ZfE

Check for updates

Getting fit for the Mathematical Olympiad: positive effects on achievement and motivation?

Franziska Rebholz (iD) · **Jessika Golle** · **Maike Tibus** ·
Evelin Ruth-Herbein · **Korbinian Moeller** · **Ulrich Trautwein**

Abstract All around the world, there are numerous academic competitions (e.g., "Academic Olympiads") and corresponding training courses to foster students' competences and motivation. But do students' competences and motivation really benefit from such courses? We developed and evaluated a course that was designed to prepare third and fourth graders to participate in the German Mathematical Olympiad. Its effectiveness was evaluated in a quasi-experimental pre- and posttest design ($N=201$ students). Significant positive effects of the training were found for performance in the academic competition (for both third and fourth graders) as well as mathematical competences as measured with a curriculum-oriented test (for fourth

Dr. Franziska Rebholz (✉)
Reutlinger Didaktik Institut, Reutlingen University, Alteburgstraße 150, 72762 Reutlingen, Germany
E-Mail: franziska.rebholz@reutlingen-university.de

Prof. Dr. Jessika Golle · Dr. Maike Tibus
Hector-Institut für Empirische Bildungsforschung, University of Tübingen,
Walter-Simon-Straße 12, 72072 Tübingen, Germany

Prof. Dr. Jessika Golle
E-Mail: jessika.golle@uni-tuebingen.de

Dr. Maike Tibus
E-Mail: maike.tibus@uni-tuebingen.de

Dr. Maike Tibus · Prof. Dr. Ulrich Trautwein
Hector-Institut für Empirische Bildungsforschung, University of Tübingen,
Europastraße 6, 72072 Tübingen, Germany
E-Mail: ulrich.trautwein@uni-tuebingen.de

Dr. Evelin Ruth-Herbein
Institute for Educational Analysis (IBBW), Heilbronner Straße 172, 70191 Stuttgart, Germany
E-Mail: evelin.ruth-herbein@ibbw.kv.bwl.de

Prof. Dr. Korbinian Moeller
Mathematics Education Center, Loughborough University, Loughborough, UK
E-Mail: K.Moeller@lboro.ac.uk

🖄 Springer

graders only). Differential effects across grade levels (with more pronounced positive effects in fourth-grade students) were observed for students' math self-concept and task-specific interest in mathematics, pointing to possible social comparison effects.

Fit für die Mathematik-Olympiade: Positive Effekte für Leistung und Motivation?

Zusammenfassung Weltweit gibt es eine Vielzahl akademischer Wettbewerbe (z. B. „Akademische Olympiaden") sowie entsprechende Begleitangebote, in denen Leistung und Motivation der Teilnehmenden gefördert werden sollen. Aber profitieren Teilnehmende wirklich von solchen Kursen? Wir haben ein solches Begleitangebot, das Lernende der dritten und vierten Klasse auf die Teilnahme an der Deutschen Mathematik-Olympiade vorbereiten soll, entwickelt und dessen Wirksamkeit in einem quasi-experimentellen Prä- and Posttest-Design evaluiert ($N = 201$ SchülerInnen). Die Ergebnisse zeigten positive Effekte auf die Leistungen in der Mathe-Olympiade für Schülerinnen und Schüler beider Klassenstufen sowie auf die mathematischen Kompetenzen der ViertklässlerInnen gemessen mit einem Curriculums-orientierten Test. Darüber hinaus wurden für das mathematische Selbstkonzept und das aufgabenspezifische Interesse der SchülerInnen in Mathematik stärkere positive Effekte bei den Viertklässlern beobachtet, was darauf hindeutet, dass soziale Vergleichsprozesse stattgefunden haben.

1 Introduction

To enrich learning environments of students who are already able to solve curriculum-based tasks, students can participate in various regional, national, and international academic competitions in different domains such as mathematics, sciences, history, or languages (Campbell et al. 2000; Forrester 2010; Rebholz and Golle 2017). To prepare students to successfully participate in such academic competitions, some authors have suggested that specific training courses be offered beforehand (Cropper 1998; Ozturk and Debelak 2008b; for examples of such courses, see Skiena and Revilla 2003; Xu 2010). Although, from a theoretical point of view, positive effects of such courses seem plausible, there is still a paucity of empirical studies that have systematically investigated the influence of such courses on students' competence and motivation.

Therefore, we developed a training course designed to prepare elementary school students to participate in the German Mathematical Olympiad—a national math competition—and investigated the effects of the course on students' success in the

competition, on students' mathematical competence as measured with a curriculum-oriented test, and on their motivation to do mathematics.

2 Theoretical background

2.1 Academic competitions as enrichment

To give students—especially those who already find it easy to solve curriculum-based tasks—the opportunity to explore and develop their learning potential, they need appropriate learning environments that challenge them (e.g., Subotnik et al. 2011). Several in- and out-of-school approaches that are aimed at accelerating or enriching the learning environments of such advanced students have been shown to be effective (see e.g., Kulik and Kulik 1987; Lubinski and Benbow 2006; Steenbergen-Hu and Moon 2010). One opportunity that is supposed to enrich students' learning environments outside of the school curriculum is participation in domain-specific academic competitions (e.g., Abernathy and Vineyard 2001; Rebholz and Golle 2017). Commonly, academic competitions are seen as one opportunity for students—especially high-achieving students—to demonstrate their skills and to experience challenging tasks that go beyond those offered by the standard curriculum (Bicknell 2008; Ozturk and Debelak 2008a, b; Riley and Karnes 1998; Wirt 2011). As such, the increase in learning opportunities is supposed to positively influence students' competence as well as motivation in the same domain (see Diezmann and Watters 2000, 2001; Wai et al. 2010).

Attending an academic competition is believed to have the potential to positively influence competence as well as motivation in the respective domain. In fact, there is some—albeit rather fragmentary—empirical support for the hypothesis that academic competitions support students' acquisition of competences and motivation. For instance, using retrospective interviews of successful participants of academic competitions, in a German sample, Lengfelder and Heller (2002) reported that over 70% of successful participants of academic competitions were in the group of top 10 students in their last year of the German Gymnasium (highest academic track in Germany). Furthermore, Campbell and Verna (2010) reported that a very high percentage (52%) of such successful participants of academic Olympiads in the US achieved a PhD degree later in their academic careers. In addition, Gordeeva et al. (2013) found that students who participated in a chemistry competition showed higher motivation to learn and achieve, as well as, for example, more efficient self-regulation strategies in comparison with students who did not participate in the competition but chose the same subject of study. Additionally, winning or losing a competition seems important in determining students' future motivation in the competition's domain (i.e., self-concept, self-efficacy, and value; see Forrester 2010).

Despite these studies, however, systematic effectiveness research with adequate control groups and appropriate standardized measures has remained scarce, but clearly needed given that participation in academic competitions is not a random process: Students attending academic competitions are likely to show high perfor-

mance levels and are highly motivated to engage in tasks related to the competition's domain already before they enter the competition (e.g., Forrester 2010). For instance, Oswald et al. (2005) found that former participants of academic competitions indicated that exploring their individual talents in a specific domain was an important reason for participating in the competition. Moreover, research is needed that does not exclusively study effects on the topic of the academic competition, but additional achievement outcomes as well. Finally, given that the participation in academic competitions is oftentimes embedded in enrichment programs, it is important to also embrace a larger perspective and focus on such preparation settings for academic competitions. We turn to this issue next.

2.2 Preparing students for academic competitions

A common practice to prepare students for the challenges in a specific academic competition is to offer a training course beforehand (Cropper 1998; Ozturk and Debelak 2008b). This preparation has two main goals: First, the effects on student learning might be amplified if some more time is spent on working through tasks that are similar to the tasks used in the academic competition. Second, a training course might increase positive effects on student motivation or inoculate against some possible negative effects of (unsuccessful) participation in academic competitions: For decades, academic competitions have been criticized for their rather competitive nature, and there is concern that students lose their motivation to engage in challenging tasks (Bicknell 2008; Fauser et al. 2007; Ozturk and Debelak 2008a, b). Although the tasks that are offered in such competitions may seem promising to appropriately challenge high-achieving students, they may also turn out to be problematic because students may not be used to dealing with the challenges. Whereas these students may perceive the tasks typically presented in school as easy, they may come in contact with more difficult tasks—even some they are not able to solve—as part of the competition. It has been argued that students should be prepared to meet the specific requirements of challenging tasks (cf. Kießwetter 2013; Nolte and Pamperien 2013). Thus, students should also be prepared for requirements of academic competitions and the competitive nature of the setting.

Often the training course is part of a larger enrichment program and, besides preparing students for the tasks of the competition and the competitive setting, it aims to enhance students' competences in the respective academic domain (for the Academic Olympiads, see e.g., Urhahne et al. 2012). Common parts of such trainings are an introduction to the academic competition and working phases in which tasks of former competitions are solved. Therewith, the training increases the learning time in a certain domain and provides the opportunity to deal more intensively with specific content. Ideally, the pedagogical preparation and accompaniment of the competition reduce the probability of possible negative effects that are due to the competitive setting. There is a variety of such training courses, but they are typically not systematically investigated in terms of their effectiveness.

Although training courses should adequately prepare students for the tasks and the competitive nature of an academic competition (e.g., Cropper 1998), it is important to realize that the composition of the preparation courses might itself have an impact on

student motivation. Typically, students are selected to participate in such a training course on the basis of teacher nominations or by successfully passing previous selection rounds (see the qualification process for the International Olympiads; for an overview, see, e.g., http://olympiads.win.tue.nl/). Thus, the average level of ability should be higher than what the participating children are used to in class (Bicknell 2008; Karnes and Riley 2005; Ozturk and Debelak 2008a; Riley and Karnes 1998). The student composition of the training course, however, might have a substantial effect on social comparison processes which, in turn, are likely to impact student motivation (for overviews, see Marsh et al. 2008; Trautwein and Möller 2016).

On the one hand, being part of the selected high-achieving group that participates in the academic competition may result in positive feelings such as pride (i.e., an assimilation effect or a *basking-in-reflected-glory effect*; e.g., Marsh et al. 2000), which in turn may result in positive effects on participants' motivation (e.g., Rinn 2007) and especially their self-concept (see Dai and Rinn 2008). However, changes in students' frame of reference when experiencing such a high-achieving group may also lead to negative effects (cf. Zeidner and Schleyer 1999). In the literature, this effect is known as the Big-Fish-Little-Pond effect (BFLPE; Marsh 1987; Marsh and Parker 1984): Of two students with the same individual ability, academic self-concept will likely be higher for the student who is placed in a low-achieving class compared with the student who is placed in a high-achieving class. In other words, being surrounded by other high-achieving peers may lead to negative social comparisons that are typically stronger than the potentially positive effects of being part of a selected group of high achievers (see Marsh et al. 2008). Negative frame-of-reference effects have been documented for academic self-concept, but also for students' interest, joy, and attainment (e.g., Fleischmann et al. 2021; Marsh et al. 2008; Trautwein et al. 2006). However, for the group of high-achieving students, Trautwein et al. (2009) also showed that such students were "less affected by the negative frame of reference effect than low-achieving students" (S. 862; see also Preckel et al. 2019).

Therewith, considering the BFLPE, the question arises how participating in a course to prepare for an academic competition influences students' motivation: Is there a risk that student motivation will decrease when they experience a learning environment characterized by higher academic achievement? As such, one of the explicit aims of any such training courses should be to prevent negative effects by stabilizing students' motivation (i.e., self-concept and value beliefs), for example, by providing individual feedback that focuses on strengths and progress (see, e.g., Nolte and Pamperien 2013).

2.3 A training course to prepare for the Mathematical Olympiad (MO)

We developed a training course called "Getting Fit for the Mathematical Olympiad" specifically addressing the abilities and challenges of high-achieving elementary school children who plan to participate in the Mathematical Olympiad (MO). In the MO, tasks are characterized by challenging open problems for which it is important to justify the solution with an appropriate notation of the solution strategy to give others (students or teachers) the chance to understand the individual solution

(strategy). To be able to solve those tasks, students need sophisticated mathematical competences, including both content-based mathematical competences—the aspect of mathematical competences that is supposed to embrace specific mathematical contents (see e.g., Blum 2012; Freudenthal 1986; Köller 2010; Kultusministerkonferenz [KMK] 2004)—and process-based mathematical competences—the aspect of mathematical competences that involves the more general aspect of mathematics such as strategies and methods (see e.g., Köller 2010; KMK 2004; National Council of Teachers of Mathematics 2000; Winkelmann et al. 2012). Therewith, the tasks of the MO differ from tasks that students of this age group are used to from normal math lessons, and can be used as motivating challenges for high-achieving students.

The course was developed as part of statewide enrichment program that provides enrichment courses in small groups on a broad range of topics with an emphasis on STEM topics (for further information, see Golle et al. 2018). Our training course aimed at (i) providing an enriched learning environment for mathematically high-achieving students and preparing them for the requirements of the MO (i.e., complex mathematical tasks), (ii) fostering students' process-based mathematical competences (especially mathematical problem solving as well as communication and arguing with mathematical contents), and (iii) sustaining students' motivation for mathematics (avoiding possible negative reference-group effects; for more about the training course, see Rebholz 2018).

The training was designed for small groups of six to 10 third- and fourth-grade students and included ten modules, each planned for a 90-min session. Contents and the didactical approaches were based on theory as well as empirical findings for the instruction of mathematically interested and high-achieving students (for more information, see e.g., Diezmann and Watters 2000; Koshy et al. 2009; Leikin 2010; McAllister and Plourde 2008; Rotigel and Fello 2016). Based on this literature, the math course incorporated four core components: (i) challenging open tasks to respond to the learning potential of the target group (e.g., Käpnick 2010); (ii) structured writing to target typical weaknesses of the target group (e.g., Seo 2015); (iii) cooperative learning to necessitate communication and argumentation among students while solving mathematical problems (e.g., Johnson and Johnson 1990; Johnson et al. 2000) and to keep students motivated (e.g., Rosenzweig and Wigfield 2016); and (iv) mathematical games to counterbalance the competitive setting of the academic competition and, therewith, also sustain students' motivation (e.g., Randel et al. 1992). All of these components and especially their combination seem promising in terms of reaching the course aims. Therefore, every course session was built upon these core components.
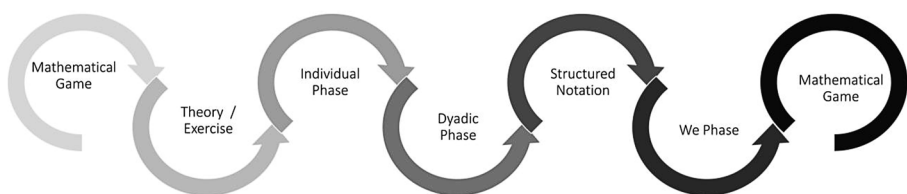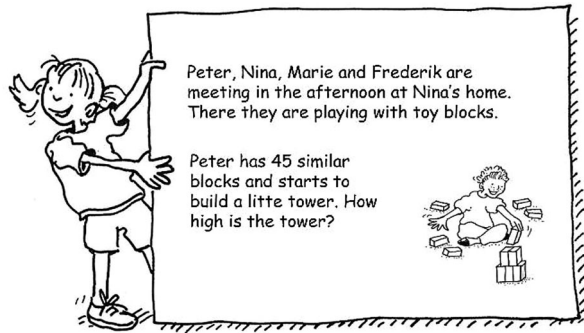

Fig. 1 Schematic illustration of the workflow of each session

In more detail, each session followed the same workflow (illustrated in Fig. 1): A mathematical game characterized the beginning of each session as a playful introduction to increase students' motivation (illustrated in Fig. 1 on the very left). Afterwards, theoretical input and exercise(s) were presented to prepare students for the contents of the module and to indicate possible solution strategies. The main parts of each session comprised three steps—based on the ideas of *think-pair-share* (see e.g., Lyman 1981), respectively, and *I–You–We* (see e.g., Gallin and Ruf 1995)—broadened by Bezold (2012) to a whole lesson: The first step was an individual problem-solving phase in which students worked on possible solution strategies by themselves (*Think*, respectively, *I*). This was followed by a dyadic problem-solving phase in which students worked with a randomly assigned partner (*Pair*, respectively, *You*). Finally, the students shared their solution and solution strategies with the whole group (*Share*, respectively, *We*). With regard to participation in the MO, a transitional phase in which the students were supposed to prepare the exchange and structured notation of their solution and solution strategy was added between the dyadic problem-solving phase and the exchange of solution.

In the individual problem-solving phase, students were supposed to apply their previously acquired knowledge from the introduction to increase their problem-solving competences (i.e., process-based competences; see, e.g., Bezold 2012). In the dyadic phase, students had to communicate their individual ideas about solving the problem and discuss different solution steps. This communication was supposed to enhance students' competences in communicating and arguing with and about mathematics (i.e., process-based competences; see e.g., Bezold 2012). Next, the dyads were required to prepare a structured transcript to clearly verbalize the arguments underlying the steps of their solutions, which again supported students' process-based competences, especially mathematical problem solving and communicating and arguing (see, e.g., Bezold 2012). Each module ended with a presentation of the mathematical problem and its proposed solution to the other students, followed by a final discussion of the arguments. Again, this component was aimed at fostering problem solving (e.g., exploring relationships, looking for patterns and structures), as well as communicating and arguing about mathematical content in order to conduct preliminary proofs (Bezold 2012).

Overall, the core of the math course was characterized by problem solving of challenging tasks that reflected the requirements of the MO and communicating and arguing about possible solutions (for more about the requirements of the MO, see Rebholz and Golle 2017). The mathematical problems the students worked on were not difficult per se (e.g., calculations did not involve numbers larger than 100); rather, the problems were challenging mostly because they consisted of open problems that allowed for multiple solution strategies and required problem solving, modeling, and reasoning. More specifically, the contents of the math course were oriented towards the previous requirements of the (German) MO for elementary school students and involved, for example, logic problems, cryptograms, cubes, combinatorics, equation-based tasks, or sequences (implemented as towers). All tasks included in the math course were custom made and could be solved in several possible ways (an example task used in the math course is shown in Fig. 2). Original tasks from previous MOs were not implemented in the math course.

Fig. 2 Typical task from the corresponding math course

Peter, Nina, Marie and Frederik are meeting in the afternoon at Nina's home. There they are playing with toy blocks.

Peter has 45 similar blocks and starts to build a litte tower. How high is the tower?

2.4 The present study

Although a variety of academic competitions as well as preparation programs exist, only a few studies have investigated the effects of any preparation. Whereas some studies have examined expectations, experiences, or educational pathways of participants who were successful in an academic competition (e.g., Campbell and Walberg 2010; Gordeeva et al. 2013; Kuech and Sanford 2014; Lengfelder and Heller 2002; Smith et al. 2019; Wirt 2011), the effects of training for academic competitions on students' performance in the respective competition and on their domain-specific competences and motivation have hardly been investigated so far.

In this study, we investigated the effectiveness of a 10-week math training course preparing elementary school students for successfully attending the Mathematical Olympiad (MO). The training was designed to support students' successful participation in the MO, to encourage their process-based competencies in mathematics, and to prevent children from negative effects of unfavorable social comparisons. We investigated three main research questions:

1. Does the training positively affect the performance in the MO?
2. Does the training for high performance in the MO also have an effect on broader mathematical competences? More precisely, does participating in the training course influence mathematical competences as measured with a curriculum-oriented test?
3. Does the training at least sustain students' math-related motivation?

To evaluate the effectiveness of the training course, we used a quasi-experimental pre- and posttest design in a natural setting (Shadish et al. 2002). In addition to a group of children who were nominated by their teachers to attend our extracurricular math course, we implemented a control group consisting of students who participated in the MO but were not nominated for the training course and did not formally train for it.

We tested the following four hypotheses: First, we expected that children participating in the training course would perform better in the respective academic competition because they were trained to solve challenging problems that contained material typically used in the MO (Hypothesis 1). Second, we hypothesized that

children participating in the training course would also show a more positive development in mathematical competences measured with a curriculum-oriented test, which are not directly linked to the MO tasks, than the children in the control group (Hypothesis 2).

In terms of the motivational outcomes (self-concept and value beliefs), the formulation of hypotheses was somewhat more complex. Although the training was designed to sustain motivation, social comparison processes need to be considered that stem from the study's specific setting in which high-achieving third and fourth graders were taught together and which may negatively affect student motivation. For this reason, we did not formulate a specific hypothesis for the average training effect on students' motivation. However, given that the math course was offered to third and fourth graders who were taught together in the same course group, we expected grade-specific effects of the training course on motivational constructs (i.e., math self-concept and value beliefs) due to social comparisons. Fourth-grade students have already experienced one more year of formal schooling; therefore, we expected that fourth-grade students would have already acquired more sophisticated mathematical competences than the third-grade students. Thus, based on findings on reference group effects as described in the BFLPE literature (e.g., Marsh 1987), different social comparison processes might be plausible for the two age groups. Whereas third graders would be placed in a learning environment where they can compare themselves with the more competent fourth-grade students (i.e., opportunities for upward comparisons), the fourth graders would be more likely to have opportunities to make downward comparisons (i.e., with third graders) in the course—but perhaps also upward comparisons (i.e., with high-achieving fourth graders). Thus, we hypothesized that effects on students' self-concept differ between third-grade and fourth-grade students (Hypothesis 3). We expected the same pattern of grade-specific results for students' math value beliefs (Hypothesis 4), in line with previous research (see e.g., Trautwein et al. 2006).

3 Method

The newly developed math training course was part of a larger enrichment program (Hector Children's Academy Program [HCAP]) in the German state of Baden-Württemberg tailored to the 10% most gifted, talented, interested, and motivated elementary school children (for more information about the HCAP, see Rothenbusch et al. 2016).

Six local HCAP sites participated in this study. The math course was taught by different instructors (mostly mathematics teachers) at each of these sites. To ensure treatment fidelity, instructors attended a 2-hr qualification session provided by the developer of the math course. As part of this qualification, all core components of the math course were introduced and explained. In addition, a scripted manual including time tables for each module and master copies of all teaching materials were given to prospective instructors.

3.1 Sample

Data were collected from 201 elementary school children (58% boys; age: $M = 9.01$, $SD = 0.43$) in Grade 3 ($N = 110$, 63 boys) or Grade 4 ($N = 91$, 54 boys). The training group was comprised of 50 children in Grades 3 and 4 (Grade 3: $N = 26$ [15 boys]; Grade 4: $N = 24$ [18 boys]) who attended the math training course "Getting Fit for the Mathematical Olympiad" in groups between 6 and 11 students. Children in the training group were from different classes and schools and had been nominated to participate in the larger extracurricular enrichment program (i.e., the HCAP) by their teachers. They chose to participate in the math course. Children in the control group were from 14 classes from six different elementary schools (six fourth-grade classes and eight third-grade classes) that hosted the HCAP. Children participating in the course belonged to different classes than those of the control group. Children in the control group attended only standard curricular mathematics classes and did not participate in the math training course.

3.2 Design and procedure

The math training course was offered at six local sites of the HCAP and took place over a 5-month period (14 sessions from November 2014 to March 2015, including ten content modules, two sessions to participate in the Mathematical Olympiad, and two sessions each for pre- and posttest). The pre- and post-testing (pre- and posttest are abbreviated as T1 and T2, respectively) of the training group were integrated into the first two and the last two sessions of the training course, respectively, for each local site. Data were collected from the children in the control group during regular classes at comparable time points (see Fig. 3). Trained research assistants and scientists administered the questionnaires and tests. The study was approved by the local ethics committee and local school authorities. Furthermore, parents provided written informed consent for their children's participation prior to the study. All participants took part in the *54th Mathematical Olympiad*—the training group as part of the math course and the control group during regular math classes.

3.3 Measures

3.3.1 Achievement outcomes

All scales and corresponding descriptive statistics (including reliabilities) are displayed in the Appendix (Tables 2 and 3, Appendix). An overview of the measures that were used is provided in Fig. 3. We assessed *Performance in the Academic Competition* using the third (most difficult) level of the German *Mathematical Olympiad* for elementary school students (for more information about the competition, see Olson 2005; Campbell and Walberg 2010; www.imo-official.org, www.moems.org). The tasks in this academic competition were grade-level specific and included (complex) word problems (an example task is shown in Fig. 4). Tasks and scoring guidelines were provided by the German Mathematical Olympiad Association. Newly developed tasks are used for the MO every year, and these tasks are released by the
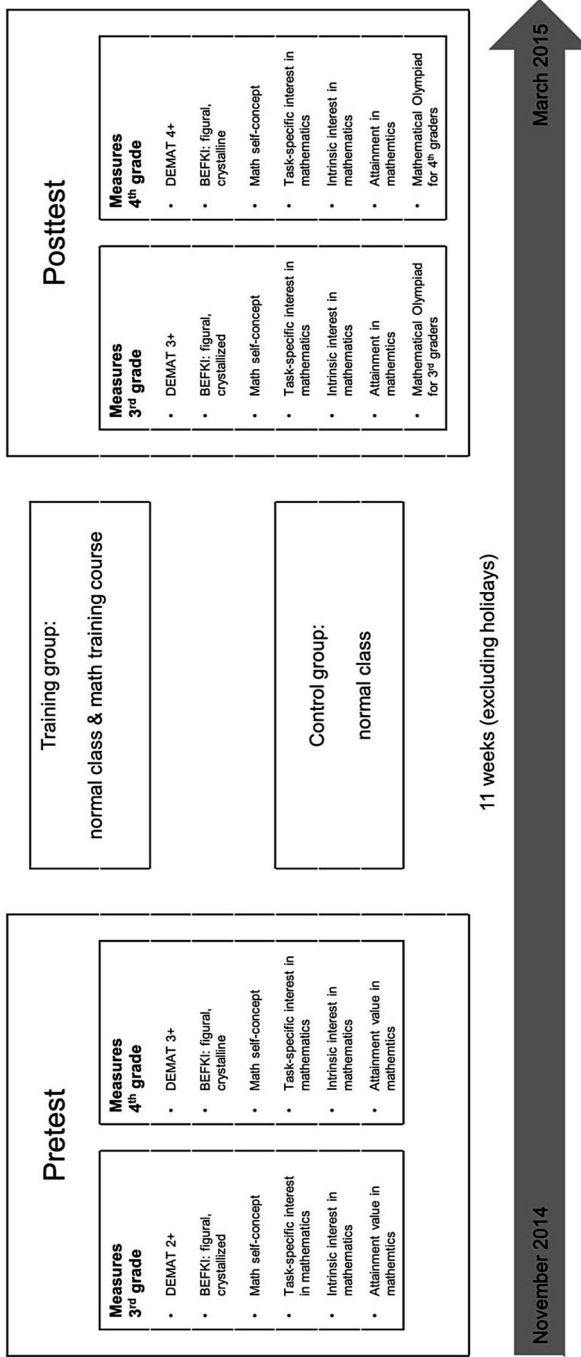
Pretest

Measures 3rd grade

- DEMAT 2+
- BEFKI: figural, crystallized
- Math self-concept
- Task-specific interest in mathematics
- Intrinsic interest in mathematics
- Attainment value in mathemtics

Measures 4th grade

- DEMAT 3+
- BEFKI: figural, crystalline
- Math self-concept
- Task-specific interest in mathematics
- Intrinsic interest in mathematics
- Attainment value in mathemtics

Training group:

normal class & math training course

Control group:

normal class

11 weeks (excluding holidays)

Posttest

Measures 3rd grade

- DEMAT 3+
- BEFKI: figural, crystallized
- Math self-concept
- Task-specific interest in mathematics
- Intrinsic interest in mathematics
- Attainment in mathemtics
- Mathematical Olympiad for 3rd graders

Measures 4th grade

- DEMAT 4+
- BEFKI: figural, crystalline
- Math self-concept
- Task-specific interest in mathematics
- Intrinsic interest in mathematics
- Attainment in mathemtics
- Mathematical Olympiad for 4th graders

November 2014

March 2015

Fig. 3 Quasi-experimental pre- and posttest design of the study and the implemented measures

Fig. 4 Translated version of a typical task from the Mathematical Olympiad (task no. 550331, 54th Mathematical Olympiad 2015/16, third grade)

Four children have the same birthday. Anton, Benita, Carla, and Dean are turning 10, 12, 14 and 16 years old, respectively, in 2016.

    a) What will be the sum of the kids' ages in 2020?
    b) In which year will the kids' ages sum to 100?
    c) Could the kids' ages ever sum to 163? Justify!

Mathematical Olympiad Society after the competition. Therewith, tasks of the competition were not known to us when developing the math course. Tasks were scored according to the scoring guidelines that were provided. The dependent variable was the *z*-standardized sum score of all correctly solved tasks for each grade level.

We assessed mathematical competences with a curriculum-oriented test at pre- and posttest by using the *German Mathematics Tests for Second, Third,* and *Fourth Grades* (i.e., pretest: DEMAT-2, Krajewski et al. 2004, $\alpha_{T1} = 0.92$; DEMAT-3, Roick et al. 2004, $\alpha_{T1} = 0.91$; posttest: DEMAT-3, Roick et al. 2004, $\alpha_{T2} = 0.82$; DEMAT-4, Gölitz et al. 2006, $\alpha_{T2} = 0.89$). These standardized instruments closely reflect the (German) core standards for the respective grade levels. Thus, different versions of the tests were necessary for the different grades at pre- and posttest. For all of these measures, we used the obtained sum score, *z*-standardized by grade level.

3.3.2 Motivational constructs

To evaluate the effects of the training course on motivation, we assessed math self-concept as well as intrinsic value, attainment value, and task-specific interest in accordance to our target group of mathematically interested and gifted students. The *math self-concept* scale consisted of six items (e.g., "I'm good at everything that has to do with mathematics," $\alpha_{T1} = 0.87$, $\alpha_{T2} = 0.92$). *Intrinsic value in mathematics* was assessed with a scale of six items (e.g., "I enjoy everything that has to do with math," $\alpha_{T1} = 0.81$, $\alpha_{T2} = 0.95$), *attainment value* with a scale of three items; (e.g., "Everything that has to do with math is important to me," $\alpha_{T1} = 0.70$, $\alpha_{T2} = 0.69$). In addition, a more activity-oriented form of interest was measured with a scale that asked for *task-specific interest* (three items; e.g., "I like to solve riddles and puzzles in computing magazines and booklets," $\alpha_{T1} = 0.77$, $\alpha_{T2} = 0.88$). All scales were based on established instruments (Arens et al. 2011; Bos et al. 2005; Gaspard et al. 2015; Ramm et al. 2006), and the respective items were answered on a 4-point scale (ranging from 1 = *not true at all* to 4 = *exactly true*, visually represented by stars that increased in size). Mean values for math self-concept, intrinsic and attainment value in mathematics, and task-specific interest in mathematics were *z*-standardized for each grade level and used as dependent variables (for more details, see Tables 2 and 3, Appendix).

3.3.3 Covariates

General cognitive abilities were assessed with the short version of the Berlin Test of Fluid and Crystallized Intelligence (BEFKI; Schroeders et al. 2016). This test includes two subscales: figural (Versions A and B) and crystallized cognitive skills.

The figural subscale consists of 16 figural items in which sequences have to be continued twice ($\alpha_{T1, A} = 0.65$, $\alpha_{T1, B} = 0.71$/$\alpha_{T2, A} = 0.68$, $\alpha_{T2, B} = 0.78$). The crystallized subscale requires students to indicate the correct answer (out of four alternatives) to 16 questions about general knowledge, for example, "What's Google?" ($\alpha_{T1} = 0.73$, $\alpha_{T2} = 0.73$). For the analyses, we used the sum score of each scale separately and z-standardized them by grade level. As there were significant differences between intervention and control groups in average age, this variable was used as an additional covariate.

3.4 Statistical analyses

All analyses were run separately for each grade level in R (R Core Team 2015). Based on the recommendations of the What Works Clearinghouse (2014), baseline equivalence of treatment and control group or group differences at pretest were analyzed by calculating Hedges's g for all measures, expect for age. Group differences for age were analyzed by computing t tests for independent samples. If there were differences between the two groups at pretest, the variables were included as co-variates in the final model. We evaluated the effectiveness of the math course with multiple linear regression analyses using the R package *lavaan* (Rosseel 2012).

In all analyses, we used cluster-robust standard errors to account for the multi-level nature of the data (i.e., students nested in local sites; McNeish et al. 2017). Predictors in our regression models were participation in the training (1 = training, 0 = control) as well as achievement in the MO, (curriculum-oriented) mathematical competence, math self-concept, intrinsic value, attainment value, and task-specific interest in mathematics. According to the standardization of the dependent variables, the multiple regression coefficient for the training variable indicated the standardized difference between the training and control groups at posttest (treatment effect) while pretest performance in mathematical competence, fluid and crystallized intelligence, math self-concept, intrinsic value, attainment value, and task-specific interest in mathematics, and age were controlled for. We controlled for pretest performance on these variables to account for initial differences between the training and control groups. Significance levels were set to an overall level of $\alpha = 0.05$. To answer Hypotheses 1 and 2 the one-tailed significance level of the training variable is reported. Because there are multiple outcome measures in our study, we corrected the p values for the training variable for multiple testing by applying the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995). Differential effects between third and fourth graders for each dependent variable were analyzed by testing the equality of the two regression coefficients of the group variables by calculating a z value as suggested by Clogg et al. (1995; see also Paternoster et al. 1998).

3.5 Missing data

In our study, the percentage of missing values was 5% at pretest (30% of the missing data occurred in the training group and 70% in the control group) and 12% at posttest (52% of the missing data occurred in the training group and 48% in the control group). There were several reasons for the missing data: 10% of the students

in the training group left the math course because the course was too long for them. One student did not attend the pre- or posttest sessions but attended all other sessions and participated in the MO. In the control group, some children (a) moved away ($n = 2$), (b) changed schools or classes ($n = 5$), or (c) did not want to participate in the MO ($n = 10$). The other missing values resulted from illness. For the implemented measures, the missing values reached a maximum of 18.0% in the training group and 22.5% in the control group. When analyzing the treatment effects, we used the full information maximum likelihood approach implemented in R to deal with missing values (Enders 2010; Graham 2009; R Core Team 2015; Rosseel 2012).

4 Results

4.1 Descriptive statistics

In a first step, we evaluated differences between the training and control groups at pretest separately for the third and fourth graders (see Tables 2 and 3; for the intercorrelations of the scales, see Tables 4 and 5, Appendix). As to be expected from the way the groups were acquired, there were statistically significant differences between the two groups, indicated by a 95% CI for Hedges's g that did not include zero (What Works Clearinghouse 2014). On average, the children in the training group showed higher (fluid and crystallized) cognitive ability, higher mathematical competence measured with a curriculum-oriented test, higher levels of math self-concept, and higher levels of intrinsic value, attainment value, and task-specific interest in mathematics (for Hedges's g, see Table 1); they were also younger than the children in the control group (with a statistically significant difference for the third graders; see Table 1). Thus, we included mathematical competence measured at pretest, fluid and crystallized cognitive ability, math self-concept, intrinsic value, attainment value, and task-specific interest in mathematics, and age as covariates in our regression models.

4.2 Effects on performance in the academic competition (Hypothesis 1)

Our first hypothesis addressed the performance in the math competition. Overall, controlling for pretest values on mathematical competence measured with a curriculum-oriented test, fluid and crystallized intelligence, math self-concept, intrinsic value, attainment value, and task-specific interest in mathematics, and age, the training group showed higher *Performance in the Mathematical Olympiad* ($B_{third\ grade} = 1.15$, $p = 0.048$; $B_{fourth\ grade} = 0.55$, $p = 0.041$; see Fig. 5 and Table 6) than the control group (details are presented in Table 6, Appendix). Thus, students participating in the math course performed between about half of a standard deviation and one standard deviation better in the MO than students who did not attend the preparatory math course, controlling for baseline differences between the two groups.

Table 1 Baseline Equivalence between the Training and Control Groups at Pretest

	Third-grade students		Fourth-grade students	
	g	95% CI	g	95% CI
Mathematical competence	−1.09	[−1.57, −0.60]	−1.02	[−1.57, −0.48]
Math self-concept	−0.75	[−1.24, −0.26]	−0.65	[−1.15, −0.15]
Intrinsic interest in math	−0.72	[−1.21, −0.23]	−0.89	[−1.40, −0.38]
Task-specific interest in math	−0.54	[−1.23, −0.05]	−0.80	[−1.31, −0.29]
Attainment value in math	−0.54	[−1.03, −0.06]	−0.72	[−1.22, −0.22]
Figural cognitive abilities	−0.07	[−0.54, 0.40]	−0.54	[−1.04, −0.05]
Crystallized cogitive abilities	−0.23	[−0.70, 0.24]	−0.71	[−1.21, −0.21]
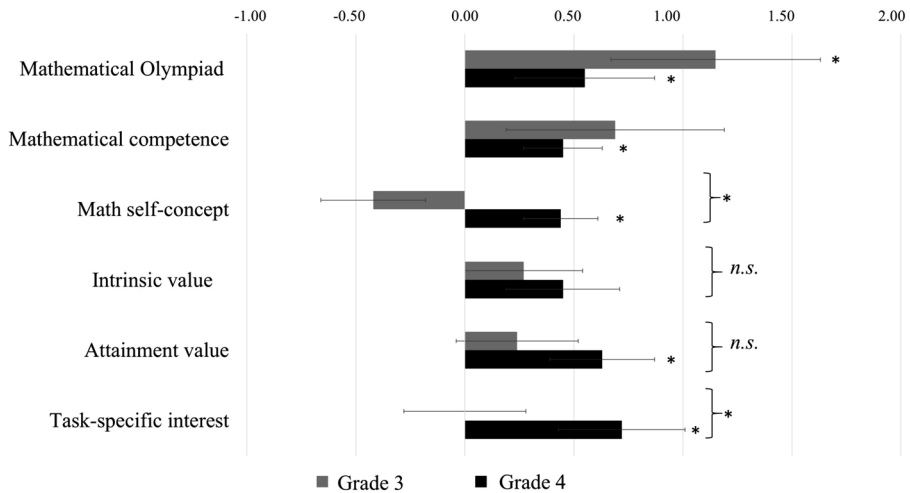
Fig. 5 Overview of treatment effects separated by grade level

4.3 Effects on mathematical competences measured with a curriculum-oriented test (Hypothesis 2)

Regarding the effects on more general *Mathematical Competences* measured with a curriculum-oriented test, controlling for pre-treatment differences in mathematical competence and the abovementioned covariates, the results indicated that children performed better in the mathematical competences measured with a curriculum-oriented test after completing the math course compared with the children in the control group, but the treatment effect was statistically significant only for the fourth graders ($B_{third\ grade} = 0.69$, $p = 0.164$; $B_{fourth\ grade} = 0.45$, $p = 0.010$; see Fig. 5 and Table 6, Appendix). Thus, fourth-grade students attending the math course scored almost half of a standard deviation higher on this measure than students who did not attend the course after controlling for baseline differences (see Table 6, Appendix).

4.4 Effects on motivation in mathematics (Hypotheses 3 and 4)

Our third and fourth hypotheses addressed differences in math self-concept, intrinsic value, attainment value, and task-specific interest in mathematics dependent on the grade level of the students. Again, in the analyses for all dependent variables (i.e., math self-concept, intrinsic value, attainment value, and task-specific interest), we controlled for the pretest values in mathematical competence, math self-concept, intrinsic value, attainment value, and task-specific interest in mathematics, fluid and crystallized general cognitive abilities, and age (see Fig. 5 and Tables 7 and 8, Appendix).

For fourth graders (but not for third graders), we observed a statistically significant treatment effect ($B_{third\ grade} = -0.42$, $p = 0.144$; $B_{fourth\ grade} = 0.44$, $p = 0.010$) on mathematical self-concept. Although, the treatment effect for third graders did not significantly differ from zero, descriptively, the treatment effects for the third and fourth graders were in different directions (i.e., they had different signs). Applying the procedure suggested by Clogg et al. (1995) for testing the equivalence of these regression coefficients, the z value calculated via the quotient of the difference of regression coefficients and the root of the sum of the standard errors' squares was statistically significant ($z = -3.78$, $p < 0.001$). Thus, as expected, a differential effect for the two grade levels was observed on math self-concept.

Moreover, we observed positive training effects on intrinsic value, attainment value, and task-specific interest in mathematics, which were statistically significant only for the fourth graders (intrinsic value: $B_{third\ grade} = 0.27$, $p = 0.232$, $B_{fourth\ grade} = 0.45$, $p = 0.041$; attainment value: $B_{third\ grade} = 0.24$, $p = 0.232$, $B_{fourth\ grade} = 0.63$, $p = 0.010$; task-specific interest: $B_{third\ grade} = 0.00$, $p = 0.499$, $B_{fourth\ grade} = 0.72$, $p = 0.041$). For more details, see Tables 7 and 8 (Appendix). There were no statistically significant training effect differences between the two grade levels for intrinsic value or attainment value ($z_{intrinsic\ value} = -0.54$, $p = 0.298$; $z_{attainment\ value} = -1.22$, $p = 0.113$). Although, the treatment effects for third graders did not significantly differ from zero, descriptively, the treatment effects for the third and fourth graders seemed to go in the same direction. However, for task-specific interest, the difference in treatment effects between third and fourth graders was statistically significant ($z_{task\text{-}specific\ interest} = -1.95$, $p = 0.026$). Thus, only for task-specific interest a differential training effect dependent on grade level was observed (favouring fourth graders).

5 Discussion

In the present study, we investigated the effectiveness of a training course that prepared third and fourth graders for the Mathematical Olympiad (MO). Our data suggested positive effects of the course on performance in the corresponding academic competition for students of both grades. The effect on mathematical competence measured with a curriculum-oriented test was statistically significant for the fourth graders, but not for the third graders. For motivational constructs, we observed no significant training effects for third graders, whereas the course increased fourth

graders' self-concept, intrinsic value, attainment value, and task-specific interest. In line with our expectations regarding social comparisons and the BFLPE in particular, effects on math self-concept and task-specific interest differed between the two grade levels. In particular, for math self-concept, the (nonsignificant) effect for the third- and the statistically significant effect for the fourth-grade students differed in direction. In the following, we will discuss results on mathematical competences and motivation in turn, before discussing some limitations of our study.

5.1 Effects on mathematical achievement

The positive effects of the training course on performance in the Mathematical Olympiad largely confirmed our expectations of the supporting effects of such courses regarding performance in the respective competition (e.g., Ozturk and Debelak 2008a, b). At first glance, the statistically significantly better performance in the MO observed for students participating in the math course might not be very surprising. However, the math course was not based exactly on the tasks of the respective MO but was rather based on information that was released after previous MOs. Moreover, students did not solve the original tasks from these former MOs in the math course, but were instructed more broadly with respect to mathematical problem solving and reasoning. This indicates that by focusing on solving (challenging) open mathematical problems, students acquired competences that were helpful for solving new mathematical problems in the MO.

Moreover, fourth graders seem to have acquired mathematical competences through participating in the training course that also helped them to solve mathematical problems of a standardized achievement test (based on the German educational standards for elementary school), the contents of which differed considerably from the contents of the math course. Thus, the focus on challenging tasks and the increase in learning opportunities due to the math training course seem to have positively affected students' mathematical competence. This finding is in line with the idea that it is possible to positively influence students' competencies by challenging them if they are already able to solve curriculum-oriented tasks (see e.g., Diezmann and Watters 2001). In addition, challenging students with complex tasks in cooperative learning settings seems to be a beneficial approach for fostering students who already show high domain-specific competence.

One possible explanation for this lack of a comparable effect for the third graders is the chosen measurement instrument. Mathematical competences measured with a curriculum-oriented test at posttest were measured by the German Mathematics Tests for Third Grades. This test is, indeed, designed to measure mathematical competences at the end of third grade and not as in our study, in the middle of the school year (i.e., March). Perhaps third grade students of both groups had not yet mastered necessary mathematical competences (such as written division) to be able to meet all the requirements of the test. In contrast, (curriculum-oriented) mathematical competences of fourth-grade students were measured by the German Mathematics Tests for Fourth Grades, which is explicitly designed to measure mathematical competences in the middle of the school year; thus, differences between the treatment and the control group may have been detectable.

5.2 Effects on motivation

The training course came with several plausible, but partly opposing effects on motivation. As described above, the course included several elements that have been shown to be motivating (e.g., Johnson et al. 2000; Randel et al. 1992). At the same time, students in the training course were brought into a new learning setting with similar high-achieving students. Students in the control group perceived the achievement of only the students in their regular math classes. In light of this complex set of potentially important factors, we did not specify a hypothesis on whether attending the course would, on average, affect student motivation. However, we expected to find differential treatment effects on students' motivation dependent on the class level. Third- and fourth-grade students were taught together, and in every course session, the dyads for the working phase were randomly assigned. Thus, there was the possibility that third and fourth graders worked on a (challenging) mathematical task together. Although the (challenging) open tasks did not go far beyond mathematical content covered by second-grade math curriculum (see KMK 2004), third graders—who had experienced one year less of formal schooling and maybe did not develop comparably elaborated mathematical competences—had more difficulties mastering the complex and challenging tasks in comparison to their fourth-grade peers. Following this thought, third graders might have been less able to use their learning potential to solve the tasks and, to some extent, might not have had the sophisticated mathematical competences that were needed to be able to understand alternative solution strategies.

In line with our expectations, we observed a difference between the training effects on self-concept and task-specific interest across third and fourth graders. In particular, the statistically significant difference in the development of students' math self-concept and task-specific interest resulted from a (descriptively) negative development of third-grade students' math self-concept and an increase in self-concept for fourth graders. This may indicate that social comparison processes for students in the two grade levels had differential consequences. Whereas third graders (who were trained in groups with higher achieving fourth graders) tended to experience a decrease in their math self-concept, the opposite seemed to occur for fourth-grade students (who were trained with lower achieving third graders) as they experienced an increase in their self-concept. In the training group, the social comparison processes were likely to be more favorable for fourth graders as compared to third graders—reflecting downwards versus upwards comparisons, respectively—which would explain the differential training effects.

In addition, the (partial) evidence for positive effects on intrinsic value, attainment value, and task-specific interest might indicate that the positive experiences of learning more about mathematics and of working in teams (even if the partner seemed more competent) stabilized students' value for mathematics. However, further research is needed to better understand the role of the various factors (including social comparison processes) that come into play in training courses for academic competitions.

5.3 Limitations and implication

There are a few points to consider when interpreting the results of our study. It is important to critically gauge the generalizability of our findings: First, it remains unclear whether the results are transferable to domains other than mathematics, to training course designs based on different core components, or to different target groups (e.g., students from secondary schools, or students with lower levels of achievement). Students in the training group showed mathematical competences that were above average. Thus, the present manuscript cannot answer the question of whether the math course "Getting Fit for the Mathematical Olympiad" would be effective for students other than those nominated for the Hector Children's Academy Program (HCAP). The nomination is done by teachers, and from previous investigations of the HCAP, we know that students nominated show above average general cognitive ability, domain-specific skills, interests, and motivation (Golle et al. 2018; Rothenbusch et al. 2016). Teachers' nominations are not fully accurate in terms of specific selection criteria (for more information about teacher nominations, see, e.g., Rothenbusch 2017). However, the nominated students (high-achieving and highly motivated) are the typical target population of academic competitions and possible training courses.

Second, the quasi-experimental research design that we used comes with some limitations. We cannot say with certainty that we were able to control for all relevant pre-treatment differences. However, we controlled for the most relevant pretreatment differences (i.e., mathematical competences, motivation, general cognitive ability) in our analyses using established measurement instruments.

Third, when interpreting the effects of our study, it is also important to consider that we can only draw conclusions on the effectiveness of the actual combination of our four core components—cooperative learning, mathematical games, challenging open tasks, and a structured notation of approaches and solutions. Indeed, all of them were aimed at fostering students' mathematical competences and motivation for mathematics and counterbalancing social comparison processes and especially the competitive environment of academic competitions. For all core components that were implemented, there are studies and theoretical considerations that corroborate these positive effects (see e.g., Hänze and Berger 2007; Johnson et al. 2000, for cooperative learning). However, which mechanisms and interactions were more or less important for the effectiveness of the course remains unclear. This could be a typical problem of multicomponent interventions or training courses, but in real-world settings multicomponent trainings are necessary to positively affect several student characteristics.

At last, we did not systematically control for treatment fidelity (Abry et al. 2015). Although, we intended to ensure high treatment fidelity through a training of the course instructors, a scripted manual, and master copies, we cannot fully eliminate differences in treatment delivery between the course instructors. In further studies, treatment fidelity should be controlled for.

6 Conclusion

To the best of our knowledge, the present study is the first we know of to systematically examine the effects of a training course for a mathematics competition on (a) achievement in the mathematics competition, (b) mathematical competences measured with a curriculum-oriented test, and (c) students' domain-specific motivation (i.e., self-concept, intrinsic value, attainment value, and task-specific interest). The results of the present study indicate that preparation courses for an academic competition can increase the success in the respective competition and positively affect students' domain-specific competences. Furthermore, students' self-concept, intrinsic value, attainment value, and task-specific interest can be affected by the training. Caution is necessary when it comes to group membership due to possible social comparisons. Our study adds to some well-established theories and discussions, for example, with regard to the consequences of social comparison processes in learners.

References

Abernathy, T. V., & Vineyard, R. N. (2001). Academic competitions in science. What are the rewards for students? *The Clearing House*, *74*(5), 269–276. https://doi.org/10.1080/00098650109599206.

Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation*, *36*(3), 320–338. https://doi.org/10.1177/1098214014557009.

Arens, A. K., Trautwein, U., & Hasselhorn, M. (2011). Erfassung des Selbstkonzepts im mittleren Kindesalter. Validierung einer deutschen Version des SDQ I. *Zeitschrift für Pädagogische Psychologie*, *25*(2), 131–144. https://doi.org/10.1024/1010-0652/a000030.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bezold, A. (2012). Förderung von Argumentationskompetenzen auf der Grandlage von Forscheraufgaben. Eine empirische Studie im Mathematikunterricht der Grandschule. *Mathematica Didactica*, *35*, 73–103.

Bicknell, B. (2008). Gifted students and the role of mathematics competitions. *Australian Primary Mathematics Classroom*, *13*(4), 16–20.

Blum, W. (2012). Einführung. In W. Blum, C. Drüke-Noe, R. Hartung & O. Köller (Eds.), *Bildungsstandards Mathematik: konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsanregungen, Fortbildungsideen* (6th edn., pp. 14–32). Berlin: Cornelsen.

Bos, W., Buddeberg, I., & Lankes, E.-M. (Eds.). (2005). *IGLU. Skalenhandbuch zur Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

Campbell, J.R., & Walberg, H.J. (2010). Olympiad studies. Competitions provide alternatives to developing talents that serve national interests. *Roeper Review*, *33*(1), 8–17. https://doi.org/10.1080/02783193.2011.530202.

Campbell, J.R., Wagner, H., & Walberg, H.J. (2000). Academic competitions and programs designed to challenge the exceptionally talented. In K.A. Heller, F.J. Mönks, R. Subotnik & R.J. Sternberg (Eds.), *International handbook of giftedness and talent* (2nd edn). Amsterdam: Elsevier.

Campbell, J.R., & Verna, M.A. (2010). Academic competitions serve the US national interests. http://files.eric.ed.gov/fulltext/ED509402.pdf [Paper presentation]. Accessed: 12. July 2022. Annual Meeting of the American Educational Research Association, Denver, CO, United States.

Clogg, C.C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, *100*(5), 1261–1293. https://doi.org/10.1086/230638.

Cropper, C. (1998). Is competition an effective classroom tool for the gifted student? *Gifted Child Today*, *21*(3), 28–31. https://doi.org/10.1177/107621759802100309.

Dai, D.Y., & Rinn, A.N. (2008). The big-fish-little-pond effect: What do we know and where do we go from here? *Educational Psychology Review*, *20*(3), 283–317. https://doi.org/10.1007/s10648-008-9071-x.

Diezmann, C.M., & Watters, J.J. (2000). Catering for mathematically gifted elementary students: learning from challenging tasks. *Gifted Child Today*, *23*(4), 14–52. https://doi.org/10.4219/gct-2000-737.

Diezmann, C.M., & Watters, J.J. (2001). The collaboration of mathematically gifted students on challenging tasks. *Journal for the Education of the Gifted*, *25*(1), 7–31. https://doi.org/10.1177/016235320102500102.

Enders, C.K. (2010). *Applied missing data analysis*. New York: Guilford.

Fauser, P., Messner, R., Beutel, W., & Tetzlaff, S. (2007). *Fordern and fördern. Was Schülerwettbewerbe leisten*. Hamburg: Körber-Stiftung.

Fleischmann, M., Hübner, N., Marsh, H., Guo, J., Trautwein, U., & Nagengast, B. (2021). Which class matters? Juxtaposing multiple class environments as frames-of-reference for academic self-concept formation. *Journal of Educational Psychology*. https://doi.org/10.31234/osf.io/7pbac. Advance online publication.

Forrester, J.H. (2010). *Competitive science events: Gender, interest, science self-efficacy, and academic major choice* (Doctoral dissertation). North Carolina: State University.

Freudenthal, H. (1986). *Didactical phenomenology of mathematical structures*. Berlin: Springer.

Gallin, P., & Ruf, U. (1995). *Sprache and Mathematik: 1.–3. Schuljahr: Ich mache das so! Wie machst du es? Das machen wir ab*. Luzern: Interkantonale Lehrmittelzentrale.

Gaspard, H., Dicke, A.-L., Flunger, B., Schreier, B., Häfner, I., Trautwein, U., & Nagengast, B. (2015). More value through greater differentiation: gender differences in value beliefs about math. *Journal of Educational Psychology*, *107*(3), 663–677. https://doi.org/10.1037/edu0000003.

Gölitz, D., Roick, T., & Hasselhorn, M. (2006). *DEMAT 4. Deutscher Mathematiktest für vierte Klassen*. Göttingen: Hogrefe.

Golle, J., Zettler, I., Rose, N., Trautwein, U., Hasselhorn, M., & Nagengast, B. (2018). Effectiveness of a "grass roots" statewide enrichment program for gifted elementary school children. *Journal of Research on Educational Effectiveness*, *11*(3), 375–408. https://doi.org/10.1080/19345747.2017.1402396.

Gordeeva, T.O., Osin, E.N., Kuz'menko, N.E., Leont'ev, D.A., & Ryzhova, O.N. (2013). Efficacy of the academic competition (Olympiad) system of admission to higher educational institutions (in chemistry). *Russian Journal of General Chemistry*, *83*(6), 1272–1281. https://doi.org/10.1134/S1070363213060479.

Graham, J.W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.

Hänze, M., & Berger, R. (2007). Cooperative learning, motivational effects, and student characteristics. An experimental study comparing cooperative learning and direct instruction in 12th grade physics Classes. *Learning and Instruction*, *17*(1), 29–41. https://doi.org/10.1016/j.learninstruc.2006.11.004.

Johnson, D. W., & Johnson, R. T. (1990). Using cooperative learning in math. In N. Davidson (Ed.), *Cooperative learning in mathematics. A handbook for teachers* (pp. 103–125). Boston: Addison-Wesley.

Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). Cooperative learning methods: a meta-analysis. https://s3.amazonaws.com/academia.edu.documents/33787421/Cooperative_Learning_Methods_A_Meta-Analysis.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1534244081&Signature=exeKEnuyQaZDpaXq4JZtrGQGG34%3D&response-content-disposition=inline%3B%20filename%3DCooperative_Learning_Methods_A_Meta-Anal.pdf. Accessed: 12. July 2022.

Käpnick, F. (2010). „Mathe für kleine Asse" – Das Münsteraner Konzept zur Förderung mathematisch begabter Kinder. In M. Fuchs & F. Käpnick (Eds.), *Eine Herausforderung für Schule und Wissenschaft. Begabungsforschung: Mathematisch begabte Kinder* (Vol. 8, 2nd edn., pp. 138–150). Münster: Lit.

Karnes, F. A., & Riley, T. L. (2005). *Competitions for talented kids*. Waco, TX: Prufrock Press.

Kießwetter, K. (2013). Können auch Grundschüler schon im eigentlichen Sinne mathematisch agieren? In H. Bauersfeld & K. Kießwetter (Eds.), *Wie fördert man mathematisch besonders befähigte Kinder? Ein Buch aus der Praxis für die Praxis* (5th edn., pp. 128–153). Offenburg: Mildenberger.

Köller, O. (2010). Bildungsstandards. In R. Tippelt & B. Schmidt (Eds.), *Handbuch Bildungsforschung* (pp. 529–548). Wiesbaden: Springer.

Koshy, V., Ernest, P., & Casey, R. (2009). Mathematically gifted and talented learners: theory and practice. *International Journal of Mathematical Education in Science and Technology*, *40*(2), 213–228. https://doi.org/10.1080/00207390802566907.

Krajewski, K., Liehm, S., & Schneider, W. (2004). *DEMAT 2+: Deutscher Mathematiktest für zweite Klassen*. Weinheim: Beltz.

Kuech, R., & Sanford, R. (2014). Academic competitions: perceptions of learning benefits from a science bowl competition. In European Scientific Institute (Ed.), *Proceedings* (2nd edn., pp. 388–394).

Kulik, J. A., & Kulik, C.-L. C. (1987). Meta-analytic findings on grouping programs. *Gifted Child Quarterly*, *36*(2), 73–77. https://doi.org/10.1177/001698629203600204.

Kultusministerkonferenz (2004). *Bildungsstandards im Fach Mathematik für den Primarbereich*. München: Luchterhand.

Leikin, R. (2010). Teaching the mathematically gifted. *Gifted Education International*, *27*(2), 161–175. https://doi.org/10.1177/026142941002700206.

Lengfelder, A., & Heller, K. A. (2002). German Olympiad studies: Findings from a retrospective evaluation and from in-depth interviews. Where have all the gifted females gone? *Journal of Research in Education*, *12*(1), 86–92.

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, *1*(4), 316–345. https://doi.org/10.1111/j.1745-6916.2006.00019.x.

Lyman, F. (1981). The responsive classroom discussion: the inclusion of all students. In A. S. Anderson (Ed.), *Mainstreaming digest* (pp. 109–113). Maryland: University of Maryland College of Education.

Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology*, *79*(3), 280–295. https://doi.org/10.1037/0022-0663.79.3.280.

Marsh, H. W., & Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, *47*(1), 213–231. https://doi.org/10.1037/0022-3514.47.1.213.

Marsh, H. W., Kong, C.-K., & Hau, K.-T. (2000). Longitudinal multilevel models of the big-fish-little-pond effect on academic self-concept: counterbalancing contrast and reflected-glory effects in Hong Kong schools. *Journal of Personality and Social Psychology*, *78*(2), 337. https://doi.org/10.1037/0022-3514.78.2.337.

Marsh, H. W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K.-T., O'Mara, A. J., & Craven, R. G. (2008). The big-fish–little-pond-effect stands up to critical scrutiny: implications for theory, methodology, and future research. *Educational Psychology Review*, *20*(3), 319–350. https://doi.org/10.1007/s10648-008-9075-6.

McAllister, B. A., & Plourde, L. A. (2008). Enrichment curriculum: essential for mathematically gifted students. *Education*, *129*(1), 40–49.

McNeish, D., Stapleton, L. M., & Silverman, R. D. (2017). On the unnecessary ubiquity of hierarchical linear modeling. *Psychological Methods*, *22*(1), 114–140. https://doi.org/10.1037/met0000078.

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston: NCTM.

Nolte, M., & Pamperien, K. (2013). Besondere mathematische Begabung im Grandschulalter – ein Forschungs- and Förderprojekt. In H. Bauersfeld & K. Kießwetter (Eds.), *Wie fördert man mathe-*

matisch besonders befähigte Kinder? Ein Buch aus der Praxis für die Praxis (5th edn., pp. 70–72). Offenburg: Mildenberger.

Olson, S. (2005). *Count down: six kids vie for glory at the world's toughest math competition*. New York, NY: Houghton Mifflin.

Oswald, F., Hanisch, G., & Hager, G. (2005). *Wettbewerbe and „Olympiaden". Impulse zur (Selbst)-Identifikation von Begabungen*. Münster: LIT.

Ozturk, M. A., & Debelak, C. (2008a). Academic competitions as tools for differentiation in middle school. *Gifted Child Today*, *31*(3), 47–53.

Ozturk, M. A., & Debelak, C. (2008b). Affective benefits from academic competitions for middle school gifted students. *Gifted Child Today*, *31*(2), 48–53. https://doi.org/10.4219/gct-2008-758.

Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, *36*(4), 859–866. https://doi.org/10.1111/j.1745-9125.1998.tb01268.x.

Preckel, F., Schmidt, I., Stumpf, E., Motschenbacher, M., Vogl, K., Scherrer, V., & Schneider, W. (2019). High-ability grouping: benefits for gifted students' achievement development without costs in academic self-concept. *Child Development*, *90*(4), 1185–1201. https://doi.org/10.1111/cdev.12996.

R Core Team (2015). *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Ramm, G. C., Adamsen, C., & Neubrand, M. (Eds.). (2006). *PISA 2003. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.

Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The effectiveness of games for educational purposes: a review of recent research. *Simulation & Gaming*, *23*(3), 261–276. https://doi.org/10.1177/1046878192233001.

Rebholz, F. (2018). *Fostering mathematical competences by preparing for a mathematical competition* (Dissertation). Tübingen: Eberhard Karls Universität. https://doi.org/10.15496/publikation-21245.

Rebholz, F., & Golle, J. (2017). Förderung mathematischer Fähigkeiten in der Grundschule – Die Rolle von Schülerwettbewerben am Beispiel der Mathematik-Olympiade. In U. Trautwein & M. Hasselhorn (Eds.), *Begabungen und Talente* (Vol. Tests und Trends, Jahrbuch der pädagogisch-psychologischen Diagnostik, N.F. Vol. 15). Göttingen: Hogrefe.

Riley, T. L., & Karnes, F. A. (1998). Mathematics + competitions = A winning formula! *Gifted Child Today*, *21*(4), 42.

Rinn, A. N. (2007). Effects of programmatic selectivity on the academic achievement, academic self-concepts, and aspirations of gifted college students. *Gifted Child Quarterly*, *51*(3), 232–245. https://doi.org/10.1177/0016986207302718.

Roick, T., Gölitz, D., & Hasselhorn, M. (2004). *Deutscher Mathematiktest für dritte Klassen (DEMAT 3+)*. Weinheim: Beltz.

Rosenzweig, E. Q., & Wigfield, A. (2016). STEM motivation interventions for adolescents: a promising start, but further to go. *Educational Psychologist*, *51*(2), 146–163. https://doi.org/10.1080/00461520.2016.1154792.

Rosseel, Y. (2012). *lavaan: An R package for structural equation modeling and more. Version 0.4-9 (BETA)*. Ghent: Ghent University.

Rothenbusch, S. (2017). *Elementary school teachers' beliefs and judgments about students' giftedness* (Doctoral dissertation). Tübingen: Eberhard Karls Universität.

Rothenbusch, S., Zettler, I., Voss, T., Lösch, T., & Trautwein, U. (2016). Exploring reference group effects on teachers' nominations of gifted students. *Journal of Educational Psychology*, *108*(6), 883–897. https://doi.org/10.1037/edu0000085.

Rotigel, J. V., & Fello, S. (2016). Mathematically gifted students: how can we meet their needs? *Gifted Child Today*, *27*(4), 46–51. https://doi.org/10.4219/gct-2004-150.

Schroeders, U., Schipolowski, S., Zettler, I., Golle, J., & Wilhelm, O. (2016). Do the smart get smarter? Development of fluid and crystallized intelligence in 3rd grade. *Intelligence*, *59*, 84–95. https://doi.org/10.1016/j.intell.2016.08.003.

Seo, B.-I. (2015). Mathematical writing: What is it and how do we teach it? *Journal of Humanistic Mathematics*, *5*(2), 133–145. https://doi.org/10.5642/jhummath.201502.12.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Skiena, S. S., & Revilla, M. A. (2003). *Texts in computer science. Programming challenges: the programming contest training manual*. Heidelberg: Springer.

Smith, K. N., Jaeger, A. J., & Thomas, D. (2019). "Science Olympiad is why I'm here": The influence of an early STEM program on college and major choice. *Research in Science Education*, *51*(4), 443–459. https://doi.org/10.1007/s11165-019-09897-7.

Steenbergen-Hu, S., & Moon, S. M. (2010). The effects of acceleration on high-ability learners: a meta-analysis. *Gifted Child Quarterly*, *55*(1), 39–53. https://doi.org/10.1177/0016986210383155.

Subotnik, R. F., Olszewski-Kubilius, P., & Worrell, F. C. (2011). Rethinking giftedness and gifted education: a proposed direction forward based on psychological science. *Psychological Science in the Public Interest*, *12*(1), 3–54. https://doi.org/10.1177/1529100611418056.

Trautwein, U., & Möller, J. (2016). Self-concept: determinants and consequences of academic self-concept in school contexts. In A. A. Lipnevich, F. Preckel & R. D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century: theory, research and practice* (pp. 187–214). Berlin: Springer. https://doi.org/10.1007/978-3-319-28606-8_8.

Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O., & Baumert, J. (2006). Tracking, grading, and student motivation: using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, *98*(4), 788–806. https://doi.org/10.1037/0022-0663.98.4.788.

Trautwein, U., Lüdtke, O., Marsh, H. W., & Nagy, G. (2009). Within-school social comparison: how students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology*, *101*(4), 853–866. https://doi.org/10.1037/a0016306.

Urhahne, D., Ho, L. H., Parchmann, I., & Nick, S. (2012). Attempting to predict success in the qualifying round of the international Chemistry Olympiad. *High Ability Studies*, *23*(2), 167–182. https://doi.org/10.1080/13598139.2012.738324.

Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: a 25-year longitudinal study. *Journal of Educational Psychology*, *102*(4), 860–871. https://doi.org/10.1037/a0019454.

What Works Clearinghouse (2014). *Procedures and standards handbook*. Washington, D.C.: US Department of Education. (Version 3.0)

Winkelmann, H., Robitzsch, A., Stanat, P., & Köller, O. (2012). Mathematische Kompetenzen in der Grandschule. *Diagnostica*, *58*(1), 15–30. https://doi.org/10.1026/0012-1924/a000061.

Wirt, J. L. (2011). *An analysis of science Olympiad participants' perceptions regarding their experience with the science and engineering academic competition* (Doctoral dissertation). South Orange, New Jersey: Seton Hall University.

Xu, J. (2010). *Lecture notes on Mathematical Olympiad courses. For junior section Vol. 2* (Mathematical Olympiad series, Vol. 8). Singapore: World Scientific Publishing Co.

Zeidner, M., & Schleyer, E. J. (1999). The big-fish–little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology*, *24*(4), 305–329. https://doi.org/10.1006/ceps.1998.0985.