ORIGINAL PAPER

The relation between creativity and students' performance on different types of geometrical problems in elementary education

Eveline M. Schoevers[1,2] · Evelyn H. Kroesbergen[1,3] · Mirjam Moerbeek[1] · Paul P. M. Leseman[1]

Abstract

Aim In the current study we aimed to investigate the relation between creativity and mathematical problem solving in the upper grades of elementary school.

Methods To examine how student's levels of general creativity were related to their performance on different types of geometrical problems, a geometry test with diverse problems was administered to a sample of 1665 Dutch students from third to sixth grade, as well as a creativity test. The geometry test consisted of four closed-ended routine problems, six closed-ended non-routine problems (related to a visual artwork) and four open-ended non-routine problems (multiple solutions problems). The Test of Creative Thinking—Drawing Production was used to measure students' creativity. Multivariate multilevel analyses were conducted to take the nested structure of the data into account.

Results The results showed that creativity was a significant predictor of students' performance on all types of geometrical problems, but most strongly associated with performance on open-ended non-routine problems, suggesting that students with higher levels of creativity perform better in solving geometry problems in general, but especially in geometry problems asking for multiple solutions.

Keywords Geometry · Elementary school · Problem solving · Creativity · Problem types

1 Introduction

In contemporary mathematics textbooks for elementary education, students predominantly have to solve routine problems in which they have to reproduce and apply a fixed solution procedure in one or two steps (Kolovou et al., 2009; van Zanten & van den Heuvel-Panhuizen, 2018). However, it is increasingly considered important that students also learn to solve problems that are not straightforward and for which they do not have a learned solution immediately available (Schoenfeld, 1983), and which, therefore, may elicit other, more complex cognitive processes, such as creative thinking (Liljedahl et al., 2016).

To promote creative thinking in mathematics, other types of problems should be offered to students, such as non-routine and open-ended problems (Carlson & Bloom, 2005; Levav-Waynberg & Leikin, 2012a; Silver, 1997). Different types of non-routine and open problems exist (Leikin, 2018), and it is an open question if all types of non-routine and open problems call upon creative thinking equally. To answer this question, we investigated the relations between students' creativity and their performance on three types of mathematical problems in the upper grades of elementary school.

The current study was conducted in the mathematical domain of geometry. Geometry can be defined as grasping the concept of space and the mathematization of space. Geometry education aims to teach students to understand, explain and predict geometric phenomena, to reason spatially, and to order and organize spatial situations. For example, it is considered important that students be able to draw a map or to reason about the effect of the height of the sun on the length of the shadow (Gravemeijer et al., 2007). Typical topics in geometry in the upper grades of elementary school are (1) 'spatial sense', including localizing, taking a standpoint and navigation, (2) 'plane and solid figures', including

✉ Evelyn H. Kroesbergen
evelyn.kroesbergen@ru.nl

1 Faculty of Social and Behavioural Sciences, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

2 Oberon Educational Resarch and Consultancy, Slachtstraat 12, 3512 BC Utrecht, The Netherlands

3 Behavioural Science Institute, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

spatial properties and relations between figures, operations, transformations and constructions, and (3) 'visualization and representation', including representations of two-dimensional and three-dimensional reality (Gravemeijer et al., 2007).

1.1 Types of mathematical problems

The opportunity for students to learn geometry in elementary school is largely determined by the geometrical content and geometrical problems presented to them in the textbooks teachers use for mathematics education (Meelissen et al., 2012; Stein et al., 2007). Different types of problems can be distinguished. Problems can be closed-ended with only one correct solution or, alternatively, open-ended with multiple solutions or interpretations (Bahar & Maker, 2015; Mihajlović & Dejić, 2015), and problems could be familiar and call upon well-established routines (routine problems) or, alternatively, require students to discover which facts, skills and procedures can be combined to solve the problem which is as such unfamiliar to them (non-routine problems; Carlson & Bloom, 2005; Schoenfeld, 1983). Most textbooks contain mainly closed-ended and routine problems and only a few open or non-routine problems (i.e., 0–8%; van Zanten & van den Heuvel-Panhuizen, 2018). By using both dimensions (open vs. closed, and routine vs. non-routine), four types of problems can be distinguished.

A typical example of a closed-ended routine geometry problem is a multiple choice test item in which students have to choose the correct fold-out of a cube or the correct picture of a block construction from a given perspective; these problems have one correct solution and call upon well-established routines (van Grootheest et al., 2011). Closed-ended non-routine problems require students to apply existing knowledge of facts and procedures to a novel problem not encountered before, for example when students are asked to draw a floor map of a still-life painting (Schoevers et al., 2020b). While this novel problem is conceptually similar to a routine block construction problem, the novel problem requires students to figure out how to apply their knowledge and skills to make the map (Gravemeijer et al., 2007). An example of an open-ended non-routine problem is a multiple solution task in which students have to compare different plane figures (e.g., isosceles triangle, right-angled triangle, square) and are invited to give multiple answers to the question of how the plane figures differ from each other. In this case, students need to recall, use and combine facts, skills, procedures and ideas in a new and meaningful way to solve the problem (Liljedahl et al., 2016; Schoevers et al., 2020b; Warner et al., 2003). Open-ended routine problems are rare, because textbooks generally contain closed-ended problems, which makes open-ended problems by definition non-routine. If a child would encounter a specific open-ended problem repeatedly, this kind of problem could become routine

for the specific child, but this is strongly dependent on the context and cannot be generalized.

1.2 Creativity and problem types

Open-ended and non-routine geometrical problems usually require more complex cognitive processing than closed-ended and routine problems and particularly require students to think creatively (Liljedahl et al., 2016). Creativity can be defined generally as a thinking process that results in a novel and meaningful product (Runco & Jaeger, 2012; Sriraman, 2005). In elementary mathematics education, this product, for example, could be an idea that is new to the student, a newly posed problem, or a solution to a problem that is novel and meaningful for a specific age group or student (e.g., Leikin, 2009). Mathematical creativity thus also depends on formerly acquired knowledge and skills in the particular domain. This means that both domain-general creativity and domain-specific knowlegde are needed to apply creativity in a specific domain (Baer & Kaufman, 2005; Willemsen et al., 2020). Accordingly, mathematical creativity has been found to consist of both domain-general creativity and domain-specific (mathematical) knowledge (Schoevers et al., 2020a).

Several researchers have argued that especially non-routine mathematical problems—both closed (i.e., with specific constraints on the solution or interpretation; Bokhove & Jones, 2018) and open problems (e.g., Kwon et al., 2006)—have the potential to elicit creative thinking. Other researchers have argued, and indeed shown, that creative thinking is elicited by tasks in which students have to provide multiple solutions for a problem (Kwon et al., 2006; Leikin, 2009, 2018; Levav-Waynberg & Leikin, 2012b; Silver, 1995). Although the role of creativity has been investigated in both open-ended and non-routine multiple solution tasks (e.g., Schoevers et al., 2020a), to the best of our knowledge no study to date systematically compared the role of creativity in solving different types of mathematical problems. Therefore, in the current study we examined whether the relations between students' domain-general creativity and their mathematical performance differ among three problem types: closed-ended routine, closed-ended non-routine and open-ended non-routine problems. Note that open-ended routine problems are not included, because it seems impossible to construct problems that are both open-ended and routine for all children (because it depends on child and context whether an open-ended problem could be regarded as routine). In our study, we adopted the basic assumption that for learning mathematics a certain level of domain general creative thinking skills could be helpful (cf. Kroesbergen & Schoevers, 2017) and thus that the association between general creativity and mathematical performance indicates the extent to which solving the mathematical problem calls upon general creative thinking processes.

1.3 Possible factors influencing the relation between creativity and mathematical performance

When studying the relation between creativity and mathematical performance, the possible effects of other factors that could at least partly explain the observed relation, should be taken into account. This holds in particular for students' visual-spatial working memory (VSWM), general mathematical ability, gender, and socio-economic status (SES). Especially in solving geometrical problems, students' VSWM is likely to be involved (Giofrè et al., 2013). VSWM refers to the ability to temporarily hold visual-spatial information activated for processing (Baddeley, 2003). When geometrical problems are related to spatial visualization and spatial sense, students may need temporarily to store and manipulate visual-spatial information. For example, a target figure, such as a block construction, needs to be stored temporarily in memory in order to mentally rotate the construction to see how it would look from another perspective. In addition, general mathematical ability should be taken into account, since basic mathematical procedures and knowledge about, for example, numbers, proportions and measurement, can be used to solve geometrical problems (Gravemeijer et al., 2007). Gender might also play a role. Although male students often score more highly on routine mathematical problems than girls (Reilly et al., 2015), girls have been found to outperform boys on multiple solution tasks (Mann, 2006), while they often also score more highly on creativity tasks (Baer & Kaufman, 2006). Finally, SES should be taken into account since a low SES is related to lower educational performance in general, and in mathematics in particular, possibly due to disadvantages in financial, cultural and social resources, and lower parental involvement in students' education (OECD, 2016; Sirin, 2005).

In addition, class level factors may influence the relation between creativity and performance, such as teachers' experience and the type of mathematical textbook that is used. Research has shown a positive effect of teachers' years of experience on students' mathematical performance (Clotfelter et al., 2007). More experienced teachers might be more effective, because they have improved their teaching performance over the years (Darling-Hammond, 2000). Furthermore, the contents of mathematical textbooks, in particular the types of problem they provide, may influence students' performance as well (van Zanten & van den Heuvel-Panhuizen, 2018).

In the present study we investigated the relations between students' independently assessed domain-general creativity and their performance on closed-ended routine, closed-ended non-routine and open-ended non-routine geometry problems. The aim of the study was to determine whether the relation between creativity and performance differed among the three problem types, which we assumed would

indicate differences between the problem types in the degree to which they call upon creative thinking. VSWM, gender, age, low SES and general mathematical ability, teachers' experience and the type of mathematical textbooks were included as covariates, in order to control for spurious relations between creativity and mathematical performance.

We hypothesized that students' creativity had stronger predictive relations with performance on non-routine than on routine problems, because the former were expected to be unfamiliar to students and, therefore, to require stronger involvement of higher-order creative thinking compared to routine problems. Moreover, as there is consensus that in addition to unfamiliarity, openness of a task puts additional demands on creative thinking, we also expected that students' domain-general creativity would have the strongest association with their performance on the open-ended non-routine task.

2 Methods

For the present purpose, pre-intervention data of a large-scale evaluation study of the Mathematics, Arts and Creativity in Education (MACE) program were used (Schoevers et al., 2020b). The MACE program was developed to support elementary schools in the Netherlands to meet the partly overlapping learning goals and objectives of the disciplines visual arts and geometry, and to promote students' creative skills in both disciplines. We focused on students in the upper grades of primary school, because they had sufficient reading and writing skills to perform the planned tasks. In the present study, only the students who took all the tests relevant to the present purpose were included.

2.1 Participants

Participants were 1665 students from grades 3 to 6, in 92 classes at 50 schools in The Netherlands. Schools were recruited by sending flyers to 428 regular elementary schools in all regions of The Netherlands; 11.68% of the schools were willing to participate. Schools differed regarding their educational vision and teaching methods, and were located in both rural and urban areas spread across The Netherlands. The sample consisted of students with low, medium and high socio-economic background, of which 48.3% were boys, and with a mean age of 10.91 years (SD = 0.93).

2.2 Instruments

2.2.1 Geometric Ability Test (GAT)

The GAT took between 20–30 min and was stopped after 30 min. The test consisted of 11 closed-ended geometry problems, of which four were routine problems, and seven

non-routine problems (see Fig. 1). All routine problems called upon spatial sense and spatial visualization. In the non-routine problems, students mainly had to reason geometrically in relation to a painting. In one problem, students had to draw a floor plan of a painting. The test started with four routine problems and ended with the non-routine art-geometry problems. Within problem type, the order of the problems was randomized and was the same for all students.

The routine problems were relatively straightforward, with clearly one correct answer. Therefore, one point was given for a correct answer and zero points for an incorrect answer. An average score for the four routine problems was calculated; a total score between 0 and 1 could be obtained. The non-routine problems related to visual arts were more complex and could also yield partially correct solutions. Therefore, two points were given for an answer with correct reasons about geometric phenomena, for example when students explained why persons or buildings in front of the painting looked bigger than in the back by referring to perspective (e.g., 'by painting smaller in the back and larger in the front). One point was given for an answer with incomplete reasoning (e.g., 'by painting smaller', without a specific explanation). Zero points were given for answers without reasoning (e.g., 'it is a photo' or 'when you paint everything is possible'). The last non-routine problem of the task was considered too difficult for grades 4 and 5 (only 9% of the students scored > 0 points) and was not included in further analyses. An average score for the non-routine

closed-ended problems was calculated based on the number of problems the students had finished.

The test–retest reliability of the GAT was acceptable ($r = 0.66$). Test–retest reliability was determined over a two-week interval, and involved 184 4th, 5th and 6th grade students. The interrater reliability ranged from sufficient to excellent for all items ($\kappa = 0.81$–$0.1.00$, ICC $= 0.67$–1.00, based on 25 tests). In line with our expectations, the internal consistency of the GAT was moderate in this study ($\alpha = 0.62$), because the GAT represents a heterogenous set of knowledge and skills.

2.2.2 Geometrical multiple solution task (GMST)

The GMST, based on the mathematical multiple solution task (Kattou et al., 2013; Schoevers et al., 2020a), took between 20 and 30 min. The task consisted of five geometry questions that were open-ended and non-routine. Students were instructed to provide multiple, but distinct and original solutions. A sample question of the test is depicted in Fig. 2.

For the scoring of the GMST, we used the scoring scheme of Leikin for creativity in the individual solution space (Leikin, 2009; Levav-Waynberg & Leikin, 2012a). Within this scheme a distinction between fluency, flexibility and originality is made. Fluency was calculated by adding the number of correct answers for each question. With the use of the scheme, each solution of a student was scored regarding flexibility (the use of a solution from a different (sub)group

Fig. 1 An example of, respectively, a routine and a non-routine art-geometry problem
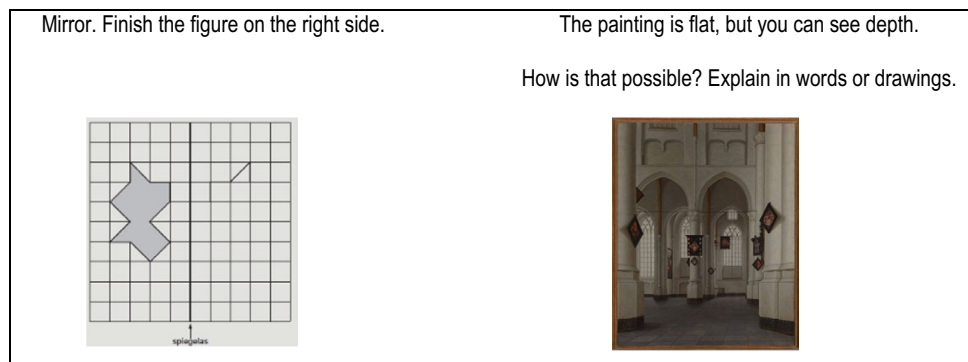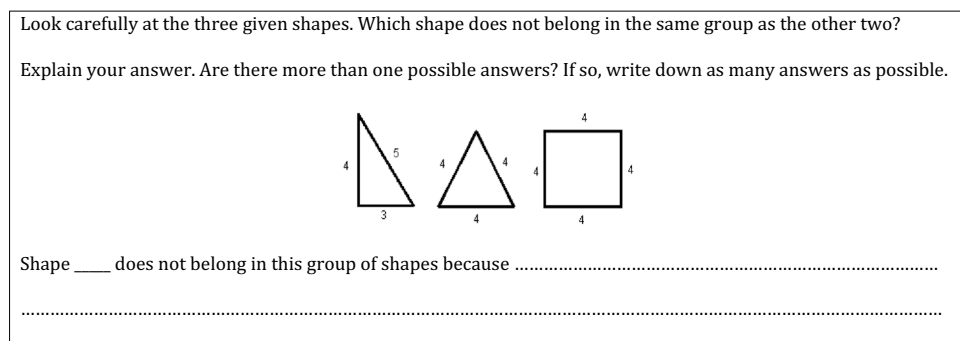

Fig. 2 A sample question of the GMST

of strategies) and originality (the infrequency or unconventionality of a used solution). Next, a final score per solution was calculated as a product of Flexibility × Originality. Afterwards, a score per question was concluded as follows: Fluency × ($\sum$ (Flexibility × Originality). Last, the scores of all questions were added into a total creativity score. In Fig. 3 we illustrated how the answers of a 6th grade student to a question of the GMST were scored.

The test–retest reliability of the GMST was good ($r = 0.84$ over a two-week interval; Schoeverset al., 2020b). The interrater reliability ranged from sufficient to excellent for all scores per solution (ICC = 0.72–0.99, based on 25 tests). The internal consistency of the GMST was moderate in this study ($\alpha = 0.68$).

2.2.3 Test for creative thinking-drawing production (TCT-DP)

Because children's domain-general creativity is also related to their mathematical creativity, we included a non-mathematical, nonverbal creativity test. The TCT-DP is based on a more holistic concept of creativity (Urban, 2004, 2005). Students have to complete a drawing using given figural fragments, such as a half circle and a half square, within 15 min (Urban & Jellen, 1996). The task requires students to apply (cycles of) divergent and convergent thinking processes to create a final drawing. Students first have to

think divergently about what they can draw starting from a (combination of) figural fragment(s) and then they have to choose what they will draw on the paper (convergent thinking process). Subsequently, they may think how they can elaborate on their drawing (divergent thinking). The TCT-DP was scored according to the guidelines in the manual (Urban & Jellen, 1996). The task was scored on fourteen aspects considered to constitute creativity, such as making connections, boundary-breaking, and unconventionality. The TCT-DP was scored by four raters and had a good interrater reliability in this study (ICC = 0.80, based on 25 tests).

2.2.4 Visual-spatial working memory

The online computerized task 'the Lion game' was administered to measure students' VSWM. The Lion game is a visual-spatial complex span task (van de Weijer-Bergsma et al., 2015). Students are presented with a 4×4 matrix containing 16 cells. In each trial, eight lions of different colors are consecutively presented at different locations in the matrix. Students have to remember the last location where a lion of a certain color has appeared, and to click on that location after the sequence has ended. The task contains 20 items at five difficulty levels with increasing WM load. The proportion of correct items was calculated. The Lion game is reported to have excellent internal consistency (Cronbach's α ranged between 0.86 and 0.90), satisfactory test–retest reliability

Fig. 3 Illustration of GMST scoring

	Flexibility	*Orginality*	Flexibility × Originality.
Shape C does not belong to this group of shapes, because it is a square and the others are triangles	*10*	*0.1*	*1*
Shape A does not belong to this group of shapes, because the sides are not the same	*10*	*1*	*10*
Shape B does not belong to this group of shapes, because if you double it you cannot make it a rectangle	*10*	*10*	*100*
Fluency × ($\sum$ (Flexibility × Originality) = 3 × (111) = 333			

($\rho = 0.71$), and good concurrent and predictive validity (van de Weijer-Bergsma et al., 2015).

2.2.5 Questionnaire for teachers

A questionnaire was used to obtain information on student, teacher and class characteristics. The educational attainment of both parents was used as an indicator of students' SES. General mathematical ability was measured with a criterion based test (Janssen et al., 2007), which covers a wide range of mathematical domains such as arithmetic operations, geometry, measurement, time, and proportions. The test has been shown to be highly reliable; the reliability coefficients range from 0.91 to 0.97 (Janssen et al., 2010). Furthermore, teachers were asked to provide information about their own gender, years of experience and the mathematic textbook(s) they used to teach mathematics.

2.3 Procedure

The data were collected in September 2017 by the first author and twelve research assistants with a bachelor's or master's degree in (special) education. Tests were administered individually in one session in a quiet classroom. Passive informed consent of parents was obtained before the start of the study for 99.2% of the students. The study was approved by the Ethical Committee of the Faculty of Social and Behavioural Sciences of Utrecht University (FETC15-083).

2.4 Analyses

Multivariate multilevel analyses were conducted in MLwiN 3.02 to take the nested structure of the data into account, and because the three outcome measures used in this study were expected to be correlated (Goldstein, 2011). A three-level model was used in which the three types of problems (level 1) were nested within students (level 2), which were nested within classes (level 3). We did not take the school level into account as a fourth level, as preliminary analyses revealed little variance (< 5%) located at this level, and no school variables were available that could be expected to be related to outcome measures. All variables used in the multivariate multilevel analyses were *z*-standardized on the grand mean.

In the first model, second- and third-level predictors were added for each outcome variable, as follows: VSWM, students' age, students' gender, low SES, general mathematical ability, creativity (level 2), teachers' years of work experience, and dummy variables representing the four most frequently used mathematical textbooks (i.e., *Wereld in Getallen (WiG), Alles Telt, Pluspunt* and *Rekenrijk*; level 3). Effects of all predictors could vary between the different outcome measures. Next, for each predictor a new model was

Table 1 Means and standard deviations of students' measurements

Measurements	*M* (SD)
Closed-ended routine problems	0.54 (0.32)
Closed-ended non-routine problems	0.61 (0.37)
Open-ended non-routine problems	957.52 (985.25)
Creativity (TCT-DP)	19.67 (8.32)
VSWM (Lion Game)	0.73 (0.15)

estimated to test whether effects were similar for all outcome measures. The order in which coefficients for each predictor were constrained to be equal was random. Based on the deviance and a chi-square test, we tested whether the new model, with common coefficients for a predictor, fitted the data. If this was true, common coefficients of a predictor were kept in the next model. This procedure led to a total of 12 models that were compared (see "Appendix 1"). Subsequently, if the effects of creativity on the three different problems were not similar, we tested if creativity differed significantly between the different problems by using contrasts. If necessary, coefficients of creativity were constrained according to the results of the contrasts in a final, 13th model.

Before the multilevel analyses were conducted, the assumptions were checked (Hox et al., 2018). With the use of SPA-ML (Moerbeek, 2015), we calculated the required sample size for a univariate model with two levels (i.e., students in classes), because the program could not calculate the required sample size for a multivariate model. With a desired power of 0.80 and expected effect size of 0.25 with 15% variance located at class level, and 85% at student level, a sample of 61 classes and 1220 students is required. Furthermore, the assumptions of linearity and absence of outliers were met for all variables. The assumption of normally distributed residuals was violated for the residuals at student level for the GMST. Therefore, robust standard errors are reported (Hox et al., 2018).

3 Results

3.1 Descriptive statistics

Teachers in this sample had on average 17.11 years of experience (SD = 10.36). Regarding the mathematical textbooks, we found that four textbooks were predominantly used for mathematics education ('Wereld in Getallen': 32.4%; 'Pluspunt: 23.0%; 'Rekenrijk': 15.7%; 'Alles Telt': 11.9%). Four other mathematical textbooks were used much less frequently and pooled in one rest category which served as reference category. Descriptive statistics for the student measures can be found in Table 1. Table 2 presents the Spearman correlations between students' creativity and

their performance on the three types of problems. Creativity was related to all types of geometrical problems, but more strongly to the multiple solution problems than to the routine problems (z = 1.73, *p* < 0.01).

3.2 Multivariate multilevel results

First, an intercept-only model was estimated to calculate the percentage of variance located at student and class level. The variance components were nearly equal for each type of problem; 15% was located on class level for the closed-ended routine problems, 14% for the closed-ended non-routine problems and 12% for the open-ended non-routine problems. Subsequently, student- and class-level predictors were added for each outcome variable. Next, for each predictor we tested in a new model whether the effects were similar for all outcome measures. We found that the effects of age, low SES, teachers' experience and the four

mathematical textbooks were similar for all three types of problems. Effects of students' VSWM, gender, mathematical achievement and creativity were different for the three types of problems (see "Appendix 1"). Subsequently, using contrasts, we found that the effect of creativity was similar for the closed-ended routine and closed-ended non-routine problems, but significantly different for the open-ended non-routine problems, $\chi^2(1) = 0.68$, *p* = 0.41. The final multivariate multilevel model can be found in Table 3.

The main question of this study was how creativity was related to the three problem types. The results show (1) that creativity is significantly related to all types of problems and (2) that creativity shows the strongest relation with the open-ended non-routine task. No difference in predictive value of creativity was found between the closed-ended routine and the closed-ended non-routine problems. These results show that students who score more highly on creativity, also perform better on all three types of mathematical problems, but also that this effect is the strongest for non-routine open-ended problems. In other words, the problem types differ in the degree to which they call upon creative thinking, with the non-routine open-ended type requiring the most creative thinking.

In addition, covariates were included to control for spuriousness. Some covariates were found to have a differential effect on the geometry problems. The results show that VSWM was a significant predictor of both types of closed-ended problems, however, no effect of VSWM was found for the open-ended problems. Children with higher working memory skills performed better on both closed-ended routine and closed-ended non-routine problems than children

Table 2 Spearman correlations between variables used in this study

	1	2	3	4
1. Creativity (TCT-DP)	–	*0.15**￼*	*0.19**￼*	*0.23**￼*
2. Closed-ended routine problems	0.17**	–	*0.37**￼*	*0.14**￼*
3. Closed-ended non-routine problems	0.20**	0.38**	–	*0.25**￼*
4. Open-ended non-routine problems	0.24**	0.19**	0.27**	–

Correlations above the diagonal are corrected for age

** *p* < .01

Table 3 Standardized fixed effects on the routine, art-geometry and multiple solution problems

	Closed-ended routine problems		Closed-ended non-routine problems		Open-ended non-routine problems	
	B [95% CI]	SE	*B* [95% CI]	SE	*B* [95% CI]	SE
Intercept	− 0.01 [− 0.09, .06]	0.04	− 0.00 [− 0.08, .08]	0.04	0.00 [− 0.07, .08]	0.04
Level 2 (student)						
VSWM	0.13 [0.07, 0.20]**	0.03	0.09 [0.04, 0.15]*	0.03	− 0.03 [− 0.08, 0.02]	0.03
Age	0.18 [0.14, 0.22]**	0.02	0.18 [0.14, 0.22]**	0.02	0.18 [0.14, 0.22]**	0.02
Gender	0.02 [− 0.04, .08]	0.03	0.16 [0.10, 0.21]**	0.03	0.11 [0.06-.16]**	0.02
Low SES	− 0.07 [− 0.10, − 0.04]**	0.02	− 0.07 [− 0.10, − 0.04]**	0.02	− 0.07 [− 0.10, − 0.04]**	0.02
General math ability	0.37 [0.31, 0.44]**	0.03	0.26 [0.19, 0.33]**	0.04	0.21 [0.14, 0.29]**	0.04
Creativity	0.09 [0.06, 0.13]**	0.02	0.09 [0.06, 0.13]**	0.02	0.17 [0.12, 0.22]**	0.03
Level 3 (class)						
Teachers' experience	0.02 [− 0.04, .08]	0.03	0.02 [− 0.04, 0.08]	0.03	0.02 [− 0.04, 0.08]	0.03
Textbook_WiG	− 0.05 [− 0.13, 0.04]	0.04	− 0.05 [− 0.13, 0.04]	0.04	− 0.05 [− 0.13, 0.04]	0.04
Textbook_Alles Telt	− 0.07 [− 0.13, − 0.02]*	0.03	− 0.07 [− 0.13, − 0.02]*	0.03	− 0.07 [− 0.13, − 0.02]*	0.03
Textbook_Pluspunt	− 0.02 [− 0.10, 0.06]	0.04	− 0.02 [− 0.10, 0.06]	0.04	− 0.02 [− 0.10, 0.06]	0.04
Textbook_Rekenrijk	0.02 [− 0.06, 0.11]	0.03	0.02 [− 0.06, 0.11]	0.03	0.02 [− 0.06, 0.11]	0.03

Robust standard errors are reported for all problems

*Significant at *p* < .01, **Significant at *p* < .001

with lower working memory skills, but such an effect was not found for the open-ended non-routine problems. General mathematical ability was related to all types of geometrical problems, but the strongest to the routine problems. This means that students with high mathematical ability scored better on all problem types. This relation was even stronger for performance on both routine problems. In contrast, gender was not related to performance on the routine problems, but the results show a significant effect for both types of non-routine tasks, on which girls performed better than boys.

4 Discussion

The present study investigated the relations between students' independently assessed domain-general creativity and their performance on three types of geometry problems: closed-ended routine, closed-ended non-routine and open-ended (multiple solution) non-routine problems. The results show that students who scored higher on the general creativity test, were also better in solving geometrical problems regardless the type of problem. This finding confirms our hypothesis and is in line with former research that also domain-general creativity is related to general mathematical ability (Kattou et al., 2013; Kroesbergen & Schoevers, 2017; Mann, 2006). The hypothesis that creativity would show the strongest relation with the open-ended non-routine task was also confirmed, but contrary to our expectations, the predictive value of creativity was the same for the closed-ended routine problems and the closed-ended non-routine problems.

The basic assumption of the current study was that the association between domain-general creativity and mathematical performance indicates the extent to which solving particular mathematical problems calls upon general creative thinking processes. Following this assumption, the current results suggest that open-ended problems trigger creative thinking of students more than closed-ended problems. This result raises the question of whether problems that trigger creative thinking could also be used to foster children's creative thinking. The results from Levav-Waynberg and Leikin (2012a) indeed showed that systematically implementing multiple solution tasks in 10th grade geometry classes increased not only students' geometrical knowledge but also their creativity.

Before drawing conclusions, alternative explanations for the current findings need to be considered. One alternative explanation is that not general creativity but general intelligence explains the pattern of findings, as creativity measures are often found to be moderately to strongly related to measures of general intelligence (e.g., Silvia, 2008). Note, however, that in the present study a measure of students' visual-spatial working memory was included as a control

covariate and that the associations of creativity with mathematical performance we found were net of the shared variance with working memory. Research has shown that working memory underlies the correlation between creativity and intelligence (Benedek et al., 2014). In addition, working memory has been found to be a stronger predictor of mathematical achievement in elementary school than intelligence (e.g., Alloway & Alloway, 2010). Therefore, explaining the pattern of findings as an effect of intelligence instead of indicating involvement of creativity seems less plausible. We cannot rule out that other generic abilities such as verbal skills, motivation or work attitude, might explain part of the shared variance (Carlson & Bloom, 2005; Elia et al., 2009). However, by including other covariates, in particular general mathematical ability, the involvement of such general processes is—at least partly—already controlled for. A specific interpretation, therefore, is more likely: especially open-ended non-routine problems call upon creative thinking.

Other findings confirm this pattern and suggest that different geometrical problem types involve different cognitive processes and problem solving strategies. VSWM and mathematical ability were relatively strong predictors of performance on both routine and non-routine closed-ended problems, but not of performance on open-ended non-routine problems. A possible explanation is that the format of the closed-ended problems is more similar to the typical mathematical problems with which the students are familiar. This aspect might trigger strategies such as the retrieval of knowledge or facts and procedures, and activating memory of previous experiences with similar problems to find the correct answer or procedure. In contrast, the multiple solution task was not only non-routine, but also open-ended. Openness of a problem seems to trigger creative thinking the most, as it stimulates students to find answers beyond what is already known and readily retrievable from memory (Kroesbergen & Schoevers, 2017; Leikin, 2018; Schoevers et al., 2020a).

Contrary to our hypotheses, the association of creativity with performance did not differ between routine geometry problems and non-routine geometry problems. We expected that the relative unfamiliarity of non-routine problems compared to routine problems would require more creative thinking and, thus, would show a stronger association between creativity and performance, which was not the case. A possible explanation is that openness of problems is more important than their familiarity. Closed-ended problems might trigger more convergent than divergent thinking processes in students, while open-ended tasks require especially divergent thinking skills. Although the closed-ended non-routine task was also intended to stimulate students to combine mathematical knowledge and procedures in a new and meaningful way, the predominantly closed-ended nature of

the task, asking for a single correct answer, may have limited the effect of the unfamiliarity.

In addition to mathematical ability and working memory, gender was also found to be differentially related to the problem types. Girls outperformed boys on both non-routine types of problems, but no differences were found on the routine tasks. It seems that unfamiliarity does trigger creative thinking in girls. This is in line with previous studies showing that girls are not advantaged at regular mathematical tasks, but do perform better on tasks that require creative thinking (Baer & Kaufman, 2006; Mann, 2006).

In interpreting the results it should be kept in mind that in the study we examined the role of creativity in only three types of geometrical problems. There are other types of open and non-routine geometrical problems (Leikin, 2018), which were not included in this study. Moreover, geometry is only one of the several domains constituting the discipline of mathematics in elementary school. The present findings can therefore not be generalized to these other types of problems or mathematical domains. Future research could extend the present approach to other types of problems and other mathematical domains. Another limitation of this study relates to our measure of creativity. The TCT-DP, although regarded as a generic, holistic measure of creativity, does not fully capture the complex multidimensional nature of creativity as it is defined nowadays (Cropley, 2010). Note, however, that this test covers both divergent and convergent thinking processes and can, for that reason, be regarded as a good single

indicator of students' domain-general creativity. Nonetheless, we recommend that multiple measures of creativity be included in future research.

To conclude, students' domain-general creativity was positively associated with their performance on three different types of geometry problems. However, students' creativity was significantly more strongly associated with performance on multiple solution problems than with performance on routine and non-routine closed-ended problems. Students with higher levels of creativity performed better in solving geometry problems in general, but especially in geometry problems asking for multiple solutions. As several covariates were included to control for spuriousness, we may cautiously conclude that the remaining effects of creativity reflect true involvement of creative thinking processes in geometrical problem solving, especially in the more open multiple solution problems. Further research should provide more evidence regarding the causal direction of the relations or the effect of including more open tasks in the curriculum on students creative skills. Other research has suggested that providing students with appropriate open-ended problems in mathematics education is probably an important way to promote creative thinking (Levav-Waynburg & Leikin, 2012a). However, longitudinal and experimental research is needed for further evidence. Nonetheless, the present study provides important first ideas on the possible involvement of creativity in mathematics that may inspire new research and curriculum development.

Appendix 1

Fixed part	Model 1		Model 2		Model 3		Model 4	
	B	SE	B	SE	B	SE	B	SE
Intercept_problem1	0.003	0.038	0.003	0.038	0.003	0.038	0.004	0.038
Intercept_problem2	− 0.001	0.040	− 0.001	0.041	− 0.001	0.040	0.001	0.039
Intercept_problem3	− 0.013	0.036	− 0.013	0.037	− 0.013	0.037	− 0.015	0.038
Visual-spatialWM_problem1	− `0.019	0.025	0.064**	0.019	− 0.020	0.026	− 0.018	0.026
Visual-spatialWM_problem2	0.090*	0.029	0.064**	0.019	0.088*	0.029	0.094*	0.030
Visual-spatialWM_problem3	0.126**	0.032	0.064**	0.019	0.127**	0.032	0.121**	0.032
Gender_problem1	0.106**	0.024	0.099**	0.025	0.106**	0.024	0.087**	0.020
Gender_problem2	0.153**	0.028	0.155**	0.028	0.154**	0.028	0.087**	0.020
Gender_problem3	0.023	0.032	0.028	0.031	0.021	0.032	0.087**	0.020
General math ability_problem1	0.213**	0.038	0.190**	0.040	0.214**	0.038	0.210**	0.038
General math ability_problem2	0.246**	0.036	0.253**	0.034	0.249**	0.036	0.234**	0.036
General math ability_problem3	0.380**	0.031	0.397**	0.030	0.376**	0.031	0.390**	0.031
Creativity_problem1	0.172**	0.027	0.171**	0.027	0.172**	0.027	0.173**	0.027

Fixed part	Model 1		Model 2		Model 3		Model 4	
	B	SE	B	SE	B	SE	B	SE
Creativity_problem2	0.106**	0.025	0.106**	0.025	0.105**	0.026	0.110**	0.026
Creativity_problem3	0.081**	0.022	0.081**	0.022	0.082**	0.022	0.078**	0.022
Age_problem1	0.173**	0.027	0.167**	0.028	0.183**	0.020	0.183**	0.020
Age_problem2	0.157**	0.033	0.158**	0.033	0.183**	0.020	0.183**	0.020
Age_problem3	0.211**	0.032	0.214**	0.032	0.183**	0.020	0.183**	0.020
LowSES_problem1	− 0.037	0.024	− 0.029	0.023	− 0.038	0.024	− 0.037	0.024
LowSES_problem2	− 0.094**	0.018	− 0.097**	0.018	− 0.096**	0.018	− 0.094**	0.018
LowSES_problem3	− 0.081**	0.022	− 0.087**	0.023	− 0.080**	0.022	− 0.081**	0.023
Teachers' experience_ problem1	− 0.005	0.041	− 0.007	0.040	− 0.005	0.041	− 0.006	0.041
Teachers' experience_ problem2	− 0.028	0.050	− 0.026	0.051	− 0.029	0.050	− 0.033	0.049
Teachers' experience_ problem3	0.064	0.040	0.068	0.042	0.066	0.041	0.070	0.042
WiG_problem1	− 0.015	0.055	− 0.020	0.056	− 0.016	0.055	− 0.017	0.054
WiG_problem2	− 0.011	0.072	− 0.010	0.073	− 0.014	0.072	− 0.017	0.070
WiG_problem3	− 0.103	0.054	− 0.100	0.057	− 0.100	0.054	− 0.097	0.056
Alles Telt_problem1	− 0.108*	0.039	− 0.116*	0.040	− 0.109*	0.039	− 0.110*	0.038
Alles Telt_problem2	− 0.003	0.052	− 0.001	0.052	− 0.003	0.053	− 0.005	0.052
Alles Telt_problem3	− 0.068	0.039	− 0.063	0.040	− 0.067	0.038	− 0.065	0.039
Pluspunt_problem1	0.008	0.047	0.009	0.049	0.008	0.047	0.006	0.047
Pluspunt_problem2	− 0.027	0.068	− 0.029	0.068	− 0.027	0.068	− 0.034	0.065
Pluspunt_problem3	− 0.048	0.055	− 0.051	0.057	− 0.049	0.054	− 0.043	0.056
Rekenrijk_problem1	0.084	0.063	0.086	0.063	0.084	0.062	0.084	0.062
Rekenrijk_problem2	0.042	0.064	0.041	0.065	0.043	0.064	0.042	0.062
Rekenrijk_problem3	-0.054	0.052	-0.056	0.054	-0.056	0.052	-0.055	0.053
Random part class level								
$\sigma^2_{\text{Constant problem1}}$	0.056		0.056		0.055		0.055	
$r_{\text{Problem1\_problem2}}$	0.017		0.013		0.015		0.015	
$\sigma^2_{\text{Constant problem2}}$	0.074		0.075		0.071		0.068	
$r_{\text{Problem3\_problem1}}$	0.006		0.004		0.006		0.006	
$\sigma^2_{\text{Constant problem3}}$	0.055		0.059		0.059		0.062	
$r_{\text{Problem3\_problem2}}$	0.032		0.035		0.031		0.031	
Random part student level								
$\sigma^2_{\text{Constant problem1}}$	0.803		0.808		0.803		0.804	
$r_{\text{Problem1\_problem2}}$	0.126		0.125		0.127		0.129	
$\sigma^2_{\text{Constant problem2}}$	0.756		0.755		0.757		0.762	
$r_{\text{Problem3\_problem1}}$	0.023		0.020		0.023		0.022	
$\sigma^2_{\text{Constant problem3}}$	0.689		0.691		0.688		0.691	
$r_{\text{Problem3\_problem2}}$	0.166		0.166		0.166		0.161	
Deviance	10,228.240		10.244.234		10.230.466		10.249.602	

Fixed part	Model 5		Model 6		Model 7		Model 8	
	B	SE	B	SE	B	SE	B	SE
Intercept_problem1	0.004	0.037	0.004	0.038	0.003	0.038	0.004	0.038
Intercept_problem2	− 0.001	0.040	− 0.001	0.041	− 0.000	0.040	− 0.003	0.041
Intercept_problem3	− 0.014	0.037	− 0.013	0.037	− 0.012	0.037	− 0.012	0.037

Fixed part	Model 5		Model 6		Model 7		Model 8	
	B	SE	B	SE	B	SE	B	SE
Visual-spatialWM_problem1	− 0.023	0.025	− 0.043	0.026	− 0.022	0.025	− 0.024	0.025
Visual-spatialWM_problem2	0.091*	0.029	0.083*	0.029	0.091*	0.029	0.090*	0.028
Visual-spatialWM_problem3	0.128**	0.033	0.154**	0.032	0.128**	0.032	0.130**	0.033
Gender_problem1	0.107**	0.024	0.120**	0.024	0.110**	0.024	0.108**	0.024
Gender_problem2	0.154**	0.028	0.160**	0.029	0.153**	0.028	0.155**	0.028
Gender_problem3	0.021	0.032	0.004	0.032	0.019	0.032	0.019	0.032
General math ability_problem1	0.212**	0.038	0.285**	0.027	0.217**	0.039	0.213**	0.038
General math ability_problem2	0.251**	0.036	0.285**	0.027	0.249**	0.036	0.253**	0.035
General math ability_problem3	0.377**	0.031	0.285**	0.027	0.373**	0.031	0.375**	0.031
Creativity_problem1	0.170**	0.027	0.163**	0.027	0.120**	0.016	0.171**	0.027
Creativity_problem2	0.106**	0.026	0.102**	0.026	0.120**	0.016	0.108**	0.025
Creativity_problem3	0.083**	0.022	0.092**	0.023	0.120**	0.016	0.080**	0.022
Age_problem1	0.183**	0.020	0.184**	0.020	0.183**	0.020	0.183**	0.020
Age_problem2	0.183**	0.020	0.184**	0.020	0.183**	0.020	0.183**	0.020
Age_problem3	0.183**	0.020	0.184**	0.020	0.183**	0.020	0.183**	0.020
LowSES_problem1	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015
LowSES_problem2	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015
LowSES_problem3	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015
Teachers' experience_problem1	− 0.005	0.040	− 0.000	0.041	− 0.011	0.041	0.017	0.031
Teachers' experience_problem2	− 0.029	0.050	− 0.028	0.052	− 0.028	0.050	0.017	0.031
Teachers' experience_problem3	0.066	0.041	0.059	0.038	0.069	0.040	0.017	0.031
WiG_problem1	− 0.014	0.055	− 0.019	0.057	− 0.015	0.056	− 0.016	0.055
WiG_problem2	− 0.016	0.072	− 0.018	0.075	− 0.015	0.072	− 0.020	0.069
WiG_problem3	− 0.101	0.055	− 0.095	0.051	− 0.100	0.053	− 0.096	0.057
Alles Telt_problem1	− 0.111*	0.039	− 0.117*	0.040	− 0.113*	0.040	− 0.111*	0.039
Alles Telt_problem2	− 0.002	0.053	− 0.005	0.055	− 0.001	0.052	− 0.003	0.051
Alles Telt_problem3	− 0.066	0.038	− 0.055	0.037	− 0.065	0.038	− 0.064	0.041
Pluspunt_problem1	0.007	0.048	0.001	0.050	0.010	0.049	0.006	0.048
Pluspunt_problem2	− 0.027	0.068	− 0.030	0.071	− 0.027	0.068	− 0.030	0.065
Pluspunt_problem3	− 0.049	0.054	− 0.040	0.050	− 0.050	0.054	− 0.045	0.056
Rekenrijk_problem1	0.082	0.063	0.073	0.063	0.082	0.063	0.082	0.063
Rekenrijk_problem2	0.046	0.064	0.042	0.066	0.046	0.063	0.047	0.064
Rekenrijk_problem3	− 0.055	0.052	− 0.044	0.051	− 0.056	0.052	− 0.057	0.053
Random part class level								
$\sigma^2_{\text{Constant problem1}}$	0.055		0.056		0.057		0.055	
$r_{\text{Problem1\_problem2}}$	0.015		0.019		0.017		0.016	
$\sigma^2_{\text{Constant problem2}}$	0.073		0.076		0.071		0.075	
$r_{\text{Problem3\_problem1}}$	0.006		0.006		0.006		0.005	
$r_{\text{Problem3\_problem2}}$	0.032		0.027		0.029		0.029	
$\sigma^2_{\text{Constant problem3}}$	0.059		0.057		0.058		0.061	
Random part student level								
$\sigma^2_{\text{Constant problem1}}$	0.805		0.809		0.806		0.805	

Fixed part	Model 5		Model 6		Model 7		Model 8	
	B	SE	B	SE	B	SE	B	SE
$r_{Problem1\_problem2}$	0.126		0.127		0.125		0.126	
$\sigma^2_{Constant\ problem2}$	0.757		0.756		0.758		0.757	
$r_{Problem3\_problem1}$	0.023		0.018		0.021		0.023	
$r_{Problem3\_problem2}$	0.166		0.165		0.167		0.166	
$\sigma^2_{Constant\ problem3}$	0.688		0.695		0.690		0.688	
Deviance	10.233.808		10.256.639		10.239.963		10.238.614	

Fixed part	Model 9		Model 10		Model 11		Model 12	
	B	SE	B	SE	B	SE	B	SE
Intercept_problem1	0.005	0.038	0.004	0.038	0.003	0.038	0.001	0.038
Intercept_problem2	− 0.002	0.041	− 0.003	0.041	− 0.000	0.041	− 0.001	0.041
Intercept_problem3	− 0.014	0.037	− 0.014	0.038	− 0.014	0.038	− 0.012	0.038
Visual-spatialWM_problem1	− 0.023	0.025	− 0.024	0.025	− 0.025	0.025	− 0.026	0.026
Visual-spatialWM_problem2	0.091*	0.028	0.091*	0.028	0.093*	0.028	0.092*	0.028
Visual-spatialWM_problem3	0.129**	0.033	0.129**	0.033	0.129**	0.033	0.130**	0.033
Gender_problem1	0.107**	0.024	0.107**	0.024	0.107**	0.024	0.107**	0.024
Gender_problem2	0.155**	0.028	0.155**	0.028	0.156**	0.028	0.156**	0.028
Gender_problem3	0.020	0.032	0.020	0.032	0.020	0.032	0.020	0.032
General math ability_problem1	0.213**	0.038	0.210**	0.038	0.211**	0.038	0.212**	0.037
General math ability_problem2	0.253**	0.035	0.256**	0.035	0.256**	0.035	0.256**	0.035
General math ability_problem3	0.374**	0.032	0.375**	0.032	0.375**	0.032	0.373**	0.032
Creativity_problem1	0.171**	0.027	0.172**	0.027	0.173**	0.027	0.174**	0.027
Creativity_problem2	0.107**	0.025	0.107**	0.025	0.105**	0.025	0.105**	0.025
Creativity_problem3	0.081**	0.022	0.081**	0.022	0.081**	0.022	0.081**	0.022
Age_problem1	0.183**	0.020	0.183**	0.020	0.183**	0.020	0.183**	0.020
Age_problem2	0.183**	0.020	0.183**	0.020	0.183**	0.020	0.183**	0.020
Age_problem3	0.183**	0.020	0.183**	0.020	0.183**	0.020	0.183**	0.020
LowSES_problem1	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015
LowSES_problem2	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015
LowSES_problem3	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015	− 0.069**	0.015
Teachers' experience_problem1	0.017	0.031	0.017	0.031	0.017	0.031	0.017	0.031
Teachers' experience_problem2	0.017	0.031	0.017	0.031	0.017	0.031	0.017	0.031
Teachers' experience_problem3	0.017	0.031	0.017	0.031	0.017	0.031	0.017	0.031
WiG_problem1	− 0.047	0.044	− 0.047	0.044	− 0.047	0.044	− 0.047	0.044
WiG_problem2	− 0.047	0.044	− 0.047	0.044	− 0.047	0.044	− 0.047	0.044
WiG_problem3	− 0.047	0.044	− 0.047	0.044	− 0.047	0.044	− 0.047	0.044
Alles Telt_problem1	− 0.126**	0.036	− 0.073*	0.030	− 0.073*	0.030	− 0.074*	0.030
Alles Telt_problem2	− 0.016	0.043	− 0.073*	0.030	− 0.073*	0.030	− 0.074*	0.030
Alles Telt_problem3	-0.041	0.037	− 0.073*	0.030	− 0.073*	0.030	− 0.074*	0.030
Pluspunt_problem1	− 0.013	0.044	0.001	0.043	− 0.020	0.040	− 0.020	0.040
Pluspunt_problem2	− 0.046	0.055	− 0.060	0.054	− 0.020	0.040	− 0.020	0.040
Pluspunt_problem3	− 0.016	0.051	− 0.023	0.051	− 0.020	0.040	− 0.020	0.040

Fixed part	Model 9		Model 10		Model 11		Model 12	
	B	SE	B	SE	B	SE	B	SE
Rekenrijk_problem1	0.066	0.061	0.078	0.060	0.073	0.060	0.022	0.042
Rekenrijk_problem2	0.033	0.057	0.021	0.056	0.030	0.056	0.022	0.042
Rekenrijk_problem3	− 0.031	0.050	− 0.038	0.049	− 0.037	0.049	0.022	0.042
Random part class level								
$\sigma^2_{\text{Constant problem1}}$	0.056		0.058		0.059		0.060	
$r_{\text{Problem1\_problem2}}$	0.016		0.014		0.013		0.013	
$\sigma^2_{\text{Constant problem2}}$	0.076		0.078		0.080		0.080	
$r_{\text{Problem3\_problem1}}$	0.004		0.003		0.003		0.000	
$r_{\text{Problem3\_problem2}}$	0.029		0.030		0.030		0.030	
$\sigma^2_{\text{Constant problem3}}$	0.062		0.062		0.062		0.066	
Random part student level								
$\sigma^2_{\text{Constant problem1}}$	0.805		0.805		0.805		0.805	
$r_{\text{Problem1\_problem2}}$	0.126		0.126		0.126		0.126	
$\sigma^2_{\text{Constant problem2}}$	0.757		0.757		0.756		0.757	
$r_{\text{Problem3\_problem1}}$	0.023		0.023		0.023		0.023	
$r_{\text{Problem3\_problem2}}$	0.166		0.166		0.166		0.166	
$\sigma^2_{\text{Constant problem3}}$	0.688		0.688		0.688		0.688	
Deviance	10.240.473		10.245.384		10.246.804		10.251.280	

Problem 1 is the GMST, problem 2 in the non-routine art-geometry problems, and problem 3 is the routine problem. *p < 0.05., ** p < 0.001

References

Alloway, T. P., & Alloway, R. G. (2010). Investigatin the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*, 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839. https://doi.org/10.1038/nrn1201

Baer, J., & Kaufman, J. C. (2005). Bridging generality and specificity: The amusement park theoretical (APT) model of creativity. *Roeper Review, 27*(3), 158–163. https://doi.org/10.1080/02783190509554310

Baer, J., & Kaufman, J. C. (2006). Gender differences in creativity. *Journal of Creative Behavior*. https://doi.org/10.1002/j.2162-6057.2008.tb01289.x

Bahar, A., & Maker, C. J. (2015). Cognitive backgrounds of problem solving: A comparison of open-ended vs. closed mathematics problems. *Eurasia Journal of Mathematics, Science and Technology Education, 11*, 1531–1546. https://doi.org/10.12973/eurasia.2015.1410a

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence, 46*, 73–83. https://doi.org/10.1016/j.intell.2014.05.007

Bokhove, C., & Jones, K. (2018). Stimulating mathematical creativity through constraints in problem-solving. In N. Amado, S. Carreira, & K. Jones (Eds.), *Broadening the scope of research on mathematical problem solving* (pp. 301–319). Springer.

Carlson, M. P., & Bloom, I. (2005). The cyclic nature of problem solving: An emergent multidimensional problem-solving framework. *Educational Studies in Mathematics, 58*, 45–75. https://doi.org/10.1007/s10649-005-0808-x

Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26*, 673–682. https://doi.org/10.1016/j.econedurev.2007.10.002

Cropley, A. (2010). Defining and measuring creativity: Are creativity tests worth using? *Roeper Review, 27*, 72–79. https://doi.org/10.1080/02783190009554069

Darling-Hammond, L. (2000). Teacher quality and student achievement. *Education Policy Analyses Archives, 8*, 1–44.

Elia, I., van den Heuvel-Panhuizen, M., & Kolovou, A. (2009). Exploring strategy use and strategy flexibility in non-routine problem solving by primary school high achievers in mathematics. *ZDM Mathematics Education, 41*, 605–618. https://doi.org/10.1007/s11858-009-0184-6

Giofrè, D., Mammarella, I. C., Ronconi, L., & Cornoldi, C. (2013). Visuospatial working memory in intuitive geometry, and in academic achievement in geometry. *Learning and Individual*

Differences, 23, 114–122. https://doi.org/10.1016/j.lindif.2012.09.012

Goldstein, H. (2011). *Multilevel statistical models*. John Wiley & Sons.

Gravemeijer, K., Figueiredo, N., Feijs, E., Van Galen, F., Keijzer, R., & Munk, F. (2007). *Measurement and geometry in upper primary school*. Sense Publishers.

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Janssen, J., Scheltens, F., & Kraemer, J. (2007). *Rekenen-Wiskunde. Handleiding [Arithmetic-Mathematics. Manual]*. Arnhem, the Netherlands: CITO

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). *Wetenschappelijke verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8 [Scientific justification of the tests LOVS Arithmetic-Mathematics for grade 1–6]*. Arnhem, the Netherlands: CITO.

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

Kolovou, A., van den Heuvel-Panhuizen, M., & Bakker, A. (2009). Non-routine problem solving tasks in primary school mathematics textbooks: A needle in a haystack. *Mediterreanean Journal for Research in Mathematics Education, 8*, 29–66.

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

Kwon, O. N., Park, J. H. J. S., & Park, J. H. J. S. (2006). Cultivating divergent thinking in mathematics through an open-ended approach. *Asia Pacific Education Review, 7*, 51–61. https://doi.org/10.1007/BF03036784

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

Leikin, R. (2018). Openness and constraints associated with creativity-directed activities in mathematics for all students. In N. Amado, S. Carreira, & K. Jones (Eds.), *Broadening the scope of research on mathematical problem solving* (pp. 387–397). Springer.

Levav-Waynberg, A., & Leikin, R. (2012a). The role of multiple solution tasks in developing knowledge and creativity in geometry. *Journal of Mathematical Behavior, 31*, 73–90. https://doi.org/10.1016/j.jmathb.2011.11.001

Levav-Waynberg, A., & Leikin, R. (2012b). Using multiple solution tasks for the evaluation of students' problem-solving performance in geometry. *Canadian Journal of Science, Mathematics and Technology Education, 12*, 311–333. https://doi.org/10.1080/14926156.2012.732191

Liljedahl, P., Santos-Trigo, M., Malaspina, U., & Bruder, R. (2016). *Problem solving in mathematics education*. Springer.

Mann, E. L. (2006). Creativity: The essence of mathematics. *Journal for the Education of the Gifted, 30*, 236–260. https://doi.org/10.4219/jeg-2006-264

Meelissen, M. R. M., Netten, A., Drent, M., Punter, R. A., Droop, M., & Verhoeven, L. (2012). *Trends in achievement in reading, mathematics and science]*. Radboud University/University of Twente.

Mihajlović, A., & Dejić, M. (2015). Using open-ended problems and problem posing activities in elementary mathematics classroom. In F. M. Singer, F. Toader, & C. Voica (Eds.), *Proceedings of the 9th international MCG conference* (pp. 34–41). Nanyang Technological University.

Moerbeek, M. (2015). *SPA-ML. A software package for power analysis of trials with multilevel data*. Department of Methodology and Statistics, Faculty of Social and Behavioural Sciences, Utrecht University. https://tinyurl.com/spaml.

OECD. (2016). *PISA 2015 results (volume I): Excellence and equity in education*. OECD Publishing. https://doi.org/10.1787/9789264266490-en

Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement : A meta-analysis of national assessment of educational progress assessments. *Journal of Educational Psychology, 107*, 645–662. https://doi.org/10.1037/edu0000012

Runco, M. A., & Jaeger, G. J. (2012). The standard definition of creativity. *Creativity Research Journal, 24*, 92–96. https://doi.org/10.1080/10400419.2012.650092

Schoenfeld, A. (1983). The wild, wild, wild, wild, wild, world of problem solving (A review of sorts). *Learning of Mathematics, 3*, 40–47. https://doi.org/10.2307/40247835

Schoevers, E. M., Kroesbergen, E. H., & Kattou, M. (2020a). Mathematical creativity: A combination of domain-general creative and domain-specific mathematical skills. *Journal of Creative Behavior, 54*, 242–252. https://doi.org/10.1002/jocb.361

Schoevers, E. M., Leseman, P. P. M., & Kroesbergen, E. H. (2020b). Enriching mathematics education with visual arts: Effects on elementary school students' ability in geometry and visual arts. *International Journal of Science and Mathematics Education, 18*, 1613–1634. https://doi.org/10.1007/s10763-019-10018-z

Silver, E. A. (1995). The nature and use of open problems in mathematics education. *ZDM Mathematics Education, 27*, 67–72.

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

Silvia, P. J. (2008). Creativity and intelligence revisited: A latent variable analysis of Wallach and Kogan (1965). *Creativity Research Journal, 20*, 34–39. https://doi.org/10.1080/10400410701841807

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417–453. https://doi.org/10.3102/00346543075003417

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

Stein, M., Remillard, J. T., & Smith, M. (2007). How curriculum influences student learning. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 319–369). Information Age Publishing.

Urban, K. K. (2004). The Test for Creative Thinking—Drawing Production (TCT-DP), the concept, application, evaluation and international studies. *Pscychology Science, 46*, 387–397.

Urban, K. K. (2005). Assessing creativity: The Test for Creative Thinking—Drawing Production (TCT-DP). *International Education Journal, 6*, 272–280.

Urban, K. K., & Jellen, H. G. (1996). *Test for creative thinking—drawing production (TCT- DP)*. Van Swets & Zeitlinger.

Van de Weijer-Bergsma, E., Kroesbergen, E. H., Prast, E. J., & Van Luit, J. E. H. (2015). Validity and reliability of an online visual–spatial working memory task for self-reliant administration in school-aged children. *Behavior Research Methods, 47*, 708–719. https://doi.org/10.3758/s13428-014-0469-8

Van Grootheest, L., Huitema, S., Erich, L., van Hijum, R., Nillesen, C., Osinga, H., & van de Wetering, M. (2011). *De wereld in getallen [The world in numbers]*. Malmberg.

Van Zanten, M., & van den Heuvel-Panhuizen, M. (2018). Opportunity to learn problem solving in Dutch primary school mathematics

textbooks. *ZDM Mathematics Education, 50*, 827–838. https://doi.org/10.1007/s11858-018-0973-x

Warner, L. B., Alcock, L. J., Coppolo, J. J., & Davis, G. E. (2003). How does flexible mathematical thinking contribute to the growth of understanding? *Proceedings Conference of the International Group for the Psychology of Mathematics Education, 4*, 371–378.

Willemsen, R. H., Schoevers, E. M., & Kroesbergen, E. H. (2020). The structure of creativity in primary education: An empirical confirmation of the Amusement Park Theory. *Journal of Creative Behavior, 54*, 857-870. https://doi.org/10.1002/jocb.411