

## **Technology-based assessment of creativity in educational context: the case of divergent thinking and its relation to mathematical achievement**

**Attila Pásztor<sup>1</sup>, Gyöngyvér Molnár<sup>2</sup>, Benő Csapó<sup>1</sup>**

<sup>1</sup>MTA-SZTE Research Group on the Development of Competencies, University of Szeged

<sup>2</sup>Institute of Education, University of Szeged

### **Abstract**

Creativity is one of the most frequently cited 21<sup>st</sup> century skills, and developing creativity is one of the most often declared goals of modern education systems. However, without easy-to-use assessment instruments available for everyday application in educational practice, systematic improvement of creativity is far from a realistic option. The aim of the present study is to explore the possibility of online assessment of divergent thinking and to contribute to the development of a reliable technology-based test. The paper also investigates the relationship between divergent thinking and mathematical achievement in different dimensions. The sample for the study was drawn from sixth-grade students (N=1,984). The computerized instrument comprising nine tasks was based on item types for divergent thinking by Torrance and by Wallach and Kogan. Our online test proved to be a reliable instrument. Based on theoretical assumptions, evidence for construct validity was provided for both the fluency-flexibility-originality and verbal-figural dimensions. Divergent thinking predicts mathematical achievement at a moderate level. The advantages of technology-based assessment made our instrument suitable for everyday school practice and large-scale assessments; however, the coding process is not yet fully automated.

### **1 Introduction**

The significant role of creativity in the 21<sup>st</sup> century is undisputed. An ever more rapid economic, social and technological development requires new and original ideas and solutions. Creativity is indispensable for success in a wide range of jobs in modern societies (Florida, 2004) and one of the most frequently mentioned 21<sup>st</sup> century skills (Binkley et al., 2012). Twenty-first century skills are described as skills which are essential to succeed in work and life in the current century, such as critical thinking, problem solving, communication, collaboration, and information and communication technology (ICT) literacy. Creativity is interconnected to other 21<sup>st</sup> century skills: solving a problem often requires creative ideas; communicating and working creatively play an important role in successful social life; and creative usage of information and

digital technologies are also essential in navigating through everyday life in the 21<sup>st</sup> century (Piiro, 2011). Thus, developing creativity is one of the most often declared goals of modern education systems (COM, 2010). From a practical perspective, one of the major obstacles to its development is the lack of easy-to-use instruments. Most existing tests are manually coded, and the coding process may involve subjective decisions. Their application is time-consuming and expensive. Without reliable measurement instruments, even the simplest training experiment is impossible, and a systematic development of creativity in an educational context requires routinely applicable assessment tools. The aim of the present study is to explore the possibilities for a technology-based assessment of creativity in regular schools and to contribute to the development of a reliable online instrument.

### **1.1 Definition and assessment of creativity: the case of divergent thinking**

Although there is an agreement about the importance of creativity, there are large numbers of diverging interpretations and views about the nature of it. Due to the different perspectives and paradigms in the research on creativity, arriving at a standard definition as a construct is a challenging enterprise. However, there are common features in different definitions, and it seems there is a consensus that creative acts result in output which is novel and has some sort of value (for more about definition problems, see Piffer, 2012; Plucker, Beghetto, & Dow, 2004; Runco & Jaeger, 2012; Simonton, 2012). Although there is a sort of agreement on these characteristics, studies conducted in the field have proved that creativity is an extremely complex phenomenon and that there are many approaches to studying it (Mayer, 1999; Runco, 2007). For example, one can focus on the creative process (cognitive factors), the individual (identifying personal traits, attitudes and behavioural correlates), the product (determining what makes a product creative) or press (attributes of creativity-fostering environments) (Plucker & Renzulli, 1999). All of these approaches have different assessment methods and highlight different aspects behind creative performance; thus, “the search for a single type of creativity assessment is misleading. There is no simple measurement of creativity” (Funke, 2009, p. 14).

Research on divergent thinking is one of the major approaches in the identification of thinking processes behind creative performance (Runco, 2011). From an educational perspective, it has been considered an indicator of creative potential (Kim, 2006; Runco & Acar, 2012). Divergent thinking was part of Guilford’s (1967) Structure of Intellect model, in which he described it as part of problem solving. Divergent thinking refers to the process of generating numerous answers or ideas for a given topic or problem. This stands in contrast to tasks that represent convergent thinking, in which only a single or a few correct solutions are possible, such as in conventional intelligence tests.

To assess divergent thinking, Guilford devised a number of tasks (Guilford, 1967), and further tests were developed based on his work which became widely used instruments in creativity research such as the Torrance Test of Creative Thinking (TTCT, Torrance, 1966) and the Wallach–Kogan Creativity Test (WKCT, Wallach & Kogan, 1965). These measurement tools usually consist of tasks with verbal- and figural-based items. In verbal-based items, both the stimuli and the responses are verbal. For example, one has to list as many unusual ways to use a book as one can think of or name all the round things. In figural-based tasks, stimuli are figural, but the response could be figural or verbal. For example, on some TTCT tasks, the respondent is expected to complete or produce drawings (figural-figural), and one has to interpret

lines or figures (figural-verbal) on WKCT instances tasks. Different types of tasks may represent different ways of thinking or strategies during task completion (see Cheung & Lau, 2010).

Different scales were suggested by Guilford (1967) to evaluate such tasks, like fluency, flexibility and originality. Fluency refers to the ability to produce numerous ideas for a given problem, and it is assessed by the number of interpretable, meaningful and relevant responses. Flexibility is described as the skill to see a problem from different approaches, and it is scored by the number of different categories implied by the responses. Originality refers to the ability to produce unique, unusual ideas, and it is usually measured by the statistical rarity of the responses in a given sample (e.g., answers given by less than 1 or 5% of the participants; for examples of different scoring techniques, see Runco & Acar, 2013). However, studies usually found highly positive correlations between the three indices of divergent thinking. Some psychometrics argued that fluency would be enough because the other two measures add only little information (e.g., Hargreaves & Bolton, 1972). On the other hand, others showed the factors can be separated (e.g., Dumas & Dunbar, 2014) and claimed that originality and flexibility are representing important aspects of creative thinking. Due to the debate others suggested alternative scoring methods for divergent thinking tests (Plucker et al., 2011; Synder et al., 2004).

## **1.2 Online assessment of divergent thinking**

Technology-based assessment is one of the most rapidly developing research areas in educational practice. The growing attention can be explained by the advantages of technology-based assessment, such as online test administration, automated scoring, improved precision, objectivity, reliability and the possibility of immediate feedback (Csapó, Ainley, Bennett, Latour, & Law, 2012). In the measurement of divergent thinking, test administration and scoring are among the major concerns: open-ended tasks generate numerous responses which are difficult to process with traditional paper-and-pencil test administration. Each answer has to be coded and scored manually. Researchers have to decipher handwriting, and data has to be digitized before performing statistical analyses. Due to these aspects of paper-and-pencil test administration, the data analysis process is extremely time-consuming and cannot be implemented effectively in everyday school practice. However, technology-based assessment of divergent thinking is still in its infancy. Only a few studies have focused on the potential for technology-based assessment of divergent thinking (Cheung & Lau, 2010; Kwon, Goetz, & Zellner, 1998; Lau & Cheung, 2010; Palaniappan, 2012; Pretz, 2008; Rosen & Tager, 2013; Villalba, 2009). Palaniappan (2012) developed an intelligent web-based Creativity Assessment System (CAS), where verbal responses were automatically scored on the basis of the database in the TTCT manual. In his pilot, he reported high correlations between scores calculated manually and by CAS for all three measures of divergent thinking (fluency, flexibility and originality). In cases where answers could not be recognized or did not fit into a category because of their novelty, the system sent them to a webpage where the researcher had to categorize them manually. Cheung and Lau (2010) developed an online assessment tool for the Wallach–Kogan Creativity Test named the e-WKCT, which is based on the standardized paper-and-pencil test of the Chinese version of the WKCT (Cheung, Lau, Chan, & Wu, 2004). They also used an automatic scoring system and conducted a large-scale study with 2,476 primary and secondary school students. The tool provided instant

feedback after test completion and an online comparison of the results to the Hong Kong norms.

These findings reveal the advantages and limitations of technology-based assessment of divergent thinking and show that technology offers a feasible solution for the problem of creating an easy-to-use instrument of creativity.

### **1.3 Creativity, divergent thinking and mathematics**

In many education systems, mathematics is the only school subject which is taught throughout the 12 years of general education. Development of early numeracy and the role of mathematics knowledge in the success of a later school career have been covered by a great number of studies. Mathematics is an integral component of large-scale international comparative studies like PISA (Programme for International Student Assessment, OECD, 2013) and TIMSS (Trends in International Mathematics and Science Study, Mullis & Martin, 2013) and national assessment systems. Based on the results of these studies, the attributes of education systems that best promote the development of mathematical knowledge and skill are well known. Thus, if we intend to place the online assessment of creativity into an educational context and choose one school subject to which its testing is related, mathematics is an obvious choice.

There are many studies which investigate the role of creativity in mathematics. One line of research explores the creative aspects of learning mathematics and the relations between creativity and mathematical performance and/or mathematical creativity (e.g., Bahar & Maker, 2011; Idris & Nor, 2010; Nadjafikhah & Yaftian, 2013; Nadjafikhah, Yaftian, & Bakhshalizadeh, 2012; Pólya, 1973, 1981; Sak & Maker, 2006; Shriki, 2010; Silver, 1997; Sriraman, 2004). Another line associates creativity with giftedness, suggesting that only a smaller group, the most talented learners of mathematics, are creative (see Leikin, 2009; Mann, 2006; Sriraman, 2005; Sriraman & Lee, 2011). In our study, we do not assume that creativity is only present in the mathematics learning and mathematical performance of gifted students; instead, we explore what role creativity plays in achievement in different dimensions of mathematics. In previous research related to mathematical creativity, divergent thinking has usually been assessed with open-ended mathematical problems, such as “Write as many original problems as you can that have an answer of 14!” (Sak & Maker, 2006, p. 283). In studies where mathematical knowledge has also been measured, correlations between mathematical knowledge and divergent thinking have ranged from .38 to .60 (Bahar & Maker, 2011; Sak & Maker, 2006), but there is a lack of research on the relation of general divergent thinking to mathematical knowledge or other dimensions of mathematics.

In the assessment of mathematics, frameworks in large-scale assessments such as PISA or TIMSS apply different classifications to describe the domain. The most common classifications group the content of assessment according to areas of mathematics as a discipline (e.g., algebra, geometry, etc.). More sophisticated frameworks distinguish different types of knowledge. Early IEA studies (International Association for the Evaluation of Educational Achievement, Husén, 1967) focused on the common disciplinary content of the curricula in participating countries, while the recent TIMSS frameworks distinguish items that measure knowing, applying and reasoning within different content domains (Mullis & Martin, 2013). The PISA frameworks also deal with content domains, but the entire assessment is application-centered, exploring how students can use their knowledge in contexts typical of modern societies (OECD, 2013).

The mathematics tests we use in our current assessment are based on a three-dimensional framework (Csapó, 2010) in which mathematical disciplinary content knowledge (MD), psychological attributes of achievement (MP) and application of mathematics knowledge (MA) are distinguished (Csikos & Csapó, 2011). The MD dimension is similar to the TIMSS *knowing* cognitive domain, as it covers mathematical concepts and procedures defined by the curriculum. MP includes mathematical reasoning, metacognition and domain-specific problem-solving strategies, while MA assesses students' abilities to apply their knowledge in new contexts and situations (what is close to the concept of *mathematical literacy* as it is interpreted in PISA).

#### **1.4 Aims and research questions**

The objective of this study is twofold. First, we explore the possibilities of a technology-based assessment of divergent thinking. Since divergent thinking items often require verbal responses, these tests are liable to language effects. In addition, norms can also be culture-biased; therefore, local versions of divergent thinking tests have been made in particular cultures (e.g., Cheung, Lau, Chan, & Wu, 2004). For these reasons, we have developed a new instrument based on item types for divergent thinking by Torrance as well as by Wallach and Kogan.

As printing and distributing creativity tests are the most expensive components of large-scale assessments and routine usage, online delivery may represent a major step towards making creativity assessment more affordable. The second impeding factor is human scoring. In this area, we endeavour to automate the scoring process, at least in large part.

We then explore some attributes and usability of the newly developed instrument. Due to the lack of research on the relationships between different factors of general divergent thinking processes and different dimensions of mathematics, we also examine the possible relationships between these domains. More specifically, we aim to answer the following research questions:

1. What are the psychometric properties of the online creativity instrument?
2. What is the relationship between creativity and mathematical achievement in different dimensions?

Regarding the second research question, it can be assumed that application and thinking dimensions of mathematics relate more strongly to divergent thinking than the content dimension. A fluent and flexible way of thinking and the ability to produce original ideas could play a significant role in the application of knowledge in new contexts and situations or in solving domain-specific problems.

## **2 Methods**

### **2.1 Participants**

The sample for the study was drawn from sixth-grade students ( $N=1,984$ , 1,005 boys and 937 girls, age  $M=12.05$ ,  $SD=0.51$ , range: 10.75–16.17 years) in primary schools in Hungary. In that system, eight years of the first phase of compulsory schooling are divided into two parts of four years each. Students in the lower grades are taught by class teachers, while different subjects are taught by specialised teachers in the upper

grades. After grade four, students usually continue their schooling in the same classes and schools, there is no tracking at this point, and all programmes are based on the same national core curriculum. School classes formed the sampling units. Altogether 97 classes from 78 primary schools participated in the study from various regions in Hungary with an average of 20.5 students taking part from each school ( $SD=9.7$ ). The distribution of students in respect of background variables (e.g., social background, academic achievement, motivation and attitudes) covered a wide range.

## 2.2 Instruments

### 2.2.1 Online measurement tool for divergent thinking

The computerized instrument of divergent thinking comprised nine tasks and was based on Torrance's (1966) and Wallach and Kogan's (1965) open-ended item types for divergent thinking. It consisted of three alternative uses tasks (match, cup and toothbrush), three instances tasks (list things which are transparent, produce light and jingle) and three picture meaning tasks (Fig. 1 and Fig. 2). Students had three minutes to provide answers for each task.

What can a **toothbrush** be used for?

Try to come up with **interesting and unusual** ways to use one.

It is **IMPORTANT** to know that you can change the toothbrush however you want to.

List as many possibilities as you can.

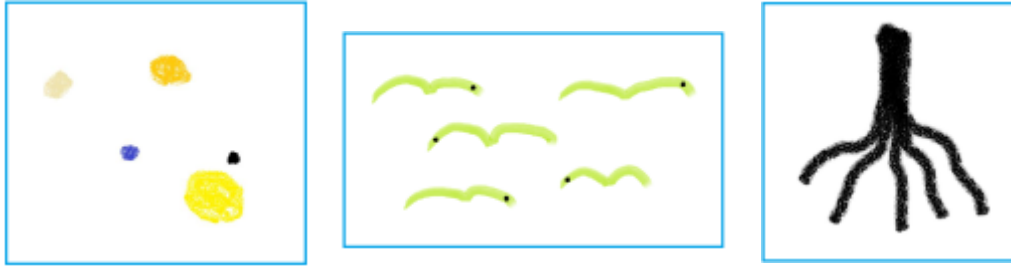
Write each of them in a different textbox. You have got 3 minutes to do this.

<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>
<input type="text"/>	<input type="text"/>	<input type="text"/>

If you have got any more ideas, you can write them here:

180  
Next

**Fig. 1** Sample item for the alternative uses and instances tasks. The original items were in Hungarian.



**Fig. 2** The three stimuli for the picture meaning tasks. The appearance of the items was similar to that of the alternative uses and instances tasks. The pictures were presented instead of the third column of textboxes on the right side of the item.

Based on the stimuli given in the tasks, two subconstructs can be distinguished within the test: verbal-verbal creativity, where both the stimuli and the responses are verbal (alternative uses and instances tasks), and figural-verbal creativity, where verbal responses should be given for visual stimuli (picture meaning tasks).

### 2.2.2 Scoring of divergent thinking answers

The answers were scored with the scales generally used in measuring divergent thinking: fluency, flexibility and originality. Since the instrument was newly developed, there was no database or test manual to score the answers (i.e., there was no information available on the relevant categories for scoring flexibility and on the rarity of an answer to measure originality). Therefore, with the participation of four raters, categories were created, all answers categorized manually and decisions made about questionable answers with regard to relevance. A two-level categorization process was used. Answers were grouped into main categories (level 1) based on given domains. For instance, in the case of *a match* (unusual uses), it can be used to build something. Within this level 1 category, many subcategories (level 2) can be formed (e.g., *to build a house* or *to build a castle* were frequent answers and formed different subcategories). There were also more rare ones; for example, *to build a model city* occurred only three times. However, these answers are still in the same general domain of building. Other answers formed a category on their own: *to mark a page in a book* was also given by only three students. It would be a plausible argument to consider *to mark a page in a book* as more original than *to build a model city*, even though the frequency of the responses is the same (i.e., three). In order to reach this aim, the number of answers within both level 1 and level 2 categories has to be taken into account in the scoring process of an individual answer. Barkóczi and Klein (1968; see also Kardos, Pléh, & Barkóczi, 1987) developed a formula which can handle this problem. The formula is the following:

$$k = \left( \frac{\frac{T-I}{T} + \frac{T-i}{T}}{2} \right)^{14} \text{ or the reformed version:}$$

$$k = \left( 1 - \frac{I+i}{2T} \right)^{14}$$

where

T = total number of responses

I = number of responses within a single domain (level 1)

i = number of responses in a subcategory (level 2)

k = originality score for an answer

The originality scores computed with this formula range between 0 and 1, so it is more sensitive to differences in originality compared to other methods that classify students into different groups based on the relative frequency of their answers (e.g., 4 points for frequency of less than 1%; 3 points for frequency of 1–2%; 2 points for 3–6% and 1 point for 7–15% (see Cropley, 1967)). Without the 14 index, the value of the formula ranges between 0.7 and 1, so it is necessary to use the whole range from 0 to 1 and increase the sensitivity of the scale. With this formula, the most original answers are those which are in a level 1 category containing few answers, such as *to mark a page in a book*. In spite of the frequency of the answer *to build a model city* being the same as that of *to mark a page in a book*, the former answer loses its original power and will receive a lower score because it is in the large level 1 category of building. An individual-level originality score was obtained by the sum of the originality values of the answers in a given task. Fluency was assessed by the number of relevant answers, and flexibility was scored by the number of level 1 categories implied by the responses.

### 2.2.3 Mathematics tests

Based on the three dimensional framework of knowledge (Csapó, 2010) a newly developed online test was applied to assess different dimensions of mathematical knowledge (Csapó & Szendrei, 2011). The three dimensions measured by the instrument were: disciplinary content knowledge (MD); the psychological dimension of mathematical achievement, that is, mathematical reasoning (MP); and the application of mathematics in practical situations, which is mathematical literacy (MA, for a sample item see Fig. 3). Each subtest consisted of four tasks, with the number of items ranging between 16 and 17 within each subtest. The reliability of the whole test (50 items) was  $\alpha=.91$  (i.e., in terms of Cronbach's alpha). The reliability of the subtests came to  $\alpha=.82$ ,  $.82$  and  $.80$ , respectively. Data administration and scoring were fully computerized, and immediate feedback was given after test completion.



The diagram below represents the distance of four students' homes from school and also the way they get to school.

**Drag** the students' names into the right cells of the table!



		Home – school distance	
		Less than 1 km	At least 1 km
The way of getting to school	On foot		
	By bicycle		

◀ Previous

Next ▶

**Fig. 3** Sample item of the mathematical test in the psychological dimension (MP). The original items were in Hungarian.

The three-dimensional measurement model for mathematics showed a good model fit (Table 1), as assumed. The preferred estimator for categorical variables, Weighted Least Squares Mean- and Variance-adjusted (WLSMV; Muthén & Muthén, 2010), was used to assess this model. Within the three-dimensional model, significant latent correlations were found between the pairs of dimensions ( $r_{MD\_MA}=.20$ ,  $r_{MD\_MP}=.57$ ,  $r_{MA\_MP}=.74$ ,  $p<.001$ ).

**Table 1**

Goodness of fit indices for testing dimensionality of mathematics.

Model	$\chi^2$	<i>Df</i>	<i>p</i>	CFI	TLI	RMSEA (95% CI)	<i>n</i>
3-dimensional	4,970.56	899	.001	.910	.905	.051 (.050–.052)	1,736
1-dimensional	6,502.81	902	.001	.876	.870	.060 (.058–.061)	1,736

*Note:* *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation;  $\chi^2$  and *df* are estimated by WLSMV.

A one-dimensional model with all three dimensions combined under one general factor was also tested. In order to test which model, the one- or the three-dimensional

model, fits the data better, a special  $\chi^2$ -difference test was carried out in Mplus. It showed that the three-dimensional model fit the data significantly better than the one-dimensional model ( $\chi^2=603.278$ ;  $df=3$ ;  $p<.001$ ). It was thus possible to distinguish the disciplinary, psychological and application factors of mathematics empirically.

### 2.3 Procedures

The online data collection was carried out with the eDia (Electronic Diagnostic Assessment; Csapó, Lőrincz, & Molnár, 2012) platform via the Internet in the schools' ICT labs. The schools had approximately two weeks to administer the tests. The divergent thinking instrument was administered first, followed by the mathematics test. Both tests took approximately 45 minutes to complete.

Regarding the divergent thinking test, calculating originality scores and counting the relevant answers and the categories used for each respondent can consume a great deal of human resources for a sample of 1,984 students. In order to address this problem, a separate online platform was developed which compared the database with the raw data and the one with the categorized answers (i.e., the test manual developed by our raters) and calculated the three indices automatically.

### 2.4 Data analyses

Confirmatory factor analyses (CFA) within structural equation modeling (SEM; Bollen, 1989) was used to test the underlying measurement model for mathematics and divergent thinking. All measurement models were computed with Mplus. Because students attend different classes, possible cluster effects had to be tested before the data analyses. Analysis of variance was applied to examine the differences of students' achievements between and within classes (Csapó, Molnár, & Kinyó, 2008). The results showed that variance between classes is significantly larger than within classes [ $F_{\text{fluency}}(96, 1,887) = 8.23$ ,  $p<.01$ ;  $F_{\text{flexibility}}(96, 1,887) = 8.80$ ,  $p<.01$ ;  $F_{\text{originality}}(96, 1,887) = 8.78$ ,  $p<.01$ ;  $F_{\text{mathematics}}(89, 1,649) = 8.88$ ,  $p<.01$ ]; thus, our data set is clustered. In order to control for potential confounding with classroom characteristics, we used the type is complex option implemented in Mplus throughout the analyses. The same procedure was also applied for mathematical scores.

In order to test the underlying measurement model for divergent thinking within Mplus, the raw scores were recoded to a five-point scale. The recoding process was necessary due to the different measurement levels and scales of the variables and it was based on the distributions (percentiles) of the achievement scores for all the items. This transformation of the scales did not change significantly the reliability of the test and subtests. Weighted Least Squares and Mean- and Variance-adjusted (WLSMV) estimation was used (Muthén & Muthén, 2010). Different fit indices, such as the Tucker–Lewis Index (TLI), the comparative fit index (CFI) and the root mean square error of approximation (RMSEA), were computed to assist in determining model fit. Nested model comparisons were conducted using a special  $\chi^2$ -difference test for the WLSMV estimator (Muthén & Muthén, 2010).

### 3 Results

#### 3.1 Psychometric properties of the online divergent thinking test

Table 2 shows the reliability coefficients (Cronbach's alpha) for the divergent thinking test and its subscales. The values are high for all three scales of divergent thinking, ranging from .81 to .92.

**Table 2**

Cronbach's alpha indices for the divergent thinking test and its subscales.

	Number of items	Cronbach's alpha
Fluency	9	.92
Verbal	6	.89
Figural	3	.87
Flexibility	9	.89
Verbal	6	.82
Figural	3	.81
Originality	9	.90
Verbal	6	.86
Figural	3	.82

The patterns of correlation presented in Table 3 provide some empirical evidence for the convergent and discriminant validity of the test: verbal fluency, flexibility and originality have considerably higher correlation values among them compared to the intercorrelation coefficients between figural and verbal subscales. The same pattern can be observed for figural fluency, flexibility and originality.

**Table 3**

Correlation coefficients between subscales of verbal (V) and figural (F) fluency, flexibility and originality.

	V-fluency	V-flexibility	V-originality	F-fluency	F-flexibility
V-flexibility	.95				
V-originality	.96	.95			
F-fluency	.76	.73	.74		
F-flexibility	.68	.71	.69	.89	
F-originality	.71	.71	.72	.93	.92

##### 3.1.1 Dimensionality of creativity

Based on the literature, a three-dimensional measurement model of divergent thinking that includes flexibility, fluency and originality was supposed. We allowed residuals of items sharing similar characteristics (i.e., figural tasks-verbal tasks) to be correlated. The three-dimensional model showed a good model fit (Table 4) according to the CFI and TLI indices. The RMSEA values were not as low as expected, but they can still be considered acceptable. Within the three-dimensional model, all three

dimensions were correlated on a latent level ( $r_{fluency\_flexibility} = .48$ ,  $r_{fluency\_originality} = .68$ ,  $r_{flexibility\_originality} = .63$ ,  $p < .001$ ).

**Table 4**

Goodness of fit indices for testing dimensionality of divergent thinking.

Model	$\chi^2$	<i>df</i>	<i>p</i>	CFI	TLI	RMSEA (95% CI)	<i>N</i>
3-dimensional	1,133.61	124	.001	.989	.979	.064 (.061–.068)	1,984
1-dimensional	1,450.87	127	.001	.985	.973	.072 (.069–.076)	1,984

*Note:* *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation;  $\chi^2$  and *df* are estimated by WLSMV.

The one-dimensional model combining the three factors under one general factor was also tested. According to the result from the special  $\chi^2$ -difference test, the three-dimensional model fit the data better than the one-dimensional model ( $\chi^2=386.01$ ;  $df=3$ ;  $p < .001$ ). It was possible to empirically differentiate the factors of flexibility, fluency and originality with respect to divergent thinking as distinguished in paper-and-pencil testing in a computer-based environment as well.

Regarding the verbal-figural distinction, two-dimensional models were also tested for all three measures of divergent thinking (Table 5). The models show a good model fit, indicating further evidence for the construct validity of the divergent thinking test.

**Table 5**

Goodness of fit indices for verbal-figural dimensions.

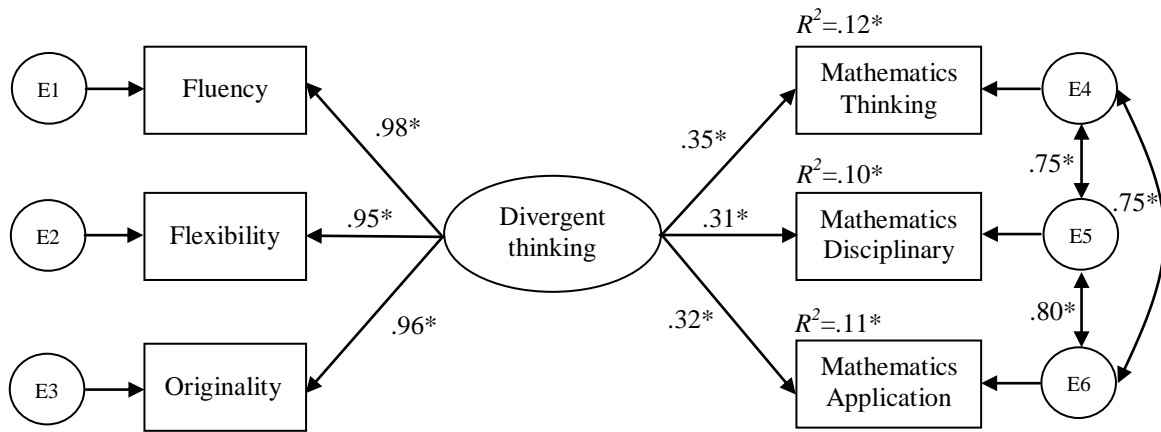
2-dimensional models (figural-verbal)	$\chi^2$	<i>df</i>	<i>p</i>	CFI	TLI	RMSEA (95% CI)	<i>N</i>
Fluency	463.66	26	.001	.953	.934	.092 (.085–.100)	1,984
Flexibility	348.22	26	.001	.946	.926	.079 (.072–.087)	1,984
Originality	333.76	26	.001	.951	.932	.077 (.084–.099)	1,984

*Note:* *df* = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker–Lewis Index; RMSEA = Root Mean Square Error of Approximation;  $\chi^2$  and *df* are estimated by WLSMV.

### 3.2 Divergent thinking and mathematical achievement

Continues factor indicators were used in SEM analysis to examine the relationships between divergent thinking and mathematical achievement. Divergent thinking as latent factor has been specified by fluency, flexibility and originality (Fig 4). Sum scores of the three divergent thinking scales were used to specify different measures of divergent thinking. We assumed that divergent thinking would predict performance in different

dimensions of mathematics but that a significant amount of variance should remain unexplained. Thus, we regressed mathematics on divergent thinking and estimated the proportion of explained variance in three dimensions of mathematics. Results showed that divergent thinking explained performance in all three dimensions of mathematics with a similar effect, but the residuals of measures of MA, MD and MP were still highly correlated ( $r=.75-.80$ ), indicating common aspects of mathematics dimensions separable from divergent thinking (see Fig 4). The model fit well (CFI=.994, TLI=.985, RMSEA=.082 [95% CI: .066-.099]).



**Fig. 4** A structural model of divergent thinking as a predictor of mathematical achievement in different dimensions. Manifest variables are depicted by rectangles and latent variables by cycles. Standardized parameter estimates are shown. ( $*p<.01$ )

## 4 Discussion

The availability of easy-to-use instruments and the examination of the relations of such skills to other domains are essential to develop creativity in the everyday school context. The aim of this paper was to explore the possibilities for a technology-based assessment of divergent thinking and to examine the relationship between the different factors of divergent thinking and different dimensions of mathematics.

### 4.1 Answering the research questions

Our online assessment instrument for divergent thinking proved to be reliable regarding the whole test and its subscales as well. Based on theoretical assumptions, evidence for construct validity was provided for both fluency-flexibility-originality and verbal-figural dimensions. Advantages of technology-based assessment, such as online test administration and automatic calculation of scoring, reduced the time and cost of the testing process. Considering these characteristics, we took the first steps to make our instrument suitable for everyday school practice and large-scale assessments. The findings indicate that online assessment may provide teachers an easy-to-use instrument for monitoring the development of students' divergent thinking and may contribute to the development of effective teaching methods.

Examining the relationships with mathematics showed that divergent thinking predicts mathematical achievement with magnitudes comparable in effect size for the three dimensions. The values are not high, and a significant amount of variance is

unexplained; however, this finding supports the claim that divergent thinking plays an important role in various aspects of mathematical performance. Regarding mathematical knowledge, we found a similar relationship to previous research results (e.g., Bahar & Maker, 2011; Sak & Maker, 2006). In addition to these findings, we showed that this relationship applies to content-general divergent thinking as well. On the other hand, the similar magnitudes of regression coefficients on all three dimensions of mathematics do not support our assumptions that the application of knowledge is an especially creative process or that divergent thinking relates more to solving domain-specific mathematical problems than mathematical knowledge does. There is no straightforward interpretation of this finding; however, a plausible reason might be that all of our mathematical items were convergent tasks, and so they were not sufficiently sensitive to address this relationship because there was no space for divergent thinking activities during task completion. It can also be assumed that there is a third factor behind the similar magnitudes: general mental abilities. The nature of possible connections between different intelligence and creativity measures that include divergent thinking is a familiar theme in the literature (e.g., Getzels & Jackson, 1962; Karwowski & Gralewski, 2013). Empirical evidence can also be found for the relation between mathematical achievement and general reasoning skills (Primi et al., 2010; Xin & Zhang, 2009).

#### **4.2 Limitations of the study and directions for further research**

With regard to technology-based assessment of divergent thinking, one of the limitations of the research at this phase is that the coding process is not yet fully automated. In the present study – due to its pioneering nature – answers had to be categorized by human raters to create a digital test manual to score the answers. This process is only required once, and, on the basis of this database, the possibilities for developing an online evaluation system can be further explored. The next stage is to develop an algorithm which is able to integrate the answers from forthcoming data administration into the existing database. It is important to highlight here that all the answers were categorized in this study; therefore, the problem of categorizing new, creative solutions did not occur. However, this problem may arise quite often in creativity research. One solution to address this issue could be a complex algorithm which can evaluate the scores for answers which did not occur in previous measurements. Another solution would be an evaluating platform where raters can easily categorize new answers. In general, results from creativity tests are mostly scored by human raters, so these kinds of online evaluating platforms could reduce testing time because raters would only have to deal with issues which cannot be handled by computers. Regarding the scoring techniques the high factor loadings of the three indices of divergent thinking can be seen on Figure 3 indicate the highly positive interdependence of the scales. Although we provided empirical evidence for separating the three factors alternative scoring methods should be explored in further research. Technology-based assessment may contribute to fulfill this endeavor with the possibility of automatized data analyzes (e.g., comparing different calculations of the scores).

Our study left many questions unanswered concerning the relation between general divergent thinking and different dimensions of mathematics. Further research should reveal how different item types (open- or closed-ended tasks) affect these relations or what the contribution of general mental abilities is in explaining mathematical

achievement with divergent thinking. In addition, it would also be worth examining relationships with other domains, such as science or reading.

It is important to note that divergent thinking is one of the major aspects of creativity research; however, it does not represent creativity. Thus, investigating other cognitive processes behind creativity and their connections to each other is also a fruitful research area. Furthermore, creativity may play an important role in other 21<sup>st</sup> century skills; it would therefore be desirable to explore the nature of their connections in order to devise innovative and effective programmes in which 21<sup>st</sup> century skills can be developed simultaneously. In order to reach this aim, it is essential to assess these skills to monitor students' progress (Mayrath, Clarke-Midura, & Robinson, 2012). With regard to creativity there are many assessment techniques and methods with great prospects but practical constraints. Overcoming those constraints often requires innovative solutions. Technology allows for the development of new assessment tools to implement interaction and multimedia elements. The automated scoring and evaluating techniques offered by technology provide feasible solutions for data processing problems. Technology that provides innovative solutions for making assessments of creativity feasible in a number of contexts contributes not only to a better understanding of the nature of creativity but promotes its development and helps creative people to find activities where their potential may best be utilized.

### Acknowledgements

This research was supported by the European Union and the State of Hungary, co-financed by the European Social Fund within the framework of the TÁMOP 3.1.9 - 11/1 - 2012 - 0001 'Developing Diagnostic Assessments' project.

### References

- Bahar, A. K., & Maker, C. J. (2011). Exploring the relationship between mathematical creativity and mathematical achievement. *Asia-Pacific Journal of Gifted and Talented Education*, 3(1), 33-48.
- Barkóczi, I., & Klein, S. (1968). Gondolatok az alkotóképességről és vizsgálatának problémáiról. [Thoughts on creativity and concerns about its assessment] *Magyar Pszichológiai Szemle*, 25, 508-515.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Martin, R., Miller-Ricci, M., & Rumble, M. (2012). Defining Twenty-First Century Skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills*. (pp. 17-66). New York: Springer. doi: 10.1007/978-94-007-2324-5
- Cheung, P. C., Lau, S., Chan, D. W., & Wu, W. Y. H. (2004). Creative potential of school children in Hong Kong: Norms of the Wallach–Kogan Creativity Tests and their implications. *Creativity Research Journal*, 16(1), 69-78. doi: 10.1207/s15326934crj1601\_7
- Cheung, P. C., & Lau, S. (2010). Gender differences in the creativity of Hong Kong school children: Comparison by using the new electronic Wallach–Kogan creativity tests. *Creativity Research Journal*, 22(2), 194-199. doi: 10.1080/10400419.2010.481522

- COM (2010). *Europe 2020: A strategy for smart, sustainable and inclusive growth*. European Commission: Brussels.
- Cropley, A. J. (1967). Creativity, intelligence, and achievement. *Alberta Journal of Educational Research*, 13, 51-58.
- Csapó, B. (2010): Goals of learning and the organization of knowledge. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*. 56. Beiheft der Zeitschrift für Pädagogik (pp. 12-27). Weinheim: Beltz.
- Csapó, B., Ainley, J., Bennett, R., Latour, T. & Law, N. (2012). Technological issues of computer-based assessment of 21<sup>st</sup>-century skills. In B. McGaw, P. Griffin, & E. Care (Eds.), *Assessment and Teaching of 21<sup>st</sup>-century Skills* (pp. 143-230). New York: Springer. doi: 10.1007/978-94-007-2324-5\_4
- Csapó, B., Lőrincz, A., & Molnár, G. (2012). Innovative Assessment Technologies in Educational Games Designed for Young Students. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in game-based learning: foundations, innovations, and perspectives* (pp. 235-254). New York: Springer. doi: 10.1007/978-1-4614-3546-4
- Csapó, B., Molnár, G., & Kinyó, L. (2008, September 16-20). *Analysis of the selectiveness of the Hungarian educational system in international context*. Paper presented at the 3rd IEA International Research Conference, Taipei, Taiwan.  
Retrieved from  
[http://www.iea.nl/fileadmin/user\\_upload/IRC/IRC\\_2008/Papers/IRC2008\\_Csapo\\_Molnar\\_etal.pdf](http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2008/Papers/IRC2008_Csapo_Molnar_etal.pdf)
- Csíkós, C., & Csapó, B. (2011). Diagnostic assessment frameworks for mathematics: Theoretical background and practical issues. In B. Csapó & M. Szendrei (Eds.), *Framework for diagnostic assessment of mathematics* (pp. 137-162). Budapest: Nemzeti Tankönyvkiadó.
- Dumas, D., & Dunbar, K. N. (2014). Understanding Fluency and Originality: A latent variable perspective. *Thinking Skills and Creativity*, 14, 56-67. doi:10.1016/j.tsc.2014.09.003
- Florida, R. (2004). *The rise of the creative class... And how it's transforming work, leisure, community and everyday life*. New York: Basic Books. doi: 10.1111/j.1467-8691.2006.00398.x
- Funke, J. (2009). On the psychology of creativity. In P. Meusburger, J. Funke, & E. Wunder (Eds.), (2009). *Milieus of creativity: An interdisciplinary approach to spatiality of creativity (Vol. 2)* (pp. 11-23). Dordrecht: Springer Science & Business Media. doi: 10.1007/978-1-4020-9877-2
- Getzels, J. W. & Jackson, P. W. (1962): *Creativity and intelligence*. London: J. Wiley.
- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- Hargreaves, D. J., & Bolton, H. (1972). Selecting creativity tests for use in research. *British Journal of Psychology*, 63, 451-462. doi: 10.1111/j.2044-8295.1972.tb01295.x
- Husén, T. (Ed.). (1967). *International study of achievement in mathematics: A comparison of twelve countries* (Vols. 1-2). Stockholm: Almqvist & Wiksell. doi: 10.1007/BF01546609
- Idris, N., & Nor, N. M. (2010). Mathematical creativity: usage of technology. *Procedia-Social and Behavioral Sciences*, 2(2), 1963-1967. doi:10.1016/j.sbspro.2010.03.264



- Kardos, L., Pléh, C., & Barkóczi, I. (1987). *Studies in creativity*. Budapest: Akadémiai Kiadó. doi: 10.1017/S0033291700008655
- Karwowski, M., & Gralewski, J. (2013). Threshold hypothesis: Fact or artifact?. *Thinking Skills and Creativity*, 8, 25-33. doi:10.1016/j.tsc.2012.05.003
- Kim, K. H. (2006). Can we trust creativity tests? A review of the Torrance Tests of Creative Thinking (TTCT). *Creativity research journal*, 18(1), 3-14. doi: 10.1207/s15326934crj1801\_2
- Kwon, M., Goetz, E. T., & Zellner, R. D. (1998). Developing a computer-based TTCT: Promises and problems. *Journal of Creative Behavior*, 32(2), 96-106. doi: 10.1002/j.2162-6057.1998.tb00809.x
- Lau, S., Cheung, P., C. (2010). Creativity assessment: Comparability of the electronic and paper-and-pencil versions of the Wallach–Kogan Creativity Tests. *Thinking Skills and Creativity*, 5(3), 101-107. doi: 10.1016/j.tsc.2010.09.004
- 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*. (Ch. 9, pp. 129-145). Rotterdam: the Netherlands: Sense Publisher.
- Mann, E. L. (2006). Creativity: The essence of mathematics. *Journal for the Education of the Gifted*, 30(2), 236-260. doi: 10.4219/jeg-2006-264
- Mayer, R. E. (1999). Fifty Years of Creativity Research. In R. J. Sternberg (Ed.), *Handbook of Creativity* (pp. 449-460). London: Cambridge University Press. doi: 10.1017/CBO9780511807916.024
- Mayrath, M., Clarke-Midura J., & D. Robinson (2012). Introduction to technology-based assessments for 21st century skills. In M. C. Mayrath, J. Clarke-Midura, D. H. Robinson, & G. Schraw (Eds.), *Technology based assessment for 21st century skills: Theoretical and practical implications from modern research*. (pp. 1-13). New York: Springer-Verlag.
- Mullis, I. V. S., & Martin, M. O. (Eds.) (2013). *TIMSS 2015 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Muthén, L. K., & B. O. Muthén (2010). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Nadjafikhah, M., & Yaftian, N. (2013). The frontage of creativity and mathematical Creativity. *Procedia-Social and Behavioral Sciences*, 90, 344-350. doi: 10.1016/j.sbspro.2013.07.101
- Nadjafikhah, M., Yaftian, N., & Bakhshalizadeh, S. (2012). Mathematical creativity: some definitions and characteristics. *Procedia-Social and Behavioral Sciences*, 31, 285-291.
- OECD (2013). *PISA 2012 assessment and analytical framework: mathematics, reading, science, problem solving and financial literacy*. Paris: OECD. doi: 10.1787/9789264190511-en
- Palaniappan, A. K. (2012). Web-based Creativity Assessment System. *International Journal of Information and Education Technology*, 2(3), 255-258. doi: 10.7763/IJiet.2012.V2.123
- Piffer, D. (2012). Can creativity be measured? An attempt to clarify the notion of creativity and general directions for future research. *Thinking Skills and Creativity*, 7(3), 258-264. doi: 10.1016/j.tsc.2012.04.009
- Piirto, J. (2011). *Creativity for 21st century skills*. Sense Publishers. Rotterdam: Sense Publisher. doi: 10.1007/978-94-6091-463-8

- Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potentials, pitfalls, and future directions in creativity research. *Educational Psychologist, 39*(2), 83-96. . doi: 10.1207/s15326985ep3902\_1
- Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. In R. J. Sternberg (Ed.), *Handbook of Creativity* (pp. 35-62). London: Cambridge University Press. doi: 10.1017/CBO9780511807916.005
- Pólya, G. (1973). *How to solve it*. Princeton, NJ: Princeton University.
- Polya, G. (1981). *Mathematical discovery*. New York: John Wiley & Sons, Inc.
- Plucker, J. A., Qian, M., & Wang, S. (2011). Is originality in the eye of the beholder? Comparison of scoring techniques in the assessment of divergent thinking. *The Journal of Creative Behavior, 45*(1), 1-22. doi: 10.1002/j.2162-6057.2011.tb01081.x
- Pretz, J. E., & Link, J. A. (2008). The creative task creator: A tool for the generation of customized, Web-based creativity tasks. *Behavior research methods, 40*(4), 1129-1133. doi: 10.3758/BRM.40.4.1129.
- Primi, R., Ferrão, M. E., & Almeida, L. S. (2010). Fluid intelligence as a predictor of learning: A longitudinal multilevel approach applied to math. *Learning and Individual Differences, 20*(5), 446-451. doi:10.1016/j.lindif.2010.05.001
- Rosen, Y., & Tager, M. (2013). Computer-based performance assessment of creativity skills: a pilot study. *Pearson Research Report*. Retrieved October 03, 2014, from <http://researchnetwork.pearson.com/wp-content/uploads/CreativityAssessmentResearchReport.pdf>
- Runco, M. A. (2007). *Creativity: Theories and themes: Research, development, and practice*. Burlington: Elsevier Academic Press.
- Runco, M. A. (2011). Divergent thinking. In M. A. Runco, & S. R. Pritzker (Eds.), *Encyclopedia of creativity (Vol. 2)* (pp. 400-403). London: Elsevier Academic Press.
- Runco, M. A., & Acar, S. (2012). Divergent thinking as an indicator of creative potential. *Creativity Research Journal, 24*(1), 66-75. doi: 10.1080/10400419.2012.652929
- Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*(1), 92-96. doi: 10.1080/10400419.2012.650092
- 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. doi: 10.1207/s15326934crj1803\_5
- Shriki, A. (2010). Working like real mathematicians: Developing prospective teachers' awareness of mathematical creativity through generating new concepts. *Educational Studies in Mathematics, 73*(2), 159-179. doi: 10.1007/s10649-009-9212-2
- Silver, E. A. (1997). Fostering creativity through instruction rich in mathematical problem solving and problem posing. *Zentralblatt für Didaktik der Mathematik, 29*(3), 75-80. doi: 10.1007/s11858-997-0003-x
- Simonton, D. K. (2012). Taking the US Patent Office criteria seriously: A quantitative three-criterion creativity definition and its implications. *Creativity Research Journal, 24*(2-3), 97-106. doi: 10.1080/10400419.2012.676974
- Snyder, A., Mitchell, J., Bossomaier, T., & Pallier, G. (2004). The creativity quotient: an objective scoring of ideational fluency. *Creativity Research Journal, 16*(4), 415-419. doi: 10.1080/10400410409534552

- Sriraman, B. (2004). The characteristics of mathematical creativity. *Mathematics Educator*, 14(1), 19-34.
- Sriraman, B. (2005). Are giftedness and creativity synonyms in mathematics? *Prufrock Journal*, 17(1), 20-36. doi: 10.4219/jsge-2005-389
- Sriraman, B., & Lee, K. (Eds.) (2011). The elements of creativity and giftedness in mathematics. Rotterdam: Sense Publishers. doi: 10.1007/978-94-6091-439-3
- Torrance, E. P. (1966). *Torrance Tests of Creative Thinking*. IL: Scholastic Testing Service, Bensenville, IL.
- Villalba, E. (2009). Computer-based Assessment and the Measurement of Creativity in Education In F. Schueremann & J. Bjornsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large scale assessment* (pp. 29-37). Brussels: European Communities. doi: 10.2788/60083
- Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction*. New York: Holt, Rinehart and Winston.
- Xin, Z., & Zhang, L. (2009). Cognitive holding power, fluid intelligence, and mathematical achievement as predictors of children's realistic problem solving. *Learning and Individual Differences*, 19(1), 124-129. doi:10.1016/j.lindif.2008.05.006