



Enriching Mathematics Education with Visual Arts: Effects on Elementary School Students' Ability in Geometry and Visual Arts

Eveline M. Schoevers¹  · Paul P. M. Leseman¹ · Evelyn H. Kroesbergen^{1,2}

Received: 21 December 2018 / Accepted: 11 September 2019 / Published online: 11 December 2019

© The Author(s) 2019

Abstract

This study evaluates the effects of the Mathematics, Arts, and Creativity in Education (MACE) program on students' ability in geometry and visual arts in the upper grades of elementary school. The program consisted of a lesson series for fourth, fifth, and sixth grade students in which geometry and visual arts were integrated, alongside with a professional development program for teachers. A quasi-experimental study was conducted in which three groups of teachers and their classes were investigated. One group of teachers taught the lesson series and followed a professional development program ($n = 36$), one group of teachers only taught the lesson series ($n = 36$), and a comparison group taught a series of traditional geometry lessons from mathematical textbooks ($n = 43$). A geometrical ability, creativity, and vocabulary test and a visual arts assignment were used in a pre- and post-measurements to test the effects of the MACE program. Results showed that students who received the MACE lesson series improved more than students who received regular geometry lessons only in geometrical aspects perceived in a visual artwork. Regarding students' understanding and explanation of geometrical phenomena and geometrical creative thinking, all students improved, but no differences between the groups were found, which implies that on these aspects the MACE program was as effective as the comparison group that received a more traditional form of geometry education.

Keywords Creativity · Visual art · Geometry · Elementary school · Teachers

Non-routine problem solving is important in elementary school mathematics (Kolovou, 2011). An example is when students have to create a paper model of a 12-sided dice. Students cannot simply apply a strategy, but have to recall, use, and combine facts,

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10763-019-10018-z>) contains supplementary material, which is available to authorized users.

✉ Eveline M. Schoevers
E.M.schoevers@uu.nl

Extended author information available on the last page of the article

skills, procedures, and ideas in a new and meaningful way to solve the problem. This requires creative and flexible thinking (Schoevers et al., 2019; Warner, Alcock, Coppolo, & Davis, 2003), which is also considered important in other disciplines in elementary education, such as visual arts (Sawyer, 2014). However, in educational practice, most teachers do not provide many opportunities for students to act creatively in mathematics (Gravemeijer, 2007; Kolovou, 2011) and visual arts (Bresler, 1999; Elfland, 1976). This might be largely due to the highly structured curriculum and mathematical textbooks that leave little room for these opportunities and may constrain creative practices that the teachers feel able and willing to engage in (Dobbins, 2009; Kolovou, 2011). The Mathematics, Arts, Creativity in Education (MACE) program was designed to change educational practice in this respect. The program aimed to teach domain-specific and overlapping learning goals of visual arts and geometry and to promote students' creative skills in both disciplines, by creating opportunities for students to act creatively in an integrated visual arts and geometry context. To reach the goals of the program, a lesson series for students was designed along with a professional development (PD). This study evaluated the intended effects of the MACE program.

Integrating Mathematics and Arts Education

The mathematical domain of geometry in education has the aim to teach students to understand and explain geometric phenomena from reality and to order and organize spatial situations (Jones, 2002; Van den Heuvel-Panhuizen & Buys, 2005), such as to draw a map or to reason about the effect of the height of the sun on the length of the shadow. This also requires students to obtain geometrical vocabulary to explain these phenomena (Buijs, Klep, & Noteboom, 2008). Furthermore, the ability to solve geometrical problems is considered important, since it is central to mathematics (Kolovou, 2011) and can be a way to construct new mathematical knowledge (Levav-Waynberg & Leikin, 2012). Problem solving requires creative thinking: students need to be able to combine known concepts, skills, procedures, and ideas from mathematics and other domains in a new way to solve the problem (Schoevers et al. 2019), which can contribute to the construction of new knowledge and deeper understanding of geometrical concepts (Levav-Waynberg & Leikin, 2012; Warner et al., 2003). Based on these core aspects of geometry education, we defined geometrical ability in this study as students' ability to understand and explain geometric phenomena, to describe these phenomena by using geometrical vocabulary, and to creatively solve geometrical problems.

Visual arts education has the aim to teach students to develop their visual-imaginative abilities by using their experiences of reality and visualizing these experiences (Braakhuis, Von Piekartz, Vogel, & De Graaf, 2012). The main aspects of visual arts education are visual art production, perception (observing, interpreting, and analyzing) and reflection (thinking and speaking about a visual art product during or after its production; Haanstra, 2014). The (cyclic) creative process is central in teaching the visual arts curriculum (Sawyer, 2014; Stichting Leerplanontwikkeling, 2015).

Thus, in both visual arts and geometry, creative processes play a central role. One of the key cognitive processes of creativity is to overcome fixation on ideas and to break

away from established mindsets. Currently, visual arts and mathematics are usually taught in separate disciplines, and students, as a consequence, rely on the highly familiar knowledge and routines specific for these domains. We hypothesize that by integrating mathematics and visual arts education, students will be stimulated to integrate different conceptual systems from both disciplines, which could activate students to create something new and meaningful (Haylock, 1987; Plucker & Zabelina, 2009).

The MACE Pedagogy

The MACE pedagogy comprises the following key features: visual arts perception, open activities, reflection, communication with peers, and a specific role of the teacher. In this section, we elaborate on these features and describe how these can enhance students' ability in both geometry and visual arts. The more formal features of the program (e.g. duration and number of lessons) are discussed in the "Methods" section.

Visual arts perception plays a role right at the start of the MACE lesson series. Artworks are discussed in a whole-class setting by using an interactive whiteboard in relation to the interdisciplinary lesson theme (e.g. in a lesson about *perspective*: Can you tell me how the photo would have looked like if the photographer had used another point of view?) and with the use of visual thinking strategies (e.g. "What's happening in this picture?," "What do you see that makes you say that?," "What more can we find?"; Housen, 2002). Students thus learn to observe and analyze the visual aspects of a piece of art, to consider the view of others, and to reflect on and discuss about possible interpretations (Hailey, Miller, & Yenawine, 2015). Educating visual arts perception also enables students to extract shapes and objects from a visual scene which, in turn, can influence their recognition and representation of visual information (Kozbelt, 2001; Tishman, MacGillivray, & Palmer, 1999). Since artworks are discussed in an interdisciplinary context, students may learn to recognize and represent the visual aspects in artworks such as "space," "shapes," and "composition" (Stichting Leerplanontwikkeling, 2018). Furthermore, visual arts perception could improve students' geometrical reasoning when they are asked to imagine how an artwork would look like if particular changes were made (Tishman et al., 1999; Walker, Winner, Hetland, Simmons, & Goldsmith, 2011).

Furthermore, within the MACE program, *open activities* are used in which students have to produce a visual artwork related to a theme at the boundaries of visual arts and geometry (e.g. perspective or symmetry). The activities contain problems for which students do not know predetermined or obvious solutions (Kolovou, 2011). Activities are "open" if they invite different solutions and if problem solving methods are open for interpretation (Silver, 1995). For example, in one of the open tasks of the MACE program, students have to make a three-dimensional representation of a two-dimensional painting by using free materials like egg boxes or paper rolls. These type of activities can stimulate students to visualize their experiences, by exploring materials, meaning, and several visual aspects (Stichting Leerplanontwikkeling, 2018). In addition, students may learn to order and organize spatial situations that require geometrical reasoning (e.g. how certain objects need to be structured to construct the three-dimensional representation of the painting). Furthermore, investigation and manipulation of the materials and thus exploration of the physical environment can help

students to form visuospatial and sensorimotor representations of geometrical structures that are embedded in the environment. This may advance their thinking during geometrical problem solving (Carbonneau, Marley, & Selig, 2013; Núñez, Edwards, & Matos, 1999).

Reflection at the end of each lesson is considered very important within this integrated pedagogy. The knowledge and skills acquired during visual arts reception and production may still be rather implicit. It is important that students learn to reflect on their (creative) process of producing an artwork, and to communicate about the thinking processes with classmates and their teacher. By clarifying what was going on and what they have learned, the implicit knowledge and skills acquired can be made explicit; reflection can extend and modify existing knowledge (Chi, De Leeuw, Chiu, & Lavancher, 1994).

Communication with peers is highly promoted in the MACE lessons. Students are stimulated to collaborate in the learning activities, and situations are created in which they are asked to communicate, discuss, and exchange their ideas. In this way, students can compare their ideas and view ideas from different standpoints, which can enhance creative thinking (Beghetto & Kaufman, 2010; Taggar, 2002). It can also increase students' learning in geometry, because they have to explain their thinking and they get feedback and encounter other points of view (Jarvis, 2001). This may enable students to reach a higher level of understanding (Van den Heuvel-Panhuizen & Drijvers, 2014) and encourage (geometrical) language development (Van Lier, 1996).

Also, *a specific role of the teacher* is required in the MACE lessons. Teachers are advised to act as a facilitator: asking questions to extend students' thinking and reasoning, instead of merely transferring knowledge (Bostic, 2011). In addition, they are stimulated to use academic geometrical vocabulary, for example in reformulating students' thinking, to improve the geometrical vocabulary of their students (Henrichs & Leseman, 2014). To stimulate creativity, teachers are advised to create an open atmosphere in which students' ideas are central, to ask open and activating questions that invite students to generate multiple answers, and to be open to these ideas (Davies et al., 2014; Schoevers et al., 2019). Although this specific role of the teacher is explained in the MACE teaching manual, also, a PD program for teacher was designed to assist teachers in implementing the MACE lesson series in educational practice.

Aim of the Study

The aim of the current study was to investigate the effects of the MACE program on students' geometrical ability and perception of visual arts in a quasi-experimental design. The comparison condition consisted of students that received regular (textbook-based) geometry and visual arts education. Compared with students who received the regular geometry lessons, students who received the MACE lessons could engage in open, hands-on, and interdisciplinary activities; had more opportunities to express, discuss, and exchange ideas with classmates and teachers; and had more time to reflect on the lesson. The presence of these aspects in the MACE lessons, relative to the absence or reduced presence of these aspects in the regular textbook-based geometry lessons, were the reasons to hypothesize that students in the MACE program would improve more in geometrical ability than students in the comparison condition.

More specifically, students were expected to better explain and understand geometric phenomena than students in the comparison group, to use more geometrical words to explain these phenomena, and to think more creatively in solving geometrical problems. Furthermore, we hypothesized that students who received the MACE program would observe and describe more geometrical aspects in visual artworks, compared with students in the comparison condition, because in regular visual arts education, visual artworks are seldom discussed, and especially not with a specific focus on the geometrical aspects (Kruiter, Hoogeveen, Beekhoven, Kieft, & Bomhof, 2016).

Methods

Participants

In this study, 2909 students in grades 3, 4, 5, and 6 situated in 121 classes and 57 schools participated. Students varied in socioeconomic background and mathematical ability. 4th, 5th, and 6th grade teachers were recruited for the MACE program by sending flyers to 428 regular elementary schools in all regions of the Netherlands. 11.68% of the schools were willing to participate. To evaluate the effect of the MACE lesson series and the MACE PD program, the participating schools were assigned to one of three conditions. Although we planned to randomly assign schools to the three conditions, this was not completely possible. For logistic reasons, schools had to be within a reasonable traveling distance of the PD training locations. Furthermore, some teachers were not available for the MACE PD program. In the first condition, the teachers followed the MACE PD program and taught the MACE lesson series to their students; in the second condition, the teachers taught the lesson series, but without following the PD program; teachers in the third condition taught regular geometry lesson from existing mathematical textbooks (but were offered to follow the MACE program after the study). Since geometry lessons are taught irregularly, spread over the school year, a lesson series was established for this study in which geometry lessons from various Dutch mathematical textbooks were combined to have the same time and intensity of geometry instruction as in the MACE program. In this way, students in the comparison group received the same geometrical content as students in the MACE program, but in a regular way. Geometry lessons from mathematical textbooks were used that were similar in geometrical content, but different from MACE lessons in pedagogy and enrichment with visual arts aspects. In the study, 10 classes and 197 students dropped out, because teachers experienced a too high workload (6.8% of the total number of students). In Table 1, the three groups are described.

The MACE Program

The MACE program consisted of a lesson series for fourth, fifth, and sixth grade students in which geometry and visual arts education were integrated along with a PD program for teachers (Keijzer, Oprins, De Moor & Schoevers, 2018). In a pilot study with 15 teachers, an initial version of the program was formatively evaluated and adjusted afterwards.

Table 1 School and class characteristics

	Experimental group 1 ^a	Experimental group 2 ^b	Comparison group
Number of schools	21	17	18
Number of classes	33	33	45
Number of students	801	811	1100
% very low SES ^c	3.2%	0.3%	3.8%
% boys	47.8%	47.8%	50.4%

^a Lesson series and PD program

^b Only lesson series

^c Low socioeconomic status (SES): at least one of the parents had elementary school as the highest completed education

Description of the MACE Lessons. The series of MACE lessons consisted of nine lessons which took each 60 – 90 minutes; five lessons related to the theme space and four to the theme patterns. Each lesson started with a whole-class introduction (15 – 25 minutes), in which students observed and discussed visual artworks in relation to the topic of the lesson. The introduction was followed by an open activity in which students created an artwork (25 – 30 minutes). During the activities, students worked mostly in (small) groups. A lesson ended with whole-class reflection (10 minutes), in which students' creative process and products were discussed as well as what students had learned.

Below we describe the MACE lesson “playing with perspective” in more detail. In Appendix 1, a short description of each MACE lesson is given. In the lesson “Playing with perspective,” the teacher starts with an introduction in which six artworks are discussed in which the artists have explored perspective and viewpoints, such as the artwork “Another World” by M.C. Escher and photo's in which photographers manipulated perspective. Questions that a teacher could ask during this introduction are stated in the manual, such as “How did the artist created this effect?,” “Can you tell something about the viewpoint of the artist and what could be the reason for this?,” and “How would the photo look like when they would have used another point of view?.” After the introduction, students make photos in which they create visual illusions by playing with perspective and point of view in groups of 3 – 4 students in- or outside the school building. After 15 – 20 minutes, students select their two best photos. At the end of the lesson, the teacher discusses the selected photos of the students and the process of making the photos in a whole-class setting. Questions that the teacher could ask are stated in the manual, such as “What effect did you want to create?,” “What did you do to create this effect?,” “What perspective did you use?,” and “Where would you stand if we would draw a map?”. Furthermore, the teacher has to ask students to reflect on what they have learned.

Description of the PD Program. The PD program for teachers consisted of five sessions (2.5 hours each), guided by experts in the field of mathematics and visual arts education. After each PD session, teachers taught one or two MACE lessons in their own schools (see also Fig. 1). The aim of the PD program was to train teachers how to

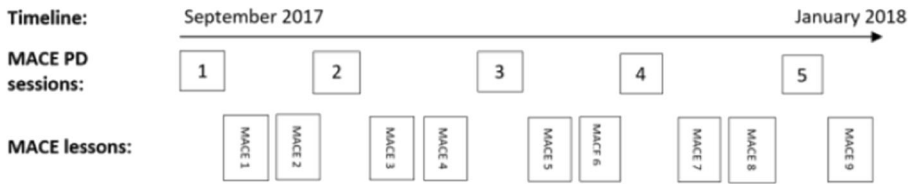


Fig. 1 An overview of the MACE program

stimulate students' creative thinking in this integrated visual arts and mathematics program. A further aim was to create a positive attitude of the teachers towards geometry, visual arts, and the integration of both. A third aim was to increase teachers' geometrical knowledge and their pedagogical content knowledge of geometry and visual arts education.

Within the PD program, active learning was considered important. Therefore, interactive methods were used in the sessions. Teachers, for example, had to experience the MACE lessons themselves, watched film fragments of other teachers, and had to design a hypothetical learning trajectory. Afterwards, teachers always had to discuss and reflect on these activities. The content of the PD program was related to the classroom practice. Furthermore, reflection on the MACE lessons was important; it could support on-going learning and encourage change.

Regular Geometry Lesson Series

A lesson series was established for the comparison group in which geometry lessons from several widely used Dutch mathematical textbooks were combined and adjusted to have the same time and intensity of geometry instruction as in the MACE lesson series. We considered mathematical textbook lessons as regular geometry lessons, because most teachers, especially in the Netherlands, use mathematical textbooks to teach the geometry curriculum (Hop, 2012). Most used Dutch mathematics textbooks mainly offer closed-ended routine problems (Van Zanten & Van den Heuvel-Panhuizen, 2018). The (type of) problems, duration, and lesson structure of these textbook-based geometry lessons were applied and used in the regular geometry lesson series for the comparison condition. The regular geometry lessons series consisted of 7 lessons, of which four related to two and three dimensionality, block constructions, floor maps, and perspective. The other three lessons were related to patterns, symmetry, and rotation. The regular geometry lessons took between 30 and 40 minutes. Each lesson started with a short whole-class introduction (5 – 10 minutes) in which the subject of the lesson was introduced. Afterwards, students independently worked on geometry problems in a workbook (15 – 20 minutes). These were mainly multiple choice problems. The lesson was completed by a short (5 minute) whole-class reflection on the lesson.

Instruments

Geometric Ability Test. The GAT measured whether students understood and could explain geometrical phenomena. The GAT was developed for this study. The test took

between 20 and 30 minutes and was stopped after 30 minutes. The test consisted of 11 geometry problems. The test started with four closed-ended routine problems that called upon spatial sense and spatial visualization (see the left sample question in Fig. 2) and ended with seven art-geometry problems in which students mainly had to reason geometrically in relation to a painting (see the right sample question in Fig. 2). As the test was specifically developed for this study, no information on item difficulty and other item characteristics was available. Therefore, within problem type, the order of the problems was randomized and the same for all students to prevent unknown order effects. Two equivalent versions (i.e. A and B) of the GAT were respectively used as pre- and posttests.

Scoring. The spatial visualization 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. The geometry problems related to visual arts were more complex and could also yield partially correct solutions. Therefore, two points were given for an answer in which students showed to be able to reason correctly about geometric phenomena (e.g. when students explained why things in front of the painting are bigger than in the back of the painting and, for example, would use the term perspective). One point was given for an answer in which the reasoning was not complete (e.g. “by painting a line down in a curve”). Zero points were given for answers in which the question was repeated or there was no reasoning about the question involved (e.g. “because when you paint you can make everything”). The last art-geometry question of the test was considered too difficult for grades 4 and 5; only 9% of the students scored one or two points (Hopkins & Antes, 1978). Therefore, this question was not used in this study. Furthermore, 27.1% of the students were able to finish the GAT in the maximum time of 30 minutes. Therefore, we calculated an average score based on the number of questions the students were able to complete. The maximum score that could be obtained was 1.6. Since items were not in an ascending order of difficulty, this was the most valid way to calculate a total score.

Reliability and Validity of the GAT. Test-retest reliability of version A is acceptable ($r = 0.66$) and good for version B ($r = 0.77$). Alternative forms reliability of the GAT is sufficient ($r = 0.80$). Furthermore, forms A and B are similar regarding difficulty of the items. Criterion validity of the GAT, as measured by the correlation between the GAT and general math ability score, was moderate ($r = 0.40 - 0.42$), but sufficient since the mathematical domain geometry covers only a small part of the general mathematical ability test. Tests at the pre- and posttests were scored by four raters. Interrater reliability (IRR) was sufficient to excellent for all items on both the pre- and posttests

Peter folds the cube. He puts the black side of the cube at the bottom. What fruit is on top of the cube?



The painting is flat, but you can see depth. How is that possible? Explain in words. You can also draw.



Fig. 2 Sample questions of the GAT

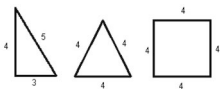
($\kappa = 0.81 - 1.00$, $ICC = 0.67 - 1.00$). In line with our expectations, the internal consistency of the GAT was not very high for both versions (version A, $\alpha = 0.62$; version B, $\alpha = 0.58$), because the GAT represents a heterogeneous set of skills.

Geometrical Vocabulary. Active geometrical vocabulary was scored if students correctly used geometrical words in the open questions of the GAT. A list of geometrical words was composed by using the learning goals of elementary school geometry education (Van den Heuvel-Panhuizen & Buys, 2005) and elementary school teacher education (Van Zanten, Barth, Faarts, Van Gool, & Keijzer, 2009). We distinguished between tier 1 and tier 2 words (Henrichs & Leseman, 2014). Domain-general academic tier 3 words were not included in this study, because they were not of interest. Tier 1 words are geometrical words used in a daily language environment (e.g. “turned” or “flat”). Tier 2 words are domain-specific academic geometrical words used in mathematics education (e.g. “vertical,” “horizontal,” “square,” and “triangle”). The complete list of words can be found in Appendix B.

Scoring. The total number of words, of tier 1 and of tier 2 words appearing in the writings, was counted for each question in the GAT. Regarding tier 2 words, a token (i.e. total number of tier 2 words used) and type score (i.e. the number of different tier 2 words used) were calculated. Since, these measures were highly correlated ($r > 0.90$), we only used the token score. Next, a ratio of the number of tier 2 geometrical words to the total number of words used was calculated for each question of the GAT. This ratio was calculated to adjust for differences in students’ wordiness. A total score for the GAT was calculated by averaging the ratio scores of the different questions. Geometrical vocabulary was scored by four raters. IRR was good in this study with regard to the total word written ($ICC > 0.99$), the number of tier 1 words ($ICC > 0.88$), and the total number of tier 2 words ($ICC > 0.86$) for both the pre- and posttests.

Geometrical Creativity Test. The domain-specific Geometrical Creativity Test (GCT) was used to measure geometrical creative thinking (see also Fig. 3). It was developed for this study and based on the Mathematical Creativity Test used in the study of Schoevers, Kroesbergen and Kattou (2018). The GCT took between 20 and 30 minutes and was stopped after 30 minutes. It consisted of four geometry questions and one problem-posing question, which were open-ended and could have multiple correct answers. Multiple solution and problem posing tasks are commonly used to measure creativity in mathematics (Leikin, Koichu, & Berman, 2009; Leikin, 2009). Students were instructed to provide multiple, but distinct solutions, which moreover had to be original. In the problem-posing questions, students were asked to pose mathematical

Look carefully at the three given shapes. Which shape does not belong in the same group as the other two? Explain your answer. Are there more than one possible answers? If so, write down as many answers as possible.



Shape ____ does not belong in this group of shapes because

.....

Fig. 3 A sample question of the GCT version A

questions based on a photo. Version A was used for the pretest, and version B was used for the posttest.

Scoring. For the scoring of the GCT, we used the scoring scheme of Leikin for creativity in the individual solution space (Leikin, 2009; Levav-Waynberg & Leikin, 2012). Within this scheme, a distinction between fluency, flexibility, and originality is made. The scheme was further elaborated in Leikin (2009); Levav-Waynberg and Leikin (2012); and Leikin et al. (2009). Fluency was calculated by adding the number of correct answers for each question. Based on the scheme, each solution was scored on flexibility and originality. Next, a final score per solution was calculated as the product of $\text{Flexibility}_i \times \text{Originality}_i$. Afterwards, a creativity score per question was computed as: $\text{Fluency}_j \times (\sum (\text{Flexibility}_i \times \text{Originality}_i))$. The sum of these creativity scores was used as total score.

Reliability and Validity of the GCT. With regard to the test-retest and alternative forms of reliability, we used the fluency score (as described by Kattou, Kontoyianni, Pitta-Pantazi, & Christou, 2013) as indicator of creativity. Test-retest reliability of versions A ($r = 0.84$) and B ($r = 0.89$) are good. Alternative forms reliability of the GAT is sufficient ($r = 0.68$). Furthermore, the difficulty of both versions was similar. In this study, the GCT was scored by seven raters. IRR was sufficient to excellent for all scores per solution on both pre- and posttests ($\text{ICC} = 0.72 - 0.99$). In line with our expectations, the internal consistency of the GCT was not very high in this study for both versions (version A, $\alpha = 0.68$; version B, $\alpha = 0.55$), because geometry comprises a heterogeneous set of knowledge and skills.

Visual Arts Assignment. In the Visual Arts Assignment (VAA), students have to write down as much as they can about a painting (see Fig. 4) in answering the following questions: “What is going on in this painting?,” “What do you see that makes you say that?,” and “What more can you find?” (Housen, 2002). The same task was used as a pre- and posttests.

Scoring. A scoring scheme for the VAA was created for this study based on literature about visual arts perception in education (see Table 2; KPC-Groep, 2000; Stichting



Fig. 4 Emmanuel De Witte—Interior with a woman at the virginal (Museum Boijmans van Beuningen, Rotterdam, the Netherlands)

Table 2 Scoring aspects of the VVA

Scoring aspect	Operationalization
a) Space	The aspect of “space”: The student expresses whether something is in front/back/in/ between/under etc., how the space is ordered, spatial constructions, filling up space (closed), encompassing space (open). Example: I see a lady cleaning in the <i>back</i> room
b) Space suggestion	Space suggestion: big-small; mirroring; depth etc. Example: I see <i>depth</i> in the painting; it is a <i>big</i> house; I see <i>3D</i> in the painting; I see that <i>through the mirror</i> .
c) Shape	Shapes (two-dimensional and three-dimensional); shape characteristics (point, angle, etc.); lines. Example: I see a <i>squared</i> mirror
d) Composition	Composition—how the whole painting is built up in pieces. For example, ordering by shape, color, or texture; rhythm, repetition of shapes; motives for decoration; pattern; ordering; balance and meaning; symmetric-asymmetric, horizontal-vertical). Example: I see a <i>pattern</i> on the tiles; the painting is <i>light in front and dark in the back</i> ; there are <i>three rooms</i> in the house.

Leerplanontwikkeling, 2015; Van Onna & Jacobse, 2008). The VVA was scored on four aspects that were related to geometry, namely “space,” “space suggestion,” “shape,” and “composition.” We scored how often each aspect occurred in the written text of the student. Furthermore, the number of words written was counted. Next, each score on each aspect was divided by the number of words written and multiplied by 100, to take the talkativeness of students into account. The pre- and posttests were scored by five raters. Interrater reliability of most aspects was good to excellent ($ICC = 0.76 - .99$).

Procedure

Data were collected in the fall of 2017 by the first author and twelve research assistants with a bachelor or master’s degree in (special) education. Before the start of the MACE program, in one session, the pretests were administered to the whole class by a research assistant who read aloud the test instructions. Posttests were administered after the MACE program in the same way as the pretests. Furthermore, teachers had to provide information about students’ age, gender, mathematical ability (based on the national mathematical ability tests), and the educational level of both students’ parents, based on information from school records. Pre- and posttests were coded afterwards by the same research assistants. Research assistants received extensive training and had to reach sufficient interrater reliability with the master coder before they were allowed to administer the pre- and posttests in the classroom, to conduct observations in the classroom, and to code part of the tests. Passive informed consent of the parents was obtained before the start of the study. 0.8% of the students did not have consent for this study.

Analyses

To take the nested structure of the data into account, multilevel analyses with three levels were conducted in HLM6 (Hox, Moerbeek, & Van de Schoot, 2018). The first level represented the repeated measurements (pre- and posttest), as nested within individuals. The second level represented the students, and the third level the classes. It was not

necessary to include the school level as a fourth level, since there was only little variance in the different pretests located at this level (1 – 5% of the variance). For each outcome variable, first a base model was created with time as a predictor. In a second model, student-level covariates were added to control for spuriousness: grade, gender (Frost, Hyde, & Fennema, 1994), SES (Crane, 1996), and general mathematical ability. It was expected that students' general mathematical ability would not have an effect on students' performance on the VAA and student's use of geometrical vocabulary, and therefore, this variable was not used as a covariate in the multilevel models related to these measures. Furthermore, it was expected that students with a low SES would score lower on the GAT, GCT, and VAA (Crane, 1996). Therefore, we controlled for low SES by using two dummy variables: students with a low SES and students with a very low SES. Students had a very low SES if elementary school was the highest completed education of at least one of the parents. Since only a small percentage of students had a very low SES, we also included students with a low SES: vocational education was the highest completed education of both parents. In a third model, third-level predictors were added as dummy variables: lesson series condition and PD program condition. Furthermore, two class-level covariates were added that were expected to influence the results of the effects of the MACE program: the number of MACE lessons given by the teacher and the number of MACE PD sessions followed by the teacher. In the fourth model, the random slope of time was added to investigate whether students' growth on an outcome variable differed per class. In the fifth model, dummy variables of condition (whether or not students had received MACE lessons and whether or not the teachers had participated in the MACE PD program) were added as predictors of the slope of time to investigate whether differences between classes could be explained by conditions.

Before the multilevel analyses were conducted, data were screened and prepared and a missing value analysis was conducted. On the pretest, between 6.6 and 9.2% of the data were missing. On the posttest, between 9.7 and 10.6% of the data were missing. Data were missing because some students were ill or for other reasons not present in the class during the test administration. Furthermore, assumptions for multilevel analysis were checked (Hox et al., 2018). The sample size was sufficient, as calculated with SPA-ML (Moerbeek, 2015), based on a desired power of 0.80 and an assumed effect size of 0.30 in a three-level model. Furthermore, the assumption of linearity was met for all variables; however, for one variable, an outlier was detected and deleted (tier 2 academic words in the GAT on student level). In addition, the assumption of normally distributed residuals was mildly violated for the residuals at the first and second level of the GCT. However, this was not considered a problem since the maximum likelihood estimator is robust against this violation with a large sample size (Hox et al., 2018). Normality of residuals was more seriously violated for the geometry words used in the GAT, and for the aspects on the VAA. Therefore, robust standard errors were used and reported (Hox et al., 2018).

Results

In Tables 3, 4, 5, and 6, the means and standard deviations of students' pre- and posttest scores on the GAT, geometrical vocabulary, GCT, and VAA are described. In Table 7,

Table 3 Descriptive statistics of students' scores on the GAT

	Experimental group 1 <i>M (SD)</i>	Experimental group 2 <i>M (SD)</i>	Comparison group <i>M (SD)</i>
GAT pretest	0.54 (0.30)	0.60 (0.28)	0.52 (0.29)
GAT posttest	0.66 (0.29)	0.73 (0.28)	0.61 (0.28)

Table 4 Descriptive statistics of ratio scores of geometrical vocabulary

	Experimental group 1		Experimental group 2		Comparison group	
	Pretest <i>M (SD)</i>	Posttest <i>M (SD)</i>	Pretest <i>M (SD)</i>	Posttest <i>M (SD)</i>	Pretest <i>M (SD)</i>	Posttest <i>M (SD)</i>
Tier 1 ^a words in the GAT	0.11 (0.06)	0.12 (0.06)	0.11 (0.06)	0.12 (0.06)	0.11 (0.07)	0.12 (0.07)
Tier 2 ^b words in the GAT	0.06 (0.09)	0.06 (0.09)	0.06 (0.08)	0.05 (0.07)	0.07 (0.10)	0.05 (0.07)
Number of written words	40.94 (26.71)	51.09 (31.02)	46.62 (27.48)	51.70 (27.56)	41.23 (23.39)	42.74 (24.54)

^a Daily geometrical words

^b Academic geometrical words

Table 5 Descriptive statistics of the GCT

	Experimental group 1 <i>M (SD)</i>	Experimental group 2 <i>M (SD)</i>	Comparison group <i>M (SD)</i>
GCT pretest	900.42 (915.66)	1032.85 (1040.36)	865.58 (919.15)
GCT posttest	912.41 (991.04)	1151.43 (1197.60)	910.06 (1022.64)

Table 6 Descriptive statistics of the VAA

Aspect	Experimental group 1		Experimental group 2		Comparison group	
	Pretest <i>M (SD)</i>	Posttest <i>M (SD)</i>	Pretest <i>M (SD)</i>	Posttest <i>M (SD)</i>	Pretest <i>M (SD)</i>	Posttest <i>M (SD)</i>
a) Space	3.10 (3.66)	4.65 (4.07)	3.53(3.66)	5.71 (4.47)	3.32 (3.45)	4.00 (4.02)
b) Space suggestion	0.77 (1.58)	1.49 (3.61)	0.97 (1.79)	1.29 (2.21)	0.86 (1.60)	0.76 (1.60)
c) Shape	0.10 (0.54)	0.19 (0.96)	0.10 (0.60)	0.17 (0.72)	0.20 (1.35)	0.14 (0.66)
d) Composition	0.11 (0.50)	0.51 (1.42)	0.06 (0.36)	0.56 (1.33)	0.10 (0.50)	0.14 (0.57)

Table 7 Spearman correlations between the measures on the pretest

	1	2	3	4	5	6	7
1. GAT	–						
2. GCT	0.28**	–					
3. Daily geometrical words	0.24**	0.09**	–				
4. Academic geometrical words	–0.14**	–0.06**	–0.18**	–			
5. VAA “space”	0.20**	0.16**	0.06**	–0.03	–		
6. VAA “space suggestion”	0.13**	0.11**	0.04	0.01	0.09**	–	
7. VAA “shape”	0.01	0.03	0.01	–0.01	0.07**	0.05*	–
8. VAA “composition”	0.08**	0.05*	0.02	0.00	0.06**	0.03	0.07*

**Significant at $p < 0.001$ (two-tailed), *significant at $p < 0.05$ (two-tailed)

Spearman correlations between all measures on the pretest are presented. The correlations, for example, indicated that the measures of geometrical ability—GAT, GCT, and daily geometrical words—are significantly positively related. However, students’ use of academic geometrical words was negatively related to the GAT, GCT, and daily geometrical words.

Multilevel analyses showed that there was a linear relation between time and students’ scores on the GAT, $t(3831) = 18.71$, $p < 0.001$. On average, students’ scores on the GAT increased between the pre- and posttests. Furthermore, the relation between time and the GAT differed per class, $\chi^2(88) = 343.50$, $p < 0.001$. These differences, however, could not be explained by the MACE lessons series ($t(90) = 1.61$, $p = 0.11$), nor by the MACE PD program ($t(90) = -0.29$, $p = 0.77$). In Table 8, only the results of the final model are presented. The results of all other models are included in Appendix 3.

With regard to geometrical vocabulary, a significant linear relationship was found between time and students’ use of daily (tier 1 $t(4083) = 4.74$, $p < 0.001$) and academic geometrical words (tier 2 $t(4083) = -4.11$, $p < 0.001$) in the GAT. The proportion of daily geometrical words increased between the GAT pre- and posttests, but decreased for academic geometrical words. The rate of improvement differed between classes (tier 1 $\chi^2(92) = 263.58$, $p < 0.001$; tier 2 $\chi^2(92) = 134.52$, $p < 0.01$); students in some classes improved more regarding their use of daily and academic geometrical words in the GAT than students in other classes. With regard to the daily geometrical words (tier1), the difference could be explained by the participation in the MACE program ($t(94) = 2.60$, $p < 0.05$), but not by participation in the MACE PD program ($t(94) = -1.42$, $p = 0.16$). When a class had received the MACE lesson series, the positive relation between time and daily geometrical words (tier 1) used in the GAT became stronger. With regard to academic geometrical words (tier 2), the difference between classes could not be explained by the MACE lesson series ($t(94) = 0.36$, $p = 0.72$), nor by participation of the teacher in the MACE PD program ($t(94) = 0.99$, $p = 0.33$). See Table 8 for the final models of the geometrical words used in the GAT.

Furthermore, multilevel analyses showed a linear relationship between time and geometrical creativity ($t(3977) = 3.04$, $p < 0.01$). On average, students’ scores on the GCT increased between the pre- and posttests. Furthermore, we found that the rate of improvement on the GCT varied between classes, $\chi^2(90) = 238.94$, $p < 0.001$. This variation, however, could not be explained neither by the MACE lesson series ($t(91) = 0.74$, $p = 0.46$) nor by the MACE PD program ($t(91) = -1.28$, $p = 0.20$; see final model in Table 8).

Table 8 Multilevel results (final models) regarding geometrical ability

Fixed part	Understanding and explanation of geometric phenomena (GAT) B (s.e.)	Daily geometrical words used in the GAT (tier1) ^a B (s.e.)	Academic geometrical words used in the GAT (tier 2) ^a B (s.e.)	Geometrical creativity (GCT) B (s.e.)
Intercept	0.37 (0.02)**	0.11 (0.00)**	0.08 (0.01)**	444.12 (66.40)**
Time	0.09 (0.02)**	0.01 (0.00)	-0.02 (0.00)**	65.84 (55.65)
Grade 3 (dummy) ^b	-	-	-	24.29 (148.03)
Grade 5 (dummy) ^b	0.10 (0.02)**	0.01 (0.00)*	0.00 (0.00)	299.18 (64.08)**
Grade 6 (dummy) ^b	0.21 (0.02)**	0.01 (0.00)**	-0.01 (0.00)*	555.55 (66.91)**
Gender ^c	0.10 (0.01)**	-0.01 (0.00)**	-0.01 (0.00)**	269.84 (34.22)**
SES low ^d (dummy)	-0.08 (0.03)**	-0.01 (0.01)*	0.01 (0.01)	-126.79 (107.48)
SES very low ^d (dummy)	-0.15 (0.03)**	-0.02 (0.01)**	0.02 (0.01)	-143.69 (106.72)
General mathematical ability	0.12 (0.01)**	-	-	223.12 (18.99)**
Lesson series condition (dummy) ^e	-0.12 (0.11)	-0.02 (0.01)	-0.01 (0.01)	343.96 (382.42)
PD program condition (dummy) ^f	-0.11 (0.08)	0.02 (0.01)	0.02 (0.02)	142.65 (289.21)
Number of MACE lessons given	0.02 (0.01)	0.00 (0.00)	0.00 (0.00)	-35.62 (44.38)
Number of PD sessions followed	0.01 (0.02)	-0.00 (0.00)	-0.00 (0.00)	-38.73 (65.30)
Time × lesson series	0.04 (0.03)	0.02 (.01)*	0.00 (0.01)	68.19 (91.77)
Time × PD program	-0.01 (0.03)	-0.01 (0.01)	0.01 (0.01)	-139.02 (108.23)
Random part				
σ_e^2	0.029	0.00320	0.00572	549,285.85
σ_{u0}^2	0.024	0.00049	0.00089	282,663.42
σ_{v0}^2	0.007	0.00016	0.00029	69,448.79
σ_{v1}	0.008	0.00048	0.00027	82,455.98

*Significant at $p < 0.05$, **significant at $p < 0.01$

^a Robust standard errors are used

^b Grade 4 is used as a reference group

^c (Male = 0, female = 1)

^d (Medium or high SES = 0, very (low) SES = 1)

^e (Received no MACE lessons = 0, received MACE lessons = 1)

^f (Received no MACE PD sessions = 0, received MACE PD sessions = 1)

Table 9 Multilevel results (final models) regarding the aspect used in the VAA

	Aspect “space” ^a	Aspect “space suggestion” ^a	Aspect “shape” ^a	Aspect “composition” ^a
Fixed part	B (s.e.)	B (s.e.)	B (s.e.)	B (s.e.)
Intercept	2.85 (0.24)**	0.61 (0.09)**	0.18 (0.05)**	0.03 (0.03)
Time	0.67 (0.23)**	-0.09 (0.09)	0.01 (0.03)	0.04 (0.03)
Grade 3 (dummy) ^b	-1.61 (0.67)*	-0.23 (0.26)	-0.11 (0.06)	-0.01 (0.06)
Grade 5 (dummy) ^b	0.10 (0.22)	0.22 (0.09)*	-0.04 (0.04)	0.05 (0.04)
Grade 6 (dummy) ^b	0.80 (0.22)**	0.25 (0.09)*	-0.05 (0.04)	0.10 (0.04)*
Gender ^c	0.56 (0.14)**	0.15 (0.06)*	0.03 (0.03)	0.03 (0.03)
SES low (dummy) ^d	-0.80 (0.30)*	-0.08 (0.13)	-0.03 (0.06)	0.03 (0.07)
SES very low(dummy) ^d	-0.92 (0.37)**	-0.05 (0.13)	-0.06 (0.05)	-0.09 (0.05)
Lesson series condition (dummy) ^e	-1.17 (0.73)	-0.27 (0.36)	-0.04 (0.13)	-0.03 (0.08)**
PD program condition (dummy) ^f	-0.63 (0.74)	-0.27 (0.22)	0.06 (0.12)	-0.10 (0.07)
Number of lessons given	0.14 (0.09)	0.05 (0.04)	0.00 (0.02)	0.03 (0.02)**
Number of PD sessions followed	0.02 (0.18)	-0.01 (0.05)	-0.01 (0.03)	0.03 (0.01)*
Time × lesson series	1.48 (0.40)**	0.42 (0.20)*		0.44 (0.09)**
Time × PD program	-0.43 (0.42)	0.27 (0.24)		-0.03 (0.13)
Random part				
σ_c^2	10.92	2.78	0.77	0.59
σ_{u0}^2	2.58	0.29	0.00	0.03
σ_{v0}^2	0.79	0.03	0.02	0.00
σ_{v1}	1.32	0.30	0.01	0.08

*Significant at $p < 0.05$, **significant at $p < 0.01$

^a Robust standard errors are used

^b Grade 4 is used as a reference group

^c (Male = 0, female = 1)

^d (Medium or high SES = 0, very (low) SES = 1)

^e (Received no MACE lessons = 0, received MACE lessons = 1)

^f (Received no MACE PD sessions = 0, received MACE PD sessions = 1)

Regarding visual arts perception, a linear relationship was found between time and the aspects “space” as used in the VAA ($t(4206) = 7.80$, $p < 0.001$), “space suggestion” ($t(4206) = 2.37$, $p < 0.05$), and “composition” ($t(4206) = 6.19$, $p < 0.001$). Students, on average, described more of these aspects in the VAA posttest compared with the VAA pretest. The rate of improvement on the aspects “space” ($\chi^2(94) = 260.57$, $p < 0.001$), “space suggestion” ($\chi^2(94) = 232.81$, $p < 0.001$), and “composition” ($\chi^2(94) = 300.95$, $p < 0.001$) differed between classes; students in some classes improved more in their use of these aspects in describing visual artworks than

students in other classes. These differences between classes could be explained by the MACE lesson series (space, $t(94) = 3.66$, $p < 0.01$; space suggestion, $t(94) = 2.14$, $p < 0.05$; composition, $t(94) = 4.88$, $p < 0.001$). However, the MACE PD program did not show such an effect (space, $t(94) = -1.02$, $p = 0.31$; space suggestion, $t(94) = 1.13$, $p = 0.26$; composition, $t(94) = -0.30$, $p = 0.76$).

Regarding the aspect of shape used in the VAA, no linear relationship with time was found ($t(4206) = 0.39$, $p = 0.70$), indicating that students, on average, did not describe this aspect more frequently in the VAA posttest compared with the VAA pretest. This relation was the same for all classes ($\chi^2(94) = 94.96$, $p = 0.45$), indicating no effect of condition. The final multilevel models regarding the aspects used in the VAA are presented in Table 9. The results of the other models can be found in Appendix 3.

Discussion

The MACE program aimed to teach the (overlapping) curriculum goals of visual arts and the mathematical domain of geometry, and to promote students' creative skills in both disciplines, by creating opportunities for students to act creatively in an integrated visual arts and geometry context. The program was evaluated in a quasi-experimental study in which students were assigned to three conditions: (1) students who received the MACE lesson series from their teachers who received a PD program, (2) students who received the MACE lesson series, and (3) students who received regular geometry lessons. Students' growth between pre- and post-measurements of geometrical ability and visual arts perception were examined to test the effect of the conditions.

Students' ability to understand and explain geometrical phenomena improved in all conditions. However, contrary to our expectations, no differences between conditions were found. A possible explanation is that the lessons of the comparison condition were actually also an intervention that improved regular geometry education. Teachers indicated that the lessons they taught were similar to regular geometry lessons, but that they enabled more interaction between students than they were used to. As a consequence, students in the comparison condition had more than in regular lesson opportunities to explain their thinking, to receive feedback, and to encounter other points of view (Jarvis, 2001), which could have enhanced their geometrical thinking (Beghetto & Kaufman, 2010; Taggar, 2002). Furthermore, more interaction between students could also evoke reflection, which might have enabled students to reach the same level of understanding as students who participated in the MACE program (Van den Heuvel-Panhuizen & Drijvers, 2014). Furthermore, in the comparison condition, the lessons were offered in a sequence, while usually the 4 to 6 geometry lessons are spread over the school year. Interestingly, our analyses showed that students in some classes significantly improved more on their ability to understand and explain geometrical phenomena than students in other classes. Differences between classes could not be explained by the type of lessons they received. This result seems to imply that for students' ability to understand and explain geometrical phenomena, the content and structure of the lesson material are of less importance than other factors. Plausible factors could be the implementation of the lesson and the quality of the teacher. For example, despite the teaching materials and its related teaching approach, some

teachers may stimulate more communication between students than other teachers, or may be better able to ask open questions that can extend students thinking, reasoning, and understanding than other teachers (Bostic, 2011). Future research should investigate these possible factors.

Students' geometrical creative thinking also improved in all conditions, but—contrary to our expectations—did not differ between the conditions. We expected that implicitly and explicitly stimulating students to act creatively and think divergently (Sawyer, 2014), especially in integrated visual arts and geometry lessons, would lead to more improvement in students' geometrical creativity compared with students who did not receive such stimulation. Although we found in a qualitative case study that students expressed more mathematically creative ideas and solutions in classroom dialogues during the MACE lessons than they did during a regular mathematics lesson (Schoevers et al., 2019), one MACE lesson per week may not be enough to bring large improvement of students' geometrical creativity.

Regarding students' use of geometrical vocabulary to explain geometric phenomena, we found a partial effect of the MACE program. Students who participated in the MACE program increased in their use of daily geometrical words proportional to the total number of words in their written explanations in the geometrical ability test more than the students who received regular geometrical lessons. However, the increase of students' daily geometrical words was not a goal of the MACE program. Instead, increase of the use of academic geometrical words was a goal, which however was not influenced by the MACE program. In fact, students in all conditions used less academic geometrical words, controlled for the number of words in the written explanations at the posttest. Furthermore, contrary to our expectations, no effect was found for the MACE PD program. Since the importance and stimulation of geometrical words was emphasized both in the manual of the lesson series and in the PD program, we expected the students' use of academic geometrical words to increase. One explanation for these results is that students used more words at the posttest to explain their answers compared with that at the pretest, but did not use more academic geometrical words. As a result, the proportion of academic geometrical words used decreased.

In contrast to the previous results, we did find that students' perception regarding geometrical aspects in visual arts changed. As expected, students who received the MACE lesson series showed more improvement in describing geometrical aspects addressed in the lesson series (i.e. space, space suggestion, and composition, but not regarding the aspect of shape) in a visual artwork compared with students in the comparison group. Since students had to observe and analyze visual artworks mainly with regard to the aspects of space and patterns in every MACE lesson, they apparently changed their recognition of visual information in visual artworks in this respect (Kozbelt, 2001; Tishman et al., 1999). The aspect of shapes played a smaller role in the lesson series, explaining why no effects were found with regard to this aspect. Participation of the teacher in the MACE PD program did not affect students' perception of the spatial aspects of visual art, probably because—as teachers indicated—the teaching manual of the MACE lesson series was elaborate and clear. For example, many sample questions related to the spatial aspects of visual art were stated in the manual. Since the teaching manual was well elaborated, the PD program probably had not much added value and did not affect students' perception of spatial aspects of visual art.

Limitations and Future Research

It is important to take the limitations of the present study into account. A first limitation is that we were not able to randomly assign the teachers and students to the different conditions. A second limitation is the used measure of geometrical creativity. Although this measure, a multiple solution task, is commonly used in the field (Leikin, 2009), it may not have been sensitive enough to do full justice to the multidimensional construct of geometrical creativity. The measure does provide information on the cognitive aspect of geometrical creativity, but other measures could be used in future research to measure other dimensions of creativity as well, such as how creativity was verbally or non-verbally expressed by students in the classroom. A third limitation is that no classroom observations were conducted in the comparison group. If these would have been conducted, classroom factors could have been investigated that could possibly explain the differences between classes in students' growth in understanding and explaining of geometric phenomena. Furthermore, another comparison group could have been used. Our current comparison group also received a lesson series with regular geometry lessons. The lessons differed in several aspects from the MACE lessons. However, we were not able to investigate differences with regard to real educational practice, in which geometry lessons are usually taught spread over the whole school year.

Conclusion

The MACE program had the aim to teach the (overlapping) curriculum goals of visual arts and the mathematical domain of geometry, and to promote students' creative skills in both disciplines by creating opportunities for students to act creatively in an integrated visual arts and geometry context. Where in regular education, geometrical concepts are directly and time-efficiently taught, students that received the MACE lesson series reached the same level of geometrical understanding with open lessons in which they were able to express themselves creatively. Although students who received the MACE lesson series did not improve more on geometrical creativity and academic geometrical vocabulary than students who followed regular education, students did improve more in their ability to perceive geometrical aspects in visual arts and their use of daily geometrical words to describe geometric phenomena.

For educational practice, the results of this study imply that teachers could use the integrated MACE lesson series instead of regular geometry lessons, in order to teach the geometry curriculum and visual arts perception. Teaching in this integrated way may save time for teachers who experience a lot of time pressure to teach the curricula. If the integrated MACE approach and pedagogy are also applied to other mathematics lessons during the week, the integrated MACE approach may also be more effective for students' creative skills. However, more research is necessary. This study is to our knowledge the first that evaluated the effectivity of integrated visual arts and geometry education with a clear theoretical framework and research design. The study and the theoretical framework can be a valuable contribution to research on interdisciplinary arts and mathematics education.

Compliance with Ethical Standards This study was approved by the Ethical Committee of the Faculty of Social and Behavioural Sciences of Utrecht University (FETC15-083).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Beghetto, R. A., & Kaufman, J. C. (2010). *Nurturing creativity in the classroom*. New York: Cambridge University Press.
- Bostic, J. D. (2011). *The effects of teaching mathematics through problem-solving contexts on sixth-grade students' problem-solving performance and representation use* (Doctoral dissertation). University of Florida, FL. Retrieved from <http://ufdc.ufl.edu/UFE0043164/00001>.
- Braakhuis, G., Von Piekartz, R., Vogel, H. & De Graaf, T. (2012). Kennisbasis Beeldend onderwijs op de Pabo [Knowledge base visual arts education at primary school teacher education]. Retrieved from <https://10voordeleraar.nl/kennisbases/publicaties>
- Bresler, L. (1999). The hybridization and homogenization of school art: Institutional contexts for elementary art specialists. *Visual Arts Research*, 25, 25–37. Retrieved from https://www.jstor.org/stable/20715983?seq=1#page_scan_tab_contents
- Buijs, K., Klep, J. & Noteboom, A. (2008). TULE - Rekenen/wiskunde [TULE – Arithmetic/Mathematics]. Enschede: SLO.
- Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, 105, 380–400. <https://doi.org/10.1037/a0031084>.
- Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, 18, 439–477. [https://doi.org/10.1016/0364-0213\(94\)90016-7](https://doi.org/10.1016/0364-0213(94)90016-7).
- Crane, J. (1996). Effects of home environment, SES and maternal test scores on mathematics achievement. *The Journal of Educational Research*, 89, 305–314. Retrieved from https://www.jstor.org/stable/27542048?seq=1#page_scan_tab_contents
- Davies, D., Jindal-Snape, D., Digby, R., Howe, A., Collier, C., & Hay, P. (2014). The roles and development needs of teachers to promote creativity: A systematic review of literature. *Teaching and Teacher Education*, 41, 34–41. <https://doi.org/10.1016/j.tate.2014.03.003>.
- Dobbins, K. (2009). Teacher creativity within the current education system: A case study of the perceptions of primary teachers. *International Journal of Primary, Elementary and Early Years Education*, 37, 95–104. <https://doi.org/10.1080/03004270802012632>.
- Elfland, A. (1976). The school art style: A functional analysis. *Studies in Art Education*, 17, 37–44. <https://doi.org/10.2307/1319979>.
- Frost, L. A., Hyde, J. S., & Fennema, E. (1994). Gender, mathematics performance, and attitudes and affect: A meta-analytic synthesis. *International Journal of Educational Research*, 21, 373–385. [https://doi.org/10.1016/s0883-0355\(06\)80026-1](https://doi.org/10.1016/s0883-0355(06)80026-1).
- Gravemeijer, K. P. E. (2007). Reken-wiskundeonderwijs anno 2007-tussen oude waarden en nieuwe uitdagingen-[Arithmetic-mathematics education in the year of 2007-between old values and new challenges-] *Panama-Post*, 26, 3–10. Retrieved from <https://www.rekenenwiskunde21.nl>.
- Haanstra, F. (2014). Nationale leerplannen en leerplankaders voor de kunstvakken [National curricula and curriculum frameworks for the arts subjects]. *Cultuur + Educatie*, 14, 8–25. Retrieved from <https://www.lkca.nl>.
- Hailey, D., Miller, A. & Yenawine, P. (2015). Understanding visual literacy: The visual thinking strategies approach. In D. M. Baylen & A. D'Alba (Eds.), *Essentials of teaching and integrating visual and media literacy* (pp. 49–73). London, UK: Springer.
- Haylock, D. W. (1987). Mathematical creativity in schoolchildren. *The Journal of Creative Behavior*, 21, 48–59. <https://doi.org/10.1002/j.2162-6057.1987.tb00452.x>.

- Henrichs, L. F., & Leseman, P. P. M. (2014). Early science instruction and academic language development can go hand in hand. The promising effects of a lowintensity teacher-focused intervention. *International Journal of Science Education*, 36, 2978–2995. <https://doi.org/10.1080/09500693.2014.948944>.
- Hop, M. (2012). *Balans van het reken-wiskundeonderwijs halverwege de basisschool 5* [balance of mathematics education halfway through primary school. *Periodic assessment 5*]. Arnhem: CITO.
- Hopkins, C. D., & Antes, R. L. (1978). *Classroom measurement and evaluation*. Itasca, IL: Peacock.
- Housen, A. C. (2002). Aesthetic thought, critical thinking and transfer. *Arts and Learning Research*, 18, 99–132. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.467.3752&rep=rep1&type=pdf>.
- Hox, J. J., Moerbeek, M., & Van de Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). New York: Routledge.
- Jarvis, D. H. (2001). *Learning between the lines: A syncretistic experiment in mathematics and visual arts education* (Unpublished master's thesis). Nipissing University, North Bay, Ontario, Canada.
- Jones, K. (2002). Issues in the teaching and learning of geometry. In: L. Haggarty (Ed.), *Aspects of teaching secondary mathematics: Perspectives on practice* (pp. 121–139). London: RoutledgeFalmer.
- Kattou, M., Kontoyianni, K., Pitta-Pantazi, D., & Christou, C. (2013). Connecting mathematical creativity to mathematical ability. *ZDM - International Journal on Mathematics Education*, 45, 167–181. <https://doi.org/10.1007/s11858-012-0467-1>.
- Keijzer, R., Oprins, B., De Moor, K., & Schoevers, E. M. (2018). Integrating visual art, geometry and creativity for primary school teachers: A pd trajectory. In M. Friman (Ed.), *EAPRIL 2017 proceedings* (pp. 52–65). Hämeenlinna, Finland.
- Kolovou, A. (2011). *Mathematical problem solving in primary school* (Unpublished doctoral dissertation). Utrecht University, Utrecht, the Netherlands.
- Kozbelt, A. (2001). Artists as experts in visual cognition. *Visual Cognition*, 8, 705–723.
- KPC-Groep (2000). *Kunstbeschouwen. Culturele en kunstzinnige vorming 2* [Visual art reception. Cultural and artistic education 2]. Den Bosch, the Netherlands.
- Kruiter, J., Hoogeveen, K., Beekhoven, S., Kieft, M. & Bomhof, M. (2016). *Rapport monitor. Cultuuronderwijs in het primair onderwijs en programma Cultuureducatie met kwaliteit* [Report monitor. Culture education in primary school and program culture education with quality]. Utrecht, the Netherlands: Sardes/Oberon.
- 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., Koichu, B., & Berman, A. (2009). Mathematical giftedness as a quality of problem-solving acts. In R. Leikin, A. Berman, & B. Koichu (Eds.), *Creativity in mathematics and the education of gifted students* (pp. 115–127). Dordrecht, the Netherlands: Sense Publishers.
- Levav-Waynberg, A., & Leikin, R. (2012). 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>.
- Moerbeek, M. (2015). SPA-ML. A software package for power analysis of trials with multilevel data. In *Department of Methodology and Statistics*. Utrecht, the Netherlands: Utrecht University. Retrieved from tinyurl.com/spaml.
- Núñez, R., Edwards, L., & Matos, J. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics*, 39, 45–65. <https://doi.org/10.1023/A:1003759711966>.
- Plucker, J. A., & Zabelina, D. (2009). Creativity and interdisciplinarity: One creativity or many creativities? *ZDM - International Journal on Mathematics Education*, 41, 5–11. <https://doi.org/10.1007/s11858-008-0155-3>.
- Sawyer, R. K. (2014). How to transform schools to foster creativity. *Teacher College Record*, 118(4). Retrieved from <http://keithsawyer.com/PDFs/TCR.pdf>.
- Schoevers, E. M., Kroesbergen, E. H., & Kattou, M. (2018). Mathematical Creativity: A Combination of Domain-general Creative and Domain-specific Mathematical Skills. *Journal of Creative Behavior*. <https://doi.org/10.1002/jocb.361>.
- Schoevers, E. M., Leseman, P. P. M., Slot, E. M., Bakker, A., Keijzer, R., & Kroesbergen, E. H. (2019). Promoting pupils' creative thinking in primary school mathematics: A case study. *Thinking Skills and Creativity*, 31, 323–334. <https://doi.org/10.1016/j.tsc.2019.02.003>.
- Silver, E. A. (1995). The nature and use of open problems in mathematics education. *ZDM - International Journal on Mathematics Education*, 27, 67–72 Retrieved from <https://eric.ed.gov/?id=EJ520599>.

- Stichting Leerplanontwikkeling. (2015). *Nieuw elan voor kunstzinnige oriëntatie in het primair onderwijs. Een praktische handreiking voor leraren* [A new impetus for arts education in primary school]. Enschede, the Netherlands: SLO.
- Stichting Leerplanontwikkeling (2018). Leerplankader kunstzinnige oriëntatie [Leather plank artistic orientation]. Retrieved from <http://kunstzinnigeorientatie.slo.nl/>.
- Taggar, S. (2002). Individual creativity and group ability to utilize individual creative resources: A multilevel model. *The Academy of Management Journal*, *45*, 315–330. <https://doi.org/10.2307/3069349>.
- Tishman, S., MacGillivray, D., & Palmer, P. (1999). *Investigating the educational impact and potential of the Museum of Modern Art's visual thinking curriculum*. Cambridge, MA: Harvard University.
- Van den Heuvel-Panhuizen, M., & Buys, K. (2005). *Young children learn measurement and geometry: A learning-teaching trajectory with intermediate attainment targets for the lower grades in primary school*. Rotterdam, the Netherlands: Sense Publishers.
- Van den Heuvel-Panhuizen, M., & Drijvers, P. (2014). Realistic mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 521–525). Dordrecht, the Netherlands: Springer. <https://doi.org/10.1007/978-94-007-4978-8>.
- Van Lier, L. (1996). *Interaction in the language curriculum: Awareness, autonomy and authenticity*. London: Longman.
- Van Onna, J., & Jacobse, A. (2008). *Laat maar zien. Een didactische handleiding voor beeldend onderwijs* [Show me. A didactical manual for visual arts education]. Houten, the Netherlands: Wolters-Noordhoff.
- Van Zanten, M., Barth, F., Faarts, J., Van Gool, A. & Keijzer, R. (2009). Kennisbasis rekenen-wiskunde voor de pabo [Mathematical knowledge base for teacher education]. Retrieved from https://10voordeleraar.nl/documents/kennisbases_pabo/kb_rekenen_wiskunde_pabo.pdf
- 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>.
- Walker, C. M., Winner, E., Hetland, L., Simmons, S., & Goldsmith, L. (2011). Visual thinking: Art students have an advantage in geometric reasoning. *Creative Education*, *2*, 22–26. <https://doi.org/10.4236/ce.2011.21004>.
- 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 of the 27th Conference of the International Group for the Psychology of Mathematics Education held jointly with the 25th conference of PME-NA*, *4*, 371–378. Retrieved from <https://eric.ed.gov/?id=ED500860>Acknowledgement.

Affiliations

Eveline M. Schoevers¹ · Paul P. M. Leseman¹ · Evelyn H. Kroesbergen^{1,2}

¹ Department of Education & Pedagogy, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, the Netherlands

² Behavioural Science Institute, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, the Netherlands