşağıdaki satıra, çarpımları 0,5 olan iki sayı yazınız.

Back Translation: In the following line, write two numbers whose product is 0.5.

Figure 1. An example of the original form, forward and back translation for the test.

Step 5. Examining the Adapted Content

It is not enough to translate the content one to one in the adaptation process, and some changes may be required to ensure cultural equivalence. Accordingly, the original version of the measurement tool content was compared with the version after the translation, and the final version of the form was created by discussing the differences found about cultural equivalence. For example, some terms and units such as inch and yard have been converted.


I have 10 coins.
I have pennies, nickels, and dimes.
The total amount is 63¢.
I have the same number of dimes as the number of pennies and nickels put together.

c. How many pennies do I have? ________
d. How many nickels do I have? ________
e. How many dimes do I have? ________

- Elimde 10 tane para var.
- Elimde 1 liralık, 5 liralık ve 10 liralıklar var.
- Elimde toplam 63 lira para var.
- Elimde 1 liralık ve 5 liralıkların toplam sayısı kadar 10 liralıklar var.

Yukarıdaki bilgilere göre aşağıdaki soruları çözünüz

c. Elimde 1 liralıklardan kaç tane var?
d. Elimde 5 liralıklardan kaç tane var?
e. Elimde 10 liralıklardan kaç tane var?

Figure 2. An example for the adaptation of the item to Turkish culture

Step 6. Application of Adapted Measurement Tool

Evidence of validity and reliability of the measuring tool, which has gone through many steps up to this point, should be made not only through expert opinion but also empirically. In this regard, it is expected that the adapted measurement tool will be implemented in the target language, in a sample that reflects the psychological structure to be measured. Thus, adapted form was applied to a similar sample to the original sample and; validity and reliability evidences obtained in Turkey sample are presented in detail in the following sections of the study.

SAMPLE

This study was carried out with 563 students in 9-12 age group and they were selected conveniently from the private and public schools in four different cities of Turkey. There were 563 students from Ankara (391 students), Karaman (147 students), Kastamonu (17 students) and Marmaris (8 students) and 308 of them were boys (54.9%) while 253 (45.1%) of them are girls. 46.4% of the participant are students attending to public schools, 38.4% to private schools and 15.3% to educational institutions where gifted students attend. Information about the students' ages and genders was collected from each of the participants included in the study, and 146 (25.9%) of the students were aged between 12 years and 12 months; 260 (46.2%) of then were aged 11 years and 12 months;130 (23.1%) of them were between the age of 10 years and 12 months, and 22 (4%) of them were between the age of 9 years and 12 months. The average age of boys and girls is approximately 11. Considering the risk of the test results being affected by age in line with the development period, it can be stated that the groups have a balanced distribution in this respect. Detailed information about the participants is presented in Table 3 and Table 4.

Table 3. Number of participants and information about their cities

City	Number of Students	Percentage
Ankara	391 students	69.4%
Karaman	147 students	26.1%
Kastamonu	17 students	3.1%
Marmaris	8 students	1.4%

Table 4. Information and percentages about participants

School Type	City	Number of Students	Percentage
Public School	Ankara	177 students	%46.4
Public School	Ankara	84 students	
Private School	Ankara	61 students	
Private School	Karaman	147 students	%38.4
Private School	Marmaris	8 students	
Science and Art Center for Gifted Students (BİLSEM)	Ankara	10 students	
Science and Art Center for Gifted Students (BİLSEM)	Kastamonu	17 students	%15.3
School for Gifted Students	Ankara	59 students	

As mentioned above, for this study, all the adaptation process phases proposed by Hambleton (2002) have been completed and the test has been applied with the sample. After this application, the reliability and validity studies of the test were conducted and the psychometric results obtained from these studies are presented below. Moreover, for this study, approval from the research ethics committee of a university was taken and the participants, whose names used as pseudonymously for their confidentiality, participated to the study voluntarily and was informed about the aims and details before the study.

3 | Findings

Using the data of the respondents who completed the test consisting of 47 items in total, the characteristics of the distribution obtained from the results of measurement tool, analysis of the test on many items, reliability and validity evidences (psychometric evidences) are presented below.

Distribution and Properties of Items

The distribution of the results obtained from the participants in the context of the total score is given in Figure 3.

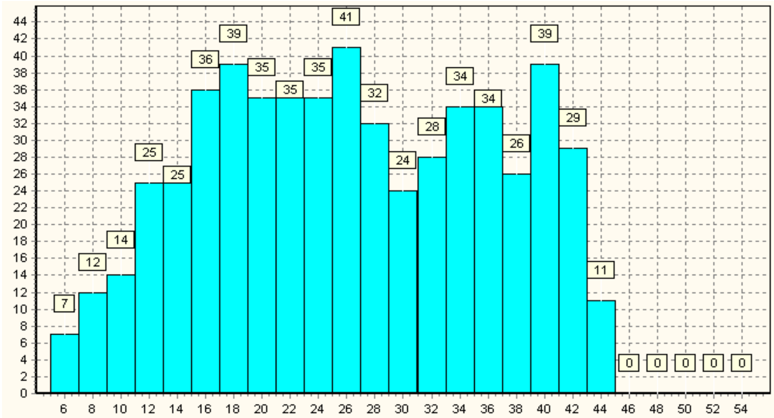


Figure 3. *Total Scores Obtained from the Participants in Tests*

When the distribution in the figure is examined, it is seen that it is slightly flattened compared to the standard normal distribution, but it shows a normal distribution pattern in accordance with its general characteristics. Other features related to the distribution are given in Table 5 below. In accordance with the results obtained, in the test consisting of 47 items, the respondents with the least number of correct answers answered 5 items and the respondents with the highest number of correct answers answered 44 items correctly. The median of the scores is 25.00 and the average is 25.75, so the distribution is very similar to the standard normal distribution. Another indicator confirming this situation is the calculation of the skewness coefficient as -0.008. Therefore, the level of skew is quite low and in contrast, since the coefficient of kurtosis is -1.006, it can be stated that the distribution is more flattened than the standard normal distribution (Pallant, 2015). In addition, it can be said that the distribution provides normal distribution properties because of the kurtosis and skewness values who are between -2 and +2 (George & Mallery, 2010).

Table 5. Variables and values related to the distribution of test items

Variable	Value
Number of Respondents	563
Highest Score Possible	47
Lowest Score	5
Highest Score	44
Median	25,00
Average	25,75
Standard Deviation	10,01
Variance	100,25
Skewness	-0,008
Kurtosis	-1,066

Properties of the items in the test; the difficulty and discrimination of the items are given in Table 6. While the test analysis examines how the test items perform in a group, item analysis examines the relationship of the items in the test with some external factors or other items in the test (Thompson & Levitov, 1985). Item analysis is a process that examines student answers with individual test items in order to evaluate the quality of these items and the entire test, and it is very important to develop and rearrange the items that will be used in the next tests (Quaigrain & Arhin, 2017). It can also be used to eliminate ambiguous, suspicious or misleading items and obtain evidence of validity for a single application test. The results obtained in the item analysis process for these purposes are presented in Table 6 as item difficulty and item discrimination index values regarding the properties of the items in the test.

Table 6. TOMAGS Item Difficulty and Item Discrimination Index Values

Item	Correct Answers	Item Difficulty	Item Discrimination Index
Q1	504	0.90	0.28
Q2	481	0.86	0.30
Q3	264	0.47	0.92
Q4	253	0.45	0.91
Q5	435	0.78	0.51
Q6	308	0.55	0.82
Q7	321	0.57	0.82
Q8	326	0.58	0.79
Q9	341	0.61	0.84
Q10	314	0.56	0.75
Q11	285	0.51	0.89
Q12	224	0.40	0.74
Q13	216	0.39	0.57
Q14	234	0.42	0.74
Q15	247	0.44	0.78
Q16	481	0.86	0.37
Q17	218	0.39	-0.61
Q18	351	0.63	0.58
Q19	105	0.19	0.42
Q20	454	0.81	0.47
Q21	309	0.55	0.81
Q22	470	0.84	-0.43
Q23	337	0.60	0.47
Q24	376	0.67	0.44
Q25	512	0.91	-0.28
Q26	512	0.91	-0.27
Q27	166	0.30	0.55
Q28	189	0.34	0.46
Q29	355	0.63	0.75
Q30	391	0.70	0.63
Q31	353	0.63	0.80
Q32	153	0.27	0.63
Q33	158	0.28	0.74
Q34	192	0.34	0.70
Q35	194	0.35	0.74
Q36	517	0.92	0.16
Q37	130	0.23	0.52
Q38	218	0.39	0.72
Q39	81	0.14	0.47
Q40	221	0.39	0.79
Q41	506	0.90	0.21
Q42	461	0.82	0.42
Q43	299	0.53	0.65
Q44	374	0.67	0.51
Q45	241	0.43	0.59
Q46	216	0.39	0.56
Q47	151	0.27	0.46

When the values in Table 6 are analyzed, it is seen that especially the discrimination of 17th, 22nd, 25th and 26th items are negative. In this regard, the correct response rate of the items in the high-performance group in terms of total score was lower than the correct response rate in the low performance group. In this respect, it can be said that 4 out of 47 items have an undesired quality in terms of discrimination.

When the items are considered as a group, the psychometric properties obtained are given in Table 7. As seen in this table, the average item difficulty was calculated as 0.548 and it is very close to 0.50. In line with the fact that the item variance gets the highest value when the item difficulty is 0.50, it is a positive feature that the average item difficulty takes a value close to 0.50 (Quaigrain & Arhin, 2017).

Table 7. General Psychometric Properties of TOMAGS Turkish Adaptation Study

Variable	Value
Number of Items Examined	47
Average of Item Difficulty	0.548
Average of Item Discrimination	0.525
Average of Double Series Correlation Coefficient	0.457
Mean Point-Double Series Correlation Coefficient	0.425
Average Number of Correct Performers of High Performs	34
Average Number of Correct Performers of Low Performers	18

In the study, three different calculations were made for item discrimination: differentiation level of correct response rate in lower and upper 27% groups, double series and point-double series correlation coefficient. The endpoint groups method can be applied to measure the discrimination power of a test item in an easy way. If the test has been applied to a large sample, the discriminative power of an item can be measured by comparing the number of high-score respondents who answered that item correctly and the number of low-score respondents who answered that item correctly. If a particular item can make a good distinction between high and low score participants, more participants in the group with the highest score will have answered that item correctly (Matlock-Hetzel, 1997). In all three methods, average item discrimination was calculated as 0.40 and above. Since the item was found to be very successful if it had a discrimination value of 0.40 and above (Crocker & Algina, 1986, p.315), it can be stated that the test can significantly differentiate the high and low performing groups in spite of the existence of four items with negative discrimination that can be shown as an exception as previously stated.

Reliability

The reliability of the achievement test, which was adapted to Turkish within the scope of this research, was tested by two different methods; Kuder-Richardson 20 Coefficient (KR-20) and Test-Halfway method. In terms of the internal consistency, these methods are KR-20 (such as true-false) which is based on correlation between results obtained from divalent measured test items (such as true-false) and Test-Halfway method that examines the results obtained from the two halves created by randomly splitting the test items (Crocker & Algina, 1986). As mentioned before, these two methods are thought to be useful and necessary in order to determine the internal consistency of the achievement test.

The KR-20 coefficient is a coefficient ranging from 0.00 to 1.00, which is used as an indicator of reliability in binary scoring (scoring as true or false) and other measurement tools (Crocker & Algina, 1986). The KR-20 coefficient, interpreted in terms of consistency between the items that make up the test, is interpreted as increasing reliability as it approaches to 1.00. In this study, the KR-20 value which was calculated according to the data obtained from 563 students was found as 0.926. According to Cortina (1993), KR-20 values of 0.90 and above indicate that the items that make up the test consistently measure the same psychological feature and there is a statistically significant relationship between them.

The Test-Halfway method is the correlation between the total scores from the two half tests created by dividing the test in one way (Crocker & Algina, 1986). Since it is a correlation coefficient, it takes values between -1.00 and 1.00. The correlation obtained belongs to one of the halves created, and the Spearman-Brown correction formula is used to calculate the reliability coefficient of the entire test, which provides evidence for internal reliability (Drost, 2011). While commenting on the reliability of the measuring tool with the test half-split method, there are different techniques regarding how to split the test, in this research, the items in the test were randomly distributed in two halves; and in this way, it was tried to prevent bias in choosing.

Correlation coefficient (r) = 0.774 between the total scores obtained from two halves which was randomly split in two groups as 24 items and 23 items. When this correlation coefficient was corrected with the Spearman-Brown formula (Spearman Brown = (2 * r) / (1 + r)), the correlation coefficient for the whole test was calculated as 0.872. In his study, Peter (1979) stated that in measurement tools developed and used in social sciences, values of 0.70 and above obtained by halfway test can be seen as satisfactory. Therefore, it can be argued that the reliability coefficients obtained by Kuder-Richardson 20 and test splitting method are satisfactory (Crocker & Algina, 1986).

Validity

The validity of a measurement tool can be defined as the degree to which the tool serves the purpose of development and in this context, the validity evidence of the TOMAGS measurement tool examined in this study was obtained as follows.

Item-Total Correlation test is applied to check the inconsistency of other behaviors measured with any item in the test and, if this is the case, to remove the item from the test. In other words, this analysis aims to eliminate unnecessary items by sieving before deciding on the factors that represent the structure (Churchill, 1979). Within the scope of this research, item-total test analysis of 47 items in the measurement tool adapted to Turkish was done and the correlation coefficients obtained are shown in Table 8 below.

Table 8. Item-Total Test Analysis Correlation Coefficients

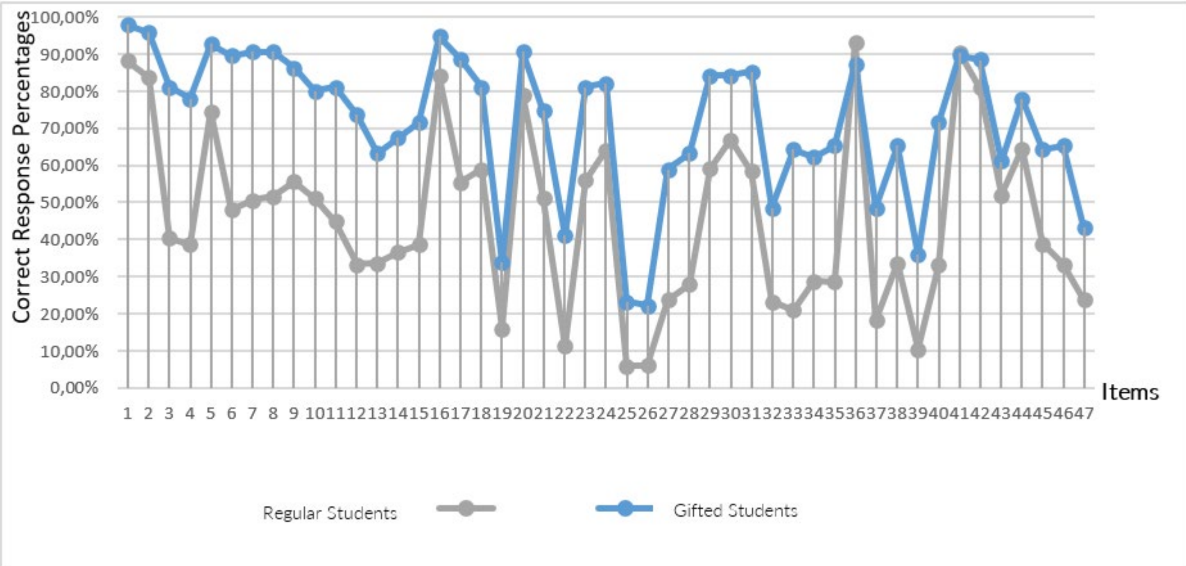
Item	Correlation Coefficient	Sig. (2-tailed)	Item	Correlation Coefficient	Sig. (2-tailed)	Item	Correlation Coefficient	Sig. (2-tailed)	Item	Correlation Coefficient	Sig. (2-tailed)
Q1	.375**	,001	Q13	.519**	,001	Q25	.486**	,001	Q37	.514**	,001
Q2	.364**	,001	Q14	.606**	,001	Q26	.465**	,001	Q38	.569**	,001
Q3	.741**	,001	Q15	.610**	,001	Q27	.502**	,001	Q39	.557**	,001
Q4	.738**	,001	Q16	.437**	,001	Q28	.442**	,001	Q40	.662**	,001
Q5	.507**	,001	Q17	.588**	,001	Q29	.606**	,001	Q41	.318**	,001
Q6	.683**	,001	Q18	.447**	,001	Q30	.524**	,001	Q42	.470**	,001
Q7	.651**	,001	Q19	.484**	,001	Q31	.650**	,001	Q43	.517**	,001
Q8	.629**	,001	Q20	.501**	,001	Q32	.605**	,001	Q44	.429**	,001
Q9	.673**	,001	Q21	.641**	,001	Q33	.691**	,001	Q45	.481**	,001
Q10	.605**	,001	Q22	.535**	,001	Q34	.599**	,001	Q46	.476**	,001
Q11	.704**	,001	Q23	.385**	,001	Q35	.642**	,001	Q47	.446**	,001
Q12	.635**	,001	Q24	.394**	,001	Q36	.260**	,001			,001

** The correlation coefficient is statistically significant. (p <.05).

As can be seen in Table 8, all 47 items are in a significant and positive relationship with the test total score variable; correlation coefficients ranged from 0.260 to 0.741. In this case, 47 items had a statistically significant relationship with the total test score; it can be said that all items measure the same psychological feature (see Churchill, 1979).

Correct Response Percentages of Students Studying at the Gifted School/BILSEM and Regular Students

The items that make up a measuring instrument are expected to differentiate those who have higher level of measured property from others, and this is considered as an indication of validity (Pierson, Kilmer, Rothlisberg & McIntosh, 2012). In this regard, the items in the TOMAGS measurement tool are expected to be answered more accurately by students diagnosed as gifted. Correct response percentages of 95 students, studying in the gifted school and in the Science and Art Centers (BİLSEM) which are institutions affiliated to the Ministry of National Education and gifted students can continue outside their school hours,



and 467 regular students are compared and analysis results are presented in Figure 4.

Figure 4. Correct Response Percentages Graph of TOMAGS Turkish Adaptation

Test Total Score and Standard Deviation Values of Gifted and Regular Students

In a test developed to distinguish gifted students, the scores of gifted and regular students were compared since the average scores of the gifted students studying at gifted school and BİLSEM are expected to be higher than other students. In this respect, the score means and standard deviations of these two groups of students from the 47-questioned measurement tool are shown in the graph in Figure 5.

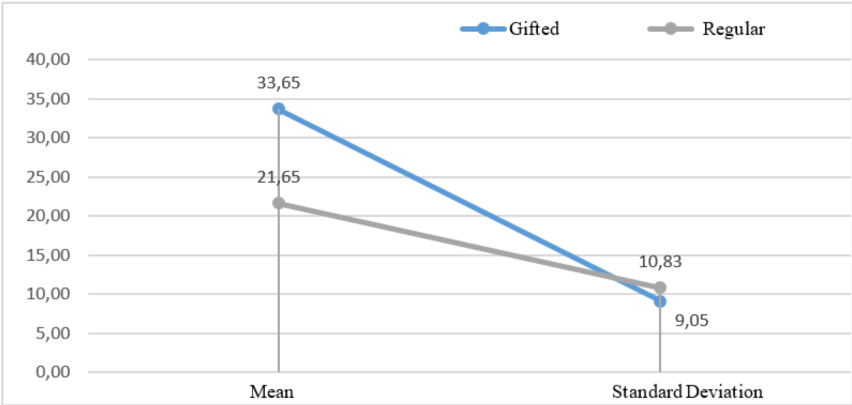


Figure 5. Total Score Mean and Standard Deviation Values of Regular and Gifted Students

As seen in the graphic above, gifted students have been more successful in terms of total score obtained from the measurement tool compared to other regular students. In addition, these students with a lower standard deviation value exhibit a more homogeneous distribution than other students.

In line with the findings, it can be stated that the measurement tool adapted to Turkish is sufficient in terms of the correlation between the items and the total score, and thus it can be concluded that it has the ability to distinguish the group that has the feature it wants to measure from the others on the basis of item and total score averages.

4 | Discussion & Conclusion

The aim of the present study is to adapt the TOMAGS to Turkish language and for this aim, adaptation stages to Turkish and Turkish culture as well as necessary validity and reliability analysis were conducted. As stated before, TOMAGS original content, which was prepared to test the limits of gifted children, can be used as a tool to identify giftedness in mathematics due to its strong validity and reliability values.

Therefore, first of all, item difficulty and item discrimination index values were calculated in accordance with the values obtained in Turkey sample of TOMAGS. The fact that the average item difficulty was found as 0.548 and it was very close to 0.50 which is the desired value, this was seen as a positive feature. In addition, the item discrimination indexes of which the assessment was made in three different ways, the correlation values of 0.40 and above and the average item discrimination as 0.525 showed that the test was at the desired level in order to distinguish the high and low performing groups significantly (Crocker & Algina, 1986, p. 315). In addition, three different evaluations were made for the item discrimination index correlation values, and these values were found to be 0.40 and above and the mean item discrimination value was 0.525. All these results showed that the test was at the desired level in order to significantly distinguish the high and low performing groups (Crocker & Algina, 1986, p. 315). However, when the individual discrimination values were examined, undesired values for the 4 items (17^{th}, 22^{nd}, 25^{th} and 26^{th}) were obtained. However, it was concluded that due to the desired result obtained at the average value, these items and values did not disrupt the general structure of the test (Crocker & Algina, 1986) and so it was decided that the results for these 4 items could be neglected. In addition, KR-20 coefficient which was calculated within the scope of reliability was found as 0.926 and this also provided a desired result as being 0.90 and above in terms of consistency of the items and reliability of the test (see Cortina, 1993). The coefficients obtained by the test halfway method ($r = 0.774$) and the coefficient obtained by the Spearman-Brown correction ($r = 0.872$) are considered as satisfactory due to being above of 0.70 (Peter, 1979) and these were provided as evidences for reliability. Within the scope of validity, the data obtained from the Item-Test Total Score Correlation showed that coefficients for the correlation change from 0.260 to 0.741 and all 47 items had a significant and positive relationship with the test total score variable. In this case, 47 items had a statistically significant relationship with the total test score; thus, it can be said that all items measure the same psychological feature. In addition, in line with the data obtained, it can be said that almost all of the items in the test provide validity by separating those who have a higher level of measured property from others.

As Dang, Weiss, and Nguyen (2013) stated, the use of intelligence or ability testing is related with the aims of diagnosing children and providing them proper educational opportunities. These children can both low and high ability students who have cognitive strengths or weakness and such kind of tests serve as a tool to aware of these students in educational environments and adapt educational content and methods to better suit the needs of these students (Kaufman, 1994). Within the scope of this adaptation study, the adaptation to Turkish language and culture, reliability and validity studies of TOMAGS which was developed for students between the ages 9-12 were carried out. By this way, a diagnosing tool that enable to find out the students who have high probability in mathematically giftedness can be obtained for Turkish language. The results of the analysis were found to be in line with the reliability and validity findings of the original TOMAGS study (Ryser & Johnsen, 1998). In other words, as with the reliability and validity values of the original study of TOMAGS, the TOMAGS–Turkish measurement tool has been found to have reliable and valid values, too. For this reason, it can be said that the TOMAGS-Turkish test can be used safely in determining the mathematical ability levels of gifted children.

In addition, findings also coincide with the findings obtained in the Turkish adaptation process of various intelligence test scales. For instance, in studies conducted by some researchers (İlhan & Çetin, 2014; Karabulut, 2012; Tatar, Tok & Saltukoğlu, 2011; Tortop & Sarar, 2018), analysis for the adaptation of the emotional intelligence, cultural intelligence and multiple intelligence scales to Turkish were reflected similar reliability and validity evidences. In addition to these, in Alma' s (2015) study, by using similar adaptation process and analysis, Gifted Rating Scale for Preschool/Kindergarten Form (GRS-P) was adapted to Turkish and so, a Turkish version of the test could be provided to the related literature and field.

In this study, adaptation study of TOMAGS- intermediate, which was mentioned in the international literature by many researchers (Crowley, 2015; Ficici & Siegle, 2008; Leader, 2008; Meehan, 2007; Ryser & Johnsen, 1998) as a reliable test developed to identify gifted students in mathematics, was carried out and it was concluded that the Turkish version of the test is reliable and valid. It is very important to determine the gifted students in mathematics and carry out studies to support their existing potential (Özdemir, 2016). Thus, it is thought that this study which aims to adopt the test that can be reliably used as a first step to identify gifted students in mathematics, might make significant contributions to the field of mathematics education and giftedness. This measurement tool can be used in achieving the goals such as identifying gifted students in potential mathematics, examining student and teacher opinions, presenting differentiated educational practices for them and evaluating their effectiveness. In addition, it is recommended that similar studies are carried out within the scope of the Turkish adaptation study of the TOMAGS 6-12 age test and the adapted test might be used in the field.

Acknowledgement

The study is part of the first researcher's dissertation study.

Statements of publication ethics

We declare that the study has no unethical problems and ethics committee approval was obtained from Middle East Technical University, Ankara (Place: ODTÜ, Date:16.06.20 Number: 286208)

Researchers' contribution rate

The authors involved in the research contributed equally.

Conflict of interest

This study has not any conflict of interest.

References

Alma, S. (2015). *Üstün yetenekliliği derecelendirme ölçekleri-Okulöncesi/Anaokulu Formu (GRS-P)'nun Türkçeye uyarlanması [Giftedness rating scales-Preschool/Kindergarten Form (GRS-P) Adaptation to Turkish]*. Doctorate Dissertation, Selçuk University, Konya, Turkey.

Austin, E. J., Saklofske, D. H., Huang, S. H., & McKenney, D. (2004). Measurement of trait emotional intelligence: Testing and cross-validating a modified version of Schutte et al.'s (1998) measure. *Personality and Individual Differences*, *36*(3), 555-562.

Ang, S., Van Dyne, L., Koh, C., Ng, K.Y., Templar, K.J., Tay, C., & Chandrasekar, N.A. (2007). Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation and task performance. *Management and Organization Review*, *3*(3), 335–71.

Ashley, R. M. (Ed.). (1973). *Activities for motivating and teaching bright children*. West Nyack, NY: Parker Publishing.

Babacan, T. & Dilci, T. (2012). Çoklu zeka ölçeği'nin türkçeye uyarlama çalışmaları [Adaptation studies of multiple intelligence scales to Turkish]. *Education Sciences*, *7*(3), 969-982.

Bar-On, R. & Parker, J. D. A. (2000). *BarOn emotional quotient inventory: Youth version*. Toronto, ON, Canada: Multi-Health system, Incorporated.

Basister, M. P. & Kawai, N. (2018). Japan's educational practices for mathematically gifted students. *International Journal of Inclusive Education*, *22*(11), 1213-1241.

Callahan, C. M. (2006). *Assessment in the classroom: The key to good instruction*. Waco, TX: Prufrock Press.

Churchill J, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 64-73.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98.

Crocker, L. & Algina, J. (1986). *Introduction to modern and classical test theory*. New York: Holt Rinehart & Winston.

Crowley, B. M, (2015). *The Effects of Problem-Based Learning on Mathematics Achievement of Elementary Students Across Time*. Western Kentucky University. Kentucky.

Dang, H. M., Weiss, B., Pollack, A., & Nguyen, M. C. (2011). Adaptation of the Wechsler intelligence scale for children-IV (WISC-IV) for Vietnam. *Psychological Studies*, *56*(4), 387-392.

Davaslıgil, Ü. (2004). Yüksek matematik yeteneğinin erken kestirimi [Early prediction of higher math ability]. In M. R. Şirin, A. Kulaksızoğlu, A.E. Bilgili (Eds.), *1. Türkiye Üstün Yetenekli Çocuklar Kongresi Seçilmiş Bildiriler Kitabı* (pp. 263-283). İstanbul: Çocuk Vakfı Yayınları.

Drost, E. A. (2011). *Validity and reliability in social science research. Education Research and Perspectives*, *38*(1), 105.

Fıçıcı, A. & Siegle, D. (2008). International teachers' judgment of gifted mathematics student characteristics. *Journal of Gifted Talented International*, *23*(1), 22-37.

Gadanidis, G., Hughes, J., & Cordy, M. (2011). Mathematics for gifted students in an arts and technology-rich setting. *Journal for the Education of the Gifted 34*(3): 397-433.

George, D., & Mallery, P. (2010). *SPSS for windows step by step. A simple study guide and reference*. GEN, Boston, MA: Pearson Education, Inc.

Glavche, M., Anevska, K., & Malcheski, R. (2019). Working with mathematically gifted students in the initial education–part one. *International Journal of Education Teacher*, *9*(18), 35-46.

Greenes, C. (1981). Identifying the gifted student in mathematics. In J., A. Hatch (Ed.), *Doing qualitative research in education settings*. Albany: State University of New York Press.

Hambleton, R. K. (2002) Adapting achievement tests into multiple languages for international assessments. In R. K. Hambleton (Ed.), *Adapting Achievement Tests into Multiple Languages for International Assessments* (pp. 58-79). Washington, DC: National Academies Press.

Hambleton, R. K., Hieronymous, A. N., & Hoover, H. D. (1987). Iowa Test of Basic Skills, Forms G and H. *Test Critiques*, *6*, 277-286.

Hannah, J., James A., Montelle C., & Nokes J. (2011). Meeting the needs of our best and brightest: curriculum acceleration in tertiary mathematics. *International Journal of Mathematical Education in Science and Technology*, *42*(3), 299-312.

House, P. A. (Ed). (1987). *Providing opportunities for the mathematically gifted, K12*. Reston, VA: National Council of Teachers of Mathematics.

İlhan, M., & Çetin, B. (2014). Kültürel Zekâ Ölçeği'nin Türkçe formunun geçerlik ve güvenirlik çalışması [Validity and reliability study of the Turkish form of the Cultural Intelligence Scale]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *29*(29-2), 94-114.

Jeong, H. W. (2010). *Teachers' perceptions regarding gifted and talented early childhood students (three to eight years of age)*. Doctorate Dissertation, Saint Louis University,

Johnson, D. T. 2000. *Teaching Mathematics to Gifted Students in a Mixed-Ability Classroom*. Reston, VA: Eric Clearinghouse.

Karabulut, A. (2012). *Duygusal zeka: Baron ölçeği uyarlaması* [*Emotional intelligence: Baron scale adaptation]*. Doctorate Dissertation, Dokuz Eylül University, İzmir, Turkey.

Kaufman, A. S. (1994). *Intelligent testing with the WISC-III*. John Wiley & Sons.

Krutetski, V. A. (1976). *The psychology of mathematical abilities in school children*. Chicago: The University of Chicago Press.

Leader, W. S. (2008). *Metacognition among students identified as gifted or nongifted using the discover assessment*. Doctorate Dissertation, The University of Arizona, Arizona.

Leikin, R. (2009). Exploring mathematical creativity using multiple solution tasks. In R. Leikin, A. Berman and B. Koichu (Eds.), *Creativity in mathematics and the education of gifted students* (pp. 129–145). Rotterdam, the Netherlands: Sense Publishers.

Lupkowski-Shoplik, A. E. & Kuhnel, A. (1995). Mathematics enrichment for talented elementary students. *Gifted Child Today Magazine*, *18*(4), 28.

Matlock-Hetzel, S. (1997). *Basic concepts in item and test analysis*. Retrieved from http://files.eric.ed.gov/fulltext/ED406441.pdf.

McClellan, J. A. & Conti, G. J. (2008). Identifying the multiple intelligences of your students. *Journal of Adult Education*, *37*(1), 13-32.

Meehan, J. M. (2007). *The role of gifted third, fourth, and fifth grade students' gender on mathematics achievement, self-efficacy, and attitude*. Doctoral dissertation, Walden University.

Ministry of National Education [MoNE]. *(2013), Ortaokul Matematik Dersi (5, 6, 7 ve 8. Sınıflar) Öğretim Programı [Middle School Mathematics Lesson (5th, 6th, 7th and 8th Grades) Curriculum]*. Ankara: MEB.

Özdemir, D. (2016). *Design and development of differentiated tasks for 5th and 6th grade mathematically gifted students*. Doctorate Dissertation, Middle East Technical University, Ankara, Turkey.

Pallant, J. (2015). *SPSS Survival Manual*. Open University Press, Berkshire.

Peter, J. P. (1979). Reliability: A review of psychometric basics and recent marketing practices. *Journal of Marketing Research*, 6-17.

Pierson, E. E., Kilmer, L. M., Rothlisberg, B. A., & McIntosh, D. E. (2012). Use of brief intelligence tests in the identification of giftedness. *Journal of Psychoeducational Assessment*, *30*, 10-24.

Pitta-Pantazi, D. P., Christou C., Kontoyianni K., & Kattou M. (2011). A model of mathematical giftedness: integrating natural, creative, and mathematical abilities, *Canadian Journal of Science, Mathematics and Technology Education*, *11*(1), 39-54.

Quaigrain, K. & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, *4*(1), 1301013.

Ryser, G.R. & Johnsen, S.K. (1998). *Test of mathematical abilities for gifted students*. Austin, TX: Pro-ed.

Sheffield, L. J. (1994). *The development of gifted and talented mathematics students and the National Council of Teachers of Mathematics Standards (Research Monograph No. 9404)*. Storrs: National Research Center on the Gifted and Talented, University of Connecticut.

Singer, F. M., Sheffield, L. J., Freiman, V., & Brandl, M. (2016). *Research on and activities for mathematically gifted students* (pp. 1-41). Springer International Publishing.

Sriraman, B. (2005). Are giftedness and creativity synonyms in mathematics?. *Prufrock Journal*, *17*(1), 20-36.

Sriraman, B. (2008). *Creativity, giftedness, and talent development in mathematics*. United States of America: Age Publishing Inc.

Sriraman, B., Haavold P., & Lee K. (2013). Mathematical creativity and giftedness: a commentary on and review of theory, new operational views, and ways forward. *Mathematics Education*, *45*(1), 215–225.

Stanley, J. C. (1976). The study of mathematically precocious youth. *Gifted Child Quarterly*, *26*, 53-67.

Stanley, J. C. (1991). An academic model for educating the mathematically talented. *Gifted Child Quarterly*, *35*(1), 36-42.

Tatar, A., Tok, S., & Saltukoğlu, G. (2011). Gözden geçirilmiş schutte duygusal zeka ölçeğinin türkçe'ye uyarlanması ve psikometrik özelliklerinin incelenmesi [Adaptation of the revised schutte emotional intelligence scale to Turkish and investigation of its psychometric properties]. *Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology*, *21*(4), 325-338.

Thompson, B. & Levitov, J. E. (1985). Using microcomputers to score and evaluate test items. *Colligate Microcomputer*, *3*, 163-168.

Thorndike, R. L. & Hagen, E. P. (1986). *Cognitive Abilities Test: Form 4*. Riverside Publishing Company.

Tortop, H. S. & Sarar, M. (2018). Üstün Yetenekliler Eğitimine İlişkin Bilgi Testi Geliştirme ve Algı Ölçeği Adaptasyon Çalışması [Knowledge Test Development and Perception Scale Adaptation Study for Gifted Education]. *Üstün Zekalılar Eğitimi ve Yaratıcılık Dergisi*, *5*(2), 67-83.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.