



## Peer tutoring and mathematics in secondary education: literature review, effect sizes, moderators, and implications for practice



Francisco Alegre<sup>a,\*</sup>, Lidon Moliner<sup>b</sup>, Ana Maroto<sup>a</sup>, Gil Lorenzo-Valentin<sup>b</sup>

<sup>a</sup> Universidad de Valladolid, Spain

<sup>b</sup> Universitat Jaume I, Spain

### ARTICLE INFO

#### Keywords:

Education  
Psychology  
Mathematics  
Review  
Peer tutoring  
Implications for practice  
Secondary education  
Effect sizes

### ABSTRACT

A literature review was undertaken to compile all data on peer tutoring in secondary education (7<sup>th</sup> to 12<sup>th</sup> grade) mathematics from existing articles. Data from 42 independent studies were included in this research. All data regarding participants' roles (fixed vs. reciprocal), participants' ages (same-age vs. cross-age), the methodological approach taken (quantitative or qualitative), the type of design for those studies that involved a quantitative approach, the variables analyzed, and the organizational matters (number of participants, duration of the program, sessions per week, and duration of the sessions) are included in the article. The effect sizes of the 42 studies were calculated and examined. The main goal of the study was to determine those variables that were moderators of effect size, that is, the variables that significantly influenced students' academic achievement outcomes. Inferential statistical analyses (Student's *t*-test and ANOVAs) were carried out for the variables. Of the 42 studies examined, 88% showed positive effect sizes with the means being close to medium (Cohen's  $d = 0.38$ ). Conclusions suggest the implementation of same-age over cross-age tutoring, during programs of fewer than 8 weeks, in sessions of less than 30 minutes is optimal for improving students' academic outcomes. Inclusion of control groups in similar future studies is recommended so effect sizes are not overestimated.

### 1. Introduction

Authors such as Miquel and Duran (2017) indicated that often students can be as helpful as teachers during learning processes for various reasons, such as using more direct speech, sharing cultural references, and having recent knowledge of areas in which their peers may need help. Peer tutoring is one way students can assist each other during learning activities. Topping et al. (2013) defined peer tutoring as students learning from and with each other in a structured way, supervised by a professional researcher or practitioner. Hence, peer tutoring may be regarded as a methodology that fosters collaboration from an inclusive perspective and may also be regarded as an alternative to traditional unidirectional teaching (Thurston et al., 2009).

Topping et al. (2017) defined two different types of tutoring based on the participants' ages: same-age and cross-age tutoring. In a same-age tutoring experience, participants are the same age, while in cross-age tutoring, they are different ages. Considering the roles played by tutoring session participants, two additional classifications of tutoring may be defined: reciprocal and fixed (Thurston and Topping, 2007). During fixed tutoring between two students, one of the pair members consistently

serves as tutor and the other as tutee, while during reciprocal tutoring situations, the students switch roles, serving half of the time as tutor and the other half as tutee. Factors such as participants' ages, experimental design, and duration and frequency of the tutoring sessions may influence the results of fixed and reciprocal experiences (Bowman-Perrott et al., 2016).

Several peer tutoring literature reviews and meta-analyses were conducted over the past few decades, some of which specifically addressed peer tutoring in mathematics. Johnson et al. (1981) analyzed the efficiency of peer tutoring for several subjects from an academic perspective; eighteen mathematics studies were examined as part of their research. The authors computed effect sizes for these studies and found that working in pairs was more effective than interpersonal competition and individual efforts; moreover, they found that cooperation between peers was also better than interpersonal competition. Davidson (1985) conducted a review of small-group learning with a special chapter dedicated to peer tutoring in mathematics. From a qualitative perspective, Davidson concluded that the helping behavior that took place during peer tutoring was highly beneficial for the students academically and socially. Years later, a review by Butler et al. (2001) examined the

\* Corresponding author.

E-mail address: [falegre@uji.es](mailto:falegre@uji.es) (F. Alegre).

benefits of teaching mathematics through peer tutoring involving students with learning disabilities, that is, students who have neurological differences in various mental processes that make it difficult for them to acquire certain skills, mostly in reading and mathematics (Penney, 2018). From a qualitative analysis, Butler et al. (2001) concluded that students with learning disabilities employed cognitive strategies successfully when peer tutoring was implemented. In addition, the authors strongly recommended using this methodology in secondary schools. Leung (2015) conducted a meta-analysis to identify the variables that influenced the academic achievement of both tutors and tutees during peer tutoring experiences; fifteen studies in mathematics were included in his study. Overall, Leung reported medium to moderate effect sizes for peer tutoring studies across several subjects. He suggested that using high achieving tutors had a positive impact on the final outcome of the peer tutoring experience. Specifically in the mathematics field, Alegre-Ansuategui et al. (2018) performed a meta-analysis on peer tutoring in mathematics from early childhood education to college. Positive effect sizes were reported for 88% of the studies when analyzing the academic achievement variable. One of the main conclusions from Alegre-Ansuategui et al. based on their research was that peer tutoring programs conducted outside of school were not as effective as those implemented during school hours. Furthermore, these authors also performed a literature review on peer tutoring in primary education in which they concluded that 91% of the studies reported positive effect sizes and that variables such as the participants' ages or the type of design (quantitative vs. qualitative) did not significantly affect the final outcome of the peer tutoring experiences (Alegre et al., 2018). As noted, most of the reviews and meta-analyses mentioned highlighted the academic benefits of peer tutoring in mathematics. In addition, promising results have also been documented for variables other than academic achievement. For example, different studies have documented improvements in students' anxiety, self-esteem, and attitudes towards mathematics as a result of peer tutoring (Galbraith and Winterbottom, 2011; Knight et al., 2018).

In spite of the abundance of existing literature on peer tutoring in mathematics, implications for practice at the secondary education level have not been addressed thoroughly. Mathematics is a core subject in secondary education (7<sup>th</sup> to 12<sup>th</sup> grade, 12–18 years old) that highly influences students' academic future (McKee and Caldarella, 2016). Cai and Hwang (2019) stated that study methodologies and students' perceptions towards mathematics play pivotal roles in the findings on the academic achievement variable. In this sense, active methodologies that allow students to participate in the learning process are encouraged (Zhao and Ding, 2019). For reasons provided, and given the proven potentiality of peer tutoring in mathematics, it is necessary to provide secondary education practitioners (teachers, researchers, and other relevant parties) with practice guidelines that may help them maximize students' academic outcomes which is the underlying purpose for conducting this meta-analysis.

## 2. Materials and methods

### 2.1. Aim and research questions

The main aim of this research was to determine variables during peer tutoring interventions in mathematics in secondary education (7<sup>th</sup> to 12<sup>th</sup> grade) that significantly impact students' academic outcomes. According to Cooper et al. (2009), a common question when analyzing educational outcomes is whether any study descriptors are associated with the magnitude of the outcome (i.e., whether they are moderators of effect size). Hence, the main goal of this study's investigation was to determine those variables that were moderators of effect size, that is, variables that significantly influenced students' academic achievement outcomes. The following variables were analyzed: (1) participating students' ages (cross-age tutoring vs. same-age tutoring); (2) students' roles (reciprocal tutoring vs. fixed tutoring); (3) students' skills (students with learning

disabilities vs. students without learning disabilities); (4) type of design (quantitative vs. qualitative); (5) type of quantitative design (pretest-posttest with control group vs. posttest only with control group, pretest-posttest without control group, and multiple baseline); (6) length of time of the tutoring program; (7) frequency of sessions; (8) length of time of the individual tutoring sessions; (9) total number of participants in the tutoring experience; and (10) type of publication. Hence, the main questions in this article referring to mathematics peer tutoring in secondary education are the following:

Q1. Does peer tutoring report considerable effect sizes for academic achievement?

Q2. Are there any statistically significant differences within the studied variables regarding effect sizes for academic achievement?

### 2.2. Literature review

Google Scholar was used to collect all studies examined in this research. The following keywords were used: "peer tutoring," "mentoring," "peer-mediated instruction," "peer-assisted learning," "mathematics," "math," "problem-solving," "arithmetic," "geometry," and "algebra." Descriptors including "secondary education" or "high school" were not included, as this may have resulted in the unnecessary exclusion of articles potentially suitable for this research. A Boolean format search (Bozzano et al., 2006) was performed combining the first four keywords with the last six keywords. The search was performed so that keywords had to appear together in the article title. The great majority of the studies collected through this first search did not address peer tutoring in mathematics and were, therefore, excluded. Prior meta-analyses and literature reviews in the field of peer tutoring mentioned previously were also consulted in order to corroborate that the keywords used in the first search had been appropriate. Articles in both English and Spanish were included. Moreover, as Cooper et al. (2009) stated, there is a great amount of "grey" or "fugitive" literature, that is, research not published in journals (e.g., doctoral dissertations and reports presented at conferences) that should also be included when performing a literature review. Indeed, having more studies increases the statistical power of meta-reviews. Hence, these types of publications were also included. Studies coming from the grey literature that did not meet quality indicators in educational research as defined by Gersten et al. (2005) and Gogolin (2016) were excluded so that sufficient experimental rigor was ensured in this review. According to both Gersten et al. and Gogolin, among other issues, every article must clearly indicate the method of data collection and the way missing data is handled or how attrition of participants occurs (if it takes place), and the instruments used to collect the data must be clearly defined. Thus, after excluding papers that did not meet these criteria, 143 articles remained to be filtered with the last 3 selection criteria, which are discussed in the following paragraph.

### 2.3. Selection criteria

As Cooper et al. (2009) indicated, the selection criteria should flow naturally from the question or objective of the research project. As such, the following selection criteria were used as a final filter for the inclusion of articles in the review.

The first selection criterion was to accept only those articles where the interaction between the tutor and tutee was direct and human-to-human (Topping, 2005). We did not consider those where one tutor helped various tutees, but looked only for tutoring dyads. Due to this criterion, 22 studies were excluded.

The second criterion was excluding from consideration any tutoring experiences that involved aid offered by parents or post-university personnel, since this type of aid comes from a person who violates the concept of "peer" in the context of tutoring (Thurston and Topping, 2007). This criterion excluded 15 additional studies.

The third criterion was that the tutees had to be secondary education students, that is, from 7th to 12th grade. Another 64 studies were

excluded due to this criterion. After applying these three criteria, 42 studies remained.

#### 2.4. Calculation of effect size

Effect sizes were calculated for those articles with enough quantitative or qualitative information. Although authors in the field of educational research differ on the suitability of including qualitative studies in meta-analyses (Bangert-Drowns and Rudner, 1991), we included qualitative studies for two main reasons. First, we did not want to exclude results from these studies, as relevant information could be omitted (Levitt et al., 2018). Second, following their recent literature review on primary education mentioned previously, Alegre et al. (2018) concluded that effect sizes for both qualitative and quantitative studies were very similar overall, so the final outcome of the review was not affected by the inclusion of the qualitative studies. For all studies in this review, Cohen's *d* (Rosenthal et al., 1994) was employed as a measure of the magnitude. This parameter was chosen since it is easy to interpret by researchers in the field (Rosnow and Rosenthal, 1996) and because it is standardized or scale free (LeCroy and Krysik, 2007). Several formulas were used, depending on the design employed in each article, when calculating Cohen's *d*. A description for each case is provided in the following paragraphs.

For those articles with a pretest-posttest control group design in which the standard deviations for the posttest or the pretest were included, the formula indicated by Fritz et al. (2012) was used. Standard deviations of the pretests were taken when possible because this is recommended, according to several authors, to avoid as much as possible the biases that may take place (Orwin, 1983). This was the process and formula used for 29 articles examined in this study, as many of them provided sufficient statistical information. If only Student's *t*-test, Fisher's exact test, or Pearson's *r* and the sample size of the control group or the experimental group were given, formulas indicated by Olive and Smith (2005) were used.

For those articles noting posttest only with control or pretest-posttest single group designs, calculations of effect sizes were performed following the indications provided by Valente and MacKinnon (2017). These authors suggested using only the standard deviation of the control group for posttest only with control designs, as it should not be influenced. In this sense, pretest standard deviations for pretest posttest single group designs are also recommended, as they should be less influenced than the posttest standard deviations.

For articles including multiple baseline or other designs, effect sizes were calculated following the indications by Hedges et al. (2013). According to these authors, effect sizes calculated for studies with a multiple baseline or analogous design may differ significantly from other studies with different experimental designs. Therefore, a statistical adjustment is needed so that no bias takes place due to differences in the experimental design. Hence, one of their proposed models for adjustment (Model MB1: Varying intercepts, fixed treatment effect, no trends) was used.

For those studies with a qualitative design, calculation of effect sizes was conducted as Onwuegbuzie (2003) suggested. This author suggested that dichotomous answers in qualitative studies ("yes" and "no," for example) can be quantified using binary numbers (0 and 1). The same theory applies to those questions in which multiple answers can be given, as many of these answers can be quantized in Likert-format scales.

#### 2.5. Effect size descriptors

Effect sizes were classified according to magnitude following the values indicated by Rosenthal (1996). Hence, if Cohen's *d* effect size is positive but not greater than 0.2, it may be considered very small. When the effect is greater than 0.2 but not greater than 0.5, it may be considered small to medium. When it is higher than 0.5 but not higher than 0.8, it can be classified as medium to large. Finally, any effect size greater

than 0.8 may be considered large.

#### 2.6. Inferential statistical analysis

In order to analyze which variables significantly influenced students' academic outcomes, inferential statistical analyses were carried out on each variable. The statistical procedure followed was the same that had been used in previous literature reviews in the field (Stenhoff & Lignugaris/Kraft, 2007; Alegre et al., 2018). In these studies, the statistics used to analyze the data were sufficient to produce the results, as differences in the effect sizes were determined for several variables, the magnitude of the effect size was clearly identified, and it was reported to what point those variables influenced the academic outcome of the peer tutoring experiences. Moreover, using the same statistical procedure is recommended, as researchers in the field may find it easier to compare results from different reviews (Entwistle et al., 2000). Hence, for those cases in which only two groups were compared, Student's *t*-test ( $p < .05$ ) was employed (De Winter, 2013). For those cases in which more than two groups were compared, analysis of variance (ANOVAs) were carried out. If the analysis reported statistically significant differences among groups, Scheffe's test was used to complement the analysis (Brown and Forsythe, 1974), using SPSS version 25.

### 3. Results

Studies obtained through the literature review are presented in Table 1 in alphabetical order and in single rows according to the first last name of the principal author. The following variables were included in this category; in parentheses are the legends used for each variable in Table 1:

- 1 Age difference among the participating students (*A*), that is, whether it is cross-age (*C*) or same-age (*S*) tutoring;
- 2 Maintenance of the student roles (*R*), that is, whether the tutoring is fixed (*F*) or reciprocal (*R*);
- 3 Skills of the participants (*D*), that is, whether disabled students are included (*Y*) or not included (*N*) in the study;
- 4 Study design (*DES*)—in this case, whether it is qualitative (*QL*) or quantitative (*QN*) and also whether it was a pretest posttest with control group (*PFCG*), a posttest only with control group (*POWC*), a pretest posttest without control group (*PPSG*), a repetition of measurements in the group itself (*MB*), or other (*O*).
- 5 Sample size, that is, number of participants in the study (*N*).
- 6 Length of the peer tutoring program (*LP*) in weeks.
- 7 Frequency of sessions, that is, number of peer tutoring sessions per week (*F*).
- 8 Length of the peer tutoring sessions (*LS*) in minutes.
- 9 Type of publication (*I*), that is, whether the study is described in a research article published in a high-indexed journal (*Y*) or not (*N*, grey literature).

Where no information could be provided, the abbreviation for *not available* (*NA*) was used.

Effect sizes for the 42 independent studies were calculated. Results are presented at the same time research questions are answered.

Q1. Does peer tutoring report considerable effect sizes for academic achievement?

A Cohen's *d* means the effect size of 0.38 with a standard deviation of 0.33 were reported. The median effect size was 0.35. Effect sizes were negative for five studies (12%), very small for eight studies (19%), small to medium for eleven studies (26%), medium to large for another eleven studies (26%), and large for seven studies (16%).

Q2. Are there any statistically significant differences within the studied variables regarding effect sizes?

The statistical analyses for the different variables are shown in Tables 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, and 14. Mean ( $\bar{x}$ ), standard

**Table 1**  
Literature review data.

Authors	A	R	D	DES	N	LP	F	LS	I
Adamson and Lewis (2017)	S	F	Y	QN & MB	10	4	5	90	Y
Alegre Ansuategui and Moliner Miravet (2017)	S	F	N	QL & PPSG	18	5	2	25	Y
Allsopp (1997)	S	R	Y	QN & PPCG	262	5	3	NA	Y
Austin (2008)	S	F	N	QL & PPSG	30	8	1	30	N
Azcoitia (1989)	C	F	Y	QN & PPSG	180	20	5	60	N
Blackbourn and Blackbourn (1993)	C	F	Y	QL & MB	2	NA	NA	NA	Y
Burley et al. (1994)	S	F	Y	QN & MB	2	3	5	20	N
Cairo and Craig (2005)	C	F	N	QN & PPSG	110	3	3	50	Y
Calhoon and Fuchs (2003)	S	R	Y	QL & PPSG	92	15	2	30	Y
Collins and Calevro (1974)	S	F	Y	QN & MB	9	12	5	30	N
Collins and Onwuegbuzie (2001)	C	F	Y	QN & PPSG	40	16	4	15	Y
Cooper (2016)	C	F	N	QN & O	78	7	NA	NA	N
Dufrene et al. (2005)	S	F	N	QN & MB	36	4	5	10	Y
Durand (2008)	S	R	N	QN & POWC	17	4	5	10	N
Early (1999)	S	F	N	QN & O	289	2	5	NA	N
Hannah (2008)	C	F	N	QN & PPSG	92	12	2	55	N
Hilo (1974)	C	F	N	QN & PPSG	184	18	5	20	N
Heintz (1975)	C	F	Y	QN & PPSG	1,370	54	3	90	Y
Hendrickson (1981)	S	F	Y	QN & O	2,386	90	NA	NA	Y
Holecek (2012)	C	F	N	QN & PPSG	6	6	5	30	Y
Ivory (2007)	C	R	Y	QL & PPSG	22	15	3	60	Y
Kane and Alley (1980)	C	F	Y	QL & PPSG	21	8	5	45	Y
Lazarus (2014)	S	R	Y	QN & PPSG	104	6	NA	NA	Y
Leal and Olivas (2014)	S	F	N	QL & O	6	NA	NA	NA	N
Mayfield and Vollmer (2007)	C	R	Y	QN & MB	2	5	5	25	Y
Mohan (1972)	C	F	N	QN & POWC	6	12	2	60	N
Mulvaney (1993)	S	F	N	QL & O	2	NA	NA	NA	Y
Murugan (2015)	S	F	N	QN & PPSG	60	12	NA	60	N
Nazzal (2002)	C	F	Y	QL & PPSG	58	6	5	25	Y
Nesselrodt and Alger (2005)	C	F	Y	QN & PPSG	71	54	3	60	Y
Novotni (1985)	C	F	Y	QN & PPSG	61	54	3	45	Y
Obidoa and Onwubolu (2013)	S	F	Y	QN & PPSG	505	54	NA	NA	Y
Pyle (2015)	S	F	Y	QL & MB	6	22	5	NA	N
Roach et al. (1983)	S	F	Y	QN & PPSG	56	8	NA	NA	N
Schneck (2010)	C	F	Y	QN & PPSG	99	54	1	40	Y
Schloss et al. (1997)	C	R	Y	QN & MB	6	4	4	30	Y
Sinha et al. (2015)	S	R	N		24	5	1	90	Y

**Table 1 (continued)**

Authors	A	R	D	DES	N	LP	F	LS	I
				QN & PPSG					
Walker (2007)	S	F	N	QL & O	18	4	3	90	Y
Worley and Naresh (2014)	S	F	N	QL & O	28	54	1	NA	Y
Zeneli et al. (2016b)	C	F	N	QN & PPSG	550	6	NA	40	Y

**Table 2**  
Cross-age vs. same-age effect sizes.

Same-age			Cross-age			Same-age vs. cross-age	
$\bar{x}$	$\sigma$	$n$	$\bar{x}$	$\sigma$	$n$	Difference (%)	Student's $t$ -test
0.61	0.29	23	0.21	0.25	19	0.30 (190.91)	3.40 ( $p < .01$ )*

**Table 3**  
Fixed vs. reciprocal effect sizes.

Fixed			Reciprocal			Fixed vs. reciprocal	
$\bar{x}$	$\sigma$	$n$	$\bar{x}$	$\sigma$	$n$	Difference (%)	Student's $t$ -test
0.33	0.31	33	0.50	0.38	9	0.17 (51.06%)	1.06 ( $p = .30$ )

deviation ( $\sigma$ ), and number of studies ( $n$ ) are included for each group. Mean differences between groups are also reported with the calculated percentage in parenthesis. Student's  $t$ -tests with their  $p$  values are reported. Degrees of freedom ( $df$ ), sum of squares ( $SS$ ), mean squares ( $MS$ ), and  $F$  ratio with its  $p$  value ( $F$ ) are reported for the ANOVAs. For those cases in which significant statistical differences were reported an asterisk was included after the  $p$  value.

The following variables reported statistically significant differences: students' ages, ages and roles combined, disability/at-risk condition, duration of the program, and duration of the sessions. No statistically significant differences were found for roles of the participants, number of participants, or type of design.

#### 4. Discussion

The global effect size reported in this manuscript is similar though somewhat lower than those reported in other reviews and meta-analyses. Authors such as Cohen et al. (1982), Leung (2018), Zeneli et al. (2018), and Alegre-Ansuategui et al. (2018) documented medium to large effect sizes in their manuscripts. This difference, according to some authors, may be attributable to a decrease in motivation and attitude towards mathematics during secondary education due to previous failure experiences (León et al., 2015). Hence, a decrease in students' attitude may affect the final outcome of the peer tutoring intervention (Topping et al., 2004; Zeneli and Tymms, 2015), which would explain the global effect size difference compared with the works of Alegre-Ansuategui et al. (2018) and Alegre et al. (2018). Regardless of these details, research has shown once again that peer tutoring usually has a positive effect on the academic performance of both the tutors and tutees (Shenderovich et al., 2016).

The fact that statistically significant differences were found when analysing cross-age vs. same-age tutoring is not consistent with prior reviews or meta-analyses mentioned in this study, as most of those reported findings that both are equally effective. Previous studies performed by Topping (1996, 2005) in primary education and college have not found any statistically significant differences between them. However, an important consideration is that, due to organizational issues, same-age tutoring usually takes place within the classroom, so usually students working in pairs previously know each other. Tymms et al. (2011) stated that confidence in each other is a key element when it

**Table 4**  
Ages and roles combined.

Age	Role	n	$\bar{x}$	SD		df	SS	MS	F (p)
Same	Fixed	15	0.62	0.29	Between Groups	3	0.87	0.29	3.75 (p = .03)*
Same	Reciprocal	6	0.60	0.33	Within groups	17	1.31	0.08	
Cross	Fixed	16	0.23	0.25	Total	20	2.17		
Cross	Reciprocal	3	0.10	-					

**Table 5**  
Scheffe's test for ages and roles comparison.

	Mean differences	Significance level
Same-age fixed vs. same-age reciprocal	0.02	.99
Same-age fixed vs. cross-age fixed	0.39	.04*
Same-age reciprocal vs. cross-age fixed	0.37	.04*

**Table 6**  
Students with learning disabilities vs. students without learning disabilities.

Students with learning disabilities			Students without learning disabilities			Students with learning disabilities vs. students without learning disabilities	
$\bar{x}$	$\sigma$	n	$\bar{x}$	$\sigma$	n	Difference (%)	Student's t-test
0.23	0.24	20	0.62	0.32	22	0.32 (170.86)	2.98 (p = .01)*

comes to peer tutoring. Hence, it is possible that students in the same-age studies had higher confidence in their peers than those in the cross-age studies, influencing the final outcome (Hodgson et al., 2015).

For the comparison of fixed vs. reciprocal tutoring, the similarity of effect size between these two types, is consistent with previous meta-analyses and reviews. As Duran and Monereo (2005) stated, the superiority of reciprocal over fixed peer tutoring has yet to be proved. Hence, the result obtained in this study when comparing same-age fixed with same-age reciprocal peer tutoring is more logical and expectable, taking into account previous research.

The comparison of effect sizes for students with learning disabilities and students without learning disabilities was also consistent with previous research. Academic outcomes for students with learning disabilities

**Table 7**  
Low vs. high sample size.

≤ 30 participants			>30 participants			Low vs. high number of participants	
$\bar{x}$	$\sigma$	n	$\bar{x}$	$\sigma$	n	Difference (%)	Student's t-test
0.42	0.50	19	0.37	0.30	23	0.05 (12.93)	0.19 (p = .86)

**Table 8**  
Low vs. high length of time of the tutoring experience.

≤ 8 weeks			>8 weeks			Low vs. high length of time of the program	
$\bar{x}$	$\sigma$	n	$\bar{x}$	$\sigma$	n	Difference (%)	Student's t-test
0.52	0.32	15	0.29	0.31	22	0.23 (80.06%)	2.04 (p = .04)*

**Table 9**  
Analysis of different categories of length of the tutoring experience.

Number of weeks	n	$\bar{x}$	$\sigma$		df	SS	MS	F (p)
≤4	7	0.63	0.49	Between Groups	3	0.33	0.11	1.02 (p = .41)
5–8	9	0.46	0.33	Within groups	17	1.84	0.11	
9–12	8	0.32	0.32	Total	20	2.17		
>12	13	0.25	0.22					

**Table 10**  
Three categories of sessions per week comparison.

Sessions per week	n	$\bar{x}$	$\sigma$		df	SS	MS	F (p)
1	2	0.48	0.31	Between Groups	2	0.39	0.19	2.67 (p = .12)
2	10	0.37	0.12	Within groups	10	0.72	0.07	
3	12	0.31	0.21	groups				
4	3	0.29	0.24	Total	12	1.11		
5	12	0.13	0.36					

**Table 11**  
Low (≤30 minutes) vs. high (>30 minutes) length of time of the tutoring session.

≤ 30 minutes			>30 minutes			Low vs. high length of time of the sessions	
$\bar{x}$	$\sigma$	n	$\bar{x}$	$\sigma$	n	Difference (%)	Student's t-test
0.55	0.28	16	0.26	0.35	19	0.30 (113.89%)	2.08 (p = .03)*

are often lower than for their non-learning-disabled peers in tutoring experiences (Kamps et al., 1994; Cook et al., 2017). Previous studies also highlighted the majority of positive effect sizes with disabled students in peer tutoring, so academic improvement can also be achieved by means of this methodology with students with disabilities (Zeneli et al., 2016a; Losinski et al., 2017).

When analyzing time-related variables, statistically significant differences were found for the duration of the sessions and the duration of the program. Those studies with programs that lasted more than 8 weeks or sessions that lasted more than 30 minutes reported lower effect sizes. Although no statistically significant differences were found when analyzing the number of sessions per week, it is evident in Table 10 that the more sessions per week, the lower the effect size will be. These results are consistent with previous theoretical and empirical studies. Meta-

**Table 12**  
PPCG design vs. POWC, PPSG, MB and other designs.

PPCG			POWC, PPSG, MB, O			PPCG vs. POWC, PPSG, MB & RM	
$\bar{x}$	$\sigma$	n	$\bar{x}$	$\sigma$	n	Difference (%)	Student's t-test
0.31	0.29	16	0.60	0.38	26	0.29 (94.19%)	2.05 (p = .04)*

**Table 13**  
Quantitative vs. qualitative designs.

Quantitative			Qualitative			Quantitative vs. qualitative	
$\bar{x}$	$\sigma$	n	$\bar{x}$	$\sigma$	n	Difference (%)	Student's t-test
0.35	0.39	30	0.44	0.38	12	0.09 (25.71%)	1.45 (p = .29)

**Table 14**  
Articles published in high-indexed journals vs grey literature articles.

High-indexed journal			Grey literature			High-indexed journal vs. grey literature	
$\bar{x}$	$\sigma$	n	$\bar{x}$	$\sigma$	n	Difference (%)	Student's t-test
0.39	0.41	26	0.36	0.37	16	0.03 (8.33%)	0.24 (p = .81)

analyses by Leung (2015) and Alegre-Ansuategui et al. (2018) reported similar effect size differences in these variables. According to several authors, if peer tutoring is too intense or too long, that is, long programs, long sessions, or too many sessions per week, students get tired and bored with this methodology. Hence, the final outcome is affected, and peer tutoring is not as effective as it should be (De Smet et al., 2010; Topping et al., 2011a).

The statistical significance found when comparing experimental designs is also supported by previous research. Meta-analyses by Zeneli et al. (2016a, b) and Leung (2019) concluded that experimental designs different from PPCG influenced the magnitude of the effect size for peer tutoring across several educational levels. They stated that the absence of a control group may overestimate the final outcome reported in peer tutoring studies.

The fact that no differences were found based on the number of participants in the study is consistent with prior meta-analyses previously mentioned. As Topping et al. (2011b) stated, the efficiency of peer tutoring should not be affected by the number of participants if the implementation of the program is properly supervised or conducted by someone with enough peer tutoring knowledge and expertise.

The comparison of effects between qualitative and quantitative studies in peer tutoring has not been studied yet. The authors of this manuscript could not find any previous reviews or meta-analyses on peer tutoring focused on this issue. Topping et al. (2000) stated that similar conclusions were drawn for post-graduate students when performing different quantitative and qualitative analysis in their peer tutoring experiences. Several authors in educational research addressing quantitative, qualitative, and mixed approaches stated that the treatment and results of qualitative research are parallel to those of quantitative research, so it would make sense to report similar effect sizes, regardless of the approach employed in the study (Bryman, 2006; Johnson and Christensen, 2008).

Effect sizes from studies published in high-indexed journals were very similar to those found in sources other than published, high-indexed journals (i.e., those found in grey literature). Hence, inclusion of the grey literature in this research was justified, as the final outcome was not influenced by the type of studies (Cheung and Slavin, 2016); the previous review on primary education performed by Alegre et al. (2018) concluded the same regarding grey literature and peer tutoring in mathematics.

#### 4.1. Limitations

Several limitations must be considered in this research when interpreting the reported conclusions and results. First, some of the designs reported in the 42 studies examined may be defined as “weak designs,” that is, studies without a control group (PPSG), with no pretest (POWC) or with multiple baselines (MB) Moss and Yeaton (2006). Hence, research validity may be affected from an experimental design perspective. The fact that these types of studies were included is justified by the fact that relevant information may have been omitted if they had been excluded (Maxwell, 2004). Furthermore, some of the 42 articles belong to what many authors consider “grey literature” (Mahood et al., 2014), that is, research not published in high-indexed journals. This issue may also compromise the validity of this research. Furthermore, although 42 is a considerable number, under no circumstances can it be regarded as a significantly large sample. Having only 42 studies with which to make effect size comparisons seriously limited this study to the point that some variables couldn't be statistically analyzed in depth. For instance, there were almost no effect sizes for studies with 1 or 4 days a week of peer tutoring sessions and there were only six same-age reciprocal and three cross-age reciprocal studies. Moreover, it is common that studies many times are only published or reported if they have positive outcomes. This must also be considered, as the limitation of the literature pool may be partially biased towards studies that show positive effects of peer tutoring (Norris, 1997). Finally, as the search of articles only included studies

published in English or Spanish, it is thought that many other studies were excluded for linguistic reasons.

#### 4.2. Recommendations for future research

Considering that the documentation of studies and meta-analyses regarding peer tutoring has increased in the last few years, repetition of this study is recommended in the future, as many more studies will be published in the coming years. The existence of a higher number of articles and the possibility of employing more rigorous filters to select the articles will facilitate the work for future studies, including meta-analyses and reviews, and result in higher experimental validity. Research on peer tutoring in less studied educational stages, such as early childhood or higher education, may also be a substantial field of study. Although primary and secondary education own the majority of studies on peer tutoring, the number of studies on peer tutoring in higher education has been increasing recently. Hence, it could be interesting to compare the outcomes in other contexts or with other educational levels. Analysis of psychological or attitudinal variables other than academic achievement is also recommended, given the promising results shown in several studies. Inclusion of studies coming from grey literature is also recommended, as overall they seem to report similar effect sizes to those studies coming from high-indexed journals.

#### 4.3. Implications for practice

Same-age tutoring is recommended over cross-age based on the high difference of effect sizes shown in this study. A shorter length for tutoring programs (less than 8 weeks), briefer sessions of less than 30 minutes, and three or fewer peer tutoring sessions per week are also suggested. Having students with learning disabilities in class should not prevent practitioners from implementing peer tutoring, as academic improvements can be found regardless of the students' learning abilities. Moreover, a PPCG experimental design is recommended for several reasons. First, it will enable a more realistic measure of the effect size, and results will not be overestimated. Also, the work will be more suitable to be included in high-indexed journals, reviews, and meta-analyses, facilitating a wider and more rigorous study in the field by future researchers. The inclusion of fixed or reciprocal tutoring, the number of participants in the study, and the approach taken (quantitative/qualitative) should not alter the academic outcomes. Although some peer tutoring scenarios seem to be more beneficial for students than others, overall, peer tutoring has proved to be beneficial most of the time. Hence, its implementation is recommended in any case.

## 5. Conclusions

Peer tutoring has been found to improve academic results in mathematics in secondary education most of the time. This methodology develops inclusive education principles and empowers cooperative learning. Interventions among same-age students, low duration of tutoring sessions (<30 minutes), low duration of tutoring programs (<8 weeks), and low frequency of sessions ( $\leq 3$  per week) are encouraged in order to maximize academic outcomes. Although this study showed greater gains for non-disabled students, practitioners should not be discouraged, as peer-tutoring has also proved to be beneficial for disabled students. Further, to ensure that results are not overestimated, a PPCG experimental design should be used. The approach taken (quantitative vs. qualitative), the number of participants, or the type of tutoring (fixed vs. reciprocal) should not significantly alter the academic outcomes. Although this study suggests implementing peer tutoring under certain conditions, practitioners should also find academic benefits in any scenario, as academic gains have been documented overall under any condition.

## Declarations

### Author contribution statement

All authors listed have significantly contributed to the development and the writing of this article.

### Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Competing interest statement

The authors declare no conflict of interest.

### Additional information

No additional information is available for this paper.

## References

- Adamson, R.M., Lewis, T.J., 2017. A comparison of three opportunity-to-respond strategies on the academic engaged time among high school students who present challenging behavior. *Behav. Disord.* 42 (2), 41–51.
- Alegre Ansuategui, F.J., Moliner Miravet, L., 2017. Emotional and cognitive effects of peer tutoring among secondary school mathematics students. *Int. J. Math. Educ. Sci. Technol.* 48 (8), 1185–1205.
- Alegre, F., Moliner, L., Maroto, A., Lorenzo-Valentin, G., 2019. Peer tutoring in mathematics in primary education: a systematic review. *Educ. Rev.* 71 (6), 767–791.
- Alegre-Ansuategui, F.J., Moliner, L., Lorenzo, G., Maroto, A., 2018. Peer tutoring and academic achievement in mathematics: a meta-analysis. *Eurasia J. Math. Sci. Technol. Educ.* 14 (1), 337–354.
- Allsopp, D.H., 1997. Using classwide peer tutoring to teach beginning algebra problem-solving skills in heterogeneous classrooms. *Remedial Special Educ.* 18 (6), 367–379.
- Austin, J., 2008. The Effects of Peer Tutoring on Fifth-Grade Students' Motivation and Learning in Math. Unpublished doctoral dissertation. State University of New York, United States.
- Azcoitia, C., 1989. Structured Peer Tutoring in Chicago's Vocational Education Program. Unpublished doctoral dissertation. Northern Illinois University, United States.
- Bangert-Drowns, R.L., Rudner, L.M., 1991. Meta-analysis in educational research. *Pract. Assess. Res. Eval.* 2 (8), 9–12.
- Blackbourn, V.A., Blackbourn, J.M., 1993. An adolescent with moderate mental disabilities tutors a 1st-grade, nondisabled child. *Teach. Except. Child.* 25 (4), 56–61.
- Bowman-Perrott, L., Mahadevan, L., Etchells, M., 2016. Assessing the academic, social, and language production outcomes of English language learners engaged in peer tutoring: a systematic review. *Educ. Treat. Child.* 39 (3), 359–388.
- Bozzano, M., Bruttomesso, R., Cimatti, A., Juntila, T., Ranise, S., van Rossum, P., Sebastiani, R., 2006. Efficient theory combination via Boolean search. *Inf. Comput.* 204 (10), 1493–1525.
- Brown, M.B., Forsythe, A.B., 1974. 372: the ANOVA and multiple comparisons for data with heterogeneous variances. *Biometrics* 719–724.
- Bryman, A., 2006. Integrating quantitative and qualitative research: how is it done? *Qual. Res.* 6 (1), 97–113.
- Burley, S., Gutkin, T., Naumann, W., 1994. Assessing the efficacy of an academic hearing peer tutor for a profoundly deaf student. *Am. Ann. Deaf* 139 (4), 415–419.
- Butler, F.M., Miller, S.P., Lee, K.H., Pierce, T., 2001. Teaching mathematics to students with mild-to-moderate mental retardation: a review of the literature. *Ment. Retard.* 39 (1), 20–31.
- Cai, J., Hwang, S., 2019. Learning to teach through mathematical problem posing: theoretical considerations, methodology, and directions for future research. *Int. J. Educ. Res.*
- Cairo III, L., Craig, J., 2005. Cross-age Tutoring Phase II—An experiment. Appalachia Educational Laboratory at Edvantia (NJ1), Louisville, KY.
- Calhoun, M.B., Fuchs, L.S., 2003. The effects of peer-assisted learning strategies and curriculum-based measurement on the mathematics performance of secondary students with disabilities. *Remedial Special Educ.* 24 (4), 235–245.
- Cheung, A.C., Slavin, R.E., 2016. How methodological features affect effect sizes in education. *Educ. Res.* 45 (5), 283–292.
- Cohen, P.A., Kulik, J.A., Kulik, C.L.C., 1982. Educational outcomes of tutoring: a meta-analysis of findings. *Am. Educ. Res. J.* 19 (2), 237–248.
- Collins, J.P., Calevro, M.J., 1974. Mainstreaming Special Education Using a Peer Tutoring System and a Minimum Objective Curriculum for Nine Eighth-Grade Students. Affiliation Orange-Washington Supervisory Union, Barre Town, VT. ERIC ED 102 788.
- Collins, K.M., Onwuegbuzie, A.J., 2001. Effect of an After-School Tutorial Program on Academic Performance of Middle School Students At-Risk. Mid-South Educational Research Association, Little Rock, AR. ERIC ED 467 167.
- Cook, S.C., Cook, B.G., Cook, L., 2017. Classifying the evidence base of classwide peer tutoring for students with high-incidence disabilities. *Exceptionality* 25 (1), 9–25.
- Cooper, H., Hedges, L.V., Valentine, J.C. (Eds.), 2009. *The Handbook of Research Synthesis and Meta-Analysis*. Russell Sage Foundation, New York, NY.
- Cooper, S.J., 2016. An Exploratory Evaluation of a Paired Maths Intervention with Secondary Aged Pupils. Unpublished doctoral dissertation. University of Nottingham, England.
- Davidson, N., 1985. Small-group learning and teaching in mathematics: a selective review of the research. *Learn. Cooperat., Cooperat. Learn.* 1 (1), 211–230.
- De Smet, M., Van Keer, H., De Wever, B., Valcke, M., 2010. Studying thought processes of online peer tutors through stimulated-recall interviews. *High. Educ.* 59 (5), 645–661.
- De Winter, J.C., 2013. Using the student's t-test with extremely small sample sizes. *Pract. Assess. Res. Eval.* 18 (10), 1–12.
- Dufrene, B.A., Noell, G.H., Gilbertson, D.N., Duhon, G.J., 2005. Monitoring implementation of reciprocal peer tutoring: identifying and intervening with students who do not maintain accurate implementation. *Sch. Psychol. Rev.* 34 (1), 74.
- Duran, D., Monereo, C., 2005. Styles and sequences of cooperative interaction in fixed and reciprocal peer tutoring. *Learn. Instr.* 15 (3), 179–199.
- Durand, S.C., 2008. Culminating Experience Action Research Projects. Using Classwide Peer Tutoring to Increase High School Math Students' Academic Performance, vol. 10. Spring, p. 182, 2007.
- Early, J.W., 1999. The Impact of Peer Tutoring on Self-Esteem and Texas Assessment of Academic Skills Mathematics Performance of Tenth-Grade Students. Unpublished doctoral dissertation. Texas A&M University, United States.
- Entwistle, N., Tait, H., McCune, V., 2000. Patterns of response to an approaches to studying inventory across contrasting groups and contexts. *Eur. J. Psychol. Educ.* 15 (1), 33.
- Fritz, C.O., Morris, P.E., Richler, J.J., 2012. Effect size estimates: current use, calculations, and interpretation. *J. Exp. Psychol. Gen.* 141 (1), 2.
- Galbraith, J., Winterbottom, M., 2011. Peer-tutoring: what's in it for the tutor? *Educ. Stud.* 37 (3), 321–332.
- Gersten, R., Fuchs, L.S., Compton, D., Coyne, M., Greenwood, C., Innocenti, M.S., 2005. Quality indicators for group experimental and quasi-experimental research in special education. *Except. Child.* 71 (2), 149–164.
- Gogolin, I., 2016. European educational research quality indicators (EERQI): an experiment. In: Ochsner, M., Hug, S.E., Daniel, H. (Eds.), *Research Assessment in the Humanities*. Springer, Cham, pp. 103–111.
- Hannah, D.C., 2008. Attitudinal Study: the Interaction of Students Taking Calculus and Prerequisite Courses while Participating in Peer tutorials. Unpublished doctoral dissertation. University of Nottingham Georgia State University, England.
- Hilo Intermediate School, 1974. Buddy-tutor Project. Final Report March-July 1974. Hawaii University.
- Hedges, L.V., Pustejovsky, J.E., Shadish, W.R., 2013. A standardized mean difference effect size for multiple baseline designs across individuals. *Res. Synth. Methods* 4 (4), 324–341.
- Heintz, P., 1975. High School Peer Tutoring (Homework Helpers) Program (School Year 1974–1975). New York City Board of Education, Brooklyn, NY.
- Hendrickson, J.M., 1981. Using low achievers to tutor reading and mathematics. In: Paper Presented at the International Reading Association Conference, Gatlinburg, TN.
- Hodgson, Y., Benson, R., Brack, C., 2015. Student conceptions of peer-assisted learning. *J. Furth. High. Educ.* 39 (4), 579–597.
- Holecek, D.L., 2012. Cross Age/cross Disability Peer Tutoring. Unpublished doctoral dissertation. University of Wisconsin-Superior, United States.
- Ivory, T.S., 2007. Improving Mathematics Achievement of Exceptional Learners through Differentiated and Peer-Mediated Instruction. Unpublished doctoral dissertation. Nova Southeastern University, Florida, United States.
- Johnson, B., Christensen, L., 2008. *Educational Research: Quantitative, Qualitative, and Mixed Approaches*. Sage, Thousand Oaks, CA.
- Johnson, D.W., Maruyama, G., Johnson, R., Nelson, D., Skon, L., 1981. Effects of cooperative, competitive, and individualistic goal structures on achievement: a meta-analysis. *Psychol. Bull.* 89 (1), 47.
- Kamps, D.M., Barbetta, P.M., Leonard, B.R., Delquadri, J., 1994. Classwide peer tutoring: an integration strategy to improve reading skills and promote peer interactions among students with autism and general education peers. *J. Appl. Behav. Anal.* 27 (1), 49–61.
- Kane, B.J., Alley, G.R., 1980. A peer-tutored, instructional management program in computational mathematics for incarcerated, learning disabled juvenile delinquents. *J. Learn. Disabil.* 13 (3), 148–151.
- Knight, R.A., Dipper, L., Cruice, M., 2018. Viva survivors—the effect of peer-mentoring on pre-viva anxiety in early-years students. *Stud. High. Educ.* 43 (1), 190–199.
- Lazarus, K.U., 2014. Effect of peer tutoring and cooperative learning instructional strategies on mathematics achievement of students with learning disabilities in Oyo State, Nigeria. *Afr. J. Psychol. Stud. Soc. Iss.* 17 (1), 96–105.
- Leal, G., Olivás, G.D.J., 2014, October. Alumnos tutores en la resolución de problemas matemáticos. Students tutors in mathematics problem solving. In: Paper presented at XIV Conferencia Interamericana de Educación Matemática, Medellín, Colombia.
- LeCroy, C.W., Krysik, J., 2007. Understanding and interpreting effect size measures. *Soc. Work Res.* 31 (4), 243–248.
- León, J., Núñez, J.L., Liew, J., 2015. Self-determination and STEM education: effects of autonomy, motivation, and self-regulated learning on high school math achievement. *Learn. Individ. Differ.* 43, 156–163.
- Leung, K.C., 2015. Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *J. Educ. Psychol.* 107 (2), 558.
- Leung, K.C., 2018. An updated meta-analysis on the effect of peer tutoring on tutors' achievement. *Sch. Psychol. Int.* 40 (2), 200–214.

- Leung, K.C., 2019. Compare the moderator for pre-test-post-test design in peer tutoring with treatment-control/comparison design. *Eur. J. Psychol. Educ.* 1–19.
- Levitt, H.M., Bamberg, M., Creswell, J.W., Frost, D.M., Josselson, R., Suárez-Orozco, C., 2018. Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology: the APA publications and communications board task force report. *Am. Psychol.* 73 (1), 26.
- Losinski, M., Sanders, S., Katsiyannis, A., Wiseman, N., 2017. A meta-analysis of interventions to improve the compliance of students with disabilities. *Educ. Treat. Child.* 40 (4), 435–463.
- Mahood, Q., Van Eerd, D., Irvin, E., 2014. Searching for grey literature for systematic reviews: challenges and benefits. *Res. Synth. Methods* 5 (3), 221–234.
- Maxwell, J.A., 2004. Causal explanation, qualitative research, and scientific inquiry in education. *Educ. Res.* 33 (2), 3–11.
- Mayfield, K.H., Vollmer, T.R., 2007. Teaching math skills to at-risk students using home-based peer tutoring. *J. Appl. Behav. Anal.* 40 (2), 223.
- McKee, M.T., Caldarella, P., 2016. Middle school predictors of high school performance: a case study of dropout risk indicators. *Education* 136 (4), 515–529.
- Miquel, E., Duran, D., 2017. Peer learning network: implementing and sustaining cooperative learning by teacher collaboration. *J. Educ. Teach.* 43 (3), 349–360.
- Mohan, M., 1972. Peer Tutoring as a Technique for Teaching the Unmotivated. State University of New York Teacher Education Research Center, Fredonia, NY.
- Moss, B.G., Yeaton, W.H., 2006. Shaping policies related to developmental education: an evaluation using the regression-discontinuity design. *Educ. Eval. Policy Anal.* 28 (3), 215–229.
- Mulvaney, M.K., 1993. Two negatives equal a positive: semiotic mediation in peer tutoring. In: Paper Presented at the Annual Meeting of the Conference on College Composition and Communication, San Diego, CA.
- Murugan, S., 2015. Effectiveness of Peer Tutoring with Student to Student Counselling in Learning Mathematics at Standard IX Level. Unpublished doctoral dissertation. Alagappa University, Tamil Nadu, India.
- Nazzal, A., 2002. Peer tutoring and at-risk students: an exploratory study. *Action Teach. Educ.* 24 (1), 68–80.
- Nesselrodt, P.S., Alger, C.L., 2005. Extending opportunity to learn for students placed at risk. *J. Educ. Stud. Placed A. T. Risk* 10 (2), 207–224.
- Norris, N., 1997. Error, bias and validity in qualitative research. *Educ. Action Res.* 5 (1), 172–176.
- Novotni, S., 1985. Peer Tutoring: A Study of its Effect on Mathematic Achievement and Attitude of Ninth Grade Math I Students of Harrisburg High School. Unpublished doctoral dissertation. University of Pennsylvania, United States.
- Obidola, M.A., Onwubolu, C.O., 2013. Remedial help in inclusive classrooms: gender differences in the enhancement of mathematics achievement of students through PAL (peer-assisted learning). *U. S. China Educ. Rev.* A 3 (3), 172–180.
- Olive, M.L., Smith, B.W., 2005. Effect size calculations and single subject designs. *Educ. Psychol.* 25 (2–3), 313–324.
- Onwuegbuzie, A.J., 2003. Effect sizes in qualitative research: a prolegomenon. *Qual. Quantity* 37 (4), 393–409.
- Orwin, R.G., 1983. A fail-safe N for effect size in meta-analysis. *J. Educ. Stat.* 8 (2), 157–159.
- Penney, C.G., 2018. Rethinking the concept of learning disability. *Can. Psychol./Psychologie Canadienne* 59 (2), 197.
- Pyle, D., 2015. The Effects of a Targeted, Unidirectional Peer Tutoring Intervention on Mathematics Outcomes for Students with Learning Disabilities in an Inclusive, Secondary Setting. Unpublished doctoral dissertation. Utah State University, United States.
- Roach, J.C., Paolucci-Whitcomb, P., Meyers, H.W., Duncan, D.A., 1983. The Comparative Effects of Peer Tutoring in Math by and for Secondary Special Needs Students. Pointer, New York, NY.
- Rosenthal, J.A., 1996. Qualitative descriptors of strength of association and effect size. *J. Soc. Serv. Res.* 21 (4), 37–59.
- Rosenthal, R., Cooper, H., Hedges, L., 1994. Parametric measures of effect size. *The Handbook of Research Synthesis*, vol. 621, pp. 231–244.
- Rosnow, R.L., Rosenthal, R., 1996. Computing contrasts, effect sizes, and counternulls on other people's published data: general procedures for research consumers. *Psychol. Methods* 1 (4), 331.
- Schloss, P.J., Kobza, S.A., Alper, S., 1997. The use of peer tutoring for the acquisition of functional math skills among students with moderate retardation. *Educ. Treat. Child.* 20 (2), 189.
- Schneck, A.J., 2010. The Effects of After-School Peer Tutoring Programs in Mathematics. Unpublished doctoral dissertation. University of Minnesota, United States.
- Shenderovich, Y., Thurston, A., Miller, S., 2016. Cross-age tutoring in kindergarten and elementary school settings: a systematic review and meta-analysis. *Int. J. Educ. Res.* 76, 190–210.
- Sinha, T., Zhao, R., Cassell, J., 2015, November. Exploring socio-cognitive effects of conversational strategy congruence in peer tutoring. In: Proceedings of the 1st Workshop on Modeling INTERPERSONAL SYNCHRONY and INFLUENCE, Seattle, WA.
- Stenhoff, D.M., Lignugaris/Kraft, B., 2007. A review of the effects of peer tutoring on students with mild disabilities in secondary settings. *Except. Child.* 74 (1), 8–30.
- Thurston, A., Topping, K.J., 2007. Peer tutoring in schools: cognitive models and organizational typography. *J. Cogn. Educ. Psychol.* 6 (3), 356–372.
- Thurston, A., Duran, D., Cunningham, E., Blanch, S., Topping, K., 2009. International online reciprocal peer tutoring to promote modern language development in primary schools. *Comput. Educ.* 53 (2), 462–472.
- Topping, K., Buchs, C., Duran, D., Van Keer, H., 2017. Effective Peer Learning: from Principles to Practical Implementation. Routledge, Abingdon, UK.
- Topping, K., Miller, D., Thurston, A., McGavock, K., Conlin, N., 2011a. Peer tutoring in reading in Scotland: thinking big. *Literacy* 45 (1), 3–9.
- Topping, K.J., Miller, D., Murray, P., Henderson, S., Fortuna, C., Conlin, N., 2011b. Outcomes in a randomised controlled trial of mathematics tutoring. *Educ. Res.* 53 (1), 51–63.
- Topping, K.J., 1996. The effectiveness of peer tutoring in further and higher education: a typology and review of the literature. *High. Educ.* 32 (3), 321–345.
- Topping, K.J., 2005. Trends in peer learning. *Educ. Psychol.* 25 (6), 631–645.
- Topping, K.J., Dehkinet, R., Blanch, S., Corcelles, M., Duran, D., 2013. Paradoxical effects of feedback in international online reciprocal peer tutoring. *Comput. Educ.* 61, 225–231.
- Topping, K.J., Kearney, M., McGee, E., Pugh, J., 2004. Tutoring in mathematics: a generic method. *Mentor. Tutoring: Partnersh. Learn.* 12 (3), 353–370.
- Topping, K.J., Smith, E.F., Swanson, I., Elliot, A., 2000. Formative peer assessment of academic writing between postgraduate students. *Assess. Eval. High. Educ.* 25 (2), 149–169.
- 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. *Sch. Eff. Sch. Improv.* 22 (3), 265–289.
- Valente, M.J., MacKinnon, D.P., 2017. Comparing models of change to estimate the mediated effect in the pretest–posttest control group design. *Struct. Equ. Model.: Multidisciplinary J.* 24 (3), 428–450.
- Walker, E.N., 2007. The structure and culture of developing a mathematics tutoring collaborative in an urban high school. *High Sch. J.* 91 (1), 57–67.
- Worley, J., Naresh, N., 2014. Heterogeneous peer-tutoring: an intervention that fosters collaborations and empowers learners: key features of an intervention peer-tutoring program highlight the cognitive and social benefits of this collaborative approach. *Middle Sch. J.* 46 (2), 26–32.
- Zeneli, M., Tymms, P., 2015. A review of peer tutoring interventions and social interdependence characteristics. *Int. J. Cross-Disciplinary Subj. Educ.* 5 (2), 2504–2510.
- Zeneli, M., Thurston, A., Roseth, C., 2016a. The influence of experimental design on the magnitude of the effect size-peer tutoring for elementary, middle and high school settings: a meta-analysis. *Int. J. Educ. Res.* 76, 211–223.
- Zeneli, M., Tymms, P., Bolden, D., 2016b. The impact of interdependent cross-age peer tutoring on social and mathematics self-concepts. *Int. J. Psychol. Educ. Stud.* 3 (2), 1–13.
- Zeneli, M., Tymms, P., Bolden, D., 2018. Interdependent cross-age peer tutoring in mathematics. *Int. J. Psychol. Educ. Stud.* 5 (3), 33–50.
- Zhao, Y., Ding, C., 2019. The association between students' mathematics knowledge and factors related to students, parents, and school: a cross-cultural comparison study. *Int. J. Educ. Res.* 93, 210–217.