TURJE
www.turje.org

2022
Volume: 11 Issue: 4

Turkish Journal of Education

https://doi.org/10.19128/turje.1037694
Research Article

Received 16.12.2021
Accepted 26.10.2022

Mathematical Creativity Test (MCT) development for middle school students

Bilge Bal-Sezerel
Anadolu University, Special Education Department, Eskisehir, Türkiye, bbal@anadolu.edu.tr
Uğur Sak
Anadolu University, Special Education Department, Eskisehir, Türkiye, usak@anadolu.edu.tr

ABSTRACT This study presents the development of a mathematical creativity test and exploration of its psychometric properties. The study was conducted in six public schools and a high ability center between 2015 and 2018. The sample of the study included 1129 middle school students. The Mathematical Creativity Test (MCT) consists of problem posing, making conjecture, and proof subtests. Each test has two items. The scores of the MCT are composed of fluency, flexibility, and creativity quotient. For construct validity, EFA yielded a 3-factor solution, namely, problem posing, making conjecture, and proof subtests. CFA confirmed the 3-factor solution, and all fit indices were found to be good. For criterion validity, one-way ANOVA for independent samples was conducted in different classes, and it showed that there was a significant difference, and Pearson's correlation coefficient was investigated between MCT scores and the report card grades of the mathematics lesson. There was a strong and positive correlation between the two variables. The internal consistency and the interrater reliability of the test scores were high.

Keywords: *Assessment of mathematical creativity, making conjecture, mathematical creativity, problem posing, proof*

Ortaokul öğrencilerine yönelik Matematiksel Yaratıcılık Testi’nin (MYT) geliştirilmesi

ÖZ Bu çalışmada matematik alanında yaratıcı olan öğrencileri tanılamak amacıyla matematiksel yaratıcılık ölçeği geliştirmek ve ölçeğin psikometrik özelliklerini ortaya koymak amaçlanmıştır. Araştırma 2015-2018 yılları arasında 5., 6., 7. ve 8. sınıf düzeyindeki 1129 öğrencinin devam ettiği MEB’e bağlı altı ortaokul ve özel yeteneklilere yönelik bir merkezde gerçekleştirilmiştir. Matematiksel Yaratıcılık Testi (MYT) üç alt ölçekten (problem oluşturma, varsayım oluşturma, kanıtlama) oluşmaktadır. Alt ölçekler ikişer maddeden meydana gelmektedir. Ölçekten akıcılık, esneklik ve yaratıcılık bölümü olmak üzere üç puan türü elde edilmektedir. Yapı geçerliğini sağlamak için açımlayıcı faktör analizi ve doğrulayıcı faktör analizi (AFA ve DFA) yapılmıştır. AFA üç faktörlü yapı önermiş, DFA ise kuramsal modeli doğrulamıştır. MYT’nin ölçüt geçerliğini ortaya koymak için yapılan bağımsız gruplar için tek yönlü ANOVA sınıflar arasında anlamlı farklılık olduğunu ve matematik dersi karne notları ile yapılan Pearson korelasyon analizi ise MYT ile korelasyonun yüksek olduğunu göstermiştir. MYT’nin iç tutarlık güvenirlik değerleri ve okuyucular arası güvenirlik değerleri de yüksek çıkmıştır.

Anahtar Sözcükler: *Kanıtlama, matematiksel yaratıcılık, matematiksel yaratıcılığın ölçümü, problem oluşturma, varsayım oluşturma*

Citation: Bal-Sezerel, B., & Sak, U. (2022). Mathematical Creativity Test (MCT) development for middle school students. *Turkish Journal of Education, 11*(4), 242-268. https://doi.org/10.19128/turje.1037694

INTRODUCTION

Creativity can be defined as the ability to present a new (original, unexpected), appropriate (practical, useful) (Sternberg & Lubart, 2009), qualified and important idea or product (Sak, 2014). The interaction between the ability, process, environment, and individual (Plucker et al., 2004) is required to present the creative idea or products. Various tools are used to evaluate the creative products that emerge through interaction. Tools such as divergent thinking tests, interest and attitude product reviews, inventories, personality inventories, and reports based on creative activity and achievements are used to identify the individual, who is an important figure of the concept of creativity (Hocevar & Bachelor, 1989). Divergent thinking tests are frequently used to determine the creative ability the individual has (Runco & Acar, 2012). Some researchers (Baer, 2012; Kaufman et al., 2008) indicate that it is inappropriate to determine the creative potential of the individual in disciplines such as mathematics and science with classical creativity tests. Kaufman and Baer (2005) emphasized that classical divergent thinking tests should be transformed into domain-specific tests by making adaptations and that domain-specific tests developed in this way would make more accurate identifications. Because, the individual needs a certain domain-knowledge in order to reveal his or her mathematical creativity (Vale et al., 2018). It is not possible to develop creative solutions in mathematics with only insight and intuition without domain-knowledge. Therefore, we need tools that require domain knowledge.

In this paper, we first discussed how to measure domain-specific mathematical creativity. Then, we mentioned making conjecture and proof skills, which are two skills that are almost not included in mathematical creativity tests in the literature but cannot be ignored within the context of originality in the field of mathematics. Finally, we revealed the psychometric properties of the mathematical creativity test, which was developed based on the Mathematical Thinking Model.

Skills That Should Be Measured in Mathematical Creativity Tests

When test development studies conducted to identify mathematical creativity are examined, it is observed that studies on the identification of creativity gained momentum especially after the 1950s (Sak et al., 2017) and skills coming to the forefront in the developed divergent thinking tests (e.g., Balka, 1974; Getzels & Jackson, 1961; Haylock, 1984; Jensen, 1973; Kattou et al., 2013; Kim et al., 2003; Lee et al., 2003; Leikin, 2009; Livne & Milgram, 2006; Prouse, 1967) were usually problem solving, problem posing (Akgül & Kahveci, 2016; Bal-Sezerel, 2019; Bicer et al., 2020; Hamid & Kamarudin, 2021; Pelczer & Rodriguez, 2011) and redefinition (Haylock, 1987) skills. Experts suggest that there is a strong correlation between the aforementioned skills and creativity (Ervynck, 1991; Fisher, 1990; Haylock, 1984; Jensen, 1973; Matlin, 1994).

The relationship between mathematical creativity and problem solving occurs at the stage of producing different acceptable answers to an open-ended problem (Haylock, 1985) and solving a problem in a variety of ways (Leikin, 2009). While producing different answers to open-ended problems, fluency has a fundamental role and many different solutions to a problem can be produced through fluent thinking. On the other hand, as the ability to solve problems in very different and unusual ways increases, so does mathematical creativity (Ervynck, 1991). Because as the creative level increases, insight comes into play and much more complex methods are used in problem solving.

According to Einstein and Infeld (1938), considering mathematical and experimental skills, problem posing is a more basic skill than problem solving. In several ways, problem posing means generation of problems and formulation (Silver & Cai, 1996) because problem posing is related to the reformulation of particular mathematical situations or the formulation of new mathematical problems. Creative thinking is also required to generate new problems or different possibilities. In parallel with the ideas of Einstein and Infeld, Charles Darwin also emphasizes that presenting a problem is a more difficult skill than solving that problem (Stoyanova, 1997). In the efficient method of learning through a continuous dialog of Socrates, the process of problem posing and answering questions triggers critical and creative thinking and enables generating new ideas (Singer et al., 2013). Thus, the problem-posing process affects problem-solving skills positively (Grundmeier, 2003). According to Pollak (1987), professional mathematicians frequently encounter ill-defined problems and situations while working in the domain, and in such cases, their ultimate goal is to generate original problems that will lead to the development of the domain. According to Silver (1994), the problem posing skill is a discriminative feature in producing creative works and determining extraordinary abilities. For example, Hadamard (1945) identifies the ability to discover important research problems as a marker of extraordinary mathematical ability. Krutetskii (1976) describes mathematical creativity in the context of problem finding, exploration, independence, and originality.

Haylock (1985) separated redefinition from problem posing and argued that it is a separate skill that can be used in the assessment of mathematical creativity. Redefinition is defined as responding to a given task in multiple, varied and unique ways by redefining the elements used in mathematics (Gontijo, 2018). When we consider the redefinition, it is useful to examine a sample item. “Redefine the numbers 16 and 36 in terms of their common properties” (Haylock, 1984, s.373). It is requested to re-express two different mathematical elements according to their common properties. However, problem-posing skill is measured in a similar way. Because in both cases, there is an adaptation to different conditions, taking into account the properties of a mathematical element. In other words, in redefinition, as in problem posing, a problem is transformed into a new problem. Therefore, it can be said that redefinition is synonymous with problem posing skill (Cohen & Stover, 1981; Leung, 1997).

Upon examining the discipline of mathematics from a wider perspective, the concepts of inductive thinking and deductive thinking are encountered. Mathematical thinking is generally based on inductive and deductive thinking styles (Rips & Asmuth, 2007). Inductive thinking is mostly referred to as the concepts of discovery or invention (Yıldırım, 2000). When the discipline of mathematics is examined through the concept of creativity, we come across the famous mathematician Henri Poincarê. Poincarê (1952) states that inductive reasoning is a fundamental skill required for mathematical discoveries. Thus, when the famous mathematicians (such as Pascal, Gauss, Euler) who came to the forefront on the history scene are examined, it is observed that they put forward various mathematical conjectures and finally proved or tried to prove these conjectures. Considering that induction is "making inferences from certain situations in reaching a general rule or generating rules to prove a general statement” (Polya, 1954, p. 10), it is understood that primarily induction takes an important place in the process of concluding the searches of mathematicians successfully. In deductive thinking, another dimension of mathematical thinking, the conjectures put forward by inductive thinking are proven by presenting various proofs (Nickerson, 2010). The mathematical proofs of a mathematician can be thought of as documents of his/her mathematical discovery. Choosing the appropriate information among a wide variety of information and using it in the right place are also correlated with creativity (Poincarê, 1952). In fact, as much as understanding relationships, insight or analogy skills require creative thinking at the inductive thinking stage, revealing a proof requires creative thinking to the same extent (Yıldırım, 2000).

Inductive and deductive ways of thinking form the basis of the mathematics education domain. They have been conceptualized as inductive reasoning and deductive reasoning. Reaching from generalizations to conjectures at the inductive reasoning stage and reaching from conjectures to proofs at the deductive reasoning stage are the ultimate goals. In the report published by National Council of Teachers of Mathematics (NCTM, 2000) for the development of the mathematics curriculum, it was stated that the mathematical process abilities (problem solving, reasoning and proof, communication, connections, and representations) of all students from primary school to the last grade of middle school should be developed. Among these standards, the item “Examining mathematical conjectures and making mathematical conjectures” (NCTM, 2000, p. 56), is also included under the heading of the reasoning and proof standard. In the same report, it is emphasized that students should be able to make conjectures and put forward the reasons for the conjectures they create in some mathematical activities (p. 197). Furthermore, it was emphasized that a few examples would not be sufficient for 3rd-5th-grade students to prove the correctness of a conjecture and they should learn that counter-examples must be given to refute a conjecture (p. 188). Moreover, in the same report, two items expressed as “Develop and evaluate mathematical arguments and proofs" and "Select and use various types of reasoning and methods of proof" (NCTM, 2000, p. 56), emphasize that the proof skill in mathematics education should be underlined and improved. According to Long et al. (2012), proof and making conjectures are important for developing critical thinking skills in mathematics education because these two skills enable students to think critically and creatively at the stage of knowledge acquisition and problem solving.

The Justification of the Mathematical Creativity Test (MCT)

According to Dunn (1975), developing mathematical creativity tests has two purposes. The first one of these purposes is to recognize the creative potential of students and to make practical decisions in the school setting. Especially if students come from diverse culture or background, creativity is an equalizing psychological construct (Kozlowski & Si, 2019). The second one is to measure achievement and try to understand how successful students are within the framework of the aims of a curriculum. NCTM (2000) states that opportunities should be provided to all students at different grade levels to think flexibly and creatively about mathematical ideas and concepts. Creativity is emphasized as follows in the published standards: "Students should regard mathematics as an exciting, useful, and creative domain" (NCTM, 2000, p. 211). The perspective offered by NCTM has affected the mathematics curricula of many nations in national and international mathematics education. For example, in the ninth Development Plan Strategy (2007-2013) prepared by the Grand National Assembly (2006) in Türkiye, the importance of education requiring quality and innovation to increase international competitiveness was emphasized. In this context, since 2007, educational objectives for the domains of different disciplines have been determined in the education" system. One of these objectives is to develop creativity in mathematics education. When the mathematics curriculum of the Ministry of National Education (MONE, 2020) is examined, facilitating creative thinking is observed to be among the objectives of mathematics education in the program. Therefore, to determine the level of the potential of a skill that is included in the curriculum objectives of the nations in an individual, firstly, identification and then education intervention are required. However, considering the identification dimension, it is observed that the number of test development studies in this domain is very few while mentioning the existence of studies on the evaluation of mathematical creativity in the international arena.

The main starting point of the developed Mathematical Creativity Test (MCT) test is to consider mathematical thinking in a holistic manner and, in this context, to put the skills that form the two thinking styles into the mathematical creativity test by placing inductive and deductive thinking in the center. Inductive thinking includes discoveries or inventions (Yıldırım, 2000). Considering that one of the basic skills of a creative individual in the domain of mathematics is the ability to make a discovery, it is also thought that there should be components for measuring these skills in divergent thinking tests. Therefore, the ability to make mathematical conjectures through inductive thinking seems to be a skill that needs to be investigated in terms of being a marker of mathematical creativity. On the other hand, in deductive thinking, another dimension of mathematical thinking, conjectures put forward by inductive thinking are proven by presenting various proofs (Nickerson, 2010). The document of a mathematical discovery is mathematical proofs. The information used at the proof stage becomes valuable by choosing the appropriate ones on the way to the solution among many pieces of information and by using this information in the right place. This process is correlated with creativity (Poincarê, 1952). Therefore, the proof skill is considered to be an important marker in determining mathematical creativity. However, in the measurement of mathematical creativity, no divergent thinking tests in which these skills were used separately or together were encountered.

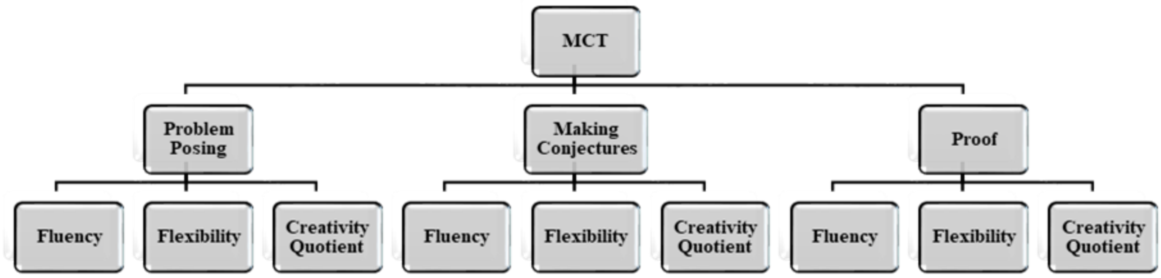
The Mathematical Creativity Test (MCT)

Framework of the MCT

The MCT was developed based on the components of the Mathematical Thinking Model (MTM) put forward by Nickerson (2010). According to Nickerson "there are many varieties of mathematical thinking" (p. 3). But he believes certain ideas (problem solving, making conjectures, proof making, and study pattern) that are exclusively descriptive of the doing maths and hold it together. These concepts form the basis of the mathematics. However, the pattern concept of the model was not included in the test. The reason why the pattern is not included in the theoretical framework is that patterns are mostly considered as tools in the proving process (Küchemann & Hoyles, 2009; Wolf, 1998). Considering the mathematical creativity tests of the problem concept, two types of problems are encountered. These are problem solving and problem posing (Silver, 1997). We preferred to evaluate student's problem posing skills. In MCT, problem solving is important ability in mathematics. According to Nickerson (2010), while the mathematician's study with patterns, they establish their problems. Then they try to solve them. Considering mathematical creativity's point of view, we need to pose problems before solving them. On the other hand, the proof concept in the model includes formal proofs used in pure mathematics. However, since the MCT is a test for middle school students, students cannot be expected to use formal proof methods. Therefore, when the concept of proof is examined from the perspective of mathematics education, it is observed that the name of the concept is changed to "informal proof" in most sources (Hersh, 1997). Furthermore, the concept is approached from the perspective of "re-examination" or "explanation" in the literature (Fosnot & Jacop, 2009). Fosnot and Jacob argued that students' proofs do not have to be formal and students should construct valid mathematical expressions through new inferences from previous mathematical expressions and conjectures based on the accepted rules, but they stated that it was necessary to explain and re-examine the correlations between mathematical expressions. Therefore, the proof concept in the MCT includes informal proofs. In MTM, the last idea is conjecture. The model suggests that they prove assumptions that mathematicians believe to be true.

After completing the test development stages of the MCT, its final form consists of a 3-component structure (problem posing, making conjectures, and proof). The items representing the components have three different creativity scores: fluency, flexibility, and creativity quotient (CQ). There are two items in each sub-test (component). The representation of the components that make up the theoretical framework of the MCT is shown in Figure 1.

Figure 1.
Theoretical Framework of the MCT



Sub-tests of the MCT and applying

The MCT is a divergent thinking test based on the paper-pencil measurement technique and designed to measure the mathematical creativity of middle school 5th, 6th, 7th, and 8th-grade students. It is sufficient for students to answer the items in the test booklet. The test can be applied as a group or individually under the supervision of a practitioner. The application of the test takes approximately one course hour. The time allocated to each item is approximately 7 minutes. The practitioner states that equal time should be allocated for each item before students start the test.

The MCT consists of 3 different sub-tests: problem posing, making conjectures, and proof. There are two items in each sub-test. Three different creativity scores (fluency, flexibility, creativity quotient) are obtained from each item. Sample items are given in Appendix-1.

Problem Posing: The problem posing items were developed based on the concept of free problem-posing situations from the ill-defined problems introduced by Stoyanova (1997). Mathematical situations are presented to students in such items. Students are asked to produce different problems using the given mathematical situations and associating them with these situations. There are a total of 2 items (squares and track) under the problem posing component of the MCT.

Squares (Item 1): It measures students' problem posing skills. It belongs to the Number Sense and Numeration strand. There is a visual in the item. Students are asked to pose more than one free problem related to the visual in the item.

Track (Item 2): It measures students' problem posing skills. It belongs to the Geometry strand. There is a visual in the item. Students are asked to pose more than one free problem related to the visual in the item.

Making Conjectures: According to Long et al. (2012), conjecture is generalizations in which the belief in its correctness is very strong. Generalizations are reached by starting from certain examples. Generalizations lead to making conjectures about the problem of interest. In such items, students are given various mathematical definitions. Students are asked to make conjectures that they think will always be correct based on these definitions and using the concepts or operations they have learned in mathematics. There are a total of 2 items (odd-even and consecutive) under the component of making conjecture of the MCT.

Odd-Even (Item 3): It measures students' ability to make conjecture. It belongs to the Algebra strand. The item contains a definition of two different number groups. Students are asked to put forward various mathematical conjectures using these number groups.

Consecutive (Item 4): It measures students' ability to make conjectures. It belongs to the Algebra strand. There is a mathematical definition in the item. Students are asked to put forward various mathematical conjectures based on this definition.

Proof: Fosnot and Jacob (2009) drew the limits of informal proof as explaining and re-examining the correlations between mathematical expressions. In these types of items, students are presented with various mathematical expressions or operations, the accuracy of which is known. Students are asked to demonstrate the accuracy of these expressions or operations with various mathematical explanations (in other words, ways). There are a total of 2 items under the proof component of the MCT.

Addition (Item 5): It measures the proof skills of students. It belongs to the Number Sense and Numeration strand. There is a mathematical equation in the item. Students are asked to prove the accuracy of the result of the equation using different mathematical methods.

Chez (Item 6): It measures the proof skills of students. It belongs to the Algebra strand. There is a visual in the item. Students are asked to prove the accuracy of the result of this visual's operation using different mathematical methods.

Scoring

The scoring system in divergent thinking tests was used in the MCT scoring method. With the responses given to the items of the test, the fluency (the total number of correct answers produced for an item), flexibility (the number of categories obtained depending on the number of different correct answers produced for an item), and creativity quotient (a numerical value obtained with a formula depending on the number of different correct answers and categories produced for an item) scores were obtained. Each correct answer to the open-ended problems was assigned as 1 point, and the wrong answer was scored as 0 points. The total fluency score of each item varies depending on the sum of the correct numbers. Snyder et al. (2004) suggested the calculation of the "creativity quotient" (CQ) in creativity tests. The formula is given below:

$$\sum_{I=1}^{I=n} log_2 (1 + u_I)$$

In the formula presented below, n represents the number of categories, and $1 \leq \imath \leq n , \imath \in \mathbb{Z}$ represents the number of similar answers in the same category.

Test development process

The test development steps -test conceptualization, test construction, test tryout, item analysis, and test revision- of Cohen and Swerdlik (2002) were taken as a basis, and the test development stages suggested by them were adapted to this study. With the adaptation and detailing of the test development process, a six-stage (test conceptualization, test construction, test tryout, pilot study, main research, item analysis, and test revision) process was followed. After the theoretical basis was determined as the Mathematical Thinking Model (MTM) of Nickerson (2010) at the conceptualization stage, a team of experienced teachers and experts in the domain of mathematics education (a total of 8 people, including 2 men and 6 women, with the length of service of 5-10 years, 2 high school teachers-3 middle school teachers-3 academicians) developed items for a total of 10 weeks (face-to-face meeting for 3 hours each week) at the construction stage. At the end of the sessions, a total of 139 items (40 problem posing items, 50 making conjecture items, and 49 proof items) were collected in the item pool. At the test construction stage, we eliminated the items in the item pool. At this stage, item selection was carried out according to the score ranking to be obtained from the evaluation form (criteria: open-endedness, measuring the targeted component, suitability with the learning domain, suitability with the target group, comprehensibility, avoiding useless concepts and information). Then MCT's items was evaluated according to these six criteria in the form. The first 55 items with the highest score among 139 items were included in the test. Then 55 items were evaluated by researchers and 2 academicians (PhD in special education), and 15 items was selected. Next, the domain experts (9 professors in mathematics education) examined the revised form comprising 15 items. Finally, the form with 12 items was obtained. Four of the 12 items represent problem posing, four represent making a conjecture, and four represent the proof component. At the test tryout stage, the researchers carried out the test to 105 middle school students (twenty-one 5th grade, twenty-eight 6th grade, twenty-nine 7th grade, twenty-seven 8th grade). At the end of the test try-out, ten items (4 problem-posing items, three making conjecture items, and three proof items) have remained on the test. In the pilot study, the researchers collected the data from 144 middle school students (34 fifth-grade, 38 sixth-grade, 37 seventh-grade, and 35 eighth-grade students). Finally, we conducted the statistical analyses and generated an answer pool. In analyzing phase, conceptually similar answers were grouped under the same categories. Because of the statistical analysis, the pilot study of the test with six-item (two-item under each component) was completed. We conducted the main research with 880 participants (103 in the center of high-ability education and 677 in the Ministry of National Education; 213 fifth-grade, 237 sixth-grade, 217 seventh-grade, 213 eighth-grade). The researchers and three practitioners applied the test in 2-course hours (40+40 min). 1-hour training was provided to the practitioners to ensure administration reliability. During the administration, the questions from the students and administration observations were recorded, then the final revisions of the items were provided by comparing the findings obtained from the data analysis. We examined the psychometric properties of the test with the data obtained from the pilot study and main research.

METHOD

Research Design and Participants

In this study, the Mathematical Creativity Test (MCT) was developed to measure the mathematical creativity of middle school students. At the stage of the test development, the cross-sectional study design among the survey design approaches was used as a research design. The cross-sectional survey approach was adopted among the survey model approaches. Cross-sectional survey approaches include studies to be carried out in a very short time on a group or sample to be taken from the population in order to obtain a judgment about the population in a population consisting of many subjects (Karasar, 2016). In determining the study group of the research, the convenience sampling was used in pilot study and main research (Büyüköztürk et al., 2007). The study was carried out with 5th, 6th, 7th and 8th-grade students studying at six public schools and a center for high ability education which provides education to special talented middle school students on weekends. Talented students were included in the main research. Therefore, we also used purposeful sampling method. A total of 1129 (540 girls, 589 boys) participants were reached in the field administration, which included the test tryout, pilot study, and main research of the study.

Instruments

Mathematic's grades

To examine the criterion validity of the test, mathematics course report card grades for the fall semester of 2017-2018 were provided from the participants. Mathematics course report card grades are graded between 0 and 100 at the middle school level.

Mathematical Creativity Test (MCT)

The other data collection tool of the study conducted in the academic periods between the fall term of 2016 and the spring term of 2018 is the MCT, which aims to measure the mathematical creativity levels of students.

Procedure

To test the criterion validity of the MCT, the correlation between mathematics report card grades and the MCT scores was examined. Therefore, in the correlation analysis conducted to determine the criterion-related validity, the connections between the mathematics course report card grades of the participants and the fluency, flexibility, and creativity quotient scores obtained from the MCT were examined. The MCT main research was carried out in the spring term, and the correlation between the students' fall term report card grades was examined.

RESULTS

Descriptive Statistics

Item analysis was performed with the data (880 students) obtained from the main research. Considering the means of the 3 types of scores (fluency, flexibity, and creativity quotient), the item with the least response based on the fluency score was found to be the 2nd item ($\bar{X}$=2.52) of the problem-posing sub-test, and the item with the most response was the 4th item ($\bar{X}$=4.37) of the making conjecture sub-test. When the items were examined in the context of the flexibility score, in other words, the number of categories, the item that produced the least different ideas was found to be the first item ($\bar{X}$=1.42) of the problem posing sub-test, and the item that produced the most different ideas was the 6th item (6) of the proof sub-test. It was found that the scores obtained in the context of the creativity quotient score were close to each other and around 2 for all items. Furthermore, the number of participants who could not respond in any type of score constitutes at most 9% of the total participants (880 participants) (min. frequency). This finding showed that 91% of the students could respond to the items.

Validity

Construct validity

For construct validity of the test, initially Exploratory Factor Analysis (EFA) was conducted with the data obtained from the pilot study. The EFA test was performed based on the fluency score obtained with the number of correct answers given to the test. Before EFA, univariate outliers and multivariate outliers in the data set, the Cook's distance, Mahalanobis distance, and centered leverage, converting the r values into z scores were checked, and 12 of the 144 observed variables exceeding critical values were excluded from the sample, and analysis was conducted with 132 observed variables. Since the *Kaiser-Meyer-Olkin* (KMO) value was .813 > .6 (Pallant, 2005), the sample size assumption was met. For the assumption of multivariate normality, *"Bartlett's test of sphericity"* ($\chi^2_{(45)} = 338.307; \mathrm{p} < .001$) was performed (Çokluk et al., 2012). Among the EFA techniques, maximum likelihood (ML) factoring was selected as the factor extraction technique (Tanaka, 1987). Kaiser's criterion, Catell's scree test (Screeplot), and parallel analysis were performed to decide on the factor number of the test (Büyüköztürk, 2011). Considering Kaiser's criterion, a one-factor structure with the total variance of 46.68% and an eigenvalue of 3.456 for 10 items, a three-factor structure in which the monotonous distribution was distorted in the scree test, and a one-factor structure in the parallel analysis (parallel analysis threshold value 1.50) emerged. At this point, the MCT theoretical framework, which is an important basis, was taken into account, and it was concluded to repeat the analysis for 3 factors. Previously, the items under the proof (1 item), making conjecture (1 item), and problem posing (2 items) components were retained from the test due to the low means of 4 items, having close factor loadings under different factors and the increase in the KMO coefficient of the test when they were retained from the test one by one, the increase in the total variance of the test, and the negative feedback from students during the administration of the test. Afterward, the maximum likelihood technique was performed with 6 items to reveal the factor design of the test, and varimax rotation was preferred among orthogonal rotation methods (Tabachnick & Fidel, 2007). The factor design of the MCT test, factor loadings of the items, descriptive statistics, and alpha reliability values are presented in Table 1.

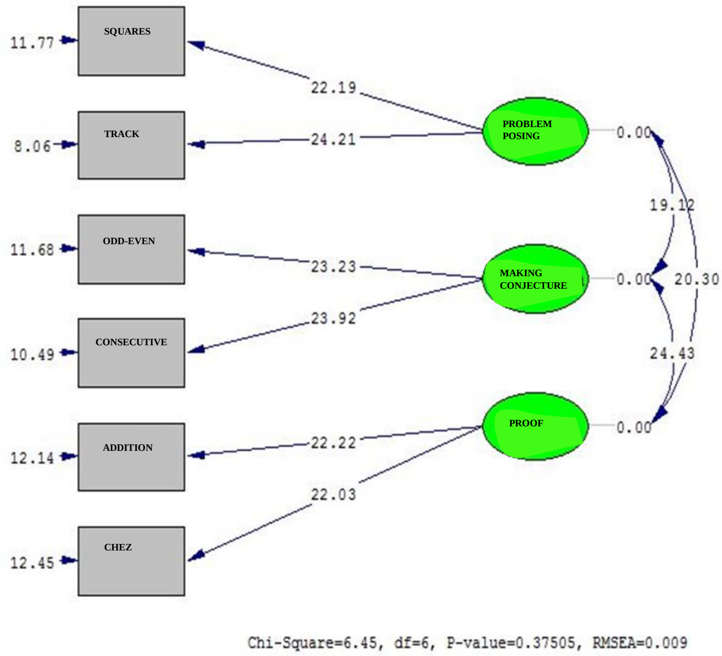
Table 1.
Descriptive Statistics of Items and Factors of MCT

Factors and Items	Variance Explained (%)	X̄	SD	Corrected item-total Correlation	Factor Loadings	Common Factor Variance (h^2)
Proof (α=0,71)						
Item1. Chez	24.17	4.12	2.275	0.42	0.52	.99
Item2. Addition		2.63	2.002	0.51	0.98	.33
Making Conjecture (α=0.65)						
Item3. Odd-Even	20.92	4.80	3.975	0.54	0.59	.67
Item4. Consecutive		3.45	3.156	0.39	0.77	.37
Problem Posing (α=0.72)						
Item5. Squares	18.26	5.05	3.223	0.55	0.46	.99
Item6. Track		4.07	2.674	0.52	0.96	.42
Total (α=0.73)	63.36	24.11	11.68			

It was observed that the attempts to prove the construct validity of the MCT supported the theoretical structure of the MCT. Moreover, gathering the items under different factors (problem posing, making conjecture, proof) was considered to be a proof that the test consists of 3 sub-tests. At the end of EFA, there were questions about the number sense and numeration (1 question) and geometry (1 question) strands in the problem posing sub-test of the MCT, and questions about the algebra (2 questions) in the making conjecture sub-test, and questions about the number sense and numeration (1 question) and algebra (1 question) strands in the proof sub-test.

The second attempt to prove the construct validity of the MCT is Confirmatory Factor Analysis (CFA). To this end, it was tested by CFA whether the MCT, developed in line with the theoretical structure based on the data obtained from the main research, verified the 3-component structure. CFA was conducted using the LISREL program. The path diagram showing parameter estimation for the theoretical model of the MCT is shown in Figure 2.

Figure 2.
Path Diagram of the MCT-CFA Results



It was found in Figure 2 that the t-values obtained for 6 items exceeded the critical value of 2.56 (Çokluk et al., 2012) and were significant at the .01 level. According to the CFA results in the path diagram, the difference between the expected and observed covariance matrices was not significant ($\chi2(6)=6.45$, $p=.375$). Furthermore, when the error variance of the observed variables was checked, it was observed that the problem posing items took values of .41 - .30, the making conjecture items took values of .39 - .35, and the proof items took values of .42 - .43. Considering the error variance values obtained from the test, it was found that the values obtained were low in parallel with the expectations of the study. In other words, it was thought that the items adequately represented the latent variable to which they belonged and must be present in the test.

After the parameter estimations made for the fit of the MCT's theoretical model and CFA model, fit indices were examined. The fit indices values ($\chi2$ /sd=1.075, RMSEA =.009, RMR=.042, SRMR=.011, NFI=.998, NNFI=1.00, CFI=1.00, GFI=.997, AGFI=.991) is observed that the acceptance levels of the goodness of fit statistics limits of good fit (Hooper et al., 2008; Kline, 2010; Jöreskog & Sörbom, 1993; Tabachnick & Fidell, 2007; Thompson, 2004). Moreover, no modification suggestions were encountered at the end of CFA. Considering the goodness of fit indices of CFA, it was observed that the theoretical model of the 3-factor structure of the test consisting of 6 items was confirmed.

For examining the internal structure of MCT, finally convergent and discriminant validity was controlled. The average variance extracted (AVE), composite reliability (CR), maximum shared variance (MSV), and average shared variance (ASV) were estimated using the factor loadings obtained from the CFA (Fornell & Larcker, 1981; Hair et al., 2010). Factor loadings ranged from .76 to .84. The AVE values of problem posing, making conjecture, and proof components of MCT were .65, .62, and .58. The CR values were .79, .77, and .73. The AVE and CR values fit convergent validity. For discriminant validity, MSV and ASV values were compared with AVE values. The MSV values were

.44, .53, and .53. The ASV values were .41, .46, and .48. All AVE was markedly higher than MSV and ASV. The indices of convergent and discriminant validity show that the MCT components accurately measure what they intend to measure.

Table 2.
Descriptive Statistics of MCT Scores in Different Grade Levels

Grade	N	Component	Score	Min.	Max.	X̄	SD
5	213	Problem Posing (PP)	Fluency	0	28.00	4.90	3.75
			Flexibility	0	8.00	2.29	1.22
			Creativity Quotient	0	12.58	3.52	2.08
		Making Conjecture (MC)	Fluency	0	31.00	6.80	4.94
			Flexibility	0	7.00	2.63	1.23
			Creativity Quotient	0	14.80	4.39	2.45
		Proof (P)	Fluency	0	16.00	4.92	3.08
			Flexibility	0	7.00	2.83	1.51
			Creativity Quotient	0	9.48	3.83	2.04
		Total	Fluency	0	67.00	16.63	10.04
			Flexibility	0	17.00	7.76	3.22
			Creativity Quotient	0	31.92	11.76	5.59
6	237	Problem Posing (PP)	Fluency	0	23.00	5.61	3.48
			Flexibility	0	11.00	2.91	1.48
			Creativity Quotient	0	16.16	4.19	2.18
		Making Conjecture (MC)	Fluency	0	25.00	6.39	4.58
			Flexibility	0	6.00	2.66	1.33
			Creativity Quotient	0	10.90	4.26	2.40
		Proof (P)	Fluency	0	14.00	5.44	2.97
			Flexibility	0	9.00	3.59	1.71
			Creativity Quotient	0	10.90	4.52	2.18
		Total	Fluency	0	48.00	17.45	9.23
			Flexibility	0	19.00	9.18	3.53
			Creativity Quotient	0	29.08	12.98	5.61
7	217	Problem Posing (PP)	Fluency	0	18.00	6.36	3.23
			Flexibility	0	8.00	3.06	1.38
			Creativity Quotient	0	11.32	4.57	1.98
		Making Conjecture (MC)	Fluency	0	31.00	8.71	5.23
			Flexibility	0	7.00	3.20	1.35
			Creativity Quotient	0	14.60	5.49	2.51
		Proof (P)	Fluency	0	16.00	6.74	2.91
			Flexibility	0	8.00	4.17	1.64
			Creativity Quotient	0	12.32	5.51	2.17
		Total	Fluency	0	60.00	21.82	9.39
			Flexibility	0	23.00	10.44	3.42
			Creativity Quotient	0	33.28	15.59	5.54
8	213	Problem Posing (PP)	Fluency	0	26.00	6.59	3.84
			Flexibility	0	10.00	3.70	1.88
			Creativity Quotient	0	13.16	5.08	2.56
		Making Conjecture (MC)	Fluency	0	36.00	9.82	6.49
			Flexibility	0	9.00	4.08	1.91
			Creativity Quotient	0	16.12	6.58	3.44
		Proof (P)	Fluency	0	20.00	6.80	3.76
			Flexibility	0	8.00	4.04	1.75
			Creativity Quotient	0	10.64	5.38	2.48
		Total	Fluency	0	65.00	23.21	11.30
			Flexibility	0	24.00	11.84	4.39
			Creativity Quotient	0	38.30	17.05	6.99

Criterion-related validity

Discrimination in different grade levels: According to Hong, and Milgram (2010) life experiences like schooling have vigorous impacts on domain-specific creativity. Some researchers concluded that years of schooling was progressively contributed to student's creativity (Haavold, 2018; Sak & Maker, 2006). Therefore, in the context of criterion-related validity, firstly, the level of discrimination of participants at different grade levels (5th, 6th, 7th, and 8th grade) by the items in the test was examined. Descriptive statistics are presented in Table 2.

It is observed that the total fluency mean scores at different grade levels increase ($\bar{X}_5$=16.63; $\bar{X}_6$=17.45; $\bar{X}_7$=21.82; $\bar{X}_8$=23.21) as the grade level increases. When the total flexibility mean scores ($\bar{X}_5$=7.76; $\bar{X}_6$=9.18; $\bar{X}_7$=10.44; $\bar{X}_8$=11.84) are examined, the means are in favor of the upper grades. The total creativity mean scores also increase linearly ($\bar{X}_5$=11.76; $\bar{X}_6$=12.98; $\bar{X}_7$=15.59; $\bar{X}_8$=17.05) with the increase in the grade level. One-way ANOVA was conducted to analyze whether this increase in different score types was significant as the grade level increased. Firstly, we tested the normality assumption of ANOVA. Skewness and kurtosis values for each score types (fluency, flexibility, CQ) were between -2 and +2. Moreover, standard z values were examined to detect the outliers and it was found that these values were between -3.3 and +3.3. We also tested Kolmogorov-Smirnov and Shapiro-Wilk values. Both of them were not significant (p>.05). Therefore, there was no violation for normality (Trochim & Donnelly, 2006). Table 3 presents the ANOVA results obtained from different grade levels.

Table 3.

ANOVA Results of MCT in Different Grade Levels

N=880		Score	Source	Sum of Squares	df	Mean Square	F	P<	η²
5-6-7-8	Problem Posing (PP)	Fluency	Between	378.024	3	126.008	9.820	.000	.033
			Within	11240.153	876	12.831			
			Total	11618.177	879				
		Flexibility	Between	215.231	3	71.744	31.209	.000	.097
			Within	2013.759	876	2.299			
			Total	2228.990	879				
		CQ	Between	276.912	3	92.304	18.849	.000	.061
			Within	4289.743	876	4.897			
			Total	4566.655	879				
5-6-7-8	Making Conjecture (MC)	Fluency	Between	1723.261	3	574.420	20.117	.000	.064
			Within	25013.121	876	28.554			
			Total	26736.382	879				
		Flexibility	Between	300.421	3	100.140	45.869	.000	.136
			Within	1912.487	876	2.183			
			Total	2212.908	879				
		CQ	Between	769.387	3	256.462	34.415	.000	.105
			Within	6527.954	876	7.452			
			Total	7297.341	879				
5-6-7-8	Proof (P)	Fluency	Between	575.498	3	191.833	18.779	.000	.060
			Within	8948.546	876	10.215			
			Total	9524.044	879				
		Flexibility	Between	236.857	3	78.952	28.723	.000	.090
			Within	2407.906	876	2.749			
			Total	2644.763	879				
		CQ	Between	399.589	3	133.196	26.871	.000	.084
			Within	4342.200	876	4.957			
			Total	4741.789	879				
5-6-7-8 Total		Fluency	Between	Within	3	2268.941	22.671	.000	.072
			Within	Between	876	100.01			
			Total	Total	879				
		Flexibility	Between	Within	3	652.752	48.452	.000	.142
			Within	Between	876	13.472			
			Total	Total	879				
		CQ	Between	Within	3	1253.034	35.302	.000	.108
			Within	Between	876	35.495			
			Total	Total	879				

It was found that the scores of the MCT components and the total scores obtained from the test created significant differences among the groups in terms of the mean fluency, flexibility, and creativity quotient scores. For determining the source of the difference between grade levels according to the achievement and non-achievement of the equality of variances, the PO-fluency and K-creativity quotient scores were examined by the Scheffe Post-Hoc test, and the other scores were examined by the Games-Howell Post-Hoc test (Huck, 2012). Table 4 presents the differences between the groups according to the results of the follow-up tests.

Table 4.
Significance Levels of Mean Difference Among Different Grade Levels

	Grade	Fluency	Flexibility	Creativity Quotient
Problem Posing (PP)	5-6	-	+	+
	5-7	+	+	+
	5-8	+	+	+
	6-7	-	+	+
	6-8	+	+	+
	7-8	-	+	+
Making Conjecture (MC)	5-6	- *	-	-
	5-7	+	+	+
	5-8	+	+	+
	6-7	+	+	+
	6-8	+	+	+
	7-8	-	+	+
Proof (P)	5-6	-	+	+
	5-7	+	+	+
	5-8	+	+	+
	6-7	+	+	+
	6-8	+	+	+
	7-8	-	-*	-*
Total	5-6	-	+	+
	5-7	+	+	+
	5-8	+	+	+
	6-7	+	+	+
	6-8	+	+	+
	7-8	-	+	+

\+ p<.05
\- p>.05
-* Reflects that the lower grades means are higher than the upper grades means.

The intergroup discrimination analysis showed that the differences between the groups were significant. When each cell in Table 4 was examined, a significant difference (p<.05) between the grades was observed in approximately 83% of the total number of cells in which the fluency, flexibility, and creativity quotient scores obtained from the sub-tests and the sum of the MCT were included (*the number of cells with a significant difference = [the number of sub-tests x the number of the combination of binary classes x the number of score types] -the number of cells without a significant difference = 6x4x3-12 = 60*). This finding obtained for 3 different score types was considered as a proof of the intergroup discrimination of the test. Moreover, the fact that the values of eta squared (Ƞ2) effect size in Table 3 were above .06 and .14 in all score types of all sub-tests, except for the fluency score type of the problem posing sub-test, it was interpreted that intergroup discrimination studies were significant at a moderate level and a high level in theory and practice (Huck, 2012).

Fit with the mathematics achievement level

In the correlation analysis performed in criterion-related validity, for revealing the connections between the mathematics course report card grades of the participants and the scores obtained from the MCT were examined via Pearson's product-moment correlations. The correlation coefficients (r_{min}=.410 and

$r_{max.}$=.485; p<.001) between the fluency, flexibility, and creativity quotient scores of the problem posing, making conjecture, and proof of the MCT and mathematics course report card grades were observed to be at a moderate level (>.30) (Cohen, 1988). Considering the total scores obtained from the MCT, the correlation between the fluency, flexibility, and creativity quotient scores and mathematics course report card grades ($r_{tot.\ flu.}$=.504, $r_{tot.\ flex.}$=.554, $r_{tot.\ cq.}$=.550; p<.001) was found to be at a high level (>.50).

Reliability

Internal consistency of the MCT

For determining the internal consistency reliability of the MCT, the Cronbach alpha coefficients were primarily examined. Table 5 presents the internal consistency analysis findings of the MCT.

Table 5.
Item-Total Correlations of MCT and Internal Consistency Analysis Results

N=880	Score	Item	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted	Cronbach Alpha Coefficient
MCT	Fluency	PO1	.607	.803	.831
		PO2	.604	.807	
		VO1	.667	.797	
		VO2	.685	.788	
		K1	.584	.812	
		K2	.572	.811	
	Flexibilty	PO1	.500	.755	.780
		PO2	.519	.750	
		VO1	.571	.736	
		VO2	.489	.756	
		K1	.551	.742	
		K2	.546	.743	
	Creativity Quotient	PO1	.591	.836	.852
		PO2	.615	.831	
		VO1	.690	.818	
		VO2	.670	.821	
		K1	.626	.829	
		K2	.647	.826	
Problem Posing Subtest	Fluency	PO1	.649		.783
		PO2	.649		
	Flexibilty	VO1	.413		.622
		VO2	.413		
	Creativity Quotient	K1	.465		.765
		K2	.465		
Making Conjecture Subtest	Fluency	PO1	.623		.809
		PO2	.623		
	Flexibilty	VO1	.682		.622
		VO2	.682		
	Creativity Quotient	K1	.451		.805
		K2	.451		
Proof Subtest	Fluency	PO1	.675		.726
		PO2	.675		
	Flexibilty	VO1	.574		.672
		VO2	.574		
	Creativity Quotient	K1	.513		.761

It was observed that the Cronbach Alpha reliability values ($\alpha_{flu.}$=.831; $\alpha_{flex.}$=.780; $\alpha_{cq.}$=.852) of the fluency, flexibility, and creativity quotient scores obtained from the overall test were above .70, which is the ideal value for the tests (Pallant, 2005). Furthermore, when an item was deleted from the test, it was observed that the fluency, flexibility, and creativity quotient scores had a negative effect on the Cronbach Alpha internal consistency coefficient. This finding was interpreted as an indicator of the consistency of the test items (Akbulut, 2010). Moreover, considering the corrected item-total correlations, the fact that the coefficients obtained were above .30 showed that the subtest to which each item belonged was correlated with the total fluency, total flexibility, and total creativity quotient scores (Field, 2009). Except for the flexibility score of the sub-tests ($\alpha_{min.}$=.622 and $\alpha_{max.}$=.672), the Cronbach Alpha values in the other score types were observed to vary in the range of $\alpha_{k\text{-}fluency}$=.726 and $\alpha_{vo\text{-}fluency}$=.809. As a result, Cronbach Alpha internal consistency analysis showed that items served the purpose of the test.

Secondly, the inter-item correlation analysis was performed to reveal the internal consistency reliability value of the MCT. Table 6 presents the correlation coefficients calculated for the MCT.

Table 6.
Inter-Item Correlation Coefficient of MCT

N=880 Score		**Items**				
		PP2	**MC1**	**MC2**	**P1**	**P2**
Fluency	PP1	.649*	.455*	.469*	.393*	.381*
	PP2		.450*	.412*	.380*	.445*
	MC1			.682*	.428*	.417*
	MC2				.481*	.438*
	P1					.574*
	P2					
Flexibility	PP1	.465*	.389*	.282*	.295*	.345*
	PP2		.394*	.300*	.344*	.353*
	MC1			.451*	.399	.368*
	MC2				.382*	.324*
	P1					.513*
	P2					
Creativity Quotient	PP1	.623*	.438*	.420*	.401*	.434*
	PP2		.494*	.433*	.410*	.449*
	MC1			.675*	.500*	.492*
	MC2				.498*	.500*
	P1					.616*
	P2					

* p<.001

It was observed that the correlation coefficients of the items in the three score types ranged from $r_{min.}$=.282 to $r_{max.}$=.675. According to the fluency score type, it was found that the correlation coefficients varied between $r_{min.}$=.380 and $r_{max.}$=.649, while the mean inter-item correlation values were r_{mean}=.479. According to the flexibility type score, it was observed that while the correlation coefficients varied between $r_{min.}$=.282 and $r_{max.}$=.513, the mean inter-item correlation values were r_{mean}=.373. According to the creativity quotient score type, it was found that while the correlation coefficients ranged from $r_{min.}$=.401 to $r_{max.}$=.675, the mean inter-item correlation values were r_{mean}=.492. Clark and Watson (1995) state that the mean inter-item correlation coefficients should be between .15 and .50 to ensure inter-item internal consistency. Furthermore, when the corrected item-total correlation values were examined, it was observed that the correlation values between the total score of the test and the fluency, flexibility, and creativity quotient scores obtained after retaining the relevant item were higher than .30. This finding showed that the items in the test exemplified similar skills (Büyüköztürk, 2011). The inter-item correlation analysis revealed that the obtained correlation coefficient values also supported the internal consistency of the MCT.

Inter-scorer reliability of the MCT

The level of consistency in the scores of the scorers was determined by inter-scorer reliability (IRR) analysis. Inter-scorer reliability means the degree of similarity of the scores obtained as a result of the evaluations made by scorers (Henning, 1993). The inter-scorer reliability study of the MCT was performed by evaluating 100 (23-5th grade, 26-6th grade, 28-7th grade, 23-8th grade) student's booklets selected by random selection from the sample also by a different scorer, apart from the one researcher. The scorer was selected from among practitioners who had previously evaluated the divergent thinking tests. The scorers rated the participants' booklets independently of each other. The intra-class correlation (ICC) analysis was conducted with the data obtained afterward. Table 7 presents the inter-scorer reliability values of the test.

Table 7.
Results of Inter-Scorer Reliability of MCT

Item	Score	r_{ICC}	F	sd	p<
Item 1 Squares	Fluency	.994	159.904	99	.001
	Flexibilty	.976	41.525	99	.001
	Creativity Quotient	.989	87.485	99	.001
Item 2 Track	Fluency	.996	280.925	99	.001
	Flexibilty	.988	82.725	99	.001
	Creativity Quotient	.994	162.133	99	.001
Item 3 Odd-Even	Fluency	.993	137.021	99	.001
	Flexibilty	.958	23.695	99	.001
	Creativity Quotient	.982	55.301	99	.001
Item 4 Consecutive	Fluency	.996	226.705	99	.001
	Flexibilty	.968	31.176	99	.001
	Creativity Quotient	.991	109.849	99	.001
Item 5 Addition	Fluency	.973	36.461	99	.001
	Flexibilty	.957	23.037	99	.001
	Creativity Quotient	.969	31.754	99	.001
Item 6 Ches	Fluency	.984	61.172	99	.001
	Flexibilty	.954	21.509	99	.001
	Creativity Quotient	.973	36.471	99	.001
Total	Fluency	.991	116.887	99	.001
	Flexibilty	.981	52.106	99	.001
	Creativity Quotient	.988	83.282	99	.001

The intra-class correlation values were observed to vary between .954 and .966 on the item basis and in the total scores of the test. According to Cicchetti and Sparrow (1990), when the intra-class correlation values are .90 and above, the inter-scorer reliability of the test is at a high level. Based on these findings, it was found that the test was reliable in terms of inter-scorer reliability.

DISCUSSION AND CONCLUSION

At the end of EFA, we found that the theoretically identified items took place under their factors and supported the 3-factor structure suggested by the theoretical framework. We observed that the total contribution of the three factors determined to the variance was 63.36%. This value is above the 40%-60% threshold, which is ideally determined in social sciences (Dunteman, 1989). That the factor loadings of the 6 items varied between between $\lambda_{min.}$=.46 and $\lambda_{max.}$=.98 showed that the items explained the factors they represented (Çokluk et al., 2012). Considering the total variance and factor loadings, we observed that the theoretically suggested 3-factor structure was also experimentally supported.

At the end of CFA, which was conducted with the data obtained from the main research to test the construct validity of the MCT, we found that the acceptance levels of the goodness-of-fit statistics were within the limits of good fit. CFA results and the indices of convergent and divergent validity showed that MCT framework theoretically represents what they intend to measure. The reason for this may be that we carried out together both statistical and practical studies in the tryout and pilot administrations of the test. In these studies, conducted in parallel between theory and practice, we repeatedly revised the test items.

In the divergent thinking tests developed for the measurement of mathematical creativity (the MCT is a paper-pencil based divergent thinking test) either the construct validity analysis of the test is not performed (Evans, 1964; Jensen, 1973) or only EFA is focused on in order to test the construct validity of the test (Balka, 1974) or only the validity proofs for predictive and convergent validity (Haylock, 1984; Leikin & Lev, 2013; Leung, 1997; Sarouphim, 1999; Singh, 1987) are presented. The reason for this may be that the test development processes are fed from different models (DeVellis, 2012) or the focus is on the items of the test. For example, Balka (1974) applied only principal component analysis (PCA) to reveal the construct validity of the Creative Ability in Mathematics Test (CAMT), which he developed to measure creative ability in the domain of mathematics and found that there was a 2-factor structure (divergent and convergent thinking), explained with 48.4% variance for problem solving. Akgül (2014) scored the test items on fluency, flexibility, and originality scores in the divergent thinking test developed to measure students' mathematical creativity and obtained the mathematical creativity score with the sum of the mentioned scores. At the end of EFA, a one-factor structure was obtained, and it was found that the total variance explained by the test items was 42%. Upon examining the studies, the variance explained by the MCT is quite high.

Examining the results of the intergroup ANOVA test to test the criterion-related validity of the MCT, we found that the scores of the sub-tests of the MCT and the total scores obtained from the overall test created significant differences between the groups in terms of fluency, flexibility, and creativity quotient score means. In the post-hoc tests and mean comparisons performed to examine between which grades the differences existed, we found that the means in all types of scores were in favor of the upper grades as the age level increased. This finding reveals that, examining the criterion-related validity studies of other mathematical creativity tests in the literature, the discrimination of the MCT is high (Bahar & Maker, 2011; Kim et al., 2003; Sak & Maker, 2006). On the other hand, we observed the means obtained from some creativity score types did not reveal significant differences between some grades (*there was no significant difference between 5th-6th, 6th-7th grades in the problem posing sub-test; between 7th-8th and 5th-6th grades in favor of 5th grades in the making conjecture sub-test, between 5th-6th, 7th-8th grades in the proof sub-test and between 5th-6th and 7th-8th grades in the overall test-See Table 4*). We can mention that there are also studies (Hall, 2009; Sarouphim, 2001) that reveal that age is not a discriminative variable in mathematical creativity. In the study carried out by Sak and Maker (2006), they observed that the mean mathematical creativity scores among primary school grades were in favor of upper grades. In the study conducted by Hall (2009) with sixth and seventh-grade students, they determined that students' mathematical creativity levels were evaluated by a test using multiple methods in problem solving and there was no difference in the multiple solutions produced by students in terms of the grade level. In the study in which Sarouphim (1999) evaluated the mathematical creativity of kindergarten, second, fourth, and fifth-grade students using the Discovering Intellectual Strengths and Capabilities (DISCOVER) assessment test, it was found that mathematical creativity did not differ depending on the students' age level.

When the literature is examined theoretically, we come across two different views. In the Componential Theory of Creativity of Amabile (1983), it is stated that domain knowledge is an indispensable component in the emergence of creativity in relation to the age variable. Simonton (1983) claims that the relationship between creativity and formal education is not linear, but has an inverted U-shaped parabolic structure. In this context, there are different opinions about the direction of the relationship between the domain-specific knowledge level and creativity. We think the main reason for differences to originate from the nature of creativity. According to Ervynck (1991), creativity does not emerge in a bell glass, but with the combination of various factors (environmental, educational, mental, etc.). Therefore, rather than explaining creativity with a single factor, it is beneficial to think of it in context.

The insignificance of the differences between the 7th and 8th grade means in the sub-tests and the fact that they are in favor of 7th graders can be made causal with the concepts called eighth-grade cliff and fourth-grade slump in the literature (Tompkins, 1994). The first reason for this is the incompatibility between the educational materials, books, and the educational content, and students' distancing from the content. The second one is the academic attitudes of course teachers. Teachers' avoidance of the domain-specific explanations in primary school adversely affects student achievement, and incomplete education in primary school causes failure in middle school (Anderson & Freebody, 1981; Sanacore & Palumbo, 2009). Anxiety about the entrance to high schools may be the second reason. Kesici and Aşılıoğlu (2017) concluded that eighth-graders experience stress to become successful in the exam called the Transition from Basic Education to Secondary Education (TEOG) and this stress adversely affects their mathematics achievement.

The contaminating factor of fluency plays a role in obtaining different findings regarding the discrimination of the MCT. Kaufman et al. (2008) state that fluency scores can have a pollutant effect in creativity studies and it should be controlled because when fluency is accepted as the correct number produced, it is observed that very similar answers also cause an inflated increase in creativity scores (Seddon, 1983). Thus, we used to eliminate the pollutant effect of fluency, the score of the creativity quotient in this study (see Snyder et al., 2004). In this context, we observed that these values, which were obtained from the fluency scores at different grade levels and were not statistically significant, created significant differences in the scores of the creativity quotient (See Table 4). Therefore, the preferred method of scoring yields consistent results.

Considering the overall test, we found that the differences between the means obtained from the overall test revealed significant differences between the grades. Especially in the results in which the differences between the grades were statistically significant, that most of the effect size (Ƞ2) values took values above .06 and above .14 was interpreted as that intergroup discrimination studies created a moderate and high level of effect in theory and practice (Huck, 2012). From this aspect, the MCT discriminates between students at different grade levels in their mathematical creativity level.

Finally, the fit of mathematical creativity with mathematics achievement was examined to test the criterion-related validity of the MCT. The correlation between the report card grades of the mathematics course and the MCT scores of students was high ($r_{tot.\ flu.}$=.504, $r_{tot.\ flex.}$=.554, $r_{tot.\ cq.}$=.550; p<.001). Studies examining the correlation between mathematical creativity and mathematics achievement have also found that there is a significant correlation between the two variables (Bicer et al., 2020, p. 255; Bahar & Maker, 2011; Kim et al., 2003). In the studies conducted, we observed that the correlation revealed between mathematics achievement and mathematical creativity ranged from $r_{min.}$=.31 to $r_{max.}$=.58. Considering the correlations obtained from the MCT, we found that the obtained coefficients of fit revealed similar findings to the literature. Considering that domain knowledge is an important component in domain-specific creativity (Amabile, 1983), the findings obtained with the MCT test also supported the hypothesis stating that mathematics achievement is an important component in mathematical creativity.

In psychometric tests, and especially in intelligence and ability tests, it should present internal consistency as a proof of the homogeneity of the test's measurement (Anastasi & Urbina, 1997). The Cronbach's Alpha reliability values ($\alpha_{flu.}$=.831; $\alpha_{flex.}$=.780; $\alpha_{cq.}$=.852) related to the fluency, flexibility, and creativity quotient scores obtained from the overall test are above .70, which is accepted as ideal (Pallant, 2005). Considering the sub-tests, the alpha reliability values of the other score types, except for only the flexibility score type ($\alpha_{problem\ p.}$=.622; $\alpha_{m.conjecture.}$=.622; α_{proof}=.672), are above .70.

We observed that the Cronbach's Alpha reliability values of mathematical creativity tests developed to determine mathematical creativity in the literature (Balka, 1974; Getzels & Jackson, 1961; Kim et al., 2003; Mann, 2009; Pham, 2014; Prouse, 1967; Sarouphim, 1999) vary between .55 and .92. According to Büyüköztürk (2011), it is sufficient for the reliability coefficient calculated for psychological tests to be .70 and above. However, the reliability coefficient of the tests to be used to select and classify individuals should be much higher. Considering both the reliability values of other tests used in the literature and the statistical acceptance level determined for the tests, the internal consistency reliability values of the MCT are quite good. Considering that the mathematical ability levels of students are determined according to the creativity quotient score type among the three types of scores obtained from the MCT, especially in the identification of students ($\alpha_{cq.}$=.852), the internal consistency reliability revealed by the test is at a high level.

In the analysis made for the inter-scorer reliability of the MCT, we observed that the intra-class correlation values in the fluency, flexibility, and creativity quotient scores obtained from 6 items in the test and the overall test was high (varied between .954 and .966). When the inter-scorer reliability analysis performed to reveal the reliability of the test in mathematical creativity tests is examined (Balka, 1974; Hall, 2009; Griffiths, 1996), it is observed that these values are between .72 and .95. When these values were examined, it was found that the MCT's inter-scorer reliability revealed parallel findings with the literature, and even the MCT's inter-scorer reliability was higher compared to other tests. One reason for this may be the high range and diversity of the answers accumulated in the answer pool during the research period. Considering that the MCT is an open-ended test and the number of answers produced by students can be much higher compared to other sciences, especially when mathematics is considered, the high representation of the sample group of the study caused the number of answers accumulated in the item pool and representation to be high. This facilitated the scorer's scoring of the test in the answers produced in different types and caused the inter-scorer assessment consistency to be at least 95%.

Suggestions

Last decades, the research trend emphasized domain-specific creativity instead of domain-general creativity (Programme for International Student Assessment [PISA], 2022). In the same direction, this study developed a test to reveal the creativity specific to the field. Besides, the given mathematical creativity tests stem from specific mathematical skills, not appropriate mathematical models. MCT test offers a much more holistic structure by adopting an MTM model thought to be directly related to mathematical creativity. This study is one of the first studies to focus on making conjecture and proof in mathematical creativity. We directly related establishing and proving assumptions in mathematics to the very nature of mathematics. Using these two skills in the test brings novelty to the field.

This study has limitations, like other studies. In this study for criterion validity, discrimination in different grade levels and math report card grades was examined. In further research, the relationship between MCT and current domain-specific mathematical creativity tests can be investigated. The items in MCT provided sample items for students for a better understanding. Further research can conduct with non-sample versions of these items. The scores students get from both different tests can reveal to what extent they affect their creativity. The skills of proof and making conjecture appear to be high-level skills. Measuring these skills at an early age can lead to earlier intervention plans. Hence, a test for the primary school level might be adopted. Also, it might be fruitful to examine the culture-specific dimensions in test development.

Acknowledgement

This research was a part of the first author's dissertation study which was supervised by the second author. This study was supported by a grant from the Anadolu University [Grant No:1605E493].

REFERENCES

Akbulut, Y. (2010). *Sosyal bilimlerde SPSS uygulamaları [SPSS applications in social sciences]*. Pasifik Ofset.

Akgül, S. (2014). *Üstün yetenekli öğrencilerin matematik yaratıcılıklarını açıklamaya yönelik bir model geliştirilmesi [A model study to examine gifted and talented students' mathematical creativity]* [Unpublished doctoral dissertation]. İstanbul Üniversitesi.

Akgul. S., & Kahveci, N. G. (2016). A study on the development of a mathematics creativity scale. *Eurasian Journal of Educational Research*, *62*, 57-76. http://dx.doi.org/10.14689/ejer.2016.62.5

Amabile, T. M. (1983). The social psychology of creativity: A componential conceptualization. *Journal of Personality and Social Psychology, 45*(2), 357-376. https://doi.org/10.1037/0022-3514.45.2.357

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Prentice-Hall.

Anderson, R., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research review* (pp. 71–117). International Reading Association.

Baer, J. (2012). Domain specificity and the limits of creativity theory. *Journal of Creative Behavior, 46*(1), 16-29. https://doi.org/10.1002/jocb.002

Bahar, A. K., & Maker, J. C. (2011). Exploring the relationship between mathematical creativity and mathematical achievement. *Asia-Pacific Journal of Gifted and Talented Education, 3*(1), 33-48.

Balka, D. S. (1974). *The development of an instrument to measure creative ability in mathematics* [Unpublished doctoral dissertation]. University of Missouri.

Bal-Sezerel, B. (2019). *Ortaokul öğrencilerinin yaratıcılığını ölçmeye yönelik matematiksel üretkenlik testinin geliştirilmesi [Development of a mathematical creativity test for creativity of middle school students]* [Unpublished doctoral dissertation]. Anadolu Üniversitesi.

Bicer, A., Chamberlin, S., & Perihan, C. (2020). A meta-analysis of the relationship between mathematics achievement and creativity. *Journal of Creative Behaviour, 55*(3), 569-590. https://doi.org/10.1002/jocb.474

Bicer, A., Lee, Y., Perihan, C., Capraro, M. M., & Capraro, R. M. (2020). Considering mathematical creative self-efficacy with problem posing as a measure of mathematical creativity. *Educational Studies in Mathematics, 105,* 457-485. https://doi.org/10.1007/s10649-020-09995-8

Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, E. A., Karadeniz, Ş. ve Demirel, F. (2017). *Bilimsel araştırma yöntemleri [Scientific research methods]* (23th ed.). Pegem Akademi.

Büyüköztürk, Ş. (2011). *Sosyal bilimler için veri analizi el kitabı [Manual of data analysis for social sciences]* (14th ed.). Pegem Akademi.

Cicchetti, D. V., & Sparrow, S. S. (1990). Assessment of adaptive behavior in young children. In J. J. Johnson, & J. Goldman (Eds.), *Developmental assessment in clinical child psychology: A handbook* (pp. 173–196). Pergamon Press.

Cohen, S. A., & Stover, G. (1981). Effects of teaching sixth-grade students to modify format variables of math word problems. *Reading Research Quarterly*, *16*(2), 175-200. https://doi.org/10.2307/747554

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Cohen, R. J., & Swerdlik, M. (2002). *Psychological testing and assessment: An introduction to test and measurement* (5th ed.). McGraw-Hill.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları [Multivariate statistics for social sciences: SPSS and LISREL applications]* (2nd ed.). Pegem Akademi.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Sage Publications.

Dunn, J. A. (1975). Tests of creativity in mathematics. *International Journal of Mathematical Education in Science and Technology,6*(3), 327-332. https://doi.org/10.1080/0020739750060310

Dunteman, G. H. (1989). *Principal component analysis: Quantitative applications in the social sciences series* (Vol. 69). Sage Publications.

Einstein, A., & lnfeld, L. (1938). *The evolution of physics.* Simon & Schuster.

Ervynck, G. (1991). Mathematical creativity. In D. Tall (Ed.), *Advanced mathematical thinking* (pp. 42-53). Kluwer.

Evans, E. W. (1964). *Measuring the ability of students to respond in creative mathematical situations at the late elementary and early junior high school level* [Unpublished Doctoral Dissertation]. University of Michigan.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Sage Publication.

Fisher, R. (1990). *"Teaching for thinking: Language and maths"and "teaching for thinking across the curriculum", chapters in teaching children to think.* Basil Blackwell.

Fornell, C. & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *American Marketing Association, 18*(1). 39-50. https://doi.org/10.2307/3151312

Fosnot, C. T., & Jacob, B. (2009). Young mathematicians at work: The role of context and models in the emergence proof. In D. A. Stykianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspectives* (pp. 102-119). Taylor & Francis.

Getzels, J. W., & Jackson, P. W. (1961). Family environment and cognitive style: A study of the sources of highly intelligent and of highly creative adolescents. *American Sociological Association, 26*(3), 351-359. https://doi.org/10.2307/2090662

Grand National Assembly (2006). *Dokuzuncu Kalkınma Planı [The Ninth Development Plan] 2007-2013. T. C. Resmi Gazete*. https://www.resmigazete.gov.tr/eskiler/2006/07/20060720-5.htm

Griffiths, S. E. (1996). *The inter-observer reliability of the DISCOVER problem-solving assessment* [Unpublished Manuscript]. University of Arizona.

Gontijo, C. H. (2018). Mathematics education and creativity: A point of view from the systems perspective on creativity: In N. Amado, S. Carreira, & K. Jones (Eds.), *Broadening the scope of research on mathematical problem solving a focus on technology, creativity and affect* (pp. 375-386). Springer.

Grundmeier, T. A. (2003). *The effects of providing mathematical problem posing experiences for K-8 pre-service teachers: Investigating teachers' beliefs and characteristics of posed problems* [Unpublished doctoral dissertation]. University of New Hampshire.

Haavold, P. Q. (2018). An investigation of the relationship between age, achievement, and creativity in mathematics. *Journal of Creative Behaviour, 54*(3), 555-566. https://doi.org/10.1002/jocb.390

Hadamard, J. (1945). *The psychology of invention in the mathematical field*. Dover Publications.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Englewood Cliffs.

Hall, L. (2009). *Problem solving and creativity: A gender and grade level comparison* [Unpublished doctoral dissertation]. Tennessee State University.

Hamid, A., & Kamarudin, N. (2021). Assessing students' mathematics achievement and mathematical creativity using mathematical creative approach: A quasi-experimental research. *Asian Journal of University Education, 17*(2), 100-112.https://doi.org/10.24191/ajue.v17i2.13399

Haylock, D. W. (1984). *Aspects of mathematical creativity children aged 11 – 12* [Unpublished doctoral dissertation]. University of London.

Haylock, D. W. (1985). High mathematical creativity in a pair of identical twins. *The Journal of Genetic Psychology, 16*(4), 547-553.

Haylock, D. W. (1987). A framework assessing mathematical creativity in schoolchildren. *Educational Studies in Mathematics, 18*(1), 59-74.

Henning, G. (1993). Issues in evaluating and maintaining an ESL writing assessment program. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 279-291). Ablex Publishing.

Hersh, R. (1997). *What is mathematics really?* Oxford University Press.

Hocevar, D., & Bachelor, P. (1989). A taxonomy and critique of measurements used in the study of creativity. In J. A. Glover, R. R. Ronning, & C. R. Reynolds (Eds.), *Handbook of creativity* (pp. 53-76). Plenum Press.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6(1)*, 53-60. https://doi.org/10.21427/D7CF7R

Huck, S.W. (2012). *Reading statistics and research* (6th ed.). Pearson.

Jensen, L. R. (1973). *The relationships among mathematical creativity, numerical aptitude, and mathematical achievement* [Unpublished doctoral dissertation]. The University of Texas.

Jöreskog, K. G., & Sörbom, D. (1993). *Lisrel 8: Structural equation modeling with the simples command language*. Scientific Software International.

Karasar, N. (2016). *Bilimsel araştırma yöntemi: Kavramlar ilkeler teknikler [Scientific research method: Concepts principles techniques]* (30th ed.). Nobel.

Kattou, M., Kontoyianni, K., Pitta-Pantazi, D., & Christou, C. (2013). Connecting mathematical creativity to mathematical ability. *ZDM Mathematics Education, 45*, 167-181. https://doi.org/10.1007/s11858-012-0467-1

Kaufman, J. C., & Baer, J. (2005). *Creativity across domains: Faces of the muse*. Lawrence Erlbaum.

Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. John Wiley & Sons.

Kesici, A., & Aşılıoğlu, B. (2017). Developing stress scale for secondary school students: Reliability and validity study. *Kastamonu Eğitim Dergisi, 25*(6), 2413-2426.

Kim, H., Cho, S., & Ahn, C. (2003). Development of mathematical creative problem solving ability test for identification of the gifted in math. *Gifted Education International, 18*(2), 164-174. https://doi.org/10.1177/026142940301800206

Kline, R. B. (2010). *Principles and practice of structural equation modeling*. Guilford Publications.

Kozlowski, J. S., & Si, S. (2019). Mathematical creativity: A vehicle to foster equity. *Thinking Skills and Creativity, 33*, 1-8. https://doi.org/10.1016/j.tsc.2019.100579

Krutetskii, V. A. (1976). *The psychology of mathematical abilities in schoolchildren*. The University of Chicago Press.

Küchemann, D., and Hoyles, C. (2009). From empirical to structural reasoning in mathematics. In D. A. Stykianou, M. L. Blanton, & E. J. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspectives* (pp. 171-203). Taylor & Francis.

Lee, K. S., Hwang, D., & Seo, J. J. (2003). A development of the test for mathematical creative problem solving ability. *Journal of the Korea Society of Mathematical Education Series D: Research in Mathematical Education, 7*(3), 163-189.

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

Leikin, R., & Lev, M. (2013). Mathematical creativity in generally gifted and mathematically excelling adolescents: What makes the difference? *ZDM-The International Journal on Mathematics Education, 45*(2), 183–197. https://doi.org/10.1007/s11858-012-0460-8

Leung, S. S. (1997). On the role of creative thinking in problem posing. *ZDM Mathematics Education, 29*(3), 81-85. https://doi.org/10.1007/s11858-997-0004-9

Livne, N. L., & Milgram, R. M. (2006). Academic versus creative abilities in mathematics: Two components of the same construct? *Creativity Research Journal, 18*(2), 199-212. https://doi.org/10.1207/s15326934crj1802_6

Long, C. T., DeTemple, D. W., & Millman, R. S. (2012). *Mathematical reasoning for elementary teachers* (6th ed.). Pearson.

Mann, E. L. (2009). The search for mathematical creativity: Identifying creative potential in middle school students. *Creativity Research Journal, 21*(4), 338-348. https://doi.org/10.1080/10400410903297402

Matlin, M. (1994). *Cognition* (3rd ed.). Hartcout Brace.

Ministry of National Education. (2020). *Milli Eğitim Bakanlığı öğretim programlarını izleme ve değerlendirme sistemi [The system of monitoring and evaluation of the curriculum of the Ministry of National Education]*. http://mufredat.meb.gov.tr/ProgramDetay.aspx?PID=329

National Council of Teachers Mathematics. (1990). *Teaching & learning mathematics in the 190's.* The National Council of Teachers Mathematics Inc.

Nickerson, R. S. (2010). *Mathematical reasoning.* Psychology Press.

Pallant, J. (2005). *SPSS survival manual* (12nd ed.). Allen & Unwin.

Pelczer, I., & Rodriguez, F. G. (2011). Creativity assessment in school settings through problem posing tasks. *The Montana Mathematics Enthusiast, 8*(1), 383-398.

Pham, L. H. (2014). *Validation of prediction relationship of creative problem-solving attributes with math creativity* [Unpublished doctoral dissertation]. St. John's University.

Programme for International Student Assessment. (2022). *PISA 2022 creative thinking framework (draft).* OECD. https:// www.oecd.org/pisa/innovation/creative-thinking/

Plucker, J. A., Beghetto, R. A., & Dow, G. (2004). Why isn't creativity more important to educational psychologist? Potential, pittfalls, and future directions in creativity research. *Educational Psychologists, 39*(2), 83-96. https://doi.org/10.1207/s15326985ep3902_1

Poincarê, H. (1952). *Science and method: Henri Poincarê*. The Modern Library.

Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *The Annals of Statistics, 15*(2), 749-779. https://doi.org/10.1214/aos/1176350373

Polya, G. (1954). *Induction and analogy in mathematics.* Princeton University Press.

Prouse, H. L. (1967). Creativity in school mathematics. *National Council of Teacher of Mathematics, 60*(8), 876-879. https://doi.org/10.5951/MT.60.8.0876

Rips, L. J., & Asmuth, J. (2007). Mathematical induction and induction in mathematics. In A. Feeney, & E. Heit (Eds.), *Inductive reasoning: experimental, developmental and computational approaches* (pp. 248-268). Cambridge University Press.

Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66-75. https://doi.org/10.1080/10400419.2012.652929.

Sak, U. (2014). *Yaratıcılık gelişimi ve geliştirilmesi [Growth and development creativity]*. Vize.

Sak, U., Ayvaz, Ü., Bal-Sezerel, B., & Özdemir, N. N. (2017). Creativity in the domain of mathematics. In J. Kaufman, V. Glăveanu, & J. Baer (Eds.) *The Cambridge handbook of creativity across domains* (pp. 276-298). Cambridge University Press.

Sak, U., & Maker, C. J. (2006). Developmental variation in children's creative mathematical thinking as a function of schooling, age, and knowledge. *Creativity Research Journal, 18*(3), 279-291. https://doi.org/10.1207/s15326934crj1803_5

Sanacore, J., & Palumbo, A. (2009). Understanding the fourth-grade slump: Our point of view. *The Educational Forum, 73*(1), 67-74. https://doi.org/10.1080/00131720802539648

Sarouphim, K. M. (1999). Discovering multiple intelligences through a performance-based assessment: consistency with independent ratings. *Exceptional Children, 65*(2), 151-161. https://doi.org/10.1177/001440299906500201

Sarouphim, K. M. (2001). DISCOVER: Concurrent validity, gender differences, and identification of minority students. *Gifted Child Quarterly, 45*(2), 130-138. https://doi.org/10.1177/001698620104500206

Seddon, G. M. (1983). The measurement and properties of divergent thinking ability as a single compound entity. *Journal of Educational Mesaurement, 20*(4), 393-402. https://doi.org/10.1111/j.1745-3984.1983.tb00216.x

Silver, E. A. (1994). On mathematical problem solving. *For the Learning of Mathematics, 14*(1), 19-28.

Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *ZDM-The International Journal on Mathematical Education, 29*(3), 75-80. https://doi.org/10.1007/s11858-997-0003-x

Silver, E. A., & Cai, J. (1996). An analysis of arithmetic problem posing by middle school students. *Journal for Research in Mathematics Education, 27*(5), 521-539. https://doi.org/10.2307/749846

Simonton, D. K. (1983). Formal education, eminence and dogmatism: The curvilinear relationship. *The Journal of Creative Behavior, 17*(3), 149-162. https://doi.org/10.1002/j.2162-6057.1983.tb00348.x

Singer, F. M., Ellerton, N., & Cai, J. (2013). Problem-posing research in mathematics education: New questions and directions. *Educational Studies in Mathematics, 83*(1), 1-7. https://doi.org/10.1007/s10649-013-9478-2

Singh, B. (1987). The development of tests to measure mathematical creativity. *International Journal of Mathematical Education in Science and Technology, 18*(2), 181-186. https://doi.org/10.1080/0020739870180203

Snyder, A., Mitchell, J., Bossomaier, T., & Pallier, G. (2004) The creativity quotient: An objective scoring of ideational fluency. *Creativity Research Journal, 16*(4), 415-420. https://doi.org/10.1080/10400410409534552

Sternberg, R. J., & Lubart, T. I. (2009). The concept of creativity: Prospects and paradigms. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 3-15). Cambridge University Press.

Stoyanova, E. (1997). *Extending and exploring student's problem solving via problem posing: A study of years 8 and 9 students involved in mathematics challenge and enrichment stages of Euler enrichment for young Australians* [Unpublished doctoral dissertation]. Edith Cowan University.

Tabachnick, B. G., & Fidel, L. S. (2007). *Using multivariate statistics* (5th ed.). Allyn & Bacon Inc.

Tanaka, J. S. (1987). How big is big enough? Sample size and goddness of fit in structural equation models with latent variables. *Child Development, 58*(1), 134-146. https://doi.org/10.2307/1130296

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* American Psychological Association.

Tompkins, G. (1994). *Teaching writing: Balancing process and product.* Macmillan.

Trochim, W. M., & Donnelly, J. P. (2006). *The research methods knowledge base* (3rd ed.). Atomic Dog.

Vale, I., Pimentel, T., & Barbosa, A. (2018). The power of seeing in problem solving and creativity: an issue under discussion. In N. Amado, S. Carreira, & K. Jones (Eds.), *Broadening the scope of research on mathematical problem solving: A focus on technology, creativity and affect* (pp. 243-272). Springer.

Wolf, R. S. (1998). *Proof, logic, and conjecture: The mathematician's toolbox.* W. H. Freeman.

Yıldırım, C. (2000). *Matematiksel düşünme [Mathematical thinking]* (3rd ed). Remzi.

APPENDICES

Appendix 1: Sample Items of the MCT

Item 1 (Squares).

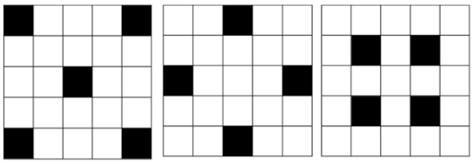


Figure I Figure II Figure III

There are 3 figures consisting of 5 rows and 5 columns above. All figures consist of unit squares. Different problems can be posed by using ***one*** or ***more of these figures***. An example is given below:

Example:* *How much more is the number of white squares in Figure 1 than the number of black squares in Figure 2?

Pose as many and different math problems as you think by using one or more figures. You will gain 1 point for each ***correct problem*** and ***more points*** for ***different correct problems***.

1.Problem.………
……
……

2.Problem.………
……
……

Item 3 (Odd-Even).

Numbers as 1, 3, 5, 7, 9, … are odd numbers.
Numbers as 0, 2, 4, 6, 8, … are even numbers.

△ represents even numbers and □ represents odd numbers. Many different mathematical expressions can be posed by using at least one of these shapes. Two examples are given below:

1.Example of Mathematical Expression	***2.Example of Mathematical Expression***
△ + □ = odd number.	□△ The units digit of two-digit natural numbers is even.
Explanation	***Explanation***
△(2) + □(3) = 5 (odd number)	The units digit of □(1)△(2) is even.

Pose as many and different math expressions that you think always may be true by using mathematical operations (+,-,x,…) as in the first example or not using mathematical operations as in the second example. You will gain 1 point for each ***correct expression*** and ***more points*** for ***different correct expression***.

Warning: *Just write a mathematical expression. Don't write an explanation as in the examples.*

1. Mathematical Expression:

……

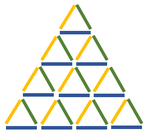
……

2. Mathematical Expression:

……

……

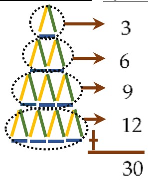
Item 6 (Chez).



In the figure above, there are 30 matchsticks. There are many ways to show that the total number of matchsticks is 30. Two examples are given below:

1.Example Method

1)Worksheet **2)Mathematical**

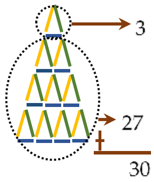


3)Explanation

The number of matchsticks in the triangles in each row is added up.

2.Example Method

1)Worksheet **2)Mathematical**

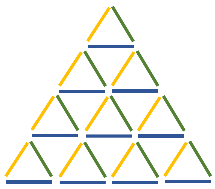


3)Explanation

The matchsticks in the first row and then the other rows are added up.

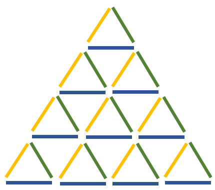
Show that total number of matchsticks is 30 in **numerous** and **different** method. While showing;
1) Draw the mathsticks in the **worksheet**,
2) Do **mathematical operations**,
3) Explain your methods, if necessary.
You will gain 1 point for each ***correct method*** and ***more points*** for ***different correct method***.

1. Method **Mathematical operation**



Explanation:...
....

2. Method **Mathematical operation**



Explanation:...
....

TÜRKÇE GENİŞLETİLMİŞ ÖZET

Geliştirilen Matematiksel Yaratıcılık Testi'nin (MYT) temel çıkış noktası matematiksel düşünmeyi bütüncül bir biçimde ele almak ve bu bağlamda tümevarımsal ve tümdengelimsel düşünmeyi merkeze oturtarak iki düşünme biçimini oluşturan becerileri matematiksel yaratıcılık ölçeğine koymaktır. Tümevarımsal düşünme içinde keşifleri ya da icatları barındırır (Yıldırım, 2000). Matematik alanında yaratıcı bireyin temel becerilerinden birinin de keşif yapabilmek olduğu göz önünde bulundurulduğunda, çoğul düşünme testlerinde bu becerilerin ölçülmesine yönelik bileşenlerin olması gerektiği de düşünülmektedir. Bu nedenle tümevarımsal düşünme yoluyla matematiksel varsayımlar oluşturma becerisi matematiksel yaratıcılığın bir işaretçisi olması açısından araştırılması gereken bir beceri olarak görünmektedir. Diğer taraftan matematiksel düşünmenin bir diğer boyutu olan tümdengelimsel düşünmede ise tümevarımsal düşünme yoluyla ortaya atılan varsayımlar çeşitli kanıtlar sunularak ispatlanır (Nickerson, 2010). Matematiksel bir keşfin belgesi matematiksel kanıtlardır. Kanıtlama aşamasında kullanılan bilgiler, pek çok bilgi parçacığı arasından çözüme giden yolda uygun olanları seçmekle ve bu bilgiyi doğru yerde kullanabilmekle değerli hale dönüşürler. İşte bu süreç yaratıcılıkla ilişkilidir (Poincarê, 1952). Bu nedenle matematiksel yaratıcılığın belirlenmesinde kanıtlama becerisi de önemli bir işaretçi olarak düşünülmektedir. Ancak matematiksel yaratıcılığın ölçümünde bu becerilerin ayrı ayrı ya da bir arada kullanıldığı herhangi bir çoğul düşünme testi ile karşılaşılmamıştır. Bu nedenle MYT temel aldığı kuramsal çerçeve açıcından alana önemli bir katkı sağlayacaktır.

MYT, Nickerson'ın (2010) ortaya koyduğu Matematiksel Düşünme Modeli'nin (MDM) bileşenleri temel alınarak geliştirilmiştir. MYT'nin ölçek geliştirme aşamaları tamamlandıktan sonra nihai hali toplam 3 bileşenli bir yapıdan (problem oluşturma, varsayım oluşturma ve kanıtlama) meydana gelmiştir. Bileşenleri temsil eden maddeler ise akıcılık (fluency), esneklik (flexibility) ve yaratıcılık bölümü (creativity quotient-CQ) olmak üzere 3 farklı yaratıcılık puanına sahiptir. Her bir alt ölçekte (bileşende) 2 adet madde yer almaktadır. MYT, ortaokul 5., 6., 7. ve 8. sınıf öğrencilerinin matematiksel yaratıcılığını ölçmek üzere tasarlanmış kâğıt-kalem ölçüm tekniğine dayalı bir çoğul düşünme testidir. Öğrencilerin test kitapçığında yer alan maddeleri yanıtlamaları yeterlidir. Ölçek bir uygulayıcı denetiminde grup şeklinde veya bireysel şekilde uygulanabilmektedir. Testin uygulanması yaklaşık bir ders saatini almaktadır. Her bir maddeye ayrılan zaman dilimi yaklaşık 7 dakikadır. Uygulayıcı, öğrenciler teste başlamadan önce her bir maddeye eşit süre ayrılması gerektiğini belirtir.

Ölçeğin geliştirilmesi aşamasında araştırma modeli olarak, tarama modeli yaklaşımları arasından kesitsel tarama yaklaşımı kullanılmıştır (Karasar, 2016). Araştırmanın çalışma grubunu belirlemede, seçkisiz olmayan örnekleme yöntemlerinden uygun örnekleme ve amaçsal örnekleme yönteminden yararlanılmıştır (Büyüköztürk vd., 2017). Çalışma, Eskişehir ili merkezinde bulunan Millî Eğitim Bakanlığı'na bağlı altı ortaokulda ve özel yetenekli ortaokul öğrencilerine hafta sonları eğitim sağlayan bir merkezde öğrenim gören 5., 6., 7. ve 8. sınıf düzeyindeki öğrenciyle gerçekleştirilmiştir. Araştırmanın ön deneme, pilot uygulama ve asıl uygulamalarını kapsayan saha uygulamasında toplam 1129 (540 kız, 589 erkek) katılımcıya ulaşılmıştır.

Ölçeğin yapı geçerliğini incelemek amacıyla pilot uygulamadan elde edilen verilerle Açımlayıcı Faktör Analizi (AFA), esas uygulamadan elde edilen verilerle ise Doğrulayıcı Faktör Analizi (DFA) yapılmıştır. Ölçeğin açıkladığı toplam varyansın %63.36 olduğu ve ölçekteki toplam 6 maddenin faktör yük değerlerinin ise $\lambda_{(min.)}$=.46 ile $\lambda_{(max.)}$=.98 arasında değerler aldığı görülmüştür. Ayrıca DFA sonunda elde edilen uyum iyiliği istatistiklerinin ($\chi 2$ /sd=1.075, RMSEA =.009, RMR=.042, SRMR=.011, NFI=.998, NNFI=1.00, CFI=1.00, GFI=.997, AGFI=.991) kabul düzeylerinin mükemmel uyum sınırları içinde olduğu bulunmuştur. MYT'nin ölçüt geçerliğini test etmek için ilk olarak sınıflararası ANOVA testi yapılmıştır. Analiz sonuçları, MYT'nin alt ölçeklerine ait puanların ve ölçeğin tamamından elde edilen toplam puanların akıcılık, esneklik ve yaratıcılık bölümü puan türü ortalamalarının sınıflar arasında anlamlı farklılıklar yarattığını ortaya koymuştur. MYT'nin ölçüt geçerliğini test etmek için ikinci olarak matematiksel yaratıcılığın matematik başarısı ile uyumu incelenmiştir. Matematik dersine ait karne notları ile MYT puanları arasındaki ilişkinin yüksek olduğu bulgulanmıştır ($r_{top.\ akı.}$=.504, $r_{top.\ esn.}$=.554, $r_{top.\ y.\ böl.}$=.550; p<.001). Ölçeğin iç tutarlık güvenirliğini ortaya koymak için yapılan güvenirlik analizlerinde, ölçeğin tamamından elde edilen akıcılık, esneklik ve yaratıcılık bölümü puanları ile ilişkili Cronbach Alpha güvenirlik değerlerinin ($\alpha_{(akı.)}$=.831; $\alpha_{(esn.)}$=.780; $\alpha_{(y.böl.)}$=.852) ideal sınırlar içerisinde yer aldığı görülmüştür. MÜT'ün iç tutarlık güvenirlik değerini ortaya koymak için ikinci olarak maddeler arası korelasyon analizleri gerçekleştirilmiştir. Maddelerin üç puan türündeki korelasyon katsayılarının $r_{min.}$=.282 ile $r_{max.}$=.675 arasında değiştiği görülmüştür. MYT'nin okuyucular arası güvenirliği için yapılan analizde ise ölçekte yer alan 6 madde ve ölçeğin tamamından elde edilen akıcılık, esneklik ve yaratıcılık bölümü puanlarında okuyucular arası sınıf içi korelasyon değerlerinin yüksek olduğu (.954 ve .966 arasında değişmektedir) görülmüştür.

Ölçeğin bahsedilen psikometrik özellikleri dikkate alındığında alana özgü yaratıcılığı belirlemeye yönelik geliştirilen MYT'nin öğrencilerin yaratıcılık düzeylerini doğru şekilde belirleyebildiği görülmektedir. Bunun yanında alanyazında yer alan matematiksel yaratıcılık ölçeklerinin genellikle matematiksel modelden değil, belli matematiksel becerilerden beslendiği görülmektedir (Sak vd. 2017). MYT ölçeği matematiksel yaratıcılıkla doğrudan ilişkili olduğu düşünülen MDM modelinden beslenerek çok daha bütüncül bir yapı sunmaktadır. Modelin içinde var olan varsayım oluşturma ve kanıtlama bileşenleri ise matematiksel yaratıcılığın ölçülmesinde ilk defa kullanılan becerilerdir. Matematikte varsayımlar oluşturmak ve bunları kanıtlamak matematiğin doğasındaki yaratımla doğrudan ilişkilidir. Ölçekte var olan bu iki becerinin kullanımı alana bir yenilik getirmektedir.

Son yıllarda araştırma eğilimi genel yaratıcılık yerine alana özgü yaratıcılık olarak görülmektedir (Programme for International Student Assessment [PISA], 2021). Aynı doğrultuda, bu çalışmada alana özgü yaratıcılığı ortaya koyabilmek için bir ölçek geliştirilmiştir. Bunun yanında alanyazında verili matematiksel yaratıcılık ölçeklerinin genellikle matematiksel modelden değil, belli matematiksel becerilerden beslendiği görülmektedir. MYT ölçeği matematiksel yaratıcılıkla doğrudan ilişkili olduğu düşünülen MDM modelinden beslenerek çok daha bütüncül bir yapı sunmaktadır. Modelin içinde var olan varsayım oluşturma ve kanıtlama bileşenleri ise matematiksel yaratıcılığın ölçülmesinde ilk defa kullanılan becerilerdir. Matematikte varsayımlar oluşturmak ve bunları kanıtlamak matematiğin doğasındaki yaratımla doğrudan ilişkilidir. Ölçekte var olan bu iki becerinin kullanımı alana bir yenilik getirmektedir.