JOURNAL OF
Creative Behavior
Publication of the Creative Education Foundation™

EVELINE M. SCHOEVERS
EVELYN H. KROESBERGEN
MARIA KATTOU

Mathematical Creativity: A Combination of Domain-general Creative and Domain-specific Mathematical Skills

ABSTRACT

Creativity is an understudied topic in elementary school mathematics research. Nevertheless, we argue that creativity plays an important role in mathematics, but that more research is needed to understand this relation. Therefore, this study aimed to investigate this relation, specifically between domain-general creativity, domain-specific mathematical creativity, and mathematical ability. Measures for these constructs were administered to 342 Dutch fourth graders. In order to examine the nature of the relation between creativity and mathematics, two competing models were tested, using Structural Equation Modeling. The results indicated that models in which general creativity and mathematical ability both predict mathematical creativity fitted the data better than models in which mathematical and general creativity predict mathematical ability. This study showed that both general creativity and mathematical ability are important to think creatively in mathematics.

Keywords: mathematical creativity, mathematical ability, domain-general creativity.

Creativity is an increasingly important aspect of personal functioning in several sectors of contemporary society (Sternberg & Lubart, 1999). Creativity seems to be related to mathematics, since it is required when a student or mathematician faces a mathematics problem for which there is no learned solution (Leikin & Pitta-Pantazi, 2013). However, due to a lack of insight into the nature of creativity, in particular the relation between domain-general and domain-specific creativity, it is not yet clear how creativity and mathematics are related. The aim of this study was therefore to investigate the relations between domain-general creativity (GC), domain-specific mathematical creativity (MC), and mathematical ability (MA) in a new and integrated way. This provides further insight into the nature of (mathematical) creativity and may inform the current literature on mathematical learning and teaching with new insights regarding the role of creativity.

Creativity is a multidimensional construct, which may or may not be domain-specific. Therefore, we will first shortly discuss the nature, definition, and measurement of creativity. With regard to the nature of creativity, some researchers argue that creativity is a domain-general ability, because creative processes are similar across domains (e.g., Plucker, 1999). Others, however, state that creativity is always related to a specific domain (e.g., mathematics), because a certain degree of knowledge or expertise within a particular content domain is required for creativity (e.g., Baer, 2012). As a result, researchers have begun to investigate domain-related creativity, like MC. Support was found for both views (e.g., Huang, Pen, Chen, Tseng, & Hsu, 2017; Jeon, Moon, & French, 2011; Plucker, 1999, 2004), suggesting that creativity may be partly domain-general and domain-specific. In this study, we have investigated both GC and MC in relation to mathematical ability.

GC is defined as "the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context" (Plucker, Beghetto, & Dow, 2004, p. 90). Measures that are often used to measure GC are, for example, the Torrance Test of Creative Thinking (TTCT; Torrance, 2008) and the Test of Creative Thinking – Drawing Production (TCT-DP; Urban & Jellen, 1996). However, it is questionable whether the TTCT and TCT-DP indeed measure GC, since domain-specific features are involved (e.g., verbal or figural features). Furthermore, these tests may not measure the whole construct of GC. Therefore, it is recommended to use different

The Journal of Creative Behavior, Vol. 54, Iss. 2, pp. 242–252 © 2018 The Authors. *The Journal of Creative Behavior* published by Wiley Periodicals, Inc. on behalf of Creative Education Foundation (CEF) DOI: 10.1002/jocb.361

measurements instead of relying on a single score when measuring students' creativity (Cropley, 2010; Kim, 2006; Treffinger, Young, Selby, & Shepardson, 2002).

MC is defined as "(a) the process that results in unusual (novel) and/or insightful solution(s) to a given problem or analogous problems, and/or (b) the formulation of new questions and/or possibilities that allow an old problem to be regarded from a new angle requiring imagination" (Sriraman, 2005, p. 24). MC is most often assessed by using a multiple-solution task, in which students can provide several solutions to a mathematical problem (Leikin, 2009).

RELATIONS BETWEEN GC, MC, AND MA

From the current literature on creativity and mathematics, it can be inferred that GC, MC, and MA are related. However, it is still ambiguous how they are related. Two different patterns can be hypothesized. First, it could be hypothesized that MA and GC both predict MC. The definition and assessment of MC suggests that MA is related to MC because a certain level of mathematical knowledge is necessary to be creative in mathematics (Sak & Maker, 2006; Weisberg, 1999). This positive relation is indeed supported by several studies (Huang et al., 2017; Mann, 2005; Sak & Maker, 2006; Weisberg, 1999). GC is expected to be related to MC since general creative processes are similar across domains (Plucker, 1999). Some studies have indeed found a connection between GC and MC (Hwang, Lee, & Seo, 2005; Jeon et al., 2011; Kattou & Christou, 2013; Kroesbergen & Schoevers, 2017). It should be noted that MC was measured by a multiple-solution tasks or math teachers' ratings of math creativity and GC by divergent thinking tasks and the TCT-DP.

Few have investigated the hypothesis that both GC and MA influence MC, simultaneously in a single study (Huang et al., 2017; Jeon et al., 2011; Kattou & Christou, 2013). However, the findings of these studies are not univocal. The study by Jeon et al. (2011) found, using regression analysis, that both GC, measured by divergent thinking, and mathematical performance predicted MC. Nevertheless, mathematical performance explained more variance (10%) in MC than GC (3%), indicating that in a structured domain like mathematics, domain knowledge is more important for MC than GC is. Kattou and Christou (2013) found similar results, although in their study MA and GC, measured by divergent thinking, were equally strong predictors of MC. In contrast, Huang et al. (2017) found that only MA, and not divergent thinking, predicted MC.

Second, it could also be argued that MC influences MA, and that MC in turn is influenced by GC, which indirectly influences MA. Recent research suggests that MC is a prerequisite for the development of high levels of MA (Kattou, Kontoyianni, Pitta-Pantazi, & Christou, 2013). Several other studies also showed that MC predicted MA: the more creative a child is in mathematics, the higher his/her performance in mathematics is (Bahar & Maker, 2011; Kattou et al., 2013; Leikin, 2007). Indeed, as an individual tries to find multiple solutions, she/he considers mathematical ideas from different perspectives, which leads to deeper mathematical knowledge (Leikin, 2007). Furthermore, it can be expected that GC influences MC (Hwang et al., 2005; Jeon et al., 2011; Kattou & Christou, 2013; Kroesbergen & Schoevers, 2017) and thus indirectly MA. It is not expected that GC is directly related to MA since these general processes may not be intrinsically related to MA (Baran, Erdogan, & Çakmak, 2011; Livne & Milgram, 2006).

RESEARCH GOALS AND HYPOTHESES

Given that there have been few previous studies and their results are mixed and inconclusive, the question remains of how GC, MC, and MA are related. To deepen the insight into these relations, we used more

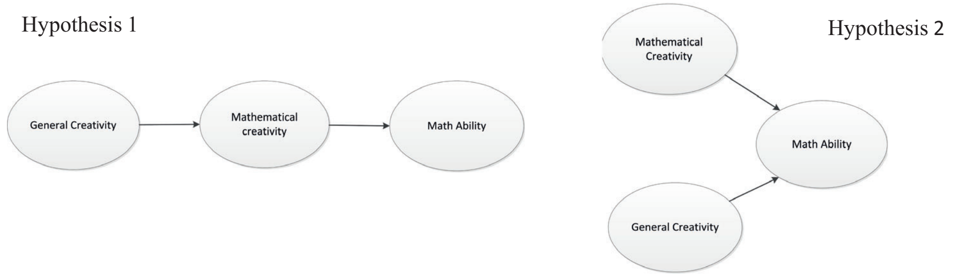

FIGURE 1. Simplified visual representations of hypotheses 1 & 2.

than one measure for GC to better estimate the latent construct; furthermore, two competing models derived from the literature were tested. We considered the following two possible competing hypotheses.

First, it was hypothesized that MC influences MA (e.g., Kattou et al., 2013) and that GC in turn directly influenced MC but only indirectly influenced MA (see Hypothesis 1, Figure 1). Second, it was hypothesized (see Hypothesis 2, Figure 1) that MC was influenced by both MA and GC (e.g., Bahar & Maker, 2011; Hong & Milgram, 2010; Jeon et al., 2011).

METHOD

PARTICIPANTS

In this study, 342 fourth-grade students participated, from 18 classes of 12 elementary schools. Twelve schools from medium- to large-sized towns in the Netherlands participated. Schools differed with regard to their policies and teaching methods used in class, and were located in various districts containing citizens of low, middle, and high socioeconomic status. Students in this sample were 50% boys and had a mean age of 9.68 years ($SD = 0.45$).

Prior to the data collection, a power analysis, performed with online software called Sloper (2015), indicated that this study required a sample size of at least 305, with an anticipated effect size of 0.15, desired power of 0.8, 10 latent variables and 34 observed variables, and a probability level of 0.05. Furthermore, Kline (2010) indicates that a minimum of 10 cases per variable are required for Structural Equation Modeling (SEM). This indicates that with 34 observed variables we have a large enough sample size for SEM (Kline, 2010; Sloper, 2015).

INSTRUMENTS

Intelligence Quotient (IQ)

Raven's Standard Progressive Matrices (SPM; Raven, 1998) was used to get an indication of the non-verbal intelligence of the students. In each of the 60 test items, the subject is asked to identify the missing element that completes a pattern. The test measures the reasoning ability of students and is a measure of non-verbal intelligence. The mean score on Raven's SPM in this study was 101.50 ($SD = 14.82$), which was based on Dutch norms (Van de Weijer-Bergsma, Kroesbergen, Prast, & Van Luit, 2014). A sample question is shown in Figure 2. With regard to predictive validity, Raven's SPM predicted mathematical ability ($r = .53$) in this study, which is comparable to other studies (e.g., Neisser et al., 1996).

Mathematical Creativity Test (MCT)

The MCT, developed by Kattou et al. (2013), was translated into Dutch for this study by following the steps for a good translation process described by Beaton, Bombardier, Guillemin, and Bos Ferraz (2000).

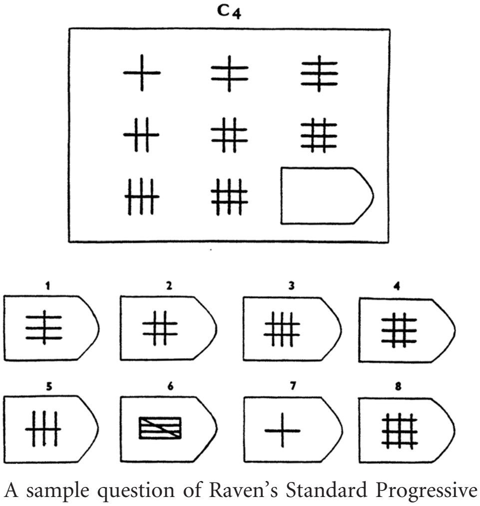

FIGURE 2. A sample question of Raven's Standard Progressive Matrices.

Look at this number pyramid. Each cell contains only one number. Each number in the pyramid can be calculated by always performing the same operation with the two numbers that appear below it. Complete the missing numbers in the pyramid, by keeping number 35 on the top cell of the pyramid. Find as many solutions as possible.

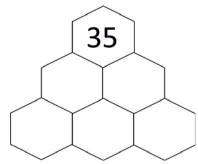

FIGURE 3. A sample question of the MCT.

The MCT took a maximum of 45 min and consisted of five questions, which were open-ended and could have multiple solutions. Students were required to provide multiple solutions, original and distinct from each other (Kattou et al., 2013). A sample question is displayed in Figure 3.

Scores were obtained for fluency (number of correct solutions), flexibility (number of different types or categories of correct solutions), and originality for each question. For detailed scoring guidelines, see Kattou et al. (2013). The internal consistency for the MCT is high (α = .80; Kline, 1999).

Tests of Domain-General Creativity

The TCT-DP (Urban & Jellen, 1996) and the Dutch version of the TTCT (Torrance, 2008) were used to measure students' general creative potential.

TCT-DP

The TCT-DP Form A was used and took 15 min. This test mirrors a more holistic concept of creativity. Students had to complete a drawing using certain figural fragments, such as a half circle and a half square, which was scored according to the guidelines (Urban & Jellen, 1996). A total score was obtained by adding the scores of 13 categories and transforming them into *z*-scores. The TCT-DP has good inter-rater reliability: α = .81–.99 for the total score and $\alpha \geq .89$ for test criteria (Urban & Jellen, 1996).

TTCT

The TTCT measures divergent thinking with words and pictures. Each activity took 10 minutes. Activities 5 (unusual uses) and 7 (just suppose) were used from the verbal test (version A). Activity 5 requires students to write down as many alternative uses for a cardboard box that they can think of. Activity 7 requires students to hypothesize about an improbable situation. Activities 2 (picture completion) and 3 (repeated lines) were used from the figural test (version A). Activity 2 requires students to draw pictures using ten incomplete figures as a starting point, to which titles are added. Activity 3 consists of three pages of sets of parallel lines, and students must draw something using these parallel lines as part of their picture. These four activities were used because they require different forms of divergent thinking. Activity 7, for example, requires more use of imagination than activity 5 (Torrance, 2008). Both tests were scored according to the guidelines (Torrance, 2008). For all activities, scores were obtained for fluency, flexibility, and originality. Additionally, for activities 2 and 3 (TTCT Figural), scores were obtained for elaboration, abstractness of titles, and resistance to premature closure. Raw scores were transformed into *z*-scores. The internal consistency was good for the TTCT Verbal in this sample (α = .75), but questionable for the TTCT Figural (α = .61; Kline, 1999).

Test of MA

Scores from the widely used standard Dutch mathematical achievement test (Janssen, Scheltens, & Kraemer, 2007) were used as a measure for MA. We used the test that was designed for grade 4. All subscales from the math test were used: "number and number relations," "mental arithmetic," "estimation arithmetic," "arithmetical operations," "geometry," "arithmetic with time and money," and "proportions, fractions and percentages." For each student, the percentage correct on the subscales was calculated. The questions on the math test are mainly math word problems. A sample question from the MA test is the following: "Cycle racers have to cycle 5 rounds of 18 kilometers. How many kilometers do they have to cycle in total?" Cronbach's alpha for the math test was excellent (α = .94) in this study (Kline, 1999).

PROCEDURE

Data were collected in the fall of 2014 by four master's students, each with a bachelor's degree in special education, supervised by the first author. Informed consent was obtained from the parents or guardians of all children involved. Information about students' age, gender, and socioeconomic status was obtained from school records. The measures of creativity were part of a larger test battery, administered in two sessions, each lasting 90 minutes. All tests were administered in a classroom setting by one or two proctors. Test instructions were read aloud. Students were not allowed to copy the work of their fellow students or to talk during test sessions.

All tests were scored by the same master's students. For the MCT, TTCT, and TCT-DP, the inter-rater reliability (IRR) was determined. For all variables, sufficient to good agreement was reached (Cicchetti, 1994). Unfortunately, no agreement was reached on the variable resistance to premature closure of the TTCT Figural activity 2 (ICC = .28). This variable was excluded from the analyses.

ANALYSES

Prior to the data analyses, assumptions for SEM were checked in SPSS Statistics (IBM corporation, 2013). Next, data were analyzed by testing different models, using SEM in M*plus* version 7.2 (Muthén & Muthén, 2012). As a result of checking our assumptions, we decided to use the MLM estimator in M*plus* to take non-normality into account (Tabachnick & Fidell, 2013).

First, separate confirmatory factor analyses (CFAs) were conducted for MA and MC. Second, an explanatory factor analysis (EFA) was conducted to test whether one latent construct for GC could be created. Third, correlations between MC, MA, and GC were computed to get insight into the existence and strength of the relations. Fourth, the two hypotheses were tested by examining the two competing models (Model 1 & 2); comparative model fit was evaluated. In Model 1, GC influenced MC and MC influenced MA. In Model 2, it was hypothesized that both MA and GC influenced MC. In these models, we controlled for IQ and gender since we expected that these variables could influence the relations between GC, MC, and MA. The covariate gender was added to the models because boys and girls score significantly different on mathematical ability tests (Preckel, Goetz, Pekrun, & Kleine, 2008) and mathematical creativity tests (Mann, 2005). Regarding the influence of gender on domain-general creativity tests, research has found inconsistent results (Baer & Kaufman, 2008). The covariate IQ was added to the models because IQ is positively related with school performance (Laidra, Pullmann, & Allik, 2007) and creativity (Kim, 2005).

RESULTS

Descriptive statistics of all variables used can be found in Appendix S1.

MA

A CFA with one factor was examined for MA, using the seven subscales of mathematical performance as observed variables to test the unidimensionality of MA. The CFA indicated that the model fitted well (CFI = 0.99; TLI = 0.99; χ^2 = 33.83, df = 14, p = .002; RMSEA = 0.06, SRMR = 0.01).

MC

A second-order CFA was conducted for MC, with the 15 subscales of MC as observed variables, and with fluency, flexibility, and originality as first-order latent variables. Covariances between the error of the abilities (fluency, flexibility, and originality) were added because the variables were highly correlated per question. This was expected because flexibility and originality scores are likely to be higher when more answers are provided (high fluency score; Torrance & Safter, 1999). Furthermore, we obtained a negative residual variance for the latent variable originality in the model. This negative variance is probably caused by outliers in the data (Bollen, 1987). Outliers were not deleted since they were deemed realistic scores. Since a negative variance is not possible, we scaled the variance to zero, which is the most closely related possible value. Results indicated that the model fitted well (CFI = .98; TLI = .98; χ^2 = 106.22, df = 73, p = .007; RMSEA = .04; SRMR = .04).

GC

For the latent variable GC, an EFA was applied using all the variables of the TTCT (Figural and Verbal) and the total score of the TCT-DP. The default setting was used (Geomin oblique rotation). An EFA was chosen because it was not clear how the TTCT and TCT-DP were related. The EFA indicated that a five-factor model would fit best with the following factors, which theoretically made sense: (a) TTCT Verbal activity 5, (b) TTCT Verbal activity 7, (c) TTCT Figural activity 2 (only fluency and originality), (d) TTCT Figural activity 3 (only fluency and originality), and (e) TCT-DP total score and TTCT Figural title and elaboration activities 2 and 3. However, not all factors were significantly correlated. After GC was added as a second-order factor, only factors 1, 2, 3, and 4 were significant indicators of GC. The insignificant factor was deleted from the model, and the negative residual variances of fluency (TTCT Verbal activity 7) and originality (TTCT Figural activity 2) were scaled to zero. This model had a good fit (CFI = .97; TLI = .96; $\chi^2 = 70.87$, $df = 33$, $p < .001$; RMSEA = .06; SRMR = .04). Theoretically, it makes sense to have factors of each activity because fluency, flexibility, and originality are often highly correlated per activity and are therefore measuring almost the same; it is more likely that flexibility and originality scores are higher when more answers are provided (a high fluency score; Torrance & Safter, 1999). Currently, each activity gives an indication of divergent thinking. In fact, the factor GC represents "generating ideas" rather than the more complex phenomenon creativity.

The other factor, measured by the TCT-DP and by "abstractness of title" and "elaboration" of the TTCT Figural activities 2 and 3, represents another measure of GC, which is, however, correlated with divergent thinking (see Table 1). This factor might be related to "deeper digging into ideas" and the "openness and courage to explore ideas" (Treffinger et al., 2002). Models 1 and 2 were examined with both measures of GC separately since adding both measures of GC in one model would reduce the power of the study.

An overview of the inter-correlations of the obtained factors and their relations with the covariates are given in Table 1, which provide an indication of the existence and strength of the relationships. Regarding the existence of the relation between gender and the other variables, we used an independent *t*-test and Mann–Whitney's *U* test. These results indicated that there are no significant differences on the factor scores of MC ($t = 1.77$, $p = .08$), MA ($t = -1.24$, $p = .22$), and IQ ($t = 1.27$, $p = .20$) for boys and girls. There are, however, significant differences in gender on "generating ideas" (GC1; $U = 8719$, $p < .01$, $r = -0.17$) and "explore and dig deeper into ideas" (GC2; $U = 7605.50$, $p < .001$, $r = -.26$), with girls scoring higher than boys.

To study how MC, GC, and MA are related, we tested two different models, namely Model 1 and Model 2. Because the EFA indicated that GC was not a unitary construct, but represented two different constructs, namely "Generating ideas (GC1)" and "Explore and dig deeper into ideas (GC2)," we decided to test Models 1 and 2 separately with GC1 and GC2. Furthermore, we added the covariates IQ and gender in the models, predicting the observed variables of MA, MC, and GC. Insignificant paths between the covariates and observed variables were deleted.

MODEL 1 & 2 WITH GC1

It was tested whether creative thinking in mathematics predicted mathematical performance, and whether divergent thinking in general (GC1) predicted MC but not MA directly (Model 1) or whether divergent thinking in general (GC1) and MA predicted MC (Model 2). In Model 1, MC was a significant predictor of MA ($r = .27$, $p < .001$) and GC1 a significant predictor of MC ($r = .35$, $p = <.01$). This model had a good

TABLE 1. Spearman Correlations between the Factor Scores of MC, GC, and Mathematical Performance and the Covariates Gender and IQ

	1	2	3	4	5
1. MC total (factor score)	–				
2. GC1 ('generating ideas')	.64*	–			
3. GC2 ('explore and dig deeper into ideas')	.55*	.64*	–		
4. MA total (factor score)	.70*	.15	.17	–	
5. IQ (Raven SPM)	.55*	.25*	.25*	.57*	–

Note. *$p < .005$ (Bonferroni correction applied (0.05/((5*4)/2)).

fit (CFI = .96; TLI = .95; χ^2 = 714.48, *df* = 481, $p < .001$; RMSEA = .04; SRMR = .06). In Model 2, GC1 was a significant predictor of MC (r = .40, p = <.01) and MA a significant predictor of MC (r = .34, $p < .001$). This model also had a good fit (CFI = .96; TLI = .96; χ^2 = 706.01, *df* = 481, $p < .001$; RMSEA = .04; SRMR = .06). Model 2 had a lower Akaike information criterion (AIC) and Bayesian information criterion (BIC; AIC = 296.65; BIC = 826.30) compared to Model 1 (AIC = 301.85; BIC = 831.49), indicating that Model 2 fitted best (Muthén & Muthén, 2012). This model is shown in Figure 4.

MODEL 1 & 2 WITH GC2

It was tested whether MC predicted MA, and whether "Explore and dig deeper into ideas" in general (GC2) predicted MC but not MA directly (Model 1) or whether "Explore and dig deeper into ideas" (GC2) and MA predicted MC (Model 2). In Model 1, MC was a significant predictor of MA (r = .38, p = <.001) and GC2 a significant predictor of MC (r = .27, p = <.01). This model had a good fit (CFI = .97; TLI = .96; χ^2 = 452.34, *df* = 330, $p < .001$; RMSEA = .04; SRMR = .06). In Model 2, GC2 was a significant predictor of MC (r = .31, p = <.001) and MA a significant predictor of MC (r = .41, p = <.001). This model also had a good fit (CFI = .97; TLI = .96; χ^2 = 448.86, *df* = 330, $p < .001$; RMSEA = .04; SRMR = .06). Model 2 had a lower AIC and BIC (AIC = −1857.89; BIC = −1383.16) compared to Model 1 (AIC = −1854.78; BIC = −1380.04), indicating that Model 2 fitted best (Muthén & Muthén, 2012). Model 2 is shown in Figure 5.

DISCUSSION

This study aimed to provide insight into the role of creativity in mathematical ability in fourth-grade students by examining the relations between MC, GC, and MA. Two competing hypotheses were tested using SEM.

A crucial first result is that GC was not a unitary construct, but consisted of two different constructs, namely "Generating ideas (GC1)" and "Explore and dig deeper into ideas (GC2)." This result suggests that either one instrument cannot capture the general measure of creativity or one of these instruments might not measure (an element of) GC. This finding highlights the importance of careful use of instruments that attempt to assess a "general creative ability." This result is in line with recommendations of other

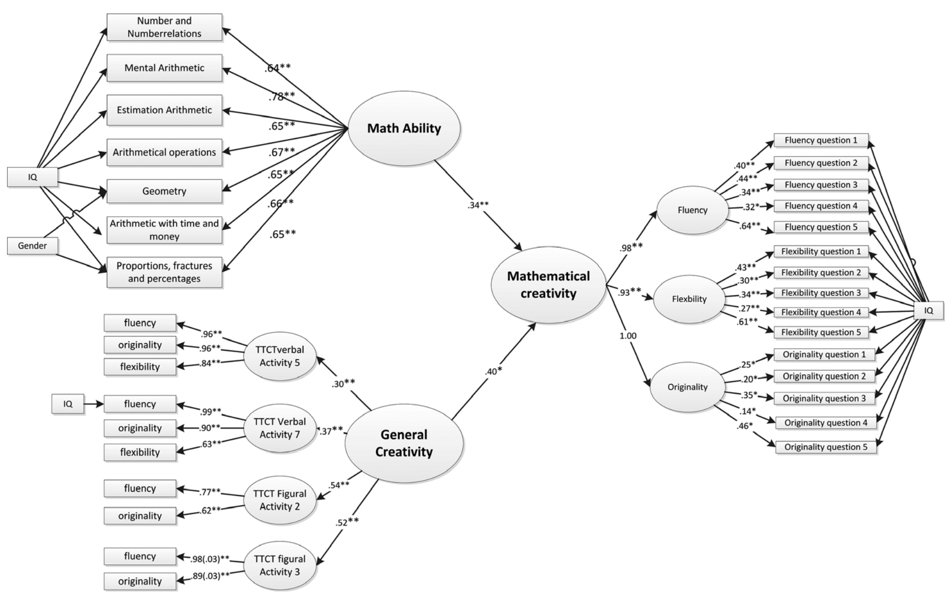

FIGURE 4. Standardized factor loadings of Model 2 (with GC representing 'generating ideas') with covariates gender and IQ. *Note.* Covariances of the observed variables of MC are not visualized and IQ is visualized multiple times in the model to make the image more clear. $^{**}p < .001$, $^{*}p < .05$.

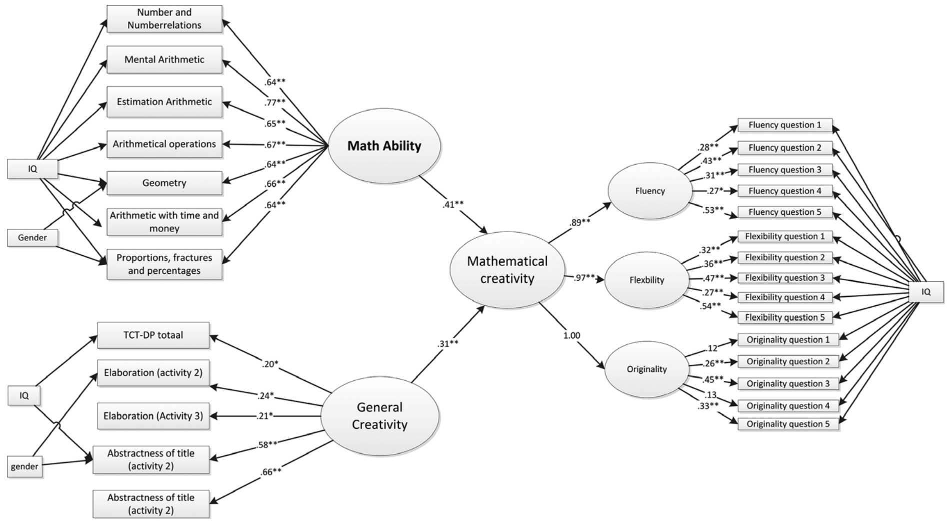

FIGURE 5. Standardized factor loadings of Model 2 (with GC representing 'explore and dig deeper into ideas') with covariates gender and IQ. *Note*. Covariances of the observed variables of MC are not visualized and IQ and gender are visualized multiple times in the model to make the image more clear. $^{**}p < .001$, $^{*}p < .05$.

researchers to use multiple measures of creativity (Cropley, 2010; Kim, 2006; Treffinger et al., 2002). Since analyses indicated that GC was not a unitary construct, relations between MC, GC1, and MA, and those between MC, GC2, and MA are discussed separately.

Regarding the relationship between GC1 ("Generating ideas"), MC, and MA, we found most support for the second hypothesis, that MA and GC1 both influence MC. Although no causal direction could be established with this study, the results suggest that divergent thinking and mathematical knowledge are almost equally important for MC. In order to think divergently, students need to combine and reorganize existing concepts to generate new concepts and ideas (Mumford, Baughman, Maher, Costanza, & Supinski, 1997). This requires cognitive flexibility, which is also an important capacity in MC. Existing mathematical concepts are combined and reorganized to generate new and multiple mathematical solutions. This result is in line with the studies of Kattou and Christou (2013) and Jeon et al. (2011), although there were some small differences regarding the strength of the predictors. For example, the variance accounted for by MA (r^2 = 10%) appeared to be larger compared to that accounted for by GC (r^2 = 3%) in the study by Jeon et al. compared to our study. Our result is in contrast to the findings of Huang et al. (2017), which showed that MA was a strong predictor of MC, but that GC was not a significant predictor of MC. These small and large differences between our study and others may be caused by the diverse measures used. We used a multiple-solution task as a measure of MC that mainly required knowledge of arithmetical operations and number relations – knowledge already mastered by most students. Therefore, GC1 ("generating ideas") may have played a slightly stronger role in MC than in MA. Huang et al. (2017) used an MC task that was rather more difficult and required a higher level of MA. For future studies, it would be interesting to study in more depth how MC and GC are related to MA, by focusing, for example, on specific task aspects.

Concerning the relation between GC2 ("Explore and dig deeper into ideas"), MC, and MA, we also found most support for the second hypothesis: MA and GC2 both influenced MC. This result is in agreement with the findings of Kroesbergen and Schoevers (2017), which similarly showed that GC (measured by the TCT-DP) was a predictor of MC. However, this is the first study that used this instrument simultaneously with measures of MC and MA in one model. This finding could therefore be significant for the literature. Contrary to our result regarding GC1, MA was a stronger predictor of MC than GC2. The reason for this may be that "Explore and dig deeper into ideas" requires both divergent and convergent thinking (Treffinger et al., 2002). Since the measurement of MC (i.e., a multiple solution task) is more closely related to divergent thinking than to convergent thinking, it may explain why GC2 had a smaller influence on MC than MA.

When interpreting the results of this study, the reader should also take into account that the sample size used in this study was rather small for SEM analyses. Therefore, it was not possible to take the multilevel structure of the data into account (Hox, 2010). We recommend that future studies do so, which would require larger samples.

In conclusion, despite the crucial finding that GC is not a unitary construct, both our results regarding the relation between GC (1 and 2), MC, and MA give more support to the hypothesis that both GC (1 and 2) and MA predict MC. The amount of influence of a component of GC (i.e., "generating ideas" or "explore and dig deeper into ideas") and MA on MC seems to depend on the instruments that are used. This result is important to take into account in (designing) research on mathematical learning and MC. Careful use of tests that attempt to measure GC is recommended.

With regard to the implication of this research for educational practice, this study suggests that in order to creatively solve mathematical problems, both mathematical knowledge and general creative thinking skills are needed. Teachers should be aware of both components when promoting students' mathematical creativity.

REFERENCES

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

Baer, J., & Kaufman, J.C. (2008). Gender differences in creativity. *The Journal of Creative Behavior*, *42*, 75–105. https://doi.org/10.1002/j.2162-6057.2008.tb01289.x.

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*, 33–48 Retrieved from http://apfgifted.org/Tpl/default/Public/Journal/2011/Mathematical_Creativity_and_Achievement.pdf.

Baran, G., Erdogan, S., & Çakmak, A. (2011). A study on the relationship between six-year-old children's creativity and mathematical ability. *International Education Studies*, *4*, 105–111. https://doi.org/10.5539/ies.v4n1p105.

Beaton, D.E., Bombardier, C., Guillemin, F., & Bos Ferraz, M. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*, 3186–3191. https://doi.org/10.1097/00007632-200012150-00014.

Bollen, K.A. (1987). Outliers and improper solutions: A confirmatory factor analysis example. *Sociological Methods and Research*, *15*, 375–384. https://doi.org/10.1177/0049124187015004002.

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*, 284–290. https://doi.org/10.1037//1040-3590.6.4.284.

Cropley, A.J. (2010). Defining and measuring creativity: Are creativity tests worth using? *Roeper Review*, *23*, 72–79. https://doi.org/10.1080/02783190009554069.

Hong, E., & Milgram, R.M. (2010). Creative thinking ability: Domain generality and specificity. *Creativity Research Journal*, *22*, 272–287. https://doi.org/10.1080/10400419.2010.503535.

Hox, J.J. (2010). *Multilevel analysis. Techniques and applications*. New York: Routledge.

Huang, P., Pen, S., Chen, H., Tseng, L., & Hsu, L. (2017). The relative influences of domain knowledge and domain-general divergent thinking on scientific creativity and mathematical creativity. *Thinking Skills and Creativity*, *25*, 1–9. https://doi.org/10.1016/j.tsc.2017.06.001.

Hwang, D.J., Lee, K.S., & Seo, J.J. (2005). Relationship between divergent thinking in mathematical and non-mathematical situations-based on the TTCT; Figural A and the MCPSAT. *Journal of Gifted/Talented Education*, *15*, 59–76 Retrieved from http://ocean.kisti.re.kr/downfile/volume/ksg/OJHHBM/2005/v15n2/OJHHBM_2005_v15n2_59.pdf.

IBM Corporation (2013). *IBM SPSS Statistics for Windows. Version 22.0*. Armonk, NY: IBM Corporation

Janssen, J., Scheltens, F., & Kraemer, J. (2007). *Rekenen-wiskunde. Handleiding. [Arithmetic-Mathematics. Manual]*. Arnhem, The Netherlands: Cito.

Jeon, K.N., Moon, S.M., & French, B. (2011). Differential effects of divergent thinking, domain knowledge, and interest on creative performance in art and math. *Creativity Research Journal*, *23*, 60–71. https://doi.org/10.1080/10400419.2011.545750.

Kattou, M., & Christou, C. (2013). Investigating the effect of general creativity, mathematical knowledge and intelligence on mathematical creativity. Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education. Retrieved from http://www.researchgate.net/publication/261286953

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

Kim, K.H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Advanced Academics*, *16*, 57–66. https://doi.org/10.4219/jsge-2005-473.

Kim, K.H. (2006). Can we trust creativity tests? A review of the Torrance Test of Creative Thinking (TTCT). *Creativity Research Journal*, *18*, 3–14. https://doi.org/10.1207/s15326934crj1801_2.

Kline, P. (1999). *The handbook of psychological testing* (2nd edn). London, UK: Routledge.

Kline, R.B. (2010). *Principals and practice of structural equation modeling*. New York: Guilford Publication.

Kroesbergen, E.H., & Schoevers, E.M. (2017). Creativity as predictor of mathematical abilities in fourth graders in addition to number sense and working memory. *Journal of Numerical Cognition*, *3*, 417–440. https://doi.org/10.5964/jnc.v3i2.63.

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, *42*, 441–451. https://doi.org/10.1016/j.paid.2006.08.001.

Leikin, R. (2007). *Habits of mind associated with advanced mathematical thinking and solution spaces of mathematical tasks*. In D. Pitta-Pantazi, & G. Philippou (Eds.), Proceedings of the fifth conference of the European Society for Research in Mathematics Education—CERME-5 (pp. 2330–2339). Retrieved from http://www.mathematik.uni-dortmund.de/~erme/CERME5b/WG14.pdf#page=112

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

Leikin, R., & Pitta-Pantazi, D. (2013). Creativity and mathematics education: The state of the art. *ZDM*, *45*, 159–166. https://doi.org/10.1007/s11858-012-0459-1.

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

Mann, E. L. (2005). Mathematical creativity and school mathematics: indicators of mathematical creativity in middle school students (Doctoral dissertation). Retrieved from http://www.gifted.uconn.edu/siegle/Dissertations/Eric%20Mann.pdf

Mumford, M.D., Baughman, W.A., Maher, M.A., Costanza, D.P., & Supinski, E.P. (1997). Process-based measures of creative problem-solving skills: IV. Category combination. *Creativity Research Journal*, *10*, 59–71. https://doi.org/10.1207/s15326934crj1001_7.

Muthén, L.K., & Muthén, B.O. (2012). *Mplus user's guide* (7th edn). Los Angeles: Muthén & Muthén.

Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A.W., Brody, N., Ceci, S.J., & Urbina, S. (1996). Intelligence: Knowns & unknowns. *American Psychologist*, *51*, 77–101. https://doi.org/10.1037/0003-066X.51.2.77.

Plucker, J. (1999). Reanalyses of student responses to creativity checklists: Evidence of content generality. *Journal of Creative Behavior*, *33*, 126–137. https://doi.org/10.1002/j.2162-6057.1999.tb01042.x.

Plucker, J. (2004). Generalization of creativity across domains: Examination of the method effect hypothesis. *Journal of Creative Behavior*, *38*, 7–12. https://doi.org/10.1002/j.2162-6057.2004.tb01228.x.

Plucker, J., 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*, 83–96. https://doi.org/10.1207/s15326985ep3902_1.

Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender differences in gifted and average-ability students comparing girls' and boys' achievement, self-concept, interest, and motivation in mathematics. *Gifted Child Quarterly*, *52*, 146–159. https://doi.org/10.1177/0016986208315834.

Raven, J.C. (1998). *Manual section 3 Standard progressive matrices*. Oxford, UK: Oxford Psychologists Press Ltd.

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*, 279–291. https://doi.org/10.1207/s15326934crj1803_5.

Sloper, D.S. (2015). A-priori Sample Size Calculator for Structural Equation Models [Software]. Retrieved from http://www.danielsoper.com/statcalc

Sriraman, B. (2005). Are giftedness and creativity synonyms in mathematics? *The Journal of Advanced Academics*, *17*, 20–36. https://doi.org/10.4219/jsge-2005-389.

Sternberg, R.J., & Lubart, T.I. (1999). The concept of creativity: Prospects and Paradigms. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 3–16). London, UK: Cambridge University Press.

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics* (6th edn). New Jersey, NJ: Pearson.

Torrance, E.P. (2008). *Torrance test of creative thinking*. Benseville, IL: Scholastic Testing Service.

Torrance, E.P., & Safter, H.T. (1999). *Making the creative leap beyond*. Buffalo, NY: Creative Education Foundation Press.

Treffinger, D. J., Young, G. C., Selby, E. C., & Shepardson, C. (2002). *Assessing creativity: A guide for educators*. Storrs, CT: University of Connecticut, The National Research Center on the Gifted and Talented. Retrieved from http://www.gifted.uconn.edu/nrcgt/reports/rm02170/rm02170.pdf

Urban, K.K., & Jellen, H.G. (1996). *Test for Creative Thinking - Drawing Production (TCT-DP)*. Lisse, The Netherlands: Swets & Zeitlinger.

Van de Weijer-Bergsma, E., Kroesbergen, E.H., Prast, E., & Van Luit, J.E.H. (2014). *Dutch percentile norms for the Raven Standard Progressive Matrices*. Internal report. Utrecht, The Netherlands: Utrecht University.

Weisberg, R.W. (1999). Creativity and knowledge: A challenge to theories. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 226–250). Cambridge, UK: Cambridge University Press.

Eveline M. Schoevers, Utrecht University

Evelyn H. Kroesbergen, Radboud University

Maria Kattou, Ministry of Education and Culture

Correspondence concerning this article should be addressed to Eveline M. Schoevers, Department of Education & Pedagogy, Utrecht University, Heidelberglaan 1, 3584CS, Utrecht, The Netherlands. E-mail: E.M.Schoevers@uu.nl

ACKNOWLEDGEMENT

The authors would like to thank all pupils and schools that participated in this study.

AUTHOR NOTE

Eveline M. Schoevers, Department of Education & Pedagogy, Utrecht University, Utrecht, The Netherlands; Evelyn H. Kroesbergen, Behavioural Science Intitute, Radboud University, Nijmegen, The Netherlands; Maria Kattou, Ministry of Education and Culture, Nicosia, Cyprus.

SUPPORTING INFORMATION

Additional supplemental material may be found online in the Supporting Information section at the end of the article:

Table S1. Means and standard deviations of the MCT.

Table S2. Means and standard deviations of the subscales of the MA task.

Table S3. Means and standard deviations of the subscales of the TTCT and final score of the TCT-DP.