Peer tutoring in middle school mathematics: Academic and psychological effects and moderators

Lidon Moliner[a]*, Francisco Alegre[a]

[a]Jaume I University, Av. Sos Baynat s/n, 12071, Castellon, Spain

Declarations of interest: none

Corresponding author:

*Corresponding author: Lidon Moliner

Mailing address: Av. Sos Baynat s/n, 12071, Castellon, Spain

Tel.: +00 34 964 72 98 06; Fax: +34 964 72 92 64

E-mail: mmoliner@uji.es

Peer tutoring in middle school mathematics: Academic and psychological effects and moderators

Abstract

In this study reciprocal peer tutoring was implemented in middle school mathematics classes for a six-month period. The effects of this methodology on students' mathematics achievements, mathematics anxiety levels, attitudes toward mathematics, and mathematics self-concepts were examined. Study participants included 768 students in grades 7 to 9 (12 to 15 years old). Statistically significant differences and positive effect sizes were reported for mathematics achievement, mathematics anxiety level, and mathematics self-concept (Hedge's g average value of 0.28). No significant statistical differences were reported for mathematics attitude. Students' repeating condition acted as significant moderators for mathematics anxiety. The main conclusion is that reciprocal peer tutoring can be very beneficial academically and psychologically for middle school mathematics students.

Keywords: peer tutoring, mathematics, self-concept, anxiety, attitude

Introduction

Peer tutoring, commonly referred to as classwide peer tutoring (CWPT) (Heron et al., 2006), is an active learning methodology in which pairs of students teach each other while they, themselves, learn by teaching (Fuchs et al., 2020; Guill et al., 2020). During PT, the more competent student (tutor) helps a less competent peer (tutee) to understand curricular content (Hoffman et al., 2020). Different types of PT can be implemented, depending on students' ages and roles. One type is reciprocal PT, during which students switch roles, going from tutor to tutee (Moliner & Alegre, 2022). Fixed PT takes place when students do not switch roles (Hajar, 2020). Depending on the students' ages, same-age tutoring (Moliner & Alegre, 2020a) or cross-age tutoring (Zeneli et al., 2016) may be arranged. Although PT can be implemented in a variety of forms, and the organizational settings (length of PT sessions, number of sessions per week, duration of the PT program, and so on) may be infinite, most PT initiatives result in successful outcomes in academic, psychological, social, and behavioral variables (Malone & Fuchs, 2014; Culver et al. 2020; Chen et al., 2021). While researchers study PT outcomes for a range of academic subjects (McMaster & Fuchs, 2016; Patchan et al., 2022), Mathematics is one of the disciplines for which PT experiences have been documented most often over the last four decades (Webb et al., 2019).

Peer Tutoring in Mathematics

The academic benefits of PT in mathematics have been documented in both recent literature reviews and meta-analyses (Alegre-Ansuategui et al., 2018; Alegre et al., 2019a; Alegre et al., 2019b). Nevertheless, several authors (Tymms et al., 2011; Shenderovich et al., 2016; Leung, 2019a) have indicated that rigorous studies with randomized controlled trials and large sample sizes are still scarce in the literature. Moreover, according to Leung (2019b) not many studies in the field address the interactions among academic and psychological variables in PT interventions in mathematics. Additionally, most studies in the literature refer

to primary education experiences, while research in upper educational levels, such as middle school, high school, or college, is not as abundant (Moliner & Alegre, 2020b).

Anxiety, attitudes and self-concept in peer tutoring

Many variables have been studied since the 80s in peer tutoring interventions. Several studies have assessed the effects of peer tutoring on mathematics anxiety, attitudes towards mathematics or mathematics self-concepts (Topping et al. 2017). According to Powell and Fuchs (2015), an increase in students' mathematics self-concepts may be expected for those students who participate in peer tutoring interventions. Moreover, receiving frequent help by a peer has frequently been linked to a reduction in students' anxiety (Topping, 2021). Given the closeness and mutual confidence between classmates, organized academic support through peer learning interventions are usually regarded as a potential tool to reduce students' anxiety (Fuchs et al., 2020). In this sense, peer tutoring interventions are also expected to have a positive impact on students' attitudes towards learning. As Tymms et al. (2018) and Topping (2022) indicate, peers may play a vital role on other students' attitudes in class. So, structured and supervised helping behaviours in educational contexts may foster students' attitudes towards learning. Although previous researchers in the field have analysed one or even two of these variables, the literature in the field that includes moderation analysis among these variables or other student related variables is still scarce (Johns & Mills, 2021, Xu et al. 2021)

Aim and Research Questions

Given these existing gaps in the literature, the aim of this study was to determine the following: 1) the effects of reciprocal PT in middle school mathematics on several academic and psychological variables, and 2) the influence of repeating condition as a moderator of one or more of these variables.

The research questions that guided this study are as follows:

1. What are the effects of reciprocal PT on middle school students' mathematics achievements, mathematics anxiety levels, attitudes toward mathematics, and mathematics self-concepts?

2. Does repeating condition moderate the effects of the reciprocal PT intervention on mathematics achievement, mathematics anxiety level, attitude toward mathematics, or mathematics self-concept?

Method

Sample

In Spain, special authorization is required in order to administer questionnaires to students under 18 years old when conducting any type of research (Berasategui et al., 2021). To this purpose, the Valencian Ministry of Education was addressed. A middle school in the Valencian region was randomly selected and authorized to undertake this study. Apart from the institutional authorization, students' parental authorization was also required. Moreover, the School Council had to approve the research, and ethical requirements provided by the Ethics Committee of the Spanish National Research Council had to be followed. After all administrative procedures had been completed, full legal authorization was secured for 770 students from grades 7 through 9 (12 to 15 years old) to participate. As two students moved to another school during the study period the final sample comprised 768 students equally distributed by grades (one third of participating students from each grade) and experimental condition (one half randomly assigned to experimental conditions and the other half to control conditions). Of these 768 students, 388 were female, and 380 were male. Students' demographic distribution was as follows: 451 (58.72%) were Hispanic, 174 (22.66%) were Rumanian, 136 (17.71%) were African, and the remaining 7 students (0.91%) were from other ethnic groups. A total of 96 students (32 students in each grade) were undertaking a grade repetition (repeaters).

Power of the Sample

StudySize 3.0 software by Creostat HB was used to determine the sample power. A sample power of .80 was determined for a sample of 768 participants when using Pearson's correlation test to report a correlation difference of 0.1 with a significance level of .05.

Intervention

Experimental Design

Authors such as Zeneli et al. (2018) suggested including control groups in studies on PT interventions, as the omission of a control group may result in an overestimation or underestimation of the outcomes. Hence, following their suggestions, students were randomly assigned to experimental or control conditions following the standards of a randomized control trial. Block randomization method was used in this research (Ye et al., 2022). This method is designed to randomize subjects into groups that result in equal sample sizes (Xia et al., 2021). All students were pretested at the end of the first trimester of the school year, prior to the beginning of the reciprocal PT intervention, and later posttested at the end of the third trimester, following the last PT session.

Mathematical Content Covered

Students were taught different mathematical content during the interventions, depending on their grade level. The main mathematical content covered during last two trimesters of the school year (period of tutoring intervention) was as follows. Students in 7th grade worked with simple and compound rules of three with direct and indirect proportionality, first-degree equations without fractions, percentages, proportional and inverse distributions, linear functions, descriptive statistics, the Laplace rule, and basic surface and volume calculations. Students in 8th grade worked with 2x2 systems of equations, equations with fractions and second-degree equations, compound probability and introduction to inferential statistics, and calculation of surfaces and volumes of irregular prisms. Students in

9th grade worked with 3x3 systems of equations, equations with grades greater than two, complex surfaces and volumes, tree diagrams, percentiles, and box diagrams.

Organization, Schedule, Frequency, and Duration of Tutoring Sessions

During the first trimester of the school year, students in both the experimental and the control groups received teacher-centered instruction (Kyttälä et al, 2022), that is, all teachers employed the one-way instructional teaching method. Teacher-centered pedagogy is based upon a model of an active teacher and a passive student (Baker et al., 2022). Students had to work individually; they could ask the teacher questions at any time, but interactions between the students themselves were limited. After the first trimester, the reciprocal PT program was implemented with students in the experimental group. This program lasted six months, which comprised the last two trimesters of the school year. Throughout these trimesters, four sessions of reciprocal PT were conducted each week during mathematics classes. The approximate time of each PT session was 25 minutes. During these last two trimesters, students in the control group continued with the teacher-centered method.

Type of Tutoring and Distribution of Students

Reciprocal and same-age PT was implemented. This type of tutoring was selected due to organizational issues. Students in the same grade tutored each other during the intervention. Pairs of students were formed following the suggestions of experienced researchers in the field (Topping, 2020). Students were ordered from highest to lowest based on their mathematics grade from the first trimester. Then, the first student was paired with the second student, the third was paired with the fourth, and so on. In this way, each student was paired with another who had a similar competency in mathematics (Thurston et al., 2019).

Teacher Training

Teachers whose students were assigned to experimental conditions received two training sessions of about one hour each. During the first session, they were instructed on the

fundamentals of reciprocal PT (De Backer et al., 2016). Teachers watched videos presented by professional researchers with expertise in the specific topics for which the class intervention took place. Elements such as correction of mistakes, patience, and positive feedback were highlighted, and issues such as the difference in the speed at which students complete their tasks and the incompatibility of pairs due to behavioural issues were addressed. During the second session, an intervention trial was performed during the teachers' classes. Teachers were given feedback during the intervention. After the trial, teachers asked several questions that had emerged during the session, which were answered in a joint meeting of all teachers. The training sessions were delivered by two qualified instructors who also served as researchers in the field.

Student Training

Students in the experimental group received two training sessions on tutoring procedures two weeks before the start of the reciprocal PT program. The same teachers who taught the students during the school year conducted these training sessions under the supervision of the researchers. Through active participation, students were asked to indicate the qualities that competent tutors and tutees should have in order to successfully tutor and be tutored. They were also told that they all shared a common goal in this process: ensuring that every classmate understood and finished the exercises and problems by the end of each tutoring session. "Pause, Prompt, and Praise" techniques were also discussed, and the importance of patience and respect was emphasized (Duran et al., 2020). Moreover, students were told that interactions between them had to be rich in mathematical content, so they should not talk about other non-related issues during the PT sessions. The importance of explaining mathematical content and procedures in different ways was also highlighted.

Teachers' Role and Distribution

Teachers facilitated the reciprocal PT sessions and supervised interactions between students. It was the teacher's responsibility to ensure that these interactions were proceeding in a respectful environment, that the work was limited to the exercises and problems covered in each session, and that effective academic help was being provided. In order to avoid any teacher effect during the intervention, each teacher was assigned approximately the same number of students in the experimental group and in the control group for the same grade.

Materials Used by Students during the Intervention

Students in both the experimental and the control groups were given the same materials during the entire school year, including during the tutoring sessions. Each grade had its own textbook, and the teachers provided worksheets and online exercises. Students were allowed to use calculators and other instruments (compass, set square, and so on) to solve the exercises and problems.

Classroom and Tutoring Dynamic

The CWPT method employed in this research is based in the DUOLOG method created by Keith Topping and other authors (Topping et al., 2011). This method was developed for working with mathematical problems and it has been used during the last two decades (Thurston et al., 2020). It consists of eight structured steps and includes discussion between students, summarising and generalising strategies. As Tymms et al. (2011) indicate, the tutor has to encourage the tutee to solve the problems while high emphasis is given on the development of metacognitive awareness of the strategies used. The distribution of time invested for each task is approximately as follows. At the beginning of each session, the teacher took approximately 15 minutes to explain new content. Then, students were given approximately 15 minutes to complete a series of exercises and/or problems related to that new content, working individually. During that time, students were allowed to ask the teacher questions about how to do the exercise or solve the problem. Meanwhile, the teacher checked

the procedures the students were following and their results, making sure that at least one of the students from each pair had the correct answers, while providing feedback. When the 15 minutes were over, a reciprocal PT session was held for about 25 minutes. Working in their assigned pairs, the students had to both check the work they had done individually by comparing their results, sharing their procedures, and asking each other questions, and also work together to solve those tasks that they had not been able to finish individually. Even if a pair of students had solved all tasks correctly, eliminating the need for tutoring, they were still required to share the procedures they had employed. When the two paired students arrived at different answers to a problem, both students tried to identify the mistake together, and the student with the correct answer had to help the other student understand how to complete the problem correctly. Although students were allowed to ask questions regarding the exercises and problems during tutoring, perseverance and individual effort were a must. If a pair of students was unable to solve a task correctly after working together, the teacher provided assistance. All students had to be able to solve the exercises and problems by themselves by the end of each tutoring session. Extra exercises and problems were given to student pairs who finished their work early. During the last 5 to 10 minutes of the session, the teacher responded to any remaining questions from students.

Data Collection and Missing Data

Data were collected in November 2018 (pretest scores, before the intervention started) and June 2019 (posttest, after the intervention). The students completed the tests during regular mathematics classes. Tests were administered by the researchers of this manuscript. SPSS Missing Value Analysis Package was used to detect all missing values (Cook, 2021). Missing data in this research was established as MCAR as the probability of a missing data value was independent of any observation in the data set (Gachau et al., 2021). Given this fact, missing data was handled as follows. Missing values were categorized by researchers

individually for each student who had one or more items with no answer. After that, incomplete tests were returned to the student for completion. Through this procedure it was ensured that every item had been answered and there was no missing data. Students were assigned a single identification code to enter on their tests in lieu of their names and were informed that their answers would be completely anonymous.

Measures against Drop Out

The main measure against drop out was a drawing for an iPad (2018) and two coupons worth 50 euros each for school materials. Students were told before the pretest that they could only participate in the drawing if they completed all tests. No students dropped out during the research.

Measures against Hawthorne-Effect

Students in the experimental group were not told that their tutoring sessions were linked to the administration of the tests or that their participation in the research differed from that of other students in that middle school. This was done to prevent a Hawthorne-effect, that is, the possibility that students may have modified an aspect of their behavior in response to their awareness of being observed. As documented in the literature, the integrity of the research could be undermined due to this effect (Tadesse et al., 2020).

Instruments

Four variables were measured in this study: the student's mathematics achievement, mathematics anxiety level, attitude toward mathematics, and mathematics self-concept. The instruments selected had been validated as rigorous, their reliability had been repeatedly tested (House, 2009; Arens et al., 2020; Wilkerson, 2021), and they were frequently used by an international audience (Taut & Rakoczy, 2016).

Students' mathematics achievements were measured using the International Association for the Evaluation of Educational Achievement (IEA) Trends in International

Mathematics and Science Study (TIMSS) mathematics achievement test (Herber et al., 2017), adapted to the Spanish curriculum. A maximum score of 10 points could be achieved: the higher the student's score, the higher his or her mathematics achievement. Recent studies refer to a reliability of .93 Cronbach's alpha for the instrument (Vesić et al., 2021; Wardat et al., 2022).

Students' mathematics anxiety levels were measured using the Mathematics Anxiety Scale for Children (MASC) developed by Chiu and Henry (1990), a four-point Likert scale test in which the higher the student's score, the higher his or her level of mathematics anxiety. Studies refer to a reliability of approximately .85 Cronbach's alpha for the instrument (Keshavarzi & Ahmadi, 2013; Hoorfar & Taleb, 2015).

Students' attitudes toward mathematics were measured using the Fennema-Sherman Mathematics Attitudes Scale (FSMAS) developed by Fennema and Sherman (1976), a five-point Likert scale assessment in which the higher the student's score, the more positive his or her attitude toward mathematics. Studies refer to a reliability of approximately .92 Cronbach's alpha for the instrument (Jon et al., 2015: Ren et al., 2016).

Finally, students' mathematics self-concepts were measured using the mathematics self-concept eight-point Likert scale developed by Marsh and Shavelson (1985): the higher the student's score, the more positive his or her mathematics self-concept. Studies refer to a reliability of approximately .88 Cronbach's alpha for the instrument (Wohlkinger et al., 2016; Yanhong et al., 2021).

Statistical Analysis

SPSS software version 25 was used to perform all analyses and calculations in this research. The Kolmogorov Smirnov test was performed to ensure normality of the data for the pretest scores in the experimental and control groups (Xiao, 2017). Chi-squared was used to address the differences in the proportions repeating condition in the groups (repeater/non-

repeater). Repeating condition (also known as "grade retention" or "non-promotion") refers to the fact that a student has repeated one or two school years at some time in his/her academic career, joining a class of younger students (Snead et al., 2022). Means and standard deviations were calculated for all variables. Cronbach's alpha was used to measure reliability of the instruments (Kalkbrenner, 2021). Students' t-tests (95% confidence level) were used to analyze the differences in the pretest scores between the control group and the experimental group. ANCOVAs were used to assess the effectiveness of the intervention on the four variables that were measured in this study. In these analysis, postest scores acted as the dependent variable with the respective pretest scores used as covariates when comparing the control group and the experimental group. Effect sizes were reported for each variable and condition (group and repeating condition). Hedge's g was used as a measure of effect size. Moderation analyses with regression were performed using the PROCESS macro for SPSS version 3.4 (Hayes & Rockwood, 2020).

Treatment Fidelity

Treatment fidelity was monitored in several ways. First, students' attendance and behavior were controlled through an online platform called Itaca. All teachers in the Valencian Region must use this platform; they are given a single username and password the first time they enter the teaching system in the Valencian Region. Among other functions, this system allows teachers to track student attendance at each session and to insert comments regarding students' behavior during their classes. Participants in both the experimental and the control groups had to attend at least 90% of the scheduled sessions. If a teacher inserted a negative comment regarding the behavior of a student in the experimental group during the PT sessions (showing absolute reluctance to help a peer), the student was dropped out of the study. Teacher performance during the tutoring interventions was supervised by one of the researchers. Each teacher participating in the study with students in the experimental group

was supervised during at least two tutoring sessions. No significant violations of the intervention protocol were reported regarding student attendance or behavior or for teacher performance.

Adequacy of Randomization

The Kolmogorov Smirnov test showed that students' scores in both the experimental and the control groups for each of the four analyzed variables followed a normal distribution ($p > .95$ for all cases). The overall proportions of females/males were .52/.48, respectively, for the experimental group and .51/.49, respectively, for the control group. These proportions did not differ significantly ($\chi2 = 0.02$, $df = 1$, $p = .99$). The overall proportions of repeaters/non-repeaters were .11/.89, respectively, in the experimental group and .13/.87, respectively, in the control group. These proportions also did not differ significantly ($\chi2 = 0.07$, $df = 1$, $p = .98$).

No statistically significant differences were reported for the pretest between the experimental and the control groups for mathematics achievement ($t = 0.36$, $df = 758$, $p = .71$), for mathematics anxiety level ($t = 0.88$, $df = 758$, $p = .37$), for attitude toward mathematics ($t = 0.79$, $df = 758$, $p = .43$), or for mathematics self-concept ($t = 0.31$, $df = 758$, $p = .76$).

<div align="center">

Results

</div>

Descriptive results for this research are reported in Tables 1 and 2. Means ($\bar{X}$), standard deviations (*SDs*) are indicated for each group and for each phase of the study. Gain scores for each group and variable and effect sizes for each variable are also reported in Table 1. Cronbach's alpha scores by group and phase of the study are reported for each variable in Table 2.

Table 1

Mean scores, SDs, gain score by group and phase of the study and global effect size for each variable

	Experimental group				Gain score	Control group				Gain Score	Effect size
	Pretest		Posttest			Pretest		Posttest			
	$\bar{X}$	SD	$\bar{X}$	SD		$\bar{X}$	SD	$\bar{X}$	SD		
Mathematics achievement	6.25	0.43	6.38	0.41	0.13	6.27	0.45	6.26	0.43	-0.01	0.29
Mathematics anxiety level	2.15	0.31	2.01	0.25	-0.14	2.19	0.41	2.08	0.29	-0.09	0.27
Attitude toward mathematics	2.92	0.49	2.95	0.51	0.03	2.87	0.52	2.94	0.62	0.07	0.02
Mathematics self-concept	3.66	0.51	3.73	0.48	0.07	3.64	0.54	3.60	0.50	-0.04	0.28

Table 2

Cronbach's alpha scores by group and phase of the study for each variable

	Experimental group		Control group	
	Pretest	Posttest	Pretest	Posttest
Mathematics achievement	.85	.87	.84	.87
Mathematics anxiety level	.88	.89	.85	.91
Attitude toward mathematics	.87	.88	.88	.85
Mathematics self-concept	.90	.92	.86	.85

Results for each research question are reported in the following sections.

Results for Research Question 1: Effects of the Reciprocal PT Intervention

Overall, the intervention effect was significant on achievement ($F_{(2,384)} = 98.23, p < .01$), mathematics anxiety level ($F_{(2,384)} = 87.45$, $p < .01$), and mathematics self-concept ($F_{(2,384)} = 56.81$, $p < .01$) when comparing control and experimental group students. No statistically significant differences were reported for attitude toward mathematics ($F_{(2,384)} = 0.07$, $p = .79$). Effect sizes by repeating condition and are shown for each variable in Table 3.

Table 3

Effect sizes for each variable by repeating condition, grade level, and overall

	Mathematics achievement	Mathematics anxiety level	Attitude toward mathematics	Mathematics self-concept
Repeaters	0.15	0.11	-0.06	0.32

Non-repeaters	0.34	0.31	0.04	0.24

Results for Research Question 2: Moderation Effects of Repeating Condition

Results showed that repeating condition acted as significant moderators for mathematics anxiety level for students in the experimental group. No other significant moderation conditions were found as shown on Table 4.

Table 4

Moderation analysis of repeating condition

	Repeating condition			
	B	SE	*t*	*p*
Mathematics achievement	0.46	2.22	1.56	.08
Mathematics anxiety level	2.18	0.13	5.77	<.01*
Attitude toward mathematics	0.24	0.36	0.43	.53
Mathematics self-concept	0.42	0.21	0.52	.66

*p<.01

Discussion

Effect Sizes and Statistical Improvements for Each Variable

The reported global effect sizes in this study are consistent with the findings of recent meta-analyses in the field. Leung (2019a, 2019b) and Alegre et al. (2019c) determined similar effect sizes in PT interventions regarding mathematics achievement. Although not specifically for mathematics, PT effects on anxiety level and self-concept were also similar to those reported in meta-analyses by Bowman-Perrott et al. (2014) and Lavrijsen and Verschueren (2020). In this sense, the reported statistical improvements as well as the percentage of students in the experimental group whose scores improved from the pretest to the posttest in these variables are consistent with the findings reported for recent studies by Alegre-Ansuategui and Moliner-Miravet (2017), Connor et al. (2018), Neu and Greer (2019), Alegre

et al. (2020), Alegre et al. (2021) and Arnándiz (2022). On the contrary, the global effect size

for attitude toward mathematics may be considered null (Polanin et al., 2016), and no

evidence of improvement was reported for this variable. Recent studies in the field by

Bergstrom and Zhang (2016) and Song et al. (2018) did not report significant improvements

regarding attitude toward mathematics. As indicated by McGowen and Davis (2020),

ambitious pedagogical practices do not seem to have a reliable and positive impact on

students' attitudes toward mathematics, although they may be effective in certain contexts.

These authors argued that students' attitudes toward mathematics may be one of the most

difficult psychological attributes to improve through learning innovations for middle school,

high school, and college students.

Repeaters vs Non-Repeaters

The fact that repeating students showed lower effect sizes than their peers who have

not repeated a grade is also consistent with recent studies in the field, including research by

Hickey and Flynn (2019), Losinski et al. (2019), and Moeyaert et al. (2020). In fact, repeating

condition acted as a significant moderator for mathematics anxiety level and nearly reached

the same classification for mathematics achievement. According to the abovementioned

authors, PT mathematics interventions are usually expected to show lower effect sizes

(although positive most of the time) for students who were already struggling academically

before implementation of the tutoring program.

Moderation analysis

the only significant finding in this study referred to repeating condition acting as

significant moderator of the mathematics anxiety variable. Regarding this issue, although not

specifically for peer tutoring, authors such as Sokolowski et al. (2019), and Lazarides and

Buchholz (2019) supported the finding that repeating condition may act as moderator

variables for mathematics anxiety in educational interventions in which cooperation among students is necessary.

Usefulness of the intervention

Although statistical significant improvements were found in three of the four analyzed variables, it must be noted that for the experimental group, in terms of percentage, overall improvements were slight (about 2% of improvements for each variable). The costs and time of intervention were notable, and, although the intervention lead to clinically significant outcomes, the slight overall change is not a strong justification to demand changes in middle-school policies taking the results from this intervention (Floden, 2020). In this sense, for instance, an increase of 2.08 % in the TIMMS scores for mathematics achievements did not mean much for the school principal or the teachers. However, this peer tutoring intervention proved to be an enjoyable and worthy alternative way of learning for middle-school students. Students improved slightly in one academic and two psychological variables at the same time they learned in a different way, helping their peers and creating an inclusive environment. Teachers and students believed it was a worthwhile intervention for several reasons. On one hand, teachers noted many students were much more enthusiastic and curious during the peer tutoring sessions. The inclusive character of the intervention, given the current legal requirements for inclusion in education in the Valencian Region (Moliner et al., 2021), was also considered an important factor by them. On the other hand, many students indicated they enjoyed helping other peers while others stated that receiving help by other peer had been really useful for them. Most of the teachers and students would be willing to participate in similar interventions in the future.

Limitations and Future Research

Although, as the authors of this research, we put forth our best efforts to follow the highest standards of a randomized controlled trial, the present study has some limitations that must be considered. First, because parental authorization was necessary for students to participate in the study, there are possible selection effects in the selection of students (Puza & Bonfrer, 2018). Furthermore, the fact that almost 6.25% of the students in the experimental group (24 out of 384) achieved a perfect score on the pretest pertaining to mathematics achievement must also be considered, as ceiling effects may have underestimated the full effect of the reciprocal PT intervention for mathematics achievement (Christofides & Manoli, 2020). That findings in this research are limited to the domain of mathematics and to middle school students must also be considered. This research focused on mathematics because it is considered one of the most important academic subjects worldwide, regardless of grade level; plus, students tend to find the subject difficult to master (Rittle-Johnson et al., 2020; Marbán et al., 2021). As a final limitation, the study sample is not representative of the population of middle school students in Spain. A more representative sample with students from different geographic and cultural contexts would have reinforced the results of this research (Arnold & Versluis, 2019; Eliyahu-Levi & Chen, 2022; Tan et al., 2021). Moreover, clustering effects must also be considered in any educational intervention (Hutchinson, 2009). They represent a major challenge and most of the times are difficult to control in educational settings (Chan et al., 2022). Given the above mentioned limitations of the sample of this research, readers must understand the fact that undesirable clustering could have affected significantly the results reported in this research.

Further research is needed in order to investigate the effects of reciprocal PT at lower or upper educational stages in mathematics (primary school, high school, or college) or with different subjects. Considerable effect sizes and statistical improvements should also be found for subjects such as chemistry, physics, technology, or biology in academic and psychological

variables, given the scientific character of the mathematics discipline and that mathematics fundamentals are needed to academically succeed in those subjects. As mentioned, further research in different contexts and cultures is necessary to confirm the promising results in this study. Since reciprocal PT does not require any specific technological support, its implementation in other contexts, such as in developing countries or with students whose families have low socioeconomic status, may be very interesting.

Conclusions

Mathematics reciprocal PT may be academically and psychologically beneficial for middle school students. Significant improvements and positive effect sizes were reported for students' mathematics achievements, mathematics anxiety levels, and mathematics self-concepts. However, the influence of reciprocal PT on students' attitude toward mathematics was null. Benefits for non-repeating students were greater than for repeating students. In this sense, repeating condition may act as a significant moderator of students' mathematics anxiety level during this type of experience. Mathematics PT effects were very similar for all grades across middle school.

Data availability statement

All data regarding this research may be accessed using the link below:

https://osf.io/ac57w/?view_only=68e60dd3886e486ea89a52e4c1c37497

References

Alegre Ansuategui, F. J., & Moliner Miravet, L. (2017). Emotional and cognitive effects of peer tutoring among secondary school mathematics students. *International Journal of Mathematical Education in Science and Technology*, *48*(8), 1185–1205. https://doi.org/10.1080/0020739X.2017.1342284

Alegre, F., Miravet, L., Lorenzo-Valentin, G., & Maroto, A. (2021). Learning statistics and probability through peer tutoring: A middle school experience. South African Journal of Education, 41(2), S1-S9, 1-9. https://doi.org/10.15700/saje.v41ns2a1861

Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin, G. (2019a). Peer tutoring in mathematics in primary education: A systematic review. *Educational Review*, *71*(6), 767–791. https://doi.org/10.1080/00131911.2018.1474176

Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin, G. (2019b). Peer tutoring and mathematics in secondary education: Literature review, effect sizes, moderators, and implications for practice. *Heliyon*, *5*(9), Article e02491. https://doi.org/10.1016/j.heliyon.2019.e02491

Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin, G. (2019c). Peer tutoring in algebra: A study in middle school. *The Journal of Educational Research*, *112*(6), 693–699. https://doi.org/10.1080/00220671.2019.1693947

Alegre, F., Moliner, L., Maroto, A., & Lorenzo-Valentin, G. (2020). Academic achievement and peer tutoring in mathematics: A comparison between primary and secondary education. *SAGE Open, 10*(2), 1-9. https://doi.org/10.1177/2158244020929295

Alegre-Ansuategui, F. J., Moliner, L., Lorenzo, G., & Maroto, A. (2018). Peer tutoring and academic achievement in mathematics: A meta-analysis. *Eurasia Journal of Mathematics, Science and Technology Education*, *14*(1), 337–354. https://doi.org/10.12973/ejmste/79805

Arens, A. K., Frenzel, A. C., & Goetz, T. (2020). Self-concept and self-efficacy in math: Longitudinal interrelations and reciprocal linkages with achievement. *The Journal of Experimental Education*, 1-19. https://doi.org/10.1080/00220973.2020.1786347

Arnándiz, O. M., Moliner, L., & Alegre, F. (2022). When CLIL is for all: Improving learner motivation through peer-tutoring in Mathematics. *System, 106, 102773*, 1-14. https://doi.org/10.1016/j.system.2022.102773

Arnold, I. J., & Versluis, I. (2019). The influence of cultural values and nationality on student evaluation of teaching. *International Journal of Educational Research*, *98*, 13–24. https://doi.org/10.1016/j.ijer.2019.08.009

Baker, S., Anderson, J., Burke, R., De Fazio, T., Due, C., Hartley, L., ... & Sidhu, R. (2022). Equitable teaching for cultural and linguistic diversity: exploring the possibilities for engaged pedagogy in post-COVID-19 higher education. *Educational Review*, 1-16. https://doi.org/10.1080/00131911.2021.2015293

Berasategi Sancho, N., Idoiaga Mondragon, N., Dosil Santamaria, M., & Picaza Gorrotxategi, M. (2021). The well-being of children with special needs during the COVID-19 lockdown: Academic, emotional, social and physical aspects. *European Journal of Special Needs Education*, 1-14. https://doi.org/10.1080/08856257.2021.1949093

Bergstrom, C., & Zhang, D. (2016). Geometry interventions for K-12 students with and without disabilities: A research synthesis. *International Journal of Educational Research*, *80*, 134–154. https://doi.org/10.1016/j.ijer.2016.04.004

Bowman-Perrott, L., Burke, M. D., Zhang, N., & Zaini, S. (2014). Direct and collateral effects of peer tutoring on social and behavioral outcomes: A meta-analysis of single-case research. *School Psychology Review*, *43*(3), 260–285.

Cantinotti, M., Désormeaux-Moreau, M., & Balbinotti, M. (2017). Mapping psychology students' perspective on group peer-tutoring in statistics. *Mentoring & Tutoring:*

Partnership in Learning, 25(4), 466–484.

https://doi.org/10.1080/13611267.2017.1403609

Chan, W., Hedges, L. V., & Hedberg, E. C. (2022). The effects of downstream clustering in longitudinal studies. The Journal of Experimental Education, 90(3), 731-759. https://doi.org/10.1080/00220973.2020.1783501

Chen, C. Y., Chang, S. C., Hwang, G. J., & Zou, D. (2021). Facilitating EFL learners' active behaviors in speaking: a progressive question prompt-based peer-tutoring approach with VR contexts. *Interactive Learning Environments*, 1-20. https://doi.org/10.1080/10494820.2021.1878232

Chiu, L. H., & Henry, L. L. (1990). Development and validation of the mathematics anxiety scale for children. *Measurement and Evaluation in Counseling and Development*, *23*(3), 121–127.

Christofides, T. C., & Manoli, E. (2020). Item count technique with no floor and ceiling effects. *Communications in Statistics-Theory and Methods*, *49*(6), 1330–1356. https://doi.org/10.1080/03610926.2018.1563165

Connor, C. M., Mazzocco, M. M., Kurz, T., Crowe, E. C., Tighe, E. L., Wood, T. S., & Morrison, F. J. (2018). Using assessment to individualize early mathematics instruction. *Journal of School Psychology*, *66*, 97–113. https://doi.org/10.1016/j.jsp.2017.04.005

Culver, K. C., Bowman, N. A., Youngerman, E., Jang, N., & Just, C. L. (2020). Promoting equitable achievement in STEM: Lab report writing and online peer review. *The Journal of Experimental Education*, 1-23. https://doi.org/10.1080/00220973.2020.1799315

De Backer, L., Van Keer, H., & Valcke, M. (2016). Eliciting reciprocal peer-tutoring groups' metacognitive regulation through structuring and problematizing scaffolds. *The*

Journal of Experimental Education, *84*(4), 804–828.

https://doi.org/10.1080/00220973.2015.1134419

Duran, D., Ribosa, J., & Sánchez, G. (2020). Peer tutoring for improvement in rhythm reading

fluency and comprehension. *International Journal of Music Education, 38*(2), 299–

312. https://doi.org/10.1177/0255761419898313

Eliyahu-Levi, D., & Chen, S. (2022). Hebrew in Tel Aviv, Beijing, Denver, and Melbourne:

assessing peer learning in a multicultural environment. *Interactive Learning*

Environments, 1-12. https://doi.org/10.1080/10494820.2022.2059519

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales:

Instruments designed to measure attitudes toward the learning of mathematics by

females and males. *Journal for Research in Mathematics Education*, *7*(5), 324–326.

https://doi.org/10.2307/748467

Floden, R. E. (2020). Philosophical issues in education policy research. In *The state of*

education policy research (pp. 3-15). Routledge, London, UK.

https://doi.org/10.4324/9781003064466

Fuchs, D., Fuchs, L. S., & Abramson, R. (2020). Peer-assisted learning strategies (PALS): A

validated classwide program for improving reading and mathematics

performance. *Student Engagement: Effective Academic, Behavioral, Cognitive, and*

Affective Interventions at School, 109-120. https://doi.org/10.1007/978-3-030-37285-

9_6

Fuchs, L., Fuchs, D., Seethaler, P. M., & Barnes, M. A. (2020). Addressing the role of

working memory in mathematical word-problem solving when designing intervention

for struggling learners. *ZDM*, *52*(1), 87-96. https://doi.org/10.1007/s11858-019-01070-

8

Gachau, S., Njagi, E. N., Owuor, N., Mwaniki, P., Quartagno, M., Sarguta, R., English, M., & Ayieko, P. (2021). Handling missing data in a composite outcome with partially observed components: Simulation study based on clustered paediatric routine data. *Journal of Applied Statistics*, 1-14. https://doi.org/10.1080/02664763.2021.1895087

Guill, K., Lüdtke, O., & Schwanenberg, J. (2020). A two-level study of predictors of private tutoring attendance at the beginning of secondary schooling in Germany: The role of individual learning support in the classroom. *British Educational Research Journal*, 46(2), 437-457. https://doi.org/10.1002/berj.3586

Hajar, A. (2020). The association between private tutoring and access to grammar schools: Voices of Year 6 pupils and teachers in south-east England. *British Educational Research Journal*, *46*(3), 459-479. https://doi.org/10.1002/berj.3587

Hayes, A. F., & Rockwood, N. J. (2020). Conditional process analysis: Concepts, computation, and advances in the modeling of the contingencies of mechanisms. *American Behavioral Scientist*, *64*(1), 19–54. https://doi.org/10.1177/0002764219859633

Herber, S. P., Quis, J. S., & Heineck, G. (2017). Does the transition into daylight saving time affect students' performance? *Economics of Education Review*, *61*, 130–139. https://doi.org/10.1016/j.econedurev.2017.07.002

Heron, T. E., Villareal, D. M., Yao, M., Christianson, R. J., & Heron, K. M. (2006). Peer tutoring systems: Applications in classroom and specialized environments. *Reading & Writing Quarterly, 22*(1), 27-45. https://doi.org/10.1080/10573560500203517

Hickey, A. J., & Flynn, R. J. (2019). Effects of the TutorBright tutoring programme on the reading and mathematics skills of children in foster care: A randomised controlled

trial. *Oxford Review of Education*, *45*(4), 519–537.

https://doi.org/10.1080/03054985.2019.1607724

Hoffman, A. S., Hamm, J. V., Lambert, K., & Meece, J. (2020). Peer-collaborator preferences

of African American and white sixth-graders in mathematics classrooms. *The Journal*

of Experimental Education, 88(4), 559-577.

https://doi.org/10.1080/00220973.2019.1615857

Hoorfar, H., & Taleb, Z. (2015). Correlation between mathematics anxiety with

metacognitive knowledge. *Procedia-Social and Behavioral Sciences, 182*, 737-741.

https://doi.org/10.1016/j.sbspro.2015.04.822

House, J. D. (2009). Mathematics beliefs and achievement of a national sample of Native

American students: Results from the Trends in International Mathematics and Science

Study (TIMSS) 2003 United States assessment. *Psychological Reports*, *104*(2), 439-

446. https://doi.org/10.2466/PR0.104.2.439-446

Hutchison, D. (2009). Designing your sample efficiently: clustering effects in education

surveys. *Educational Research, 51*(1), 109-126.

https://doi.org/10.1080/00131880802704806

Johns, C., & Mills, M. (2021). Online mathematics tutoring during the COVID-19 pandemic:

Recommendations for best practices. *Primus*, *31*(1), 99-117.

https://doi.org/10.1080/10511970.2020.1818336

Jong, C., Hodges, T. E., Royal, K. D., & Welder, R. M. (2015). Instruments to Measure

Elementary Preservice Teachers' Conceptions: An Application of the Rasch Rating

Scale Model. Educational Research Quarterly, 39(1), 21-48.

Kalkbrenner, M. T. (2021). Alpha, omega, and H internal consistency reliability estimates:

Reviewing these options and when to use them. *Counseling Outcome Research and*

Evaluation, 1-12. https://doi.org/10.1080/21501378.2021.1940118

Keshavarzi, A., & Ahmadi, S. (2013). A comparison of mathematics anxiety among students by gender. *Procedia-Social and Behavioral Sciences, 83*, 542-546. https://doi.org/10.1016/j.sbspro.2013.06.103

Kyttälä, M., Björn, P. M., Rantamäki, M., Lehesvuori, S., Närhi, V., Aro, M., & Lerkkanen, M. K. (2022). Assessment conceptions of Finnish pre-service teachers. *European Journal of Teacher Education*, 1-19. https://doi.org/10.1080/02619768.2022.2058927

Lavrijsen, J., & Verschueren, K. (2020). Student characteristics affecting the recognition of high cognitive ability by teachers and peers. *Learning and Individual Differences, 78*, Article 101822. https://doi.org/10.1016/j.lindif.2019.101820

Lazarides, R., & Buchholz, J. (2019). Student-perceived teaching quality: How is it related to different achievement emotions in mathematics classrooms? *Learning and Instruction*, *61*, 45–59. https://doi.org/10.1016/j.learninstruc.2019.01.001

Leung, K. C. (2019a). An updated meta-analysis on the effect of peer tutoring on tutors' achievement. *School Psychology International*, *40*(2), 200–214. https://doi.org/10.1177/0143034318808832

Leung, K. C. (2019b). Compare the moderator for pre-test-post-test design in peer tutoring with treatment-control/comparison design. *European Journal of Psychology of Education*, *34*, 685–703. https://doi.org/10.1007/s10212-018-00412-6

Losinski, M. L., Ennis, R. P., Sanders, S. A., & Nelson, J. A. (2019). A meta-analysis examining the evidence-base of mathematical interventions for students with emotional disturbances. *The Journal of Special Education*, *52*(4), 228–241. https://doi.org/10.1177/0022466918796200

Malone, A. S., & Fuchs, L. S. (2014). Comparing the contribution of teacher versus tutor ratings of inattentive behavior in predicting mathematics achievement. *Remedial and Special Education, 35*(6), 378-386. https://doi.org/10.1177/0741932514539855

Marbán, J. M., Palacios, A., & Maroto, A. (2021). Enjoyment of teaching mathematics among pre-service teachers. *Mathematics Education Research Journal,* 33(3), 613-629. https://doi.org/10.1007/s13394-020-00341-y

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist*, 20(3), 107–123. https://doi.org/10.1207/s15326985ep2003_1

McMaster, K. L., & Fuchs, D. (2016). Classwide intervention using peer-assisted learning strategies. In *Handbook of response to intervention* (pp. 253-268). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7568-3_15

McGowen, M. A., & Davis, G. E. (2020). Spectral analysis of concept maps of high and low gain undergraduate mathematics students. *The Journal of Mathematical Behavior, 55,* Article 100686. https://doi.org/10.1016/j.jmathb.2019.01.002

Moeyaert, M., Klingbeil, D. A., Rodabaugh, E., & Turan, M. (2020). Three-level meta-analysis of single-case data regarding the effects of peer tutoring on academic and social-behavioral outcomes for at-risk students and students with disabilities. *Remedial and Special Education*. https://doi.org/10.1177/0741932519855079

Moliner, L., & Alegre, F. (2022). Attitudes, beliefs and knowledge of mathematics teachers regarding peer tutoring. *European Journal of Teacher Education*, 45(1), 93-112. https://doi.org/10.1080/02619768.2020.1803271

Moliner, L., & Alegre, F. (2020a). Effects of peer tutoring on middle school students' mathematics self-concepts. *PLOS One, 15*(4), Article e0231410. https://doi.org/10.1371/journal.pone.0231410

Moliner, L., & Alegre, F. (2020b). Peer tutoring effects on students' mathematics anxiety: A middle school experience. *Frontiers in Psychology, 11*, Article 1610. https://doi.org/10.3389/fpsyg.2020.01610

Moliner, O., Lozano, J., Aguado, T., & Amiama, J. (2021). Building inclusive and democratic schools in Spain: strategies for mobilising knowledge on inclusive education through participatory action research. *International Journal of Inclusive Education*, 1-18. https://doi.org/10.1080/13603116.2021.1956604

Neu, J. A., & Greer, R. D. (2019). Fifth graders learn math by observation faster when they observe peers receive corrections. *European Journal of Behavior Analysis*, *20*(1), 126–145. https://doi.org/10.1080/15021149.2019.1620044

Patchan, M. M., Rambo-Hernandez, K. E., Deitz, B. N., & McNeill, J. (2022). Using peer assessment to improve middle school mathematical communication. *The Journal of Educational Research*, 1-15. https://doi.org/10.1080/00220671.2022.2074948

Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research*, *86*(1), 207–236. https://doi.org/10.3102/0034654315582067

Powell, S. R., & Fuchs, L. S. (2015). Peer-assisted learning strategies in mathematics. In *The power of peers in the classroom: Enhancing learning and social skills* (pp. 188-223). Guilford Press, New York, NY.

Puza, B., & Bonfrer, A. (2018). A series of two-urn biased sampling problems. *Communications in Statistics-Theory and Methods*, *47*(1), 80–91. https://doi.org/10.1080/03610926.2017.1300282

Ren, L., Green, J. L., & Smith, W. M. (2016). Using the Fennema-Sherman mathematics attitude scales with lower-primary teachers. *Mathematics Education Research Journal, 28*(2), 303-326. https://doi.org/10.1007/s13394-016-0168-0

Rittle-Johnson, B., Farran, D. C., & Durkin, K. L. (2020). Marginalized students' perspectives on instructional strategies in middle-school mathematics classrooms. *The Journal of Experimental Education*, 1-18. https://doi.org/10.1080/00220973.2020.1728513

Shenderovich, Y., Thurston, A., & Miller, S. (2016). Cross-age tutoring in kindergarten and elementary school settings: A systematic review and meta-analysis. *International Journal of Educational Research*, *76*, 190-210. https://doi.org/10.1016/j.ijer.2015.03.007

Snead, S. L., Walker, L., & Loch, B. (2022). Are we failing the repeating students? Characteristics associated with students who repeat first-year university mathematics. *International Journal of Mathematical Education in Science and Technology*, 53(1), 227-239. https://doi.org/10.1080/0020739X.2021.1961899

Song, Y., Loewenstein, G., & Shi, Y. (2018). Heterogeneous effects of peer tutoring: Evidence from rural Chinese middle schools. *Research in Economics*, *72*(1), 33–48. https://doi.org/10.1016/j.rie.2017.05.002

Tadesse, T., Gillies, R. M., & Manathunga, C. (2020). Shifting the instructional paradigm in higher education classrooms in Ethiopia: What happens when we use cooperative learning pedagogies more seriously? *International Journal of Educational Research*, *99*, Article 101509. https://doi.org/10.1016/j.ijer.2019.101509

Tan, S. C., Looi, C. K., Cheung, Y. L., Chung, S. H., Lim, S. J., & Wong, W. H. (2021). Designing and evaluating a mobile peer tutoring application: a cultural historical activity theory approach. *Interactive Learning Environments*, 1-12. https://doi.org/10.1080/10494820.2021.1878232

Taut, S., & Rakoczy, K. (2016). Observing instructional quality in the context of school evaluation. *Learning and Instruction*, *46*, 45–60. https://doi.org/10.1016/j.learninstruc.2016.08.003

Thurston, A., Cockerill, M., & Craig, N. (2019). Using cooperative learning to close the reading attainment gap for students with low literacy levels for grade 8/year 9

students. *International Journal of Educational Research*, *94*, 1–10. https://doi.org/10.1016/j.ijer.2019.02.016

Thurston, A., Roseth, C., Chiang, T. H., Burns, V., & Topping, K. J. (2020). The influence of social relationships on outcomes in mathematics when using peer tutoring in elementary school. *International Journal of Educational Research Open*, *1*, 100004, 1-8. https://doi.org/10.1016/j.ijedro.2020.100004

Topping, K. (2021). Peer Assessment: Channels of Operation. *Education Sciences*, *11*(3), 91, 1-20. https://doi.org/10.3390/educsci11030091

Topping, K. J. (2022). Digital peer assessment in school teacher education and development: a systematic review. *Research Papers in Education*, 1-27. https://doi.org/10.1080/02671522.2021.1961301

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). *Effective peer learning: From principles to practical implementation*. Routledge, London, UK. https://doi.org/10.4324/9781315695471

Topping, K. J., Miller, D., Murray, P., Henderson, S., Fortuna, C., & Conlin, N. (2011). Outcomes in a randomised controlled trial of mathematics tutoring. *Educational Research*, *53*(1), 51-63. https://doi.org/10.1080/00131881.2011.552239

Topping, K. J. (2020). A theoretical model of intergenerational tutoring: Advancing the field. *Journal of Intergenerational Relationships, 18*(1), 88-105.. https://doi.org/10.1080/15350770.2019.1646182

Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K., & Miller, D. (2011). Improving attainment across a whole district: School reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement*, *22*(3), 265-289. https://doi.org/10.1080/09243453.2011.589859

Tymms, P., Merrell, C., & Bailey, K. (2018). The long-term impact of effective teaching. *School Effectiveness and School Improvement*, *29*(2), 242-261. https://doi.org/10.1080/09243453.2017.1404478

Vesić, D., Džinović, V., & Mirkov, S. (2021). The role of absenteeism in the prediction of math achievement on the basis of self-concept and motivation: TIMMS 2015 in Serbia. Psihologija, 54(1), 15-31. https://doi.org/10.2298/PSI190425010V

Wardat, Y., Belbase, S., & Tairab, H. (2022). Mathematics teachers' perceptions of Trends in International Mathematics and Science Study (TIMSS)-related practices in Abu Dhabi Emirate schools. *Sustainability, 14*(9), 5436. https://doi.org/10.3390/su14095436

Webb, N. M., Franke, M. L., Ing, M., Turrou, A. C., Johnson, N. C., & Zimmerman, J. (2019). Teacher practices that promote productive dialogue and learning in mathematics classrooms. *International Journal of Educational Research*, *97*, 176–186. https://doi.org/10.1016/j.ijer.2017.07.009

Wilkerson, J. B. (2021). Cultivating a productive-disposition toward mathematics by engaging in service-learning. *PRIMUS*, *31*(8), 869-882. https://doi.org/10.1080/10511970.2020.1776803

Wohlkinger, F., Bayer, M., & Ditton, H. (2016). Measuring self-concept in the NEPS. In Methodological issues of longitudinal surveys (pp. 181-193). Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-11994-2_11

Xia, F., George, S. L., Ning, J., Li, L., & Huang, X. (2021). A signature enrichment design with Bayesian adaptive randomization. *Journal of applied statistics, 48*(6), 1091-1110. https://doi.org/10.1080/02664763.2020.1757048

Xiao, Y. (2017). A fast algorithm for two-dimensional Kolmogorov–Smirnov two sample tests. *Computational Statistics & Data Analysis*, *105*, 53–58. https://doi.org/10.1016/j.csda.2016.07.014

Xu, Y., Chen, C. C., Spence, C., Washington-Nortey, M., Zhang, F., & Brown, A. (2021). Supporting young Spanish speaking English learners through teacher scaffolding and reciprocal peer tutoring. *Early Child Development and Care*, 1-13. https://doi.org/10.1080/03004430.2021.1874944

Yanhong, L., Zhenrong, S., & Yiping, Z. (2021). Self-esteem and interpersonal difficulties of college students: Influences of self-control and self-concept. *Journal of Psychology in Africa*, 31(6), 615-621. https://doi.org/10.1080/14330237.2021.2001950

Ye, T., Shao, J., Yi, Y., & Zhao, Q. (2022). Toward better practice of covariate adjustment in analyzing randomized clinical trials. *Journal of the American Statistical Association*, 1-13. https://doi.org/10.1080/01621459.2022.2049278

Zeneli, M., Tymms, P., & Bolden, D. (2016). The impact of interdependent cross-age peer tutoring on social and mathematics self-concepts. *International Journal of Psychology and Educational Studies, 3*(2), 1-13. https://doi.org/10.17220/ijpes.2016.02.001

Zeneli, M., Tymms, P., & Bolden, D. (2018). Interdependent cross-age peer tutoring in mathematics. *International Journal of Psychology and Educational Studies*, *5*(3), 33–50. https://doi.org/10.17220/ijpes.2018.03.004