EJAES

JOURNAL OF ADVANCED EDUCATION STUDIES

İleri Eğitim Çalışmaları Dergisi

7(1): 20-44, 2025

THE MATHEMATICAL CREATIVITY MEASUREMENT TOOL: THE CASE OF BAYESIAN EXPLORATORY AND CONFIRMATORY FACTOR ANALYSIS

Özgül YAYLA[1]
Dilara BAKAN KALAYCIOĞLU[2]

Geliş Tarihi/Received: 25.02.2025 Elektronik Yayın / Online Published: 25.06.2025
DOI: 10.48166/ejaes.1646254

ABSTRACT
This study aimed to develop a measurement tool to assess the mathematical creativity of high school students. The study group consisted of 500 students enrolled in high schools located in the central districts of Ankara province during the 2022–2023 academic year. A survey design was employed for the research. To measure students' mathematical creativity, a mathematical creativity test was administered under the supervision of the researcher, with a 40-minute time limit. Test scoring was also conducted by the researcher. To examine the construct validity of the scores obtained from the measurement tool, Bayesian exploratory and confirmatory factor analyses were performed. The findings indicated that, in the first-order confirmatory factor analysis, the measurement tool could be defined by five factors named generating equality, generating relationship, generating equation, creating equal area, and generating triangle. The second-order confirmatory factor analysis defined it as a single factor called mathematical creativity. The test's internal consistency was examined by calculating the Cronbach's alpha (0,650), stratified Cronbach's alpha (0,598), and McDonald's omega coefficients (0,890). It is demonstrated by findings that the psychometric features of the mathematical creativity measurement tool were sufficient.
Keywords: Bayesian explatory and confirmatory factor analysis, reliability, validity

[1] Teacher, Gazi University, Institute of Educational Sciences, Ankara, Turkey, e-mail: ozgul_06yayla@hotmail.com, ORCID:0000-0002-2013-5178

[2] Prof. Dr., Gazi University, Faculty of Education, Measurement and evaluation in education, Ankara, Turkey, e-mail: dilarabakan@gmail.com, ORCID: 0000-0003-1447-6918

ÖZET

Bu araştırmanın amacı, lise öğrencilerinin matematiksel yaratıcılıklarını ölçmeye yönelik bir ölçme aracı geliştirmektir. Araştırmanın çalışma grubunu 2022-2023 eğitim-öğretim yılında Ankara ilinin merkez ilçelerinde öğrenim göre toplam 500 lise öğrencisi oluşturmaktadır. Araştırmada tarama modeli kullanılmıştır. Öğrencilere matematiksel yaratıcılıklarını belirlemek amacıyla matematiksel yaratıcılık testi 40 dakikalık süre verilerek araştırmacı gözetiminde uygulanmıştır. Matematiksel yaratıcılık testi için puanlama araştırmacı tarafından yapılmıştır. Ölçme aracından elde edilen puanların yapı geçerliğinin incelenmesi amacıyla bir Bayes açımlayıcı ve doğrulayıcı faktör analizi gerçekleştirilmiştir. Bulgular; ölçme aracının birinci düzey doğrulayıcı faktör analizinde eşitlik üretme, ilişki üretme, denklem üretme, eş alan oluşturma ve üçgen üretme olarak adlandırılan beş faktörle, ikinci düzey doğrulayıcı faktör analizinde ise matematiksel yaratıcılık olarak adlandırılan faktörle tanımlanabildiğini göstermiştir. Testin iç tutarlılığı ise, Cronbach alfa (0,650), tabakalı Cronbach alfa (0,598) ve McDonald omega (0,890) katsayıları hesaplanarak incelenmiştir. Geliştirilen matematiksel yaratıcılık ölçme aracının psikometrik özelliklerinin yeterli olduğu bulgularla ortaya konulmuştur.

Anahtar Kelimeler: Bayes açımlayıcı ve doğrulayıcı faktör analizi, güvenirlik, geçerlik

1. INTRODUCTION

In a rapidly changing world where technological and scientific developments change social networks and individuals' lives, creativity is needed to adapt to change and sustain these developments. Creativity, which exists in every individual but is open to development with experiences and learning experiences, can be expressed as the ability of an individual to solve a problem by establishing a connection between what they have learned as a result of their learning experience and to put forward a new, original thought or product by using these connections (Güleryüz, 2000). The necessity for creative individuals and the development of reliable tools for measuring and evaluating creativity is increasingly recognized.

However, defining and assessing creativity remains a complex challenge (Runco, 1993). The literature identifies various types of creativity across domains, including linguistic, scientific, and mathematical creativity (Kanlı, 2019). In this context, it is essential to determine the specific fields in which creativity manifests, develop domain-specific definitions, and research to enhance creativity within these fields. Among these, mathematical creativity is a key area of focus in domain-specific creativity. Given the interconnectedness of mathematics with other disciplines and cognitive processes, fostering mathematical creativity can significantly contribute to success in mathematics and other fields. Therefore, understanding the concept of mathematical creativity and recognizing its importance is crucial. This understanding should guide efforts to assess individuals' levels of mathematical creativity and implement strategies to develop their creative potential (Türkan, 2010).

Tall (2002) defines mathematical creativity as the ability to develop thinking in problem-solving or structures, considering the unique logical-deductive nature of mathematics and the suitability of the concepts created to integrate with the essence of what is important in mathematics. Chamberlin and Moon (2005) define mathematical creativity as the ability to use mathematics-specific thinking processes to solve non-routine problems and generate original solutions to simulated or real applied problems using mathematical modelling.

Mathematical creativity is related to the ability of individuals to generate solutions to the problems they face not only with known methods but also with original and innovative approaches. Individuals with mathematical creativity have the potential to contribute to social progress in fields such as science, engineering, and technology with their analytical thinking, problem-solving solving, and linking abstract concepts. At this point, measuring mathematical creativity with reliable methods can help identify individuals with high creative thinking potential and help them receive the education that will reveal their potential. Therefore, measuring mathematical creativity with reliable and valid methods is essential.

The three dimensions used in the literature to measure mathematical creativity and their explanations are as follows:

- **Fluency:** It is defined as the ability to come up with many ideas that can answer a problem. For instance, writing different mathematical prossesses whose results are five. Creative individuals can generate a large number of ideas for solving a problem.
- **Flexibility:** It is defined as the ability to approach a problem from different perspectives, see various situations, generate ideas in different categories, and approach an event from different perspectives. The more ideas are put forward to examine the problem from different angles, the higher the flexibility. Creative individuals show solutions to the problem from different angles.
- **Originality:** It refers to being unique in thought and action. The fewer individuals come up with an idea, the more original it is considered. Creative individuals create original thoughts (Atasoy, Kadayıfçı, & Akkuş, 2007).

In the literature, most studies on measuring mathematical creativity focus on primary (Kattou, Kontoyianni, Pitta-Pantazi, & Christou, 2011; Schoevers, Kroosbergen, & Kattou, 2018) and secondary school students (Akgül & Kahveci, 2016; Balka, 1974; Evans, 1964; Kartono & Rusilowati, 2019; Mann, 2005; Prouse, 1964; Sahliawati & Nurlaelah, 2020; Sezerel, 2019; Siswono, 2011; Taşkın, 2016; Türkan, 2010; Zainudin, Subali, & Jailani, 2019). Since the studies above are at the level of primary and secondary school, the items used in the measurement tools are limited to basic mathematical knowledge. Some of the studies on mathematical creativity were conducted to determine the variables predicting mathematical creativity (Acar, Tanıdık, Uysal, Myers, & Inetas, 2022; Alkan, 2014; Suherman & Vidakovich, 2024; Tyagi, 2015). While research on creativity skills in mathematics exists, studies specifically examining the mathematical creativity of high school students remain limited.

Mettler (1976) adapted two items from Evans' 1964 test for seventh-grade students and applied them to ninth-grade algebra students. Similarly, Leikin and Lev (2007) used a two-item instrument with limited multiple solutions to assess the mathematical creativity of tenth and eleventh-grade high school students. However, neither study was analyzed to determine the factor structure of the obtained scores. Unlike Leikin and Lev's study, the present research includes items with unlimited response possibilities. Factor analyses were conducted using the Bayesian method to ensure the validity of the obtained scores. This study aims to develop a valid and reliable measurement tool to assess high school students' mathematical creativity.

2. METHOD

The survey model was used in the study. Survey models are research approaches that aim to describe the past or present situation as it exists. The event, person, or object that is the subject of the survey is tried to be defined as it is within its conditions (Kuzu, 2013).

2.1. Study Group

The pilot scheme of the study was carried out with 311 students. Descriptive statistics of the pilot scheme are given in Table 1.

Table 1. Descriptive Statistics of The Pilot Scheme

Student admission status	Number of students taking form A	Number of students taking form B	Number of students taking form C	Total
Without exam	24	25	22	71
With exam	30	29	29	88
Without exam	18	17	20	55
With exam	32	33	32	97
Total	104	104	103	311

The study group included 547 ninth-grade students from Ankara. However, the booklets of 47 participants who did not sign the consent form or left it incomplete were excluded from the analysis.

Purposive sampling was used to select the sample for the study group. Purposive sampling is also called knowledge-based sampling. The researcher selects a sample based on the experience or knowledge of the group to be sampled (Airasian, 2000). Descriptive statistics for the study group are given in Table 2.

Table 2. Descriptive Statistics of The Study Group

Variable	Category	*f*	%
Gender	Female	251	50.2
	Male	248	49.6
High school type	Anatolian high school	300	60
	Science high school	200	40
Status of being a BİLSEM student	Yes	39	7.8
	No	461	92.2
	Total	500	100

As seen in Table 2, 50.2% of the students are female and 49.6% are male. Sixty percent of the students are studying at anatolian and 40% in science high schools. Of these students, 7.8% are continuing their education in BİLSEM centres.

2.2. Data Collection Tool

First, the Mathematical Creativity Test (MCT) was developed to determine the mathematical creativity of high school students. While some of the items were adapted from the items created by researchers aiming to measure mathematical creativity in the literature (Akar & Karaduman, 2021; Evans, 1964; Haylock, 1984, 1987; Kanlı, 2019; Mann, 2005; Price, 2006; Prouse, 1964), some of them were developed by the researcher.

In order to develop the measurement tool for data collection, 32 open-ended items were initially prepared and reviewed by three experts. The experts included a mathematics teacher with a master's degree and 24 years of experience, a research assistant with a doctorate in measurement and evaluation, and a mathematics teacher with nine years of experience who is currently pursuing a master's degree. Based on their feedback, eight items were eliminated to form the final tool. The remaining 24 items were distributed into Forms A, B, and C, ensuring that each form contained eight items and that the mathematical content was balanced across the forms.

These forms were used for the pilot study. The pilot study was conducted in high schools that admit students with and without entrance exams under the researcher's supervision. Students were given 40 minutes to complete the paper-and-pencil test. Form A was administered to 104 students, Form B to 104 students, and Form C to 103 students. It was observed that students struggled to complete all eight items within the allotted time.

After the pilot study, responses were evaluated based on three dimensions: fluency, flexibility, and originality. Fluency was scored by awarding one point for each correct answer. Flexibility was scored by assigning one point for each correct answer in different categories. Originality was assessed based on the rarity of responses: six points were awarded if the answer was given by fewer than 5% of participants, three points if given by 5–10%, and zero points if given by more than 10%. These raw scores were calculated for each dimension. The questions are presented in the appendix.

In all items, it was found that students from high schools accepting students with an exam performed better in fluency, flexibility, and originality than those from high schools without an exam. In the pilot application, the average percentage of unanswered items was significantly higher in non-selective schools (64.25%) compared to selective schools (34.47%). Based on these findings, it was decided that the final implementation would be conducted only in high schools with an exam and Science and Arts Centers (Bilim ve Sanat Merkezi, BİLSEM).

For the final test, six open-ended items were selected to ensure content validity: two from numbers, two from algebra, and two from geometry. These items were chosen based on the lowest percentage of unanswered responses and the highest scores achieved by students in the pilot study. The test was administered to 547 students in total contained the students of Anatolian and Science high schools that admit students through entrance exams, and the students of BİLSEM, with a 40-minute completion time under the researcher's supervision.

2.3. Data Analysis

The item difficulty index refers to the proportion of correct responses to an item. This index is denoted by "p" ranges between 0 and 1. The closer the index is to 0, the more difficult the item ; the closer it is to 1, the easier the item is. For open-ended items, the difficullty index is calculated by taking the difference between the lowest score assigned to the item and the average score received and then dividing this difference by the range obtained by subtracting the lowest score from the highest score assigned to the item (Karakaya, 2022).

The item discrimination index is an indicator that provides insightin to how well an item distinguishes betweet individuals with the high and low scores in relation to the measured construct. When calculating the idiscrimination indices for open-ended items, the following procedure is followed: The scores obtained from the responses to the relevant item are ranked from highest to lowest. The top 27% of scorers form the upper group, and the bottom 27% form the lower group. The difference between the average scores of the upper and the lower groups is then calculated and divided by the range between the highest and lowest scores received for the item (Karakaya, 2022).

Each of the six open-ended items constituting the MCT was scored based on fluency, flexibility, and originality dimensions. The scores for the first item were labeled "fl1" for fluency, "fx1" for flexibility, and "or1" for originality, with the same coding applied to the other items. This resulted in a dataset with 18 scores. However, in the factor analysis, since it was determined that the fluency, flexibility, and originality scores of an item in the MCT prevented factorization, the scores of the relevant item were removed, and since the fluency and flexibility scores of the fourth and fifth items were equal due to their structure (fl4=fx4 and fl5=fx5), fx4 and fx5 scores were not included in the factor analysis. Therefore, the factor analysis was conducted over 13 points.

The fluency, flexibility, and originality scores for each item were summed to calculate total creativity scores on an item basis. Before conducting the factor analysis, assumptions such as outliers, missing data, multicollinearity, linearity, normality, and sample adequacy were checked. To identify univariate and multivariate outliers, z-scores and Mahalanobis distance D^2 values were used (Şen, 2023). A total of 95 outliers were identified, having z-scores outside the ±3 range and significant D^2 values, and were removed from the dataset, leaving 405 participants. Missing data was checked, and no missing data was found.

Multicollinearity was assessed by checking for correlations above 0.90 (Çokluk, Şekercioğlu, & Büyüköztürk, 2012). Some scores showed correlations exceeding 0.90, which was expected since the MCT includes three different scores per item. The linearity assumption was evaluated through a scatter plot matrix, where ellipses would indicate linearity (Çokluk, Şekercioğlu, & Büyüköztürk, 2012). The scatter diagram matrix was examined, and it was determined that very few of the distributions had ellipse images. However, it is expected that the scores obtained from the items in MCT differ from each other; therefore, there is no linearity. For instance, a student who scores high on the fluency of an item may score low on the originality of the same item, or a student who scores high on the flexibility of an item may score very low on the flexibility of another item. Since the Bayesian method does not require a normality assumption (Alkış, 2016), it was not checked. The adequacy of the sample was determined using the Kaiser-Meyer-Olkin (KMO) statistic, where a value between 0.5 and 0.7 is considered sufficient (Can, 2016; Şen, 2023). The KMO value for this study was 0.637, confirming that the sampling adequacy assumption was met. Bartlett's test of sphericity results were calculated (χ^2 = 6031.811, df = 120, and $p < 0.05$).

Fabrigar, Wegener, MacCallum, and Strahan (1999) recommend that when the sample size is large enough, the data should be randomly split into two halves: one half for Exploratory Factor Analysis (EFA) and the other for Confirmatory Factor Analysis (CFA). In this study, the dataset of 405 participants, after removing outliers, was randomly divided into two groups by SPSS: one group with 211 participants and another with 194 participants. The first data set with 211 participants was subjected to Bayesian EFA. If a theoretical structure requires a relationship between factors, it is recommended to use oblique rotation methods (Can, 2016; Koğar & Koğar, 2023). Therefore, the GEOMIN oblique rotation method was applied during the Bayesian EFA. For the second half, with 194 participants, first and second-level CFA was conducted using the Bayesian estimation method. CFA with the Bayesian estimation method allows parameter estimation based on a priori information through a large number of iterations and produces much more realistic results compared to traditional CFA methods (Depaoli, 2021).

In general, in model-data fit, a Posterior Predictive P (*PPP*) value equal to 0.50 is an indicator of an excellent fit, while a value greater than 0.05 is an indicator of an acceptable fit (Asparouhov & Muthén, 2021). In addition, in a perfect-fit model, the 95% confidence interval (credible-interval-CI) calculated for the difference in χ^2 values should contain a value of zero (Wang & Wang, 2020). The Deviance Information Criterion (DIC) and Bayes Factor (BF) are used to decide on the final solution by comparing different models. A model with a smaller DIC value indicates a better model-data fit than a larger DIC value (Asparouhov, Muthén & Morin, 2015). The Bayes factor is based on the comparison of two different hypotheses: H_0 and H_1. Accordingly, a BF value greater than 100 demonstrates that (H_1) is supported (Jeffreys, 1961; as cited in Koğar & Koğar, 2023).

To evaluate the reliability of the scores obtained from the MCT, Cronbach's alpha, stratified Cronbach's alpha, and McDonald's omega coefficients were calculated. Cronbach's alpha assumes that the measurement tool is unidimensional and composed of parallel items, characterized by equal means, standard deviations, covariances, and factor loadings (Uysal & Sarıça, 2018). A high Cronbach's alpha coefficient supports the assumption of unidimensionality; however, a low coefficient should not be automatically interpreted as an indication of low reliability, particularly in cases where the measurement tool encompasses multiple constructs (Başol, 2015; Şencan, 2005). In practice, achieving parallelism among items is often challenging (Tekindal, 2017), and Cronbach's alpha, as a measure of internal consistency, may be insufficient for assessing the overall reliability of multidimensional instruments. Considering that the items within the MCT exhibit varying means, standard deviations, and covariances, the scale is more appropriately classified as a congeneric measurement model (Yurdugül, 2006). In congeneric measurements, Cronbach's alpha is known to underestimate the true reliability of the scores (Uysal & Sarıça, 2018). Therefore, McDonald's omega coefficient was also computed, specifically designed for congeneric measurements and estimates reliability as the ratio of true score variance to total variance (McDonald, 1985; 1999). Furthermore, stratified Cronbach's alpha was calculated to provide a more accurate reliability estimate in the context of multidimensional congeneric measurement structures (Soysal, 2023). The use of these complementary reliability coefficients offers a more robust and comprehensive evaluation of the reliability of the MCT scores.

Item analyses were conducted using IBM SPSS 25.0 software, reliability of the scores using IBM SPSS 25.0, and R psych package (Revelle, 2025), and EFA and CFA analyses using the Bayesian method using Mplus 7.0 (Muthén & Muthén, 2012).

3. FINDINGS

3.1. Item Analysis

The raw scores for fluency, flexibility, and originality for each item were summed to calculate the overall creativity score for each item (e.g., the creativity score for item 1 was calculated as fl1 + fx1 + or1). Item analyses were conducted according to classical test theory (CTT). These item-based creativity scores and item analyses are presented in Table 3.

Table 3. Item Mean, Standard Deviation, Minimum, Maximum, Difficulty, and Discrimination Values

Item no	$\overline{X}$	Ss	Minimum	Maximum	p	disc
1	14.44	12.98	0	89	0.18	0.34
2	10.62	10.58	0	64	0.21	0.42
3	21.70	16.38	0	159	0.13	0.21
4	7.98	10.56	0	79	0.11	0.29
5	7.19	15.57	0	95	0.08	0.28
6	7.35	16.68	0	161	0.05	0.18

As shown in Table 3, the mean scores range from 7.19 to 21.70, the standard deviations range from 10.56 to 16.68, the item difficulty indices range from 0.05 to 0.21, and the item discrimination indices range from 0.18 to 0.42. These results suggest that the items were generally challenging for the students. When the item difficulty indices are analyzed, it is seen that the most difficult item for the students is item 6. When the discrimination index of this item is analyzed, it is seen that item six is less discriminative than the other items.

3.2. Findings Regarding Construct Validity

In the dataset of 211 participants, multiple analyses were conducted to identify the most suitable factor structure. It was found that the data could be factorized after removing the three scores related to item 6 (fl6, fx6, or6). The only item in the booklet in which no sample solution was given to students was item six. Students may not have received sufficient points because they had difficulty answering this item. For this reason, item six may not have worked in the factor analysis. The results of the Bayesian EFA are presented in Table 4.

Table 4. Bayesian EFA Results

Number of factors	Posterior Predictive P (*PPP*)	95% Confidence Interval between observed and replicated chi-square values difference (CI)		Deviance (DIC)	BF
1	0.000	1265.60	1340.71	14059.21	
2	0.000	730.67	801.25	-588924.84	>1000000
3	0.000	332.68	407.18	-596665.04	0.00
4	0.000	118.08	194.31	-600579.38	0.00
5	0.106	-15.46	63.87	-603107.45	>1000000

When Table 4 is examined, the model-data fit of the five-factor solution is at an acceptable level ($PPP = 0.106 > 0.05$), and the CI contains zero value (-15.464 63.874), which indicates a perfect fit. In the five-factor solution, the DIC value is -603107.457. This DIC value is better as it is smaller than the DIC value of one, two, three, and four-factor solutions. Moreover, the BF value is calculated as > 1000000 as a result of the comparison of the four-factor model (H_0) and the five-factor model (H_1), and this value is greater than 100, which is the evidence that H_1 is supported. In other words, the five-factor solution is better. Since it was determined that the five-factor structure was the best solution according to the Bayesian EFA result in the measurement tool, the factor loadings, Eigenvalues, and residual variances for only the five-factor solution are as in Table 5.

Table 5. Bayesian EFA Results of the Five-Factor Structure

Variables	Factors					Residual variances
	1	2	3	4	5	
fl1	.861*					.252
fx1	.992*					.010
or1	.663*					.541
fl2		.914*				.162
fx2		.997*				.006
or2		.730*				.435
fl3			.365*			.829
fx3			.958*			.061
or3			.659*			.548
fl4				.850*		.252
or4				.945*		.103
fl5					.994*	.007
or5					.996*	.007
Eigenvalues	2.688	2.482	2.265	1.660	1.451	

*at 0.05 significance level

As shown in Table 5, the MCT consists of five factors. The variables fl1, fx1, and or1 for the first factor, fl2, fx2, and or2 for the second factor, and so on, all have statistically significant factor loadings. Additionally, all factor loadings, except for fl3, are greater than 0.40.

The residual variance values for each variable vary between 0.006 and 0.829. The residual variance values of variables that are well differentiated between factors and have high factor loadings are also low (Koğar & Koğar, 2023). As a result, Bayesian EFA in MCT, the best solution was reached in a five-factor structure. The first factor was called "generating equality (gey)", the second factor was as "generating relationship (gr)", the third factor as "generating equation (gen)", the fourth factor as "generate equal area (gea)" and the fifth factor as "generating triangle (gt)". Factor loadings ranged between 0.365 and 0.997.

In the next stage of the study, the results obtained from Bayesian EFA were tested with first- and second-order Bayesian CFA to examine the model's construct validity. The *PPP* value calculated for the holistic view of the first-order factor structure was found to be 0.211, which indicates that the model is at an acceptable level. The first-order Bayesian CFA results are presented in Table 6.

Table 6. First Order Bayesian CFA Results

Generating equality	Factor loading %95 GA [Lower-Upper]	Generation relations	Factor loading %95 GA [Lower-Upper]	Generation equation	Factor loading %95 GA [Lower-Upper]	Generate equal area	Factor loading %95 GA [Lower-Upper]	Generating triangles	Factor loading %95 GA [Lower-Upper]
fl1	.87 [.82-.91]	fl2	.91 [.87-.95]	fl3	.31 [.19-.43]	fl4	.95 [.83-.1.0]	fl5	1.00 [.99-1.0]
fx1	.98 [.94-1.0]	fx2	.98 [.94-1.0]	fx3	.97 [.88-1.0]	or4	.88 [.80-.99]	or5	1.00 [.99-1.0]
or1	.87 [.54-.72]	or2	.59 [.48-.68]	or3	.70 [.61-.78]				

When Table 6 is examined, it is seen that the factor loadings (standardized Beta coefficients) ranged between 0.87 and 0.98 for the items constituting the equality generation factor; between 0.59 and 0.98 for the items constituting the relationship generation factor; between 0.31 and 0.97 for the items constituting the equation generation factor; between 0.88 and 0.95 for the items constituting the generate equal area factor; and 1.00 for the items constituting the triangle generation dimension ($p < 0.001$). The diagram of the first-order Bayesian CFA is presented in Figure 1.

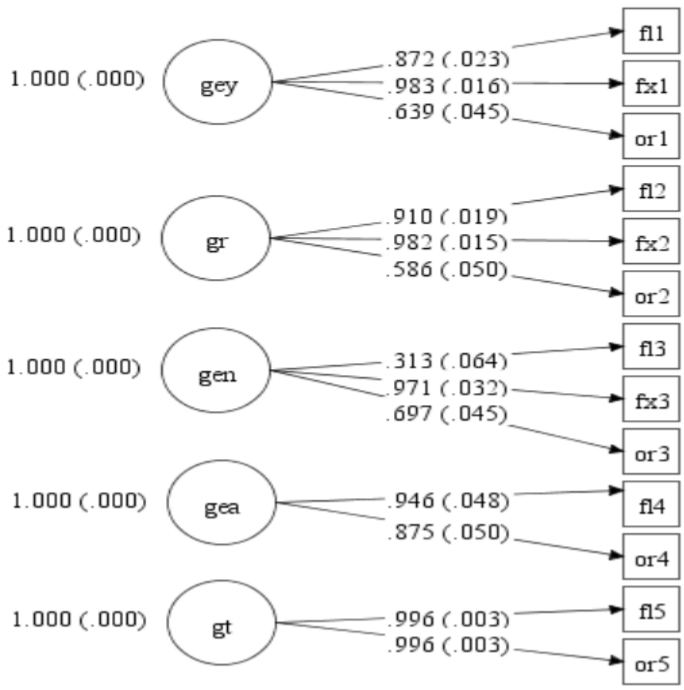


Figure 1. First-order Bayesian CFA diagram

Note: gey: generating equality; gr: generating relations; gen: generating equation; gea: generating equal area; gt: generating triangle

The *PPP* value calculated for the holistic view of the second-order Bayesian CFA factor structure was found to be 0.218, which indicates that the model is at an acceptable level. The second-order Bayesian CFA results are given in Table 7.

Table 7. Second-Order Bayesian CFA Results

Generating equality	Factor loading %95 GA [Lower-Upper]	Generation relations	Factor loading %95 GA [Lower-Upper]	Generation equation	Factor loading %95 GA [Lower-Upper]	Generate equal area	Factor loading %95 GA [Lower-Upper]	Generating triangles	Factor loading %95 GA [Lower-Upper]
fl1	.86 [.81-.91]	fl2	.91 [.87-.91]	fl3	.25 [.11-.39]	fl4	.95 [.83-.1.0]	fl5	1.00 [.99-1.0]
fx1	.98 [.94-1.0]	fx2	.98 [.94-.1.0]	fx3	.97 [.87-.1.0]	or4	.87 [.80-.99]	or5	1.00 [.99-1.0]
or1	.63 [.53-.71]	or2	.58 [.47-.67]	or3	.69 [.60-.78]				

As can be seen in Table 7, the factor loadings (standardized Beta coefficients) range between 0.63 and 0.98 for the items constituting the equality generation factor, between 0.58 and 0.98 for the items constituting the relationship generation factor, between 0.25 and 0.97 for the items constituting the equation generation factor, between 0.87 and 0.95 for the items constituting the generate equal area factor, and 1.00 for the items constituting the triangle generation dimensions ($p < 0.001$). The diagram of the second-level Bayesian CFA is shown in Figure 2.

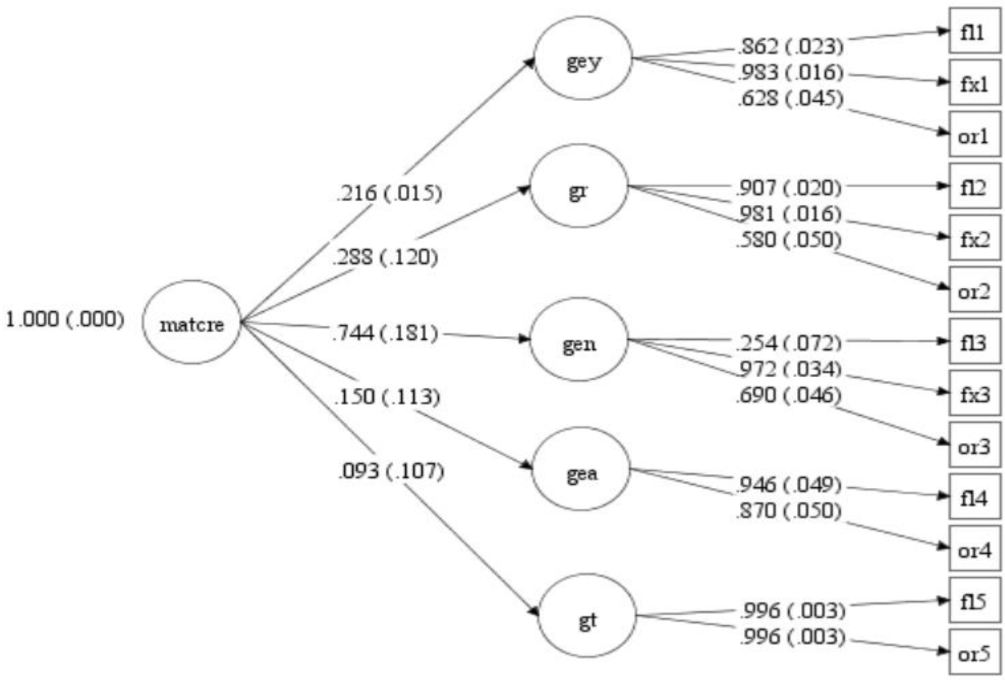


Figure 2. The second-order Bayesian CFA

Note: matcre: mathematical creativity; gey: generating equality; gr: generating relations; gen: generating equation; gea: generating equal area; gt: generating triangle

Figure 2 shows that the first-order factors are predicted by the second-order factors, with values ranging from 0.09 to 0.74.

3.3. Findings Regarding Internal Consistency Reliability

The analyses were conducted using a data set of 405 participants with 13 scores for internal consistency reliability. The stratified Cronbach's alpha coefficient was found to be 0.598. Given that the mathematical creativity items are measured on different scales, the standardized Cronbach's Alpha was computed, yielding a value of 0.650. Tavakol and Dennick (2011) state that a the acceptable values of alpha, ranging from 0.70 to 0.95, and explains that a low value of alpha could be due to a low number of questions, or heterogeneous constructs. In addition to Cronbach's alpha coefficient, McDonald's omega coefficient was also calculated and found to be 0.890. McDonald's omega coefficient above 0.70 indicates that the measurement tool has a good level of reliability (Demir, Bektaş, Demir & Bektaş, 2020). Accoding McDonald's omega coefficient it can be concluded that the MCT has a good level of reliability.

4. DISCUSSION, CONCLUSION, AND RECOMMENDATIONS

This study was conducted to develop an instrument to measure high school students' mathematical creativity skills. When the descriptive statistics of the items were analyzed, it was seen that the highest average of the total creativity scores obtained from the items was in equation generation, and the lowest average was in triangle generation. It can be argued that students showed the best performance in producing equations and the lowest performance in producing triangles. The highest performance was in algebra, and the lowest performance was in geometry.

It was determined that the developed measurement tool has a five-factor structure. This result differs from the results of Balka (1974) and Sezerel (2019). Balka (1974) found that the six-item measurement tool the researcher developed to determine creative ability in mathematics had a two-factor structure. Sezerel (2019) found that the scale had a three-factor structure in which the researcher developed a six-item mathematical productivity test to measure the creativity of middle school students.

According to Bayesian factor analyses, a five-factor structure was observed in the EFA, while the first and second-level CFA results supported the model-data fit of the five-factor structure.

In the study, the stratified Cronbach's alpha reliability coefficient of MCT was found to be 0.598, the Cronbach's alpha reliability coefficient was found to be 0.650, and the McDonald's omega reliability coefficient was found to be 0.890. Akgül and Kahveci (2016) found the reliability coefficient of the scale they developed to measure the mathematical creativity of middle school students to be 0.80, and Türkan (2010) found the reliability coefficient of the test developed to measure the mathematical creativity of sixth, seventh, and eighth-grade students as 0.78. Similarly, Balka (1974) found the reliability of the mathematical creativity measurement tool to be as 0.72. CFA was not reported in these studies. The low number of items in the MCT may have caused Cronbach's alpha coefficients to be low.

This study was conducted in high schools and BİLSEM centers in the central districts of Ankara province. For this reason, the findings to be obtained from samples including a large number of students studying in different provinces of Turkiye may reveal more comprehensive findings regarding the validity and reliability of the MCT. A similar study can be conducted to include high schools that accept students without a central exam. In addition, in future studies, defining the relationship between the concept of "mathematical creativity" and factors such as mathematics achievement and mathematics may contribute to expanding the theoretical framework of mathematical creativity.

Another limitation of the study is that the items in the MCT cover the number, geometry, and algebra sub-learning areas of mathematics. The study was administered in 40 minutes, which is the duration of one lesson. Since the items were open-ended and did not have a limited number of solutions, it was observed that the students needed a long time to answer. A similar study can be conducted by preparing a measurement tool that includes more items covering different sub-learning areas of mathematics, such as data, counting, and probability, and taking into account the test duration accordingly.

Another limitation of the study is that rater errors could not be calculated, because the scoring was done by a single rater. A similar study can be conducted with more than one rater doing the scoring.

MCT does not provide insights into students' general abilities or creativity beyond their mathematical creativity. Therefore, the findings of this study should be interpreted with caution in future research.

Despite the summarized limitations, it is possible to say that the MCT is a valid, reliable, and useful test that can be used to assess the mathematical creativity of high school students.

Acknowledgement

The authors would like to thank Gazi University Academic Writing Application and Research Center for proofreading the article.

REFERENCES

Acar, S., Tanıdık, H., Uysal, R., Myers, D., & Inetas, B. (2022). Socio-economic status and creativity: A meta-analysis. *The Journal of Creative Behavior, 57*(1), 138-172.

Airasian, P. (2000). *Educational research*. USA: Prentice Hall.

Akar, Ş., Ş. & Karaduman, G., B. (2021). *Özel yeteneklilerde matematik öğretimi ve matematiksel yaratıcılığın desteklenmesi.* Ankara: Pegem Akademi.

Akgül, S., & Kahveci, N. G. (2016). A study on the development of a mathematics creativity scale. *Eurasian Journal of Educational Research, 62*, 57-76.

Alkan, R. (2014). *Genel yaratıcılık, matematiksel yaratıcılık ve akademik başarı arasındaki ilişkilerin incelenmesi.* Doktora Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Alkış, N. (2016). Bayes yapısal eşitlik modellemesi: kavramlar ve genel bakış. *Gazi İktisat ve İşletme Dergisi*, *2*(3), 105-116.

Asparouhov, T., & Muthén, B. (2021). Advances in Bayesian Model Fit Evaluation for Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 28*(1), 1-14.

Asparouhov, T., Muthén, B., & Morin, A. J. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management, 41*(6), 1561-1577.

Atasoy, B., Kadayıfçı, H. & Akkuş, H. (2007). Öğrencilerin çizimlerinden ve açıklamalarından yaratıcı düşüncelerinin ortaya konulması. *Türk Eğitim Bilimleri Dergisi, 5*(4), 679-700.

Balka, D. S. (1974). The development of an instrument to measure creative ability in mathematics. PhD Thesis, University of Missouri, Columbia.

Başol, G. (2015). *Eğitimde ölçme ve değerlendirme.* Ankara: Pegem Akademi.

Can, A. (2016). *SPSS ile bilimsel araştırma sürecinde nicel veri analizi.* Ankara: Pegem Akademi.

Chamberlin, S. A., & Moon, S. M. (2005). Model-eliciting activities as a tool to develop and identify creatively gifted mathematicians. *The Journal of Secondary Gifted Education, 17*(1), 37-47.

Çokluk, Ö., Şekercioğlu, G. & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları.* Ankara: Pegem Akademi.

Demir, D., Bektaş, M., Demir, S., & Bektaş, İ. (2020). Psychometric properties of the Turkish version of the addiction-like eating behavior scale for university students. *Current Psychology, 40*(6), 2590-2598.

Depaoli, S. (2021). *Bayesian structural equation modeling*. New York: The Guilford Press.

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. PhD Thesis, University of Michigan, Education, Theory, and Practice, Michigan.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Güleryüz, H. (2000). *Eğitim programlarının dili ve yaratıcı öğrenme.* Ankara: Pegem Akademi.

Haylock, D. W. (1984). Aspect of mathematical creativity in children aged 11-12. Doctoral Dissertation, University of London, Centre for Science and Mathematics Education, Chelsea College, London, Great Britain.

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

Kanlı, E. (2019).*Yaratıcılık ve alan uygulaması*. Ankara: Nobel Yayınları.

Karakaya, İ. (2022). *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi*. Pegem Akademi, Ankara.

Kartono, K., & Rusilowati, A. (2019). Development of assessment instruments for mathematical creative thinking ability on junior high school students. *Journal of Research and Educational Research Evaluation, 8*(1), 84-90.

Kattou, M., Kontoyianni, K., Pitta-Pantazi, D., & Christou, C. (2011). On the comparison between mathematically gifted and non-gifted students' creative ability. Retrieved from https://www.academia.edu/search?q=ON%20THE%20COMPARISON%20BETWEEN%20MATHEMATICALLY%20GIFTED%20AND%20NON-GIFTED%20STUDENTS%E2%80%99%20CREATIV%CE%95%20%CE%91BILITY.

Koğar, E. Y., & Koğar, H. (2023). *Mplus ve R ile ileri düzey psikometri uygulamaları.* Ankara: Pegem Akademi.

Kuzu, A. (2013). Açıköğretim Fakültesi Okul Öncesi Öğretmenliği Lisans Programı Bilimsel Araştırma Yöntemleri, https://www.academia.edu/16901153/AOF_Bilimsel_Arastirma_Yontemleri_Kitabi.pdf sayfasından erişilmiştir.

Leikin, R., & Lev, M. (2007, July). Multiple solution tasks as a magnifying glass for observation of mathematical creativity. In *Proceedings of the 31st International Conference for the Psychology of Mathematics Education* (Vol. 3, pp. 161-168).

Mann, E. L. (2005). Mathematical creativity and school mathematics: Indicators of mathematical creativity in middle school students. PhD Thesis, University of Connecticut, Connecticut.

McDonald, R. P. (1985). *Factor analysis and related methods*. New York: Psychology Press.

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, New Jersey: Lawrence Erlbaum Associates Publishers.

Mettler, M. A. (1976). The relationships between mathematical creativity and achievement among ninth grade algebra students. Master's Thesis, California State College, Bakersfield.

Muthén, L. K., & Muthén, B. O. (2012). *Mplus (Version 7.0) [Computer Software]*. Author.

Price, A. (2006). *Creative maths activities for able students: Ideas for working with children aged 11 to 14*. London: Sage Publications.

Prouse, H. L. (1964). The construction and use of a test for the measurement of certain aspects of creativity in seventh-grade mathematics. PhD Thesis, State University of Iowa, Iowa.

Revelle W. (2025). psych: Procedures for Psychological, Psychometric, and Personality Research. Northwestern University, Evanston, Illinois. R package version 2.5.3, https://CRAN.R-project.org/package=psych.

Runco, M. A. (1993). Divergent thinking, creativity, and giftedness. *Gifted Child Quarterly, 37*(1), 16-22.

Sahliawati, M., & Nurlaelah, E. (2020, February). Mathematical creative thinking ability in middle school students. In *Journal of Physics: Conference Series* (Vol. 1469, No. 1, p. 012145). IOP Publishing.

Schoevers, E. M., Kroesbergen, E. H., & Kattou, M. (2018). Mathematical creativity: A combination of domain-general creative and domain-specific mathematical skills. *The Journal of Creative Behavior, 54*(2), 242-252.

Sezerel, B., B. (2019). *Ortaokul öğrencilerinin yaratıcılığını ölçmeye yönelik matematiksel üretkenlik testinin geliştirilmesi.* Doktora Tezi, Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü, Eskişehir.

Siswono, T. Y. E. (2011). Level of students' creative thinking in classroom mathematics. *Educational Research and Reviews, 6*(7), 548.

Soysal, S. (2023). Çok boyutlu test yapılarında alfa, tabakalı alfa ve omega güvenirlik katsayılarının karşılaştırılması. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi, 5(1),* 213-236. https://doi.org/10.38151/akef.2023.51

Suherman, S., & Vidákovich, T. (2024). Relationship between ethnic identity, attitude, and mathematical creative thinking among secondary school students. *Thinking Skills and Creativity, 51*, 101448.

Şen, S. (2023). *Mplus ile yapısal eşitlik modellemesi uygulamaları.* Ankara: Nobel Yayınları.

Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.

Tall, D. (2002). *Advanced mathematical thinking*. USA: Kluwer Academic Publishers.

Taşkın, D. (2016). *Üstün yetenekli tanısı konulmuş ve konulmamış öğrencilerin matematikte yaratıcılıklarının incelenmesi: bir özel durum çalışması.* Doktora Tezi, Karadeniz Teknik Üniversitesi Eğitim Bilimleri Enstitüsü, Trabzon.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education, 2,* 53.

Tekindal, S. (2017). *Okullarda ölçme ve değerlendirme yöntemleri*. Ankara: Nobel Yayınları.

Türkan, Y. (2010). *Matematiksel üretkenlik testi (MÜT)'nin ilköğretim 6., 7. ve 8. sınıflar düzeyinde psikometrik özelliklerinin incelenmesi.* Yüksek Lisans Tezi, Anadolu Üniversitesi Eğitim Bilimleri Enstitüsü, Eskişehir.

Tyagi, T. K. (2015). A study to examine the relationship between mathematical creativity and mathematical problem-solving performance. *Journal of Indian Education, 40*(3), 80-89.

Uysal, İ., & Sarıça, S. (2018). Çizgi filmlerin ilkokul öğrencilerinin duyuşsal özelliklerine etkisine yönelik bir ölçek geliştirme. *İlköğretim Online*, *17*(3), 1302-1316.

Wang, J., & Wang, X. (2020). *Structural equation modeling applications using Mplus* (2nd edition). John Wiley & Sons, UK.

Yurdugül, H. (2006). Paralel, eşdeğer ve konjenerik ölçmelerde güvenirlik katsayılarının karşılaştırılması. *Ankara Üniversitesi Eğitim Fakültesi Dergisi, 39*(1), 15-37.

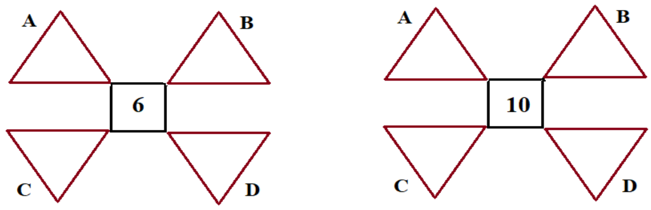
Zainudin, M., Subali, B., & Jailani. (2019). Construct validity of the mathematical creativity instrument: First-order and second-order confirmatory factor analysis. *International Journal of Instruction, 12*(3), 595-614. https://doi.org/10.29333/iji.2019.12336a.

APPENDIX Items of Mathematical Creative Test

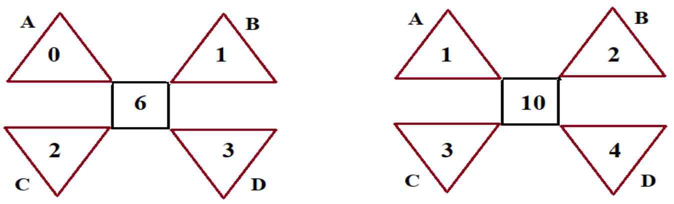
Item 1) Obtain equations such as 2.4 = 8 using the numbers 0, 2, 4, 6, 8. However, do not write an equation such as 3.4 = 12 because the number 1 in this equation is not among the numbers that can be used. When obtaining equations;

- Not every number has to be used in every equation.
- Numbers can only be used once in every equation.

Item 2)



Write numbers inside the triangles so that they are related to the numbers in the squares given above. State the rule for the relationship you have obtained, provided that the relationship is the same for both shapes. Create as many different relationships as possible. There is no need to write the same numbers for the same letters. The letters are added only to make it easier to express the rule. An example is given below.

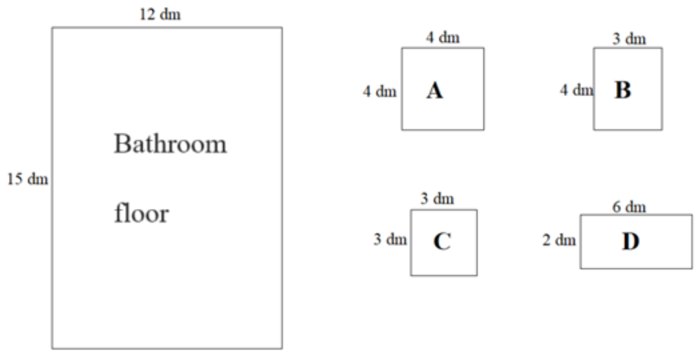


Rule

A+B+C+D

Item 3) Generate as many different equations as possible that include one unknown and the value of the unknown is 4. (For example: x + 5 = 9).

Item 4)

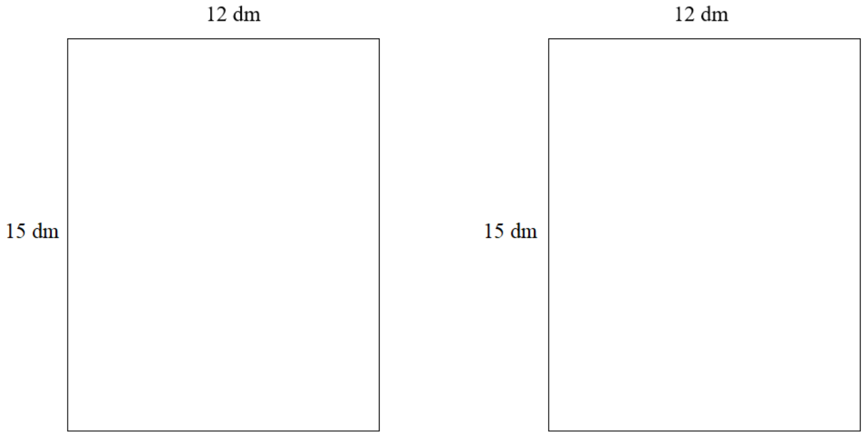


Above is a bathroom floor with sides of 12 dm and 15 dm and tiles of types A, B, C and D, two of which are square and two are rectangular, with given side lengths. The bathroom floor will be laid using these tiles. While laying the bathroom floor;

- At least two different types of tiles should be used.

- There should be no gaps on the bathroom floor.
- Tiles should not overlap and tiles should not break.

Show how you laid the bathroom floor by following these conditions and writing the letters of the tiles you used on the figures below.



Item 5) If you want to form a triangle by joining three sticks with lengths of a units, b units and c units end to end, the length of any one of these sticks must be shorter than the sum of the lengths of the other two and longer than the absolute value of their difference. So;

$|b - c| < a < b + c$

$|a - c| < b < a + c$

$|a - b| < c < a + b$ should be

20 cm

Cem will cut the 20 cm long stick into pieces and join any three of the pieces end to end to obtain as many different triangles as possible. What triangles can Cem obtain? An example is given below.

4 pieces obtained from the stick	The sides of the triangle formed are 4, 5, 6
4, 5, 5, 6	$|5 - 4| < 6 < 5 + 4$

Item 6) Expressions that contain at least one unknown and an operation are called algebraic expressions. Create as many different algebraic expressions as possible that correspond to the algebraic expression x + 4 in its simplest form.

GENİŞLETİLMİŞ TÜRKÇE ÖZET

MATEMATİKSEL YARATICILIK ÖLÇME ARACI: BAYES YÖNTEMİ İLE AÇIMLAYICI VE DOĞRULAYICI FAKTÖR ANALİZİ ÖRNEĞİ

GİRİŞ

Teknolojik ve bilimsel gelişmelerin sosyal ağları ve bireylerin hayatlarını değiştirdiği, hızla değişen dünyada, değişime uyum sağlamak ve bu gelişmeleri sürdürmek için yaratıcılığa ihtiyaç vardır. Her bireyde var olan ancak deneyimlerle ve öğrenme yaşantıları ile gelişmeye açık olan yaratıcılık, bireyin öğrenme yaşantısı neticesinde öğrendikleri arasında bağlantı kurarak karşılaştığı bir problemi çözebilmesi, bu bağlantıları kullanarak ortaya yeni, özgün bir düşünce ya da ürün koyabilmesi olarak ifade edilebilir (Güleryüz, 2000). Yaratıcı bireylere ve bireylerin yaratıcılığının güvenilir bir şekilde ölçülüp değerlendirilmesine olan ihtiyaç gün geçtikçe artmaktadır.

Matematiğin diğer alanlarla ve düşünce şekilleriyle olan ilişkisi nedeniyle matematiksel yaratıcılık, matematikte olduğu kadar diğer alanlardaki başarıya da doğrudan etkisi olacak bir kavramdır. Bu amaçla kavramın içeriğinin anlaşılıp değerinin ortaya konması sonucu bireylerdeki matematiksel yaratıcılık seviyelerinin belirlenip var olan potansiyelleri geliştirici çalışmaların yapılması üzerine yoğunlaşılmalıdır (Türkan, 2010).

Tall (2002), matematiksel yaratıcılığı, matematiğin kendine özgü mantıksal-tümdengelimli doğasını ve oluşturulan kavramların matematikte önemli olanın özüyle bütünleşmeye uygunluğunu dikkate alarak, problem çözme veya yapılarda düşünmeyi geliştirme yeteneği olarak tanımlamaktadır.

Matematiksel yaratıcılığa sahip olan bireylerin analitik düşünme, problem çözme ve soyut kavramlar arasında bağlantı kurma becerileriyle bilim, mühendislik, teknoloji gibi alanlarda toplumsal ilerlemeye katkı sağlama potansiyelleri bulunmaktadır. Bu noktada matematiksel yaratıcılığın güvenilir yöntemlerle ölçülmesi, yaratıcı düşünme potansiyeli yüksek olan bireyleri belirlemesine yardımcı olarak potansiyellerini ortaya çıkaracak şekilde eğitimi almalarına yardımcı olabilir. Bu nedenle, matematiksel yaratıcılığın güvenilir ve geçerli yöntemlerle ölçülmesi önemlidir.

Matematiksel yaratıcılığın ölçülmesinde literatürde kullanılan üç boyut ve bunların açıklamaları şu şekildedir:

- Akıcılık: Bir probleme yanıt olabilecek birçok fikir ortaya koyabilme olarak tanımlanmaktadır. Örneğin, sonucu 5 olan farklı işlemler yazmak gibi. Yaratıcı bireyler problemin çözümüne dair fazla sayıda düşünce üretebilir.

- Esneklik: Bir probleme farklı bakış açılarıyla yaklaşabilme, çeşitli durumları görebilme, değişik kategorilerde fikir üretme, bir olaya farklı perspektiflerden yaklaşabilme olarak tanımlanmaktadır. Ortaya konulan fikirler, problemi ne kadar değişik açılardan inceliyorsa esnekliğin o derece yüksek olduğu anlamına gelmektedir. Yaratıcı bireyler probleme değişik açılarda çözüm yolları gösterirler.
- Orijinallik: Düşünce ve eylemde kendine has olma durumunu ifade etmektedir. Ortaya konan fikir ne kadar az sayıda bireyin aklına geliyorsa o derece orijinal kabul edilir. Yaratıcı bireyler orijinal düşünceler oluştururlar (Atasoy, Kadayıfçı ve Akkuş, 2007).

Literatürde matematiksel yaratıcılığı ölçmeyi amaçlayan çalışmaların büyük kısmı ilkokul (Kattou, Kontoyianni, Pitta-Pantazi ve Christou, 2011; Schoevers, Kroosbergen ve Kattou, 2018) ve ortaokul öğrencilerine yöneliktir (Akgül ve Kahveci, 2016; Balka, 1974; Evans, 1964; Kartono ve Rusilowati, 2019; Mann, 2005; Prouse, 1964; Sahliawati ve Nurlaelah, 2020; Sezerel, 2019; Siswono, 2011; Taşkın, 2016; Türkan, 2010; Zainudin, Subali ve Jailani, 2019). Matematiksel yaratıcılık üzerine yapılan çalışmaların bir kısmı da matematiksel yaratıcılığı yordayan değişkenlerin belirlenmesi amacıyla gerçekleştirilmiştir (Acar, Tanıdık, Uysal, Myers ve Inetas, 2022; Alkan, 2014; Suherman ve Vidakovich, 2024; Tyagi, 2015). Literatürde matematiksel alanda yaratıcılık becerilerini ortaya koymaya yönelik çalışmalar bulunmakta ancak lise grubu öğrencilerinin matematiksel alandaki yaratıcılık becerilerini ortaya koymaya yönelik sınırlı sayıda çalışma bulunmaktadır. Çalışmada elde edilen puanların geçerliği belirlenmek üzere Bayes yöntemiyle faktör analizleri gerçekleştirilmiştir. Çalışmanın lise seviyesindeki öğrencilerin matematiksel alandaki yaratıcılık becerilerini bilim ve sanat merkezi (BİLSEM), Anadolu lisesi ve fen lisesi öğrencilerinden elde edilen veriler ile belirleme amacıyla alana katkı sağlaması beklenmektedir.

YÖNTEM

Araştırmada tarama modeli kullanılmıştır. Ankara ilinde öğrenim gören dokuzuncu sınıf öğrencilerinden 547 öğrenci araştırmanın çalışma grubunu oluşturmuştur. Çalışma grubunda katılımcı onam formunu işaretlemeyen, uygulama esnasında yarım bırakan 47 katılımcının kitapçığı araştırmaya dahil edilmemiştir. Örneklem seçiminde amaçlı örnekleme kullanılmıştır. Amaçlı örnekleme aynı zamanda bilgiye dayalı örnekleme olarak da adlandırılır. Araştırmacı, örnekleme alınacak grubun deneyimine veya bilgisine dayanarak örneklem seçer (Airasian, 2000).

Öncelikle, lise öğrencilerinin matematiksel yaratıcılıklarını belirlemeye yönelik bir matematiksel yaratıcılık testi (MYT) geliştirilmiştir. MYT geliştirilirken maddelerin bir kısmı literatürde matematiksel yaratıcılığı ölçmeyi amaçlayan araştırmacıların oluşturduğu maddelerden uyarlanırken (Akar ve Karaduman, 2021; Evans, 1964; Haylock, 1984; 1987; Kanlı, 2019; Mann, 2005; Price, 2006; Prouse, 1964), bir kısmı da araştırmacı tarafından geliştirilmiştir.

Veri toplamada kullanılacak ölçme aracı için açık uçlu 32 madde hazırlanmış, sonra üç uzmanın görüşüne sunulmuş ve uzmanların görüşleri doğrultusunda sekiz madde çıkarılarak 24 maddeye indirilmiştir. Bu 24 madde, uzman görüşleri dikkate alınarak her formda sekiz madde ve dengeli matematiksel içerik bulunacak biçimde form A, B ve C'ye yerleştirilmiştir. Bu üç ayrı form sınavla ve sınavsız öğrenci alan liselerde kâğıt-kalem testi şeklinde öğrencilere 40 dakikalık süre verilerek araştırmacı gözetiminde uygulanmıştır. Öğrencilerin sekiz maddeyi 40 dakikalık sürede yetiştirmede zorlandıkları gözlenmiştir. Pilot uygulama sonrasında maddeler akıcılık, esneklik ve orijinallik boyutlarına göre puanlanmıştır. Akıcılık boyutunda puanlama yapılırken doğru yanıtların her birine bir puan, esneklik boyutunda puanlama yapılırken farklı kategorilerdeki her doğru yanıt için bir puan verilmiştir. Orijinallik boyutunun puanlaması yapılırken ise verilen doğru yanıt, grubun %5'inden daha az öğrenci tarafından yazılmışsa altı puan, %5'ten fazla ancak %10'dan az öğrenci tarafından yazılmışsa üç puan, %10'dan fazla öğrenci tarafından yazılmışsa sıfır puan verilmiştir. Böylelikle akıcılık, esneklik ve orijinallik boyutlarının hepsinde ham puanlar hesaplanmıştır. Tüm maddelerde akıcılık, esneklik ve orijinallik boyutlarında sınavla öğrenci alan liselerde öğrenim gören öğrencilerin sınavsız öğrenci alan liselerde öğrenim gören öğrencilere oranla daha iyi bir performans gösterdiği tespit edilmiştir. Bu nedenle, nihai uygulamanın sadece sınavla öğrenci alan liselerde ve bilim ve sanat merkezi (BİLSEM) kurumlarında yapılmasının daha uygun olacağına karar verilmiştir. Nihai uygulama için altı açık uçlu maddenin seçiminde, kapsam geçerliğini sağlamak açısından iki tane sayılar, iki tane cebir ve iki tane de geometri alanlarından madde bulunmasına karar verilmiştir. Nihai uygulama yalnızca sınavla öğrenci alan liselerde ve BİLSEM kurumlarında öğrenim gören 547 öğrenci üzerinde, araştırmacı gözetiminde öğrencilere yine 40 dakika süre verilerek uygulanmıştır.

MYT'yi oluşturan altı açık uçlu maddenin her biri akıcılık, esneklik ve orijinallik boyutlarına göre puanlanmıştır. Madde analizleri, Bayes açımlayıcı ve doğrulayıcı faktör analizleri yapılmış ayrıca puanların güvenirliğini belirlemek amacıyla Cronbach alfa, tabakalı Cronbach alfa ve McDonald omega güvenirlik katsayıları hesaplanmıştır.

Madde analizleri, IBM SPSS 25.0 programı, puanların güvenirlikleri IBM SPSS 25.0 ve R psych paketi (Revelle, 2025), Bayes yöntemiyle AFA ve DFA analizleri ise Mplus 7.0 (Muthén & Muthén, 2012) programı kullanılarak gerçekleştirilmiştir.

SONUÇ ve ÖNERİLER

Bu araştırma lise öğrencilerinin matematiksel yaratıcılık becerilerini ölçmeyi amaçlayan bir ölçme aracı geliştirmek için yapılmıştır. Maddelerin betimsel istatistikleri incelendiğinde maddelerden elde edilen toplam yaratıcılık puanlarında en yüksek ortalamanın denklem üretmede en düşük ortalamanın üçgen üretmede olduğu görülmüştür. Buna göre öğrenciler en iyi performansı denklem üretmede, en düşük performansı üçgen üretmede gösterdikleri söylenebilir. En yüksek performans cebir alanında gerçekleşirken en düşük performans geometri alanında gerçekleşmiştir.

Geliştirilen ölçme aracının beş faktörlü bir yapıya sahip olduğu belirlenmiştir. Bu sonuç Balka (1974) ve Sezerel'in (2019) sonuçlarıyla farklılık göstermektedir. Balka (1974), matematikte yaratıcı yeteneği belirlemek için geliştirdiği altı maddelik ölçme aracının iki faktörlü bir yapıya sahip olduğunu, Sezerel (2019) ise, ortaokul öğrencilerinin yaratıcılığını ölçmeye yönelik olarak altı maddeden oluşan matematiksel üretkenlik testini geliştirdiği araştırmasında ölçeğin üç faktörlü bir yapıya sahip olduğunu tespit etmiştir.

Bayes faktör analizlerine göre, AFA'da beş faktörlü bir yapı gözlenirken ardından yapılan birinci ve ikinci düzey DFA sonuçlarının beş faktörlü yapının model-veri uyumunu desteklediği gözlenmiştir.

Araştırmada MYT'nin Cronbach alfa güvenirlik katsayısı 0.650 tabakalı Cronbach alfa katsayısı 0.598, olarak ve McDonald omega güvenirlik katsayısı ise 0.890 bulunmuştur. Akgül ve Kahveci (2016) ortaokul öğrencilerinin matematiksel yaratıcılıklarını ölçmek için geliştirdikleri ölçeğin güvenirlik katsayısını 0.80 olarak, Türkan (2010) altıncı, yedinci ve sekizinci sınıf öğrencilerinin matematiksel yaratıcılıklarını ölçmek için geliştirdiği testin güvenirlik katsayısını 0.78 olarak, benzer biçimde Balka (1974), geliştirdiği matematiksel yaratıcılık ölçme aracının güvenirliğini 0.72 olarak bulmuştur. Bu çalışmalarda DFA raporlanmamıştır. MYT'deki madde sayısının az olması Cronbach alfa katsayılarının düşük çıkmasına neden olmuş olabilir.

Bu araştırma Ankara ilinin merkez ilçelerindeki sınavla öğrenci alan liselerde ve BİLSEM'lerde yapılmıştır. Bu nedenle, Türkiye'nin farklı illerinde öğrenim gören çok sayıda öğrencinin yer alacağı örneklemlerden elde edilecek bulgular, MYT'nin geçerlik ve güvenirliğine yönelik daha kapsamlı bulgular ortaya koyabilir. Benzer bir araştırma sınavsız öğrenci alan liseleri de kapsayacak şekilde yapılabilir. Ayrıca gelecekte yapılacak çalışmalarda 'matematiksel yaratıcılık' kavramının öğrencinin matematik başarısı, matematiksel gibi faktörlerle olan ilişkisinin tanımlanması matematiksel yaratıcılığın kuramsal çerçevesinin genişletilmesine katkı sağlayabilir.

Çalışmanın bir başka sınırlılığı MYT'deki maddelerin matematiğin sayılar, geometri ve cebir alt öğrenme alanlarını kapsamasıdır. Çalışma, bir ders süresi olan 40 dakikada uygulanmıştır. Maddeler açık uçlu olduğundan ve maddelerin sınırlı sayıda çözümü olmadığından öğrencilerin yanıtlamak için uzun süreye gereksinim duyma eğiliminde oldukları gözlenmiştir. Benzer bir araştırma matematiğin veri, sayma ve olasılık gibi farklı alt öğrenme alanlarını da kapsayan daha fazla sayıda maddenin yer aldığı ve buna uygun olarak test süresinin göz önünde bulundurulduğu bir ölçme aracı hazırlanarak yapılabilir.

Çalışmanın bir diğer sınırlılığı ise puanlamanın tek bir puanlayıcı tarafından yapılması nedeniyle puanlayıcı hatalarının incelenememesidir. Benzer bir çalışma ölçme aracından elde edilen verilerin birden fazla puanlayıcının puanlamayı yapması ve puanlayıcı hatalarının analiz edilmesi ile gerçekleştirilebilir.

Özetlenen sınırlılıklarına rağmen, MYT'nin lise seviyesindeki öğrencilerin matematiksel yaratıcılığının değerlendirilmesinde kullanılabilecek özgün, geçerli, güvenilir ve kullanışlı bir test olduğunu söylemek mümkündür.