EURASIA Journal of Mathematics, Science and Technology Education
ISSN: 1305-8223 (online) 1305-8215 (print)
OPEN ACCESS
2018 14(1):337-354
DOI:10.12973/ejmste/79805

Peer Tutoring and Academic Achievement in Mathematics: A Meta-Analysis

Francisco Jose Alegre-Ansuategui [1*], Lidón Moliner [2], Gil Lorenzo [2], Ana Maroto [1]

[1] Universidad de Valladolid, Valladolid, SPAIN

[2] Universitat Jaume I, Castello de la Plana, Comunitat Valenciana, SPAIN

Received 22 July December 2017 ▪ Revised 30 August 2017 ▪ Accepted 7 October 2017

ABSTRACT

A meta-analysis of findings from 50 independent studies of peer tutoring programs in Mathematics at multiple educational stages showed that 88% of these programs have positive effects on the academic performance of the participants (Hedge's g = 0.333). Some of the variables to be taken into account when developing a peer tutoring experience were analyzed. Results showed that variables such as the ages of the participants, roles, skills of the tutees (disabled or at academic risk vs non-disabled and not at academic risk), length of the sessions and frequency were not significant moderators of the academic achievement. Variables such as educational stage, design of the study, duration of the program, level of knowledge of the tutors, time of the day (school time vs out of school time) and sample size turned out to be significant moderators. Results are discussed and proposals for future research are suggested.

Keywords: academic achievement, mathematics, meta-analysis, multiple educational stages, peer tutoring

JUSTIFICATION

One of the top specialists at the international level in the area of peer tutoring is Keith Topping. He offers two differentiated positions regarding this method. The first of these is an archaic definition where the tutoring student is seen as a substitute for the professor. This refers to a linear conception of the transmission of knowledge: a fixed role defined by skills and capacities, where the more competent teaches the less competent: "the more capable students help less capable students to learn through cooperative work in pairs or small groups, carefully organized by a professional teacher" (Topping, 1996).

Along these lines, Damon and Phelps (1989) note that classmates with more knowledge and/or ability, after a process of training or skills development, and within the framework of an asymmetric relationship that is externally planned by a professional, provide help and support to other students with less knowledge and/or ability through cooperation in pairs or in small groups. The second definition that Topping (1996) offers, and the most recent, refers to the fact that people in similar social groups, who are not professors, help others to learn, and themselves learn by teaching.

He later adds that they help interactively, intentionally and systematically (Topping, 2000). Duran, Torró and Vila (2003) see peer tutoring as a method of cooperative learning based on the creation of pairs, having an asymmetric relationship, derived from the tasks of their respective roles, tutor and tutee, where both students have a common, shared objective: the acquisition or improvement of some curricular capacity, which is acquired by means of an interaction planned by the teacher.

For Boud (1988) mutual learning is the road to superseding independence and advancing towards interdependence. Students learn by expressing their ideas to others through participation in activities where they can learn from their peers. The emphasis is on the process of learning, and providing emotional support more than aid with the specific learning task. It is an educational practice in which the student interacts with other students in order to reach educational goals. Therefore, peer tutoring is an alternative kind of teaching-learning where the

✉ falegre@uji.es **(*Correspondence)** ✉ mmoliner@uji.es ✉ valentin@uji.es ✉ amaroto@am.uva.es

Contribution of this paper to the literature

- This study evaluates the effects of peer tutoring in students' academic achievement in Mathematics including also grey literature.
- Sixteen research questions regarding variables that may influence students' academic achievement (type of peer tutoring, length of the program, frequency of sessions…) are answered and discussed.
- Indications for future studies are suggested.

student adopts an attitude of commitment to thinking, reasoning and exchanging knowledge (Luca and Clarkson, 2002).

Peer tutoring in Mathematics has been implemented for more than 25 years. Several authors such as John Fantuzzo, Lynn Fuchs or Keith Topping have contributed to this field. In regards to the reviews done in this field, the first one took place in 1989 (Britz et al, 1989) and the second one in 2005 (Robinson et al., 2005). Since then, no review has been done in peer tutoring in Mathematics. Leung (2015) performed a meta-analysis in peer tutoring and academic achievement, but only included 17 studies of Mathematics and didn't take into account the so called grey literature (Cooper, 2009). Meta-analysis is a methodology that allows making systematic reviews of empirical research and identify relationships among variables and outcomes (Durlak, 1995; Rosenthal, 1984). Therefore, it is necessary to do review of all research of peer tutoring in Mathematics and to perform a meta-analysis in order to try to determine which variables are determinant to the academic achievement.

AIM AND RESEARCH QUESTIONS

The aim in the present investigation was to synthesize previous research on peer tutoring in mathematics, evaluate the effects of peer tutoring on academic achievement, and identify the crucial determinants of the effectiveness of peer tutoring in mathematics. Therefore, we will analyze the following variables: types of tutoring (fixed as opposed to reciprocal and same-age as opposed to cross-age), skills of the tutees (disabled or at academic risk vs non-disabled and not at academic risk), level of knowledge of the tutors (low, medium, high), duration of the sessions, frequency with which they occur, total duration of the program, structuring of the pairs (whether pairing is done randomly or following some pattern), stage at which the tutoring is carried out (preschool and kindergarten, primary, secondary or university), time of the day that the tutoring occurs (during school hours or at other times), design employed, indexation of the studies and training of the participants. Thus, the principal questions in this research project are the following:

Table 1. Research questions

Q1. Is cross-age tutoring more effective than same-age?
Q2. Is fixed tutoring more effective than reciprocal
Q3. Is fixed same-age tutoring more effective than reciprocal same-age?
Q4. At what educational stage are the best results obtained?
Q5. When students at academic risk or with functional diversity participate, are better results obtained than with non-disabled students or those not at academic risk?
Q6. Do there exist significant differences in the results due to the design employed (pretest-posttest control group (PPCG) as opposed to posttest-only with control (POWC) and pretest-posttest single group (PPSG))?
Q7. In the studies where the duration of the sessions was greater than 30 minutes, were better results obtained than in those which lasted less than 30 minutes?
Q8. In those studies with a greater frequency of sessions per week (more than twice a week) are better results obtained than with those that had less frequent sessions (1 or 2 times per week)
Q9. Did the longer-term studies (more than 15 weeks) obtained better results than those with a shorter duration (15 or fewer weeks)?
Q10.Were there any differences between the studies as a result of the level of knowledge of the tutors?
Q11. Did the studies carried out during school time obtain better results than those that were carried out extra-scholastically?
Q12. Do studies with structured peer tutoring (in which students are not paired randomly) obtain better results than those which are not structured?
Q13. Are there significant differences in the results of the studies depending on the date they were published?
Q14. Do the studies with larger samples (higher than 60) obtain better results than those carried out with smaller sample sizes?
Q15. Do the studies published in indexed journals and those not published in indexed journals present important differences in the results obtained?
Q16. In the studies where the participants received exhaustive training, were better results obtained than in other studies?

METHODOLOGY

Search Procedure

In order to be able to respond to the review questions, we carried out a search in order to find articles related to peer tutoring in mathematics. We did not restrict the search by publication date. We performed the search employing the following databases: ERIC (Educational Resources Information Center), Google Scholar and APA PsycNet (American Psychological Association). The search descriptors included the following 10 terms: "peer tutoring," "mentoring," "peer-mediated instruction," "peer-assisted learning," "mathematics," "math," "problem-solving", "arithmetic," "geometry" and "algebra." The first four terms were combined with the last six so that each search only included two of these terms, that is, one of the first four terms and one of the last six terms resulting in a total of 24 different combinations. In the case of Google Scholar we made it a search criterion that the keywords had to appear together in the article title, since a search with just the keywords "peer tutoring" and "mathematics" returned more than 78500 results.

In order to corroborate the appropriateness of these search terms, we contrasted the references found in the databases with those published in the highest-impact journals where we knew that research projects of this type were presented, such as *American Educational Research Journal; European Journal of Psychology of Education; Educational Research; Educational Studies; Higher Education; Learning and Instruction; Mentoring & Tutoring; Partnership in learning; Remedial and Special Education; Research in Education; Support for Learning; The International Journal of Higher Education and Educational Planning*. In addition, we consulted certain websites specialized in research into peer tutoring, including Peer Tutoring Resource Center and GRAI (Research Group about Peer Learning). In addition, we compared the articles with other prior meta-analyses in fields related to that of this study (Kroesbergen & Van Luit, 2003; Bowman-Perrott et al., 2013; Cohen et al., 1982; Leung et al., 2005; Rohrbeck et al., 2003; Gersten et al., 2009; Kunsch & Jitendra 2007). In addition, we also made a comparison with the bibliographic reviews concerning peer tutoring in mathematics (Robinson et al., 2005; Britz, 1989) and others concerning peer tutoring in general (Topping, 1996; Delquadri et al., 1986). We included articles in both English and Spanish. Including only articles in English and Spanish may have weakened the external validity of the present article due to a tendency among writers in those languages to publish mostly articles with experiences whose results have significant positive effects (White, 2009; Kroesbergen & Van Luit, 2003).

In addition, as Cooper et al. (2009) note, there is a great deal of grey literature (also called "fugitive" literature) that has not been published in journals. This literature includes doctoral theses and reports presented at conferences. According to this author, the inclusion of this kind of research results provides a much more representative sample of the studies that have been carried out. Thus, with the purpose of obtaining the most representative sample possible, not only have we included articles published in journals, but have also included doctoral theses, reports presented at conferences and all other publications having sufficient experimental rigor to form part of a meta-analysis.

Selection Criteria

Through searches in the databases indicated above, we obtained a total of 1145 bibliographic references, including repeated articles. The majority of the references found with the keywords "mentoring" and "problem-solving" did not deal with peer tutoring in mathematics, and as a result we excluded them from this meta-analysis. In addition, many articles were not based on empirical analysis or were not analyses at all, but were manuals or reviews. Thus, after excluding the articles just mentioned, 124 articles remained to be included in the meta-analysis, to be filtered with the last 5 selection criteria that we discuss in the following paragraphs.

The first selection criterion was that we only accepted those articles where the interaction between the tutor and tutee was direct and human-to-human. We did not consider those where one tutor helped various tutees, but looked only for tutoring dyads. We excluded all research that viewed interactions between a computer and students as tutoring, treating the computer as a simulation of a student. We did not exclude those where the interactions occurred through a computer, where students were in simultaneous interaction with one another. In this case the computer was treated simply as a communication tool. 14 studies were excluded due to this criterion.

The second criterion is that we did not consider to be peer tutoring any aid offered by parents or post-university personnel, since we considered that this type of aid comes from a person who violates the concept of "peer" in the context of tutoring. 13 studies were excluded due to this criterion.

The third criterion is that the research must have dealt with scholastic progress in mathematics in a quantitative manner, taking the different study variables into account, since this is the object of study of this meta-analysis. We did not take into account those studies that analyzed other variables, like self-concept or the attitude of the student towards the class or program, if they did not separately study the academic performance of the student in mathematics. 21 studies were excluded due to this criterion.

The fourth criterion, strongly linked to the previous one, is that we had to be able to extract quantitative data about the academic performance of the students based on the research in question. The various data regarding academic performance had to have a comparative pattern, so that we could carry out an analysis of the improvement of the students as a result of their participation in the peer tutoring program (pretest and posttest and/or a control group). The article had to include one of the following items: the statistical parameter employed for the analysis together with its level of significance, numeric data on the basis of which we could perform a comparison, or graphics on the basis of which we could interpret the numeric results. In no case did we include those articles where the quantitative information was insufficient or was omitted. 14 studies were excluded due to this criterion.

The fifth criterion indicates that, for statistical reasons, those studies with fewer than 3 students were also excluded. Although there is not minimum number of subjects for a study to be included in meta-analysis, publication bias regarding small studies has been widely discussed (Nuijten et al., 2015; Ledgerwood & Sherman, 2012; Bakker et al., 2012) and previous meta-analysis in education also defined this criterion (Kroesbergen & Van Luit, 2003). 12 studies were excluded due to this criterion.

Calculation of Effect Size

As a measure of effect size we employed Hedges' g (Hedges, 1982). We chose this parameter since it produces less bias than other measures such as Cohen's d (Cumming, 2013). For the calculation of Hedges' g and of the standard error, the process followed and the formulas employed are those described in what follows:

1- For the case where the study is a PPCG (pretest posttest control group) and provides the pretest and posttest measures of both the control and experimental groups and the standard deviations in the pretest, we carried out the following calculations, as Lakens (2013) indicates:

$$g = j \times d \tag{1}$$

$$d = \frac{[(M_{post,T} - M_{pre,T}) - (M_{post,C} - M_{pre,C})]}{SD_{pooled}} \tag{2}$$

where $M_{post,T}$, $M_{post,C}$, $M_{pre,T}$ and $M_{pre,C}$ are the posttest and pretest measures for the experimental group (T) and control (C) and in addition,

$$SD_{pooled} = \sqrt{\frac{(n_T - 1)SD^2_{pre,T} + (n_C - 1)SD^2_{pre,C}}{n_T + n_C - 2}} \tag{3}$$

$$j = 1 - \frac{3}{4(n_T + n_C - 2) - 1} \tag{4}$$

where n_T and n_C are the sizes of the experimental group and control group respectively, and $SD^2_{pre,T}$ and $SD^2_{pre,C}$ are the standard deviations squared of the experimental and control groups.

We took the standard deviations of the pretests because there are various authors that hold that this formula is the best when it comes to avoiding as much as possible the possible biases that could occur (Morris, 2007). By taking them from the pretest we avoided bias due to the treatment carried out during the experimental phase (Glass et al., 1981).

In addition, for the calculation of standard error (SE_g) we do not assume that the standard deviations are equal, and we calculate following Borenstein et al. (2009):

$$V_D = \frac{n_T + n_C}{n_T n_C} + \frac{d^2}{2(n_T + n_C)} \tag{5}$$

$$V_g = j^2 \times V_D \tag{6}$$

$$SE_g = \sqrt{V_g} \tag{7}$$

This is the process followed for the majority of the studies included in the current meta-analysis, since a very high percentage of them are of a PPCG type, and provide all the necessary statistical data.

2- For the cases of PPCG studies where only the Student's t and the size of the experimental and control groups were provided, we used the formula proposed by Thalheimer and Cook (2002).

$$d = t\sqrt{(\frac{n_T + n_C}{n_T n_C})(\frac{n_T + n_C}{n_T + n_C - 2})} \tag{8}$$

and we proceed to calculate g and SE_g as we indicated above.

3- For the case of the PPCG studies that include only the data from the Fisher's exact test and not the median squared errors (MSE), together with the size of the experimental and control groups, we used the formula proposed by Thalheimer and Cook (2002):

$$d = \sqrt{F(\frac{n_T + n_C}{n_T n_C})(\frac{n_T + n_C}{n_T + n_C - 2})} \tag{9}$$

and we proceeded to calculate the g and SEg as indicated in section 1.

4- For the cases of the PPCG studies which only provided Pearson's r and the sample sizes of the control and experimental groups, we calculated d with the formula proposed by Becker (2000):

$$d = \frac{2r}{\sqrt{1 - r^2}} \tag{10}$$

and we proceeded to calculate the g and SEg as indicated in section 1.

5- For some studies, although the design was PPCG, certain information necessary for the calculation of g and SEg was omitted. The information that was most often lacking was the standard deviations of the pretest and posttest. The first thing we did in this type of case was attempt to contact the authors of the study. In two cases it was possible to obtain data that was not present in the article, after contacting the authors (Galia (2015) and Tella (2013)). In the article by Nesselrodt & Alger (2005), given that the authors only provided the standard deviations of the posttests, we used them instead of those of the pretests, as is typically recommended for the calculations of g and SEg, assuming homogeneity of variance (Morris & Deshon, 2002) since it was impossible to contact the authors. In the cases where, after having attempted to contact the authors, we did not obtain sufficient information to calculate the effect sizes, we substituted an estimation of those values (in all the cases of the standard deviations). Thus, just as Cooper et al. (2009) indicate in the section dealing with "missing data," the non-inclusion of those studies in the meta-analysis meant assuming that those studies were omitted in a random way. Thus, we believe that it is appropriate to include those studies through some procedure. We follow the indications of Higgins & Green (2008).

Thus, for the case of studies which provide the standard error, we estimated the standard deviation with the formula:

$$SDpooled = \frac{SE}{\sqrt{\frac{1}{N_E} + \frac{1}{N_C}}} \tag{11}$$

where SE is the standard error and Ne and Nc are the size of the sample of the control group and experimental group, respectively.

On the other hand, for the case of studies in which only the Student's t and the differences of means were provided, with no standard deviation, we calculated the standard error with the formula

$$SE = \frac{(M_{post,T} - M_{pre,T}) - (M_{post,C} - M_{pre,C})}{t} \tag{12}$$

And once we had obtained SD we could use the previous formula to calculate the pooled SD and thus, on the basis of the preceding formulas, calculate g and SEg in each case.

6- In addition, in the meta-analysis we carried out there is a high percentage of studies (25%), whose design is not PPCG, but rather PPSG, POWC or MB/RM (Multiple Baseline / Repeated Measures). As Borenstein et al. (2009) indicate in their section "Including different designs in the same study," from a statistical perspective the g parameter has the same meaning no matter what the study design is. Nonetheless, *while there are no technical barriers to using studies with different designs in the same analysis, there may be a concern that these studies could differ in substantive ways as well*. Although there are some perspectives that support the inclusion of these studies and others that do not, it is undeniable that one must be cautious when carrying out this process (Morris & Deshon, 2002; Rosenthal & Rubin, 1982; Carlson & Schmidt, 1999). Thus, in this research project, we opted to include POWC and PPSG studies in the meta-analysis and to analyze the design variable in order to see if it is a moderating variable. With the analysis of the design variable we were surer of being able to clarify whether the inclusion of this kind of study was appropriate.

For the calculation of the effect sizes of the studies with this kind of design, we chose to adopt the formulas presented in Carlson & Schmidt (1999). For the difference of means in the denominator of the d parameters in POWC studies, we took the difference between the mean of the experimental group and of the control, while for the standard deviations we took those that were available for the control and experimental groups, and thus we obtained pooled SD using equation (3). We must bear in mind that, for this type of study, it was common that they indicated that in previous observations there were no significant differences detected in prior measurements of the students, although they did not provide the data. For the case of the PPSG studies, we adopted as a difference of means in the denominator of d the difference of means of the posttest and pretest, and as standard deviations we took those available for the pretest and posttest for the calculation of the pooled SD according to equation (3).

In addition, we note that for any type of design—PPCG, POWC or PPSG—if we obtained a value of g higher than 1.2 we eliminated that study and did not include it in the meta-analysis, since it was considered unrealistic, and its inclusion could statistically alter the results, since in 90% of the cases no values of g superior to 1 were obtained.

In the case that the students in a study took various tests for different mathematical contents/skills, we calculated the Hedges' g and the SEg for each test individually, and later we developed a fixed effects model to obtain the mean g and the SEg for that study. This is because if we had considered each test as a study, we would have had an unreal estimation of the size of the global effect (Rosenthal & Rubin, 1982, Swanson et al., 1996). In some studies, more than one posttest was taken. We decided to only employ the results of the first posttest for the calculation of the effect size. Thus, we defined a single effect size for each study, unless in that study we could analyze, separately and with sufficient information, two of the distinct variables to be studied mentioned in the section "Research Questions."

Codification of the Studies

We present the studies in the following tables, by alphabetical order according to the first last name of the principal author. We distinguish three categories for codifying the studies. The first of them is the *identification and design of the study*. We included the following variables in this category:

1- Year of publication.

2- Study design (**DES**). In this case we indicated whether it is a PPCG, a PPSG or a POWC.

3- Type of publication (**I**): whether it is a case of a study published in a journal (Y) or whether it is a study unpublished in any journal (N).

The second category is the *study description* and includes the following variables

4- Sample size (**N**). In this section we included only the students in the control and experimental groups in the calculation of the effect size.

5- Maintenance of the student roles (**R**), that is, whether the tutoring is fixed (F) or reciprocal (R).

6- Differences of age among the participating students (**A**), that is, whether we are dealing with cross-age (C) or same-age (S) tutoring.

7- Stage at which peer tutoring is carried out (**E**), that is, whether it is carried out at preschool or kindergarten level (C), primary (P), secondary (S) or at the university (U).

8- Skills of the participants (**D/R**), which is, whether they are disabled or at-risk students (Y) or not (N).

9- Tutor achievement (**TA**), that is, whether the tutors have a high level of knowledge in the topic they are explaining (H), average (A) or low (L).

The third category is *development of the intervention* and includes the following variables:

10- Length of each session (**LS**).

11- Frequency of the sessions (days a week are indicated, **FS**).

12- Total duration of the peer tutoring program (**LP**).

13- Time of the day (**ST/OS**), i.e. whether it was during school time (ST) or outside school time (OST).

14- Structuring of the peer tutoring (**S**), i.e. whether the students were paired randomly (N) or whether on the contrary some procedure was used to first order them and then pair them (Y).

15- Training (**T**), i.e. for how much time were the participants trained before the implementation of the peer tutoring program.

Statistical Analysis

For the statistical analysis of the data, in the first place, we used a spreadsheet program to calculate the g and SEg values for each study, applying the formulas indicated previously. Later we introduced the g and SEg data into the program Comprehensive Meta-Analysis version 3.0 (Borenstein, Hedges, Higgins and Rothstein, 2015) together with the experimental and control group sizes, since the program gives the option of entering them. In the case that the study included different tests for the students (Calhoon & Fuchs 2003, Fuchs et al. 2008, Nesselrodt & Alger 2005), we first developed a fixed effects model with these tests in the software package, obtaining a mean g and a mean SEg in a global manner for each one of the studies, which would be the values that we used in the definitive meta-analysis. Once we had the effect size for each individual study, we had the software perform an analysis via a fixed effects model and a mixed effects model. Just as Meca (2010) has indicated, with the fixed effects model one assumes that all the studies are estimated to have the same population effect size, so that the possible variability between the effect sizes will be due only to a sampling error. On the other hand, the mixed effects model, as Borenstein et al (2009) indicated, refers to the creation of a model based on a model of random effects between groups and a model of fixed effects over all the groups. As Marín-Martínez and Sánchez-Meca (2010) indicate, a *mixed-effects model shall be conducted when a meta-analysis aims to search for moderator variables that can explain the variability in the effect-size estimates.* Thus, if the fixed effects model is shown to be homogeneous, that will indicate that this model is grounded and that the error is due only to a sampling error. On the contrary if one obtains a heterogeneous condition in the fixed effects model, the mixed effects model will have to be analyzed. This tendency to use both models is in consonance with current trends in this type of research project (Leung, 2015).

To analyze the heterogeneity of the different groups in the meta-analysis we carried out the Q test. If this test reaches statistical significance it will indicate that the effect sizes are heterogeneous and, therefore, the mean effect size represents it adequately (Meca (2010)). The statistic Q is complemented with the statistic I^2 of heterogeneity, so that 25%, 50% and 75% can be interpreted as low, medium and high heterogeneity respectively (Borenstein et al., 2009).

Thus we analyzed, on the one hand, the effect sizes obtained in each model between and within the groups (g) and, on the other hand, we calculated homogeneity via the Q_B statistic between groups and the Q_W statistic within the groups (Hedges & Olkin, 1985). We carried out this analysis for each of the variables mentioned previously. A significant Q_B indicates that the variable of the group is a significant moderator and that the mean effect size differs between subgroups. On the other hand, the statistic Q_W was used to test the effect between groups, with a non-significant Q_W indicating that the studies could be grouped into homogeneous subgroups (Leung, 2015).

In addition to all this, we also employed a Forest Plot, a graphical representation that shows, for each individual study in the meta-analysis, its effect size together with the confidence limits in a visual, numeric form. Just as Meca (2010) notes, it permits a global view of the results together with the degree of heterogeneity existing between the individual effect sizes. The software package we used provides a Forest Plot with all the values.

RESULTS

The 50 studies finally included in the meta-analysis showed a global Hedges' g value of 0.33, with a standard error of 0.03 (see **Table 2** and **Figure 1**). The global heterogeneity value was Q = 127.07, with $p < 0.001$ and an I^2 value equal to 61.439. The test results for effect moderators were as follows (see **Table 3**), although for the mixed effects model, none of the studied variables provided significant results. In the fixed effects model there were different variables that provided significant results. Using the results of the fixed effects model, we present the responses to the research questions in **Table 1**.

Table 2. Codification of the studies, Hedges' g and standard error

Study	N	A	R	E	D/R	DES	LS	FS	LP	TA	ST/OST	S	I	T	g	SEg
Alegre (2014)	18	S	F	S	N	PPSG	30 min	1	10 weeks	A	ST	Y	N	1 session, 20 minutes	0.679	0.335
Allsopp, D. (1997)	262	S	R	S	Y/N	PPCG	NA	2 to 4	5 weeks	A	ST	Y	Y	4 sessions, 40 minutes	0.139	0.124
Azcoitia (1989)	180	C	F	S	Y	PPCG	60 min	5	20 weeks	H	OST	N	N	Yes, not concise	0.318	0.256
Bar-Eli & Raviv (1982)	60	C	F	P	N	PPCG	50 min	3	4 months	L	ST	N	Y	No training	0.696	0.260
Beine-Smith (1991)	30	C	F	P	Y	PPCG	30 min	5	4 weeks	H	ST	N	Y	No training	0.128	0.428
Brush (1997)	63	S	R	P	N	PPCG	50 min	1	11 weeks	A	ST	N	Y	Yes, not concise	0.402	0.264
Cairo & Carig (2005)	111	C	F	P	N	PPCG	50 min	3	3 weeks	H	ST	N	N	Before every session	0.050	0.188
Calhoon & Fuchs (2003)	92	S	R	S	Y	PPCG	30 min	2	15 weeks	L	ST	Y	N	4 sessions, 50 minutes	0.441	0.194
Dennis, L. (2013)	105	C	F	P	Y	POWC	30 min	1	1 trimester	L	ST	Y	N	1 session < 50 min.	0.697	0.205
Durand, S. (2007)	17	S	R	S	N	POWC	10 min	5	4 weeks	L	ST	N	N	Yes, not concise	0.969	0.490
Fantuzzo (1992)	32	S	R	P	Y	POWC	45 min	2 or 3	5 months	L	ST	N	Y	4 to 5 sessions, 45 min.	0.476	0.251
Fantuzzo et al (1995)	26	S	R	P	Y	PPCG	45 min	2	10 weeks	L	ST	N	Y	2 sessions, 45 minutes	0.732	0.709
Fuchs 1997 I (no disability)	12	S	R	P	N	PPCG	35 min	2	18 weeks	A	ST	Y	Y	5 sessions, 30 min. each	0.327	0.228
Fuchs 1997 II (disability)	19	S	R	P	Y	PPCG	35 min	2	18 weeks	A	ST	Y	Y	5 sessions, 30 min. each	0.339	0.338
Fuchs et al (2001)-II	15	S	R	C	Y	PPCG	20 min	2	15 weeks	L	ST	Y	Y	10 min. prior to activity	0.376	0.492
Fuchs et al (2001)-I	233	S	R	C	N	PPCG	20 min	2	15 weeks	A	ST	Y	Y	10 min. prior to activity	0.345	0.163
Fuchs et al (2002)	323	S	R	P	N	PPCG	30 min	3	16 weeks	A	ST	Y	Y	10 min., each week	0.166	0.111
Fuchs et al (2008)	35	C	F	P	Y	PPCG	30 min	3	12 weeks	H	ST	N	Y	2 weeks	0.631	0.104
Galia, M. (2015)	50	S	F	U	Y	PPCG	1 hour	3	3 months	H	OST	N	N	No training	0.882	0.292
Ginsburg-Block and Fantuzo (1997)	40	S	R	P	Y	PPCG	45 min	2	10 weeks	L	ST	N	Y	2 sessions, 45 min. each	0.659	0.328
Greenwood (1991)	37	S	F	P	Y	POWC	30 min	5	1 year	L	ST	Y	N	3 sessions, 30 min. each	0.180	0.176
Grimm (2004)	32	S	R	P	N	PPSG	NA	NA	12 weeks	A	ST	N	N	Yes, but very brief	-0.193	0.248
Hannah, D (2008)	92	C	F	S	N	PPCG	55 min	2	12 weeks	H	OST	N	N	2 weeks, 4 sessions	0.523	0.169
Hawaii (1974)	184	C	F	S	N	PPCG	20 min	5	18 weeks	H	OST	N	N	Yes, but concise	0.303	0.148
Heintz (1975)	1370	C	F	S	Y	PPSG	2 hours	2 to 4	1 school year	H	OST	N	N	1 session	1.023	0.490
Heller (1992)	54	S	R	P	Y	PPCG	45 min	2	1 school year	L	ST	N	N	4 sessions, 45 min. each	0.942	0.283
Henson et al (2009)	70	S	R	U	N	PPCG	1 to 2 h	1	1 semester	A	ST	N	N	1 session, not concise	0.448	0.239
Hugger (2012)	37	S	R	P	N	PPCG	30 min	2	12 weeks	A	ST	Y	N	5 sessions 30 minutes	-0.189	0.324
Ivory, T. (2007)	22	C	R	S	Y	PPCG	1 h	3	15 weeks	L	ST	N	N	No training	0.006	0.412
Jştendra (2012)	115	C	F	P	Y	PPCG	30 min	5	12 weeks	H	OST	N	N	No training	0.538	0.156
Kane & Alley (1980)	21	C	F	S	Y	PPSG	45 min	5	8 weeks	L	ST	N	Y	20 min., 5 days	-0.044	0.319
Lazarus, k. (2014)	104	S	R	S	Y	PPCG	NA	NA	6 weeks	L	ST	N	N	Yes, very brief	0.559	0.198
Mahan (2010)	21	S	F	P	Y	PPCG	30 min	5	6 weeks	H	ST	Y	N	3 sessions, 30 min.	0.584	0.453
McKethan (1982)	53	C	F	U	Y	PPCG	50 min	2	9 weeks	H	OST	N	N	5 sessions, 75 minutes	0.227	0.212
Murugan (2015)	60	S	F	S	N	PPCG	1 h	NA	3 monhs	H	OST	N	N	30 min per session	0.986	0.27
Nazzal, A. (2002)	58	C	F	S	Y	PPGC	25 min	5	6 weeks	L	OST	N	N	1 session, very brief	0.322	0.261
Nesselrodt & Alger (2005)	71	C	F	S	Y	PPCG	1 h	3	346 sessions	H	OST	N	Y	90 minutes, 2/3 sessions	-0.006	0.136
Novotni (1985)	61	C	F	S	Y	PPCG	45 min	3	One year	H	ST	N	N	10 sessions	0.057	0.253
Obidoa & Onwubolu (2013)	505	S	F	S	Y	PPCG	NA	NA	1 years	H	ST	Y	N	No training	0.346	0.112
Philips et al (1993)	30	S	R	P	N	PPCG	30 min	≥2	6 weeks	A	ST	N	Y	No training	0.372	0.380
Roach et al (1983)	56	S	F	S	Y	PPCG	NA	NA	2 months	L	ST	N	N	30 minutes, 5 days	0.436	0.180
Schneck, A. (2010)	99	C	F	S	Y	PPCG	40 min	1 to 2	One semester	H	OST	N	N	1 hour	-0.168	0.217
Shamir et al (2006)	54	C	F	P	N	PPCG	25 min	1	7 weeks	H	ST	Y	Y	1 session, 5 minutes	0.638	0.195
Sharpley (1983)	171	C	F	P	N	PPCG	30 min	4	5 weeks	A	ST	N	Y	N/A	0.860	0.160
Sprinthall & Scott (1989)	30	C	F	P	N	PPCG	20 min	3	15 weeks	H	ST	N	Y	Yes, very brief	0.765	0.369
Tella (2013)	118	S	R	P	N	PPCG	NA	5	8 weeks	A	ST	N	N	No training	0.996	0.198
Topping et al (2011)	562	B	B	P	N	PPCG	30 min	1 or 3	15 weeks	A	ST	Y	Y	1 session, < 100 min.	0.025	0.118
Tsuei (2014)	4	S	R	P	Y	PPSG	30 min	1	1 year	L	ST	Y	Y	No training	1.167	0.681
Vazquez, A. (1995)	5	C	F	P	Y	PPSG	30 min	5	7 weeks	H	ST	N	N	No training	0.956	0.610
Xu et al (2001)	1224	C	F	U	N	POWC	NA	NA	4 months	A	OST	N	N	No training	-0.177	0.107

*Details of all the abbreviations are indicated in the section "Codification of the studies". When there was no information available for a variable in a study the "NA" abbreviation is used.

Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Statistics for each study						
Alegre (2014)*	0,679	0,335	0,112	0,022	1,336	2,027	0,043
Allsopp, D. (1997)*	0,139	0,124	0,015	-0,103	0,381	1,124	0,261
Azcoitia (1989)*	0,318	0,256	0,066	-0,184	0,820	1,242	0,214
Bar-Eli & Raviv (1982)*	0,696	0,260	0,068	0,186	1,206	2,677	0,007
Beirne-Smith (1991)*	0,128	0,429	0,184	-0,712	0,969	0,299	0,765
Brush (1997)*	0,402	0,264	0,070	-0,115	0,919	1,523	0,128
Cairo & Carig (2005)*	0,050	0,188	0,035	-0,318	0,418	0,266	0,790
Calhoon & Fuchs (2003)*	0,441	0,194	0,038	0,061	0,821	2,273	0,023
Dennis, L. (2013)*	0,697	0,205	0,042	0,295	1,099	3,400	0,001
Durand, S. (2007)*	0,969	0,490	0,240	0,009	1,929	1,978	0,048
Fantuzzo (1992)*	0,476	0,251	0,063	-0,016	0,968	1,896	0,058
Fantuzzo et al (1995)*	0,732	0,709	0,502	-0,657	2,121	1,033	0,302
Fuchs 1997 I (no disability)*	0,327	0,228	0,052	-0,120	0,774	1,434	0,152
Fuchs 1997 II (disability)*	0,339	0,338	0,114	-0,323	1,001	1,003	0,316
Fuchs et al (2001- II *	0,376	0,492	0,242	-0,588	1,340	0,764	0,445
Fuchs et al (2001) - I*	0,345	0,163	0,027	0,026	0,664	2,117	0,034
Fuchs et al (2002)*	0,166	0,111	0,012	-0,052	0,384	1,495	0,135
Fuchs et al (2008) *	0,631	0,104	0,011	0,427	0,835	6,067	0,000
Galia, M. (2015)*	0,882	0,292	0,085	0,310	1,454	3,021	0,003
Ginsburg-Block and Fantuzo (1997)*	0,659	0,328	0,108	0,016	1,302	2,009	0,045
Greenwood (1991)*	0,180	0,176	0,031	-0,165	0,525	1,023	0,306
Grimm (2004)*	-0,193	0,248	0,061	-0,678	0,292	-0,778	0,436
Hannah, D. (2008)*	0,523	0,169	0,029	0,192	0,854	3,095	0,002
Hawaii (1974)*	0,303	0,148	0,022	0,013	0,593	2,047	0,041
Heintz (1975)*	0,969	0,490	0,240	0,009	1,929	1,978	0,048
Heller (1992)*	0,942	0,283	0,080	0,387	1,498	3,326	0,001
Henson et al (2009)*	0,448	0,239	0,057	-0,020	0,916	1,874	0,061
Hugger (2012)*	-0,189	0,324	0,105	-0,823	0,445	-0,585	0,558
Ivory, T. (2007)*	0,006	0,412	0,170	-0,802	0,814	0,015	0,988
Jitendra (2012)*	0,538	0,156	0,024	0,232	0,844	3,449	0,001
Kane & Alley (1980)*	-0,044	0,319	0,102	-0,669	0,581	-0,138	0,890
Lazarus, k. (2014)*	0,559	0,198	0,039	0,171	0,947	2,823	0,005
Mahan (2010)*	0,584	0,453	0,205	-0,304	1,472	1,289	0,197
McKethan (1982)*	0,227	0,212	0,045	-0,189	0,643	1,071	0,284
Murugan(2015)*	0,986	0,270	0,073	0,456	1,516	3,649	0,000
Nazzal, A. (2002)*	0,322	0,261	0,068	-0,190	0,834	1,234	0,217
Nesselrodt & Alger (2005)*	-0,006	0,136	0,018	-0,273	0,261	-0,044	0,965
Novotni (1985)*	0,056	0,253	0,064	-0,440	0,552	0,221	0,825
Obidoa & Onwubolu (2013)*	0,346	0,112	0,013	0,126	0,566	3,089	0,002
Phillips et al (1993)*	0,372	0,380	0,144	-0,373	1,117	0,978	0,328
Roach et al (1983)*	0,436	0,180	0,032	0,083	0,789	2,422	0,015
Schneck, A. (2010)*	-0,168	0,217	0,047	-0,593	0,257	-0,774	0,439
Shamir et al (2006)*	0,638	0,195	0,038	0,256	1,020	3,272	0,001
Sharpley et al (1983)*	0,860	0,160	0,026	0,547	1,174	5,376	0,000
Sprinthall & Scott (1989)*	0,765	0,369	0,136	0,042	1,488	2,073	0,038
Tella (2013)*	0,996	0,198	0,039	0,608	1,384	5,030	0,000
Topping et al (2011)*	0,025	0,118	0,014	-0,206	0,256	0,212	0,832
Tsuei (2014)*	1,167	0,681	0,464	-0,168	2,502	1,714	0,087
Vazquez, A. (1995)*	0,956	0,610	0,372	-0,240	2,152	1,567	0,117
Xu et al (2001)*	-0,177	0,107	0,011	-0,387	0,033	-1,654	0,098
	0,333	0,028	0,001	0,278	0,388	11,887	0,000

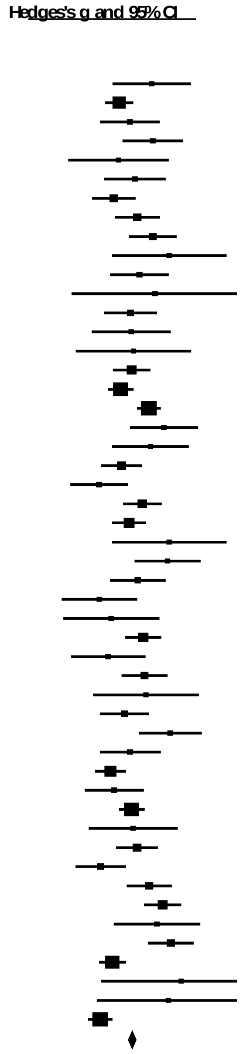

Figure 1. Forest plot.

*Mean effect size indicated at the bottom of the forest plot.

Table 3. Test results for effects moderators

		Fixed effect – Significant and nonsignificant moderator				Mixed effect – nonsignificant moderator		
Variable	**k**	**Q_B**	**Q_{Wg}**	**g**	**95% CI**	**Qb**	**g**	**95% CI**
Year of publication		1.54	125.52***	0.33	[0.28, 0.39]	0.26	0.39	[0.29, 0.48]
After year 2000	28		96.70***	0.31	[0.24, 0.37]		0.36	[0.24, 0.49]
Before year 2000	22		28.81	0.39	[0.29, 0.48]		0.42	[0.26, 0.57]
Design		14.49***	112.56***	0.33	[0.28, 0.39]	1.44	0.38	[0.29, 0.48]
PPSG or POWC	40		24.32***	0.07	[-0.07, 0.21]		0.25	[0.01, 0.49]
PPCG	10		88.24***	0.38	[0.32, 0.44]		0.41	[0.31, 0.51]
Publication		0.02	127.05***	0.33	[0.28, 0.39]	0.00	0.39	[0.39, 0.49]
Not indexed	29		79.63***	0.33	[0.26, 0.41]		0.39	[0.39, 0.52]
Indexed	21		47.42***	0.34	[0.26, 0.42]		0.38	[0.38, 0.54]
Size of the sample		15.25***	111.83***	0.33	[0.28, 0.39]	2.98	0.38	[0.29, 0.48]
> 60 students	22		72.20***	0.26	[0.19, 0.32]		0.31	[0.19, 0.43]
≤ 60 students	28		28.53	0.49	[0.39, 0.58]		0.48	[0.33, 0.62]
Fixed vs Reciprocal		0.02	111.81***	0.35	[0.30, 0.41]	0.01	0.40	[0.30, 0.49]
Fixed	28		84.83***	0.36	[0.29, 0.42]		0.40	[0.26, 0.53]
Reciprocal	21		34.98*	0.35	[0.25, 0.41]		0.39	[0.25, 0.54]
Same-age vs cross-age		0.29	119.54***	0.35	[0.30, 0.41]	0.37	0.40	[0.30, 0.50]
Cross-age	23		77.52***	0.34	[0.26, 0.42]		0.37	[0.22, 0.51]
Same-age	26		42.03**	0.37	[0.29, 0.45]		0.43	[0.29, 0.57]
Educational level		16.16***	110.91***	0.33	[0.28, 0.39]	1.94	0.38	[0.29, 0.47]
Kindergarten	2		0.00	0.35	[0.05, 0.65]		0.35	[0.05, 0.65]
Elementary	26		64.63***	0.42	[0.34, 0.50]		0.45	[0.31, 0.60]
Secondary	18		30.20*	0.30	[0.21, 0.39]		0.32	[0.19, 0.45]
University	4		16.08***	0.06	[-0.11, 0.22]		0.30	[-0.15, 0.75]
Length of the session		1.41	81.45***	0.38	[0.31, 0.44]	0.42	0.40	[0.30, 0.50]
> 30 minutes	21		38.10***	0.33	[0.22, 0.43]		0.37	[0.22, 0.52]
≤ 30 minutes	22		43.36***	0.41	[0.32, 0.48]		0.44	[0.30, 0.56]
Frequency of sessions		0.94	100.40***	0.37	[0.30, 0.43]	0.33	0.40	[0.30, 0.50]
> 2 times a week	24		64.92***	0.39	[0.31, 0.47]		0.39	[0.31, 0.47]
≤ 2 times a week	20		35.48**	0.33	[0.23, 0.43]		0.33	[0.23, 0.43]
Length of the program		14.59***	112.48***	0.33	[0.28, 0.39]	2.422	0.38	[0.09, 0.48]
> 15 weeks	21		47.57***	0.22	[0.14, 0.30]		0.30	[0.16, 0.44]
≤ 15 weeks	29		64.91***	0.43	[0.36, 0.51]		0.45	[0.32, 0.60]
When PT takes place		11.30***	115.78***	0.33	[0.28, 0.39]	2.09	0.38	[0.29, 0.48]
Out of school time	11		34.66***	0.18	[0.08, 0.29]		0.26	[0.08, 0.45]
During school time	39		81.12***	0.39	[0.33, 0.46]		0.42	[0.31, 0.53]
Assignment of pairs		0.242	127.01***	0.33	[0.28, 0.39]	0.349	0.39	[0.30, 0.49]
Not structured	33		92.42***	0.35	[0.27, 0.42]		0.41	[0.29, 0.54]
Structured	17		34.59***	0.32	[0.23, 0.40]		0.35	[0.18, 0.51]
Disability / At risk		3.17	121.29***	0.34	[0.29, 0.40]	0.03	0.39	[0.29, 0.49]
No	21		78.27***	0.30	[0.21, 0.37]		0.36	[0.24, 0.53]
Yes	28		43.02*	0.40	[0.32, 0.48]		0.40	[0.26, 0.54]
Tutor achievement		14.01***	113.06***	0.33	[0.29, 0.48]	2.81	0.38	[0.29, 0.48]
Average	15		57.22**	0.21	[0.12, 0.46]		0.28	[0.12, 0.45]
High	19		41.33**	0.39	[0.25, 0.55]		0.40	[0.25, 0.55]
Low	16		14.51	0.48	[0.31, 0.68]		0.49	[0.31, 0.68]
Training		6.24*	105.36***	0.32	[0.27, 0.38]	0.47	0.38	[0.29, 0.47]
> 2 hours or sessions	14		26.25*	0.43	[0.33, 0.53]		0.43	[0.26, 0.60]
≤ 2 times a week	34		79.11*	0.27	[0.21, 0.34]		0.36	[0.24, 0.47]
Same-age Fixed vs Same-age Reciprocal		0.11	0.12*	0.39	[0.30, 0.47]	0.03	0.44	[0.32, 0.55]
Same-age Fixed	6		7.39	0.41	[0.26, 0.56]		0.45	[0.23, 0.68]
Same-age Reciprocal	19		29.24*	0.38	[0.28, 0.47]		0.43	[0.29, 0.57]

* p ≤ 0.05 , ** p ≤ 0.01 , *** p ≤ 0.001

Q1. The ages of the participating students turned out not to be a significant moderator for academic performance, since we obtained effect sizes that were very similar for cross-age and same-age tutoring (g = 0.34 and g=0.37) with a very low heterogeneity (Q_B = 0.29).

Q2. The role of the participants was also not a significant moderator, with practically identical effect sizes for fixed and reciprocal tutoring (g= 0.36 and 0.35) with a very low heterogeneity (Q_B = 0.02).

Q3. The role of the students within same-age tutoring relationships was also not a significant moderator, with very similar effect sizes for fixed same-age and reciprocal same-age tutoring (g= 0.41 and g = 0.38) with a very low heterogeneity (Q_B = 0.11).

Q4. The educational stage at which the studies were carried out did turn out to be a significant moderator, with a very low effect size for the interventions carried out at university (g = 0.06) as opposed to those carried out in preschool/kindergarten and primary and secondary levels (g = 0.35, g = 0.42 and g = 0.30, respectively) with an elevated heterogeneity (Q_B = 16.16, p = 0.001). An additional comparative study that only included the primary and secondary educational stages also showed that the educational stage was a significant moderator, showing a moderate heterogeneity (Q_B = 4.14, p = 0.042)

Q5. The disability or academic risk of the students did not turn out to be a significant moderator, although it was close, since the effect sizes for students with and without disability or at academic risk were quite different (g= 0.29 and g=0.40), with a moderate heterogeneity close to being significant (Q_B = 3.17, p = 0.075).

Q6. The type of design did turn out to be a significant moderator, with effect sizes that were very different for the PPSG and POWC types on the one hand, and for the PPCG on the other (g= 0.074 and g=0.38) with a high heterogeneity that was highly significant (Q_B = 14.50 and p<0.001).

Q7. Session length did not turn out to be a significant moderator, with similar effect sizes for peer tutoring sessions longer than 30 minutes and less than or equal to 30 minutes (g = 0.33 and g = 0.41) and a low heterogeneity (Q_B = 1.41). A later study comparing sessions longer than 40 minutes with others that were shorter or equal in length also did not show significant results.

Q8. The frequency with which the sessions were held was also not a significant moderator, showing that for programs with more than 2 sessions per week effect sizes similar to those with 1 or 2 sessions were obtained (g = 0.39 and g = 0.33) with a low heterogeneity (Q_B = 1.41).

Q9. The duration of the programs was a significant moderator, with effect sizes that were much lower (g = 0.22 and g = 0.43) for long-duration programs (more than 15 weeks) than for those of short duration (15 weeks or more), showing a high heterogeneity (Q_B = 14.59 and p<0.001).

Q10. The level of knowledge of the tutors was a significant moderator, showing large differences between the effect sizes of tutors with a medium knowledge level (g= 0.21) and those with a high or low level of knowledge (g = 0.39 and g = 0.48) with an elevated heterogeneity (Q_B = 14.01 and p=0.001). The comparison between tutors with high and low level of knowledge did not turn out to be significant and had a low heterogeneity (Q_B = 1.22 and p=0.269).

Q11. The time of the day in which the tutoring took place also turned out to be a significant moderator, with much lower effect sizes in the programs that took place outside school hours (g = 0.18) than in those that took place during the school day (g = 0.39), showing an elevated heterogeneity (Q_B = 11.30 and p=0.001).**Q12.** The structuring of the students in the program did not turn out to be a significant moderator, with effect sizes similar in the cases in which there had been a structuring of the participating students and those where there had not (g = 0.35 and g = 0.32), with a low heterogeneity (Q_B = 0.24).

Q13. The year of publication was also not a significant moderator, with effect sizes of g = 0.31 and g = 0.39 for the studies published before and after the year 2000, demonstrating a low heterogeneity (Q_B = 1.54).

Q14. The sample size did turn out to be a significant moderator, with lower effect sizes for studies with more than 60 students (g = 0.26) than for studies with 60 or fewer students (g = 0.49), with an elevated heterogeneity (Q_B = 14.59 and p<0.001).

Q15. The fact that a study was published in a journal or not did not turn out to be a significant moderator, with very similar effect sizes for both non-indexed and indexed publications (g = 0.33 and g = 0.34), with a very low heterogeneity (Q_B = 0.02).

Q16. Training was a significant moderator, showing how studies in which the participants had received more than two hours or two sessions of training obtained greater effect sizes than those in which the participants had not received this training (g = 0.43 and g = 0.27), with a moderate heterogeneity (Q_B = 6.24 and p=0.013).

DISCUSSION

In this section we will analyze each of the research questions posed in this article.

The fact that we did not encounter significant differences between cross-age and same-age tutoring is in agreement with other prior research projects about peer tutoring (Scruggs & Osguthrope, 1985) and also with other previous meta-analyses (Leung, 2015), showing that age has practically no influence on the effectiveness of the tutoring (Q1). In addition, the fact that the effectiveness of fixed or reciprocal tutoring was nearly identical corresponds with findings of previous studies (Menesses & Gresham, 2009; Leung 2015). All of this indicates that peer tutoring, independently of its typology, is effective and that the differences in effectiveness between the different types of peer tutoring is minimal (Q2 and Q3).

The educational level of the participants was a significant moderator, showing that research in primary and secondary education gives better results than those obtained from university or preschool/kindergarten contexts (Q4). Nevertheless, although this is in consonance with Leung's (2015) meta-analysis, the lower number of surveys available in this research project for the stages of higher education and preschool/kindergarten means that the interpretation of this result ought to be extremely cautious.

In addition, the fact that student disability or academic risk was not significant correlated with previous research, such as that of Osguthrope & Scruggs (1986) and Byrd (1990), which indicates that peer tutoring is effective regardless of the academic level of the students (Q5). This is another item of evidence that shows how inclusive practices can be developed in the ordinary classroom that attend to and take into account the diversity of the student body. In this way, schools can avoid segregation that makes educational inclusion difficult.

In regards to study design, the high degree of statistical significance invites us to keep in mind the publications of Morris & DeShon (2002) and Morris (2007) (Q6). In them the authors emphasize that one must exercise caution when combining studies with different designs in a single meta-analysis. We should bear in mind, in addition, that of the 8 articles excluded for having an effect size higher than 1.2, three were PPCG, another three were PPSG and the other two POWC. Of them, only three articles had an effect size greater than 2. Greenfield & McNeil (1987) (PPSG, g = 3.961), Mohan (1972) (POWC, g = 3.209) and Menesses & Gresham (2009) (PPCG, g = 2.239). Therefore, we should keep in mind the importance of design when carrying out research on peer tutoring, with the most recommended design being PPCG because of its more experimental character and because it shows more controlled effect sizes. When using this type of design we assure ourselves of having a more rigorous control of all the variables that are involved in the process. However, we should be careful with the control groups, assuring that they do not benefit from the experiences of learning. One should therefore ensure an appropriate intervention for control groups once the study has been completed.

On the other hand, the fact that neither the frequency nor the duration of the sessions was a significant moderator, but the total duration of the program was, coincides with the results obtained in the meta-analysis carried out by Leung (2015) (Q7, Q8 and Q9). Thus, it has been demonstrated that with short interventions one can obtain very effective results.

In regards to the level of the tutors (a significant moderator), the fact that the high and low level tutors obtained effect sizes much higher than those of the middle level, in addition to coinciding with the results obtained by Leung (2015) also coincides with what we find in the articles of Gaustad (1993) and Lippit (1976) (Q10). According to these researchers, "Tutors who have struggled academically may be more patient and understanding than those who haven't," "empathy contributes greatly to low-achievers" and that "tutors often 'pick up on things teachers weren't able to' because they had experienced similar problems a few years earlier." This, then, is more evidence that peer tutoring is a cooperative method of learning that facilitates the inclusion of students who encounter barriers to learning and participation.

The fact that the tutoring programs that took place outside school time gave significantly lower results in comparison with those that took place during school hours is related to the meta-analysis carried out by Lauer et al. (2006) for programs outside the school hours of reading and mathematics (Q11). In that study, we see that although the results were significant, the effect sizes are low, which would indicate that the time at which the peer tutoring was carried out was a significant moderator.

On the other hand, the fact that the structuring of the pairs was not a significant moderator is in conflict with various authors who hold that it is important (Q12). Nevertheless, the results obtained in this meta-analysis are in agreement with those of Leung (2015), according to which the effect sizes among the structured and non-structured peer tutoring programs are very similar. Therefore, it is clear how the help of a peer is a critical factor for this method. For we understand people of a similar social group who are not teachers or educational professionals, and who help others to learn and themselves learn in teaching (Topping, 1998).

The fact that the date of publication was not a significant moderator indicates that the publication of results and their magnitude has not been in any way affected by the passage of time (Q13). In addition, the fact that the sample size was a significant moderator (Q14) can be explained from the logical point of view of control of the students. When the group of students is smaller, control of interventions is easier and the peer tutoring proceeds adequately and according to theory. Nevertheless, if the number of participants is very high, it may be impossible to effectively control the intervention, which produces worse results for student experiences.

The fact that the type of publication was not a significant moderator (publications in indexed journals vs. in non-indexed sources) contradicts the results of previous meta-analyses (Cohen et al., 1982 and Smith 1980), although it is backed up by the publications of more recent authors who support the inclusion of so-called grey literature, cf. Cooper et al., (2009) (Q15). Therefore, on the basis of the results of this study, we would recommend the inclusion of non-indexed publications when performing meta-analyses of peer tutoring. Grey literature is transforming the spread of scientific knowledge and its vehicles of transmission, making available studies of great interest for the scientific community.

The fact that training showed up as a significant moderator (Q16) agrees with previous studies (Fuchs et al., 1994), although it contradicts the results obtained by Leung (2015). Various authors, such as Gordon (2005), have emphasized the necessity for adequate training in order to be a successful peer tutor.

In regards to the fact that some of the moderators above mentioned were significant for the fixed effects model but not for the mixed effects, this suggests that these moderators could be obtained if different participants were selected from same population included in the present study, but couldn`t be found on all peer tutoring interventions beyond the studies included in the present investigation (Leung, 2015).

The findings in this study offer several implications for practitioners and policy-makers. From the results it is shown that it is important to take a PPCG design when evaluating this type of experiences. Besides, very long or intense peer tutoring programs don't necessarily imply better results. Same-age vs cross-age and fixed peer tutoring vs reciprocal peer tutoring is not a major issue as all types show similar effectiveness. Primary and secondary levels are regarded as the best educational stages to perform this type of experiences. Students with disabilities or at risk can participate in these type of programs showing great results. Furthermore, low-achieving students can also perform as great tutors if they are properly assigned to a peer tutoring task. Selecting high-achieving students as tutors is recommended, but the potentiality of low level students must be taken into account as they show the greatest results. Caution when performing peer tutoring programs outside the school is recommended and it is suggested to perform them during school time. Employing a very big sample of participants may not be such a great idea. Training the students at least two hours before the implementation of the program is recommended.

LIMITATIONS

The results of this study are promising and can help in the implementation of peer tutoring in mathematics. Nonetheless, they should be interpreted with caution, keeping in mind a series of limitations. In the first place, one must consider that, despite the fact that 50 studies is more than sufficient to carry out a meta-analysis, there are some experiences that we have not been able to include in this research project due to certain restrictions (language, restrictions imposed on access...). In addition, in some subgroups, such as in the case of the educational level of the students, the number of studies is very small, and it would be helpful if those groups were expanded in future studies. We should also bear in mind that the criteria for excluding certain articles in order not to have an excessively high effect size was subjective, and as a result we do not know what would have occurred if those studies had ultimately been added to this meta-analysis, as well as those which could not be accessed for various reasons (studies in languages other than Spanish and English, inaccessible studies…). Furthermore, although the inclusion of articles with models other than PPCG ended up not influencing the sample, since design was not a significant moderator, there are doubts about the experimental quality of studies with PPSG and POWC designs. On the other hand, the fact that we included articles from the grey literature in this study meant that the methodological quality of the studies was not filtered by top-level experts, although lower-ranking experts did perform filtering.

Looking towards future research projects, we recommend the repetition of this meta-analysis once there has been documentation of a greater number of peer tutoring experiences in mathematics. This will make it possible to apply stronger filters on experimental and methodological quality, assuring a greater reliability of the results, while also making it possible to study the appropriateness of including or not including grey literature in the meta-analysis. The study of certain variables, such as the educational level of the students, will be strengthened by the inclusion of a greater number of studies. Furthermore, one will be able to discuss, with a much more empirical character, the results obtained for variables such as sample size, since currently the absence of studies that deal with these issues means that discussion of them depends more on the subjective opinion of the author than on a well-founded theoretical base.

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

*Alegre, F. J. (2014) La incidencia de la tutoría entre iguales fija en las variables cognitivas y emocionales del alumnado. Una experiencia en el área de matemáticas. *Universitat Jaume I.*

*Allsopp, D. H. (1997). Using classwide peer tutoring to teach beginning algebra problem-solving skills in heterogeneous classrooms. *Remedial and Special Education*, *18*(6), 367-379. doi:10.1177/074193259701800606

*Azcoitia, C. (1989). Structured peer tutoring in Chicago's vocational education program. *Northern Illinois University.*

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7(6), 543-554*. doi:10.1177/1745691612459060

*Bar-Eli, N., & Raviv, A. (1982). Underachievers as tutors. *The Journal of Educational Research*, *75*(3), 139-143. doi:10.1080/00220671.1982.10885370

Becker, L. A. (2000). Effect size (ES). *University of Colorado Colorado Springs*.

*Beirne-Smith, M. (1991). Peer tutoring in arithmetic for children with learning disabilities. *Exceptional Children*, *57*(4), 330-337.

Boud, D. (1988). Moving towards autonomy. *Developing student autonomy in learning*, *2*, 17-39.

Borenstein, M., Cooper, H., Hedges, L., & Valentine, J. (2009). Effect sizes for continuous data. *The handbook of research synthesis and meta-analysis*, *2*, 221-235.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis* (pp. i-xxix). John Wiley & Sons, Ltd.

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School Psychology Review*, *42*(1), 39.

Britz, M. W. (1989). The Effects of Peer Tutoring on Mathematics Performance: A Recent Review. *BC Journal of Special Education*, *13*(1), 17-33.

*Brush, T. (1997). The effects on student achievement and attitudes when using integrated learning systems with cooperative pairs. *Educational Technology Research and Development*, *45*(1), 51-64. doi:10.1007/BF02299612

Byrd, D. E. (1990). Peer tutoring with the learning disabled: A critical review. *The Journal of Educational Research*, *84*(2), 115-118. doi:10.1080/00220671.1990.10886001

*Cairo III, L., & Craig, J. (2005). Cross-Age Tutoring Phase II--An Experiment. *Appalachia Educational Laboratory at Edvantia (NJ1)*.

*Calhoon, 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 and Special Education*, *24*(4), 235-245. doi:10.1177/07419325030240040601

Carlson, K. D., & Schmidt, F. L. (1999). Impact of experimental design on effect size: Findings from the research literature on training. *Journal of applied psychology*, *84*(6), 851.

Cohen, P. A., Kulik, J. A., & Kulik, C. L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American educational research journal*, *19*(2), 237-248. doi:10.3102/00028312019002237

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation.

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Damon, W., & Phelps, E. (1989). Critical distinctions among three approaches to peer education. *International journal of educational research*, *13*(1), 9-19. doi:10.1016/0883-0355(89)90013-X

Delquadri, J., Greenwood, C. R., Whorton, D., Carta, J. J., & Hall, R. V. (1986). Classwide peer tutoring. *Exceptional children*, *52*(6), 535-542.

*Dennis, L. M. G. (2013). Multiplication Achievement and Self-Efficacy in Third-and Fifth-Grade Students: Effects of Cross-Age Peer Tutoring and Skill Training. *The University of Arizona*.

Duran, D., Torró, J., & Vilà, J. (2003). *Tutoria entre iguals: un mètode d'aprenentatge cooperatiu per a la diversitat: de la teoria a la pràctica*. Institut de Ciències de l'Educació, Universitat Autònoma de Barcelona.

*Durand, S. C. (2008). Using classwide peer tutoring to increase high school math students' academic performance. *Culminating Experience Action Research Projects, Volume 10, Spring 2007*, 182.

Durlak, J. A. (1995). Understanding meta-analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics (pp. 319–352). Washington, DC: American Psychological Association.*

*Fantuzzo, J. W., Davis, G. Y., & Ginsburg, M. D. (1995). Effects of parent involvement in isolation or in combination with peer tutoring on student self-concept and mathematics achievement. *Journal of Educational Psychology*, *87*(2), 272.

*Fantuzzo, J. W., King, J. A., & Heller, L. R. (1992). Effects of reciprocal peer tutoring on mathematics and school adjustment: A component analysis. *Journal of Educational Psychology*, *84*(3), 331.

Fuchs, D., Fuchs, L. S., Thompson, A., Svenson, E., Yen, L., Al Otaiba, S., ... & Saenz, L. (2001). Peer-Assisted Learning Strategies in Reading Extensions for Kindergarten, First Grade, and High School. *Remedial and Special Education*, *22*(1), 15-21. doi:10.1177/074193250102200103

*Fuchs, L. S., Fuchs, D., & Karns, K. (2001). Enhancing kindergartners' mathematical development: Effects of peer-assisted learning strategies. *The Elementary School Journal*, 495-510. doi:10.1086/499684

Fuchs, L. S., Fuchs, D., Bentz, J., Phillips, N. B., & Hamlett, C. L. (1994). The nature of student interactions during peer tutoring with and without prior training and experience. *American Educational Research Journal*, *31*(1), 75-103. doi:10.3102/00028312031001075

*Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *The Elementary School Journal*, 223-249. doi:10.1086/461863

Fuchs, L. S., Fuchs, D., Powell, S. R., Seethaler, P. M., Cirino, P. T., & Fletcher, J. M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly*, *31*(2), 79-92. doi:10.2307/20528819

*Fuchs, L. S., Fuchs, D., Yazdian, L., & Powell, S. R. (2002). Enhancing first-grade children's mathematical development with peer-assisted learning strategies. *School Psychology Review*, *31*(4), 569.

*Fuchs, L. S., Seethaler, P. M., Powell, S. R., Fuchs, D., Hamlett, C. L., & Fletcher, J. M. (2008). Effects of preventative tutoring on the mathematical problem solving of third-grade students with math and reading difficulties. *Exceptional Children*, *74*(2), 155-173.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. Review *of Educational Research*, *79*(3), 1202-1242. doi:10.3102/0034654309334431

*Galia, M. L. D. (2015) Cross-age peer tutoring as intervention to cultural minority students with learning difficulties in mathematics. *Journal of Teaching and Education.*

Gaustad, J. (1993). *Peer and cross-age tutoring*. University of Oregon.

*Ginsburg-Block, M., & Fantuzzo, J. (1997). Reciprocal peer tutoring: An analysis of" teacher" and" student" interactions as a function of training and experience. *School Psychology Quarterly*, *12*(2), 134. doi:10.1037/h0088955

Glass, G. V., Smith, M. L., & McGaw, B. (1981). *Meta-analysis in social research*. Sage Publications, Incorporated.

Gordon, E. E. (2005). *Peer tutoring: A teacher's resource guide*. R&L Education.

Greenfield, S. D., & McNeil, M. E. (1987). Improving Math Performance through a Peer Tutoring Program.

*Greenwood, C. R. (1991). Classwide peer tutoring: Longitudinal effects on the reading, language, and mathematics achievement of at-risk students. *Reading, Writing, and Learning Disabilities*, *7*(2), 105-123. doi:10.1080/0748763910070203

*Grimm, S. R. (2004). An Analysis and Synthesis of Research on Classroom Peer Tutoring Methods that Benefit Intermediate Math Students. *Cerdaville University.* doi:10.15385/tmed.2004.1

*Hannah, D. C. (2008). Attitudinal study: The interaction of students taking calculus and prerequisite courses while participating in peer tutorials. *Georgia State University.*

*Hawaii (1974). Buddy-Tutor Project. Hilo Intermediate School. Final Report, March-July, 1974. *Hawaii University.*

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *Psychological bulletin*, *92*(2), 490. doi:10.1037/0033-2909.92.2.490

Hedges, L., & Olkin, I. (1985). Statistical methods in meta-analysis. *New York: Academic Press. Hedges, LV, & Pigott, TD (2001). The power of statistical tests in meta-analysis. Psychological Methods*, *6*, 203-217. doi:10.2307/1164953

*Heintz, P. (1975). High School Peer Tutoring (Homework Helpers) Program; School Year, 1974-1975.

*Heller, L. R. (1992). Reciprocal peer tutoring and parent partnership: Does parent involvement make a difference? *University of Pennsylvania***.**

*Henson, L. D., Hagos, L. C., & Villapando, R. A. (2009). The effectiveness of reciprocal peer tutoring (RPT) on the academic performance of students in mathematics. *Proceedings of the 2nd International Conference of Teaching and Learning (ICTL 2009). INTI University College, Malaysia*

Higgins, J. P., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions* (Vol. 5). Chichester: Wiley-Blackwell.

*Hugger, K. A. (2014). Evaluating the effects of Peer-Assisted Learning Strategies (PALS) in mathematics plus an anxiety treatment on achievement and anxiety of third grade students. *University of Southern Maine.*

*Ivory, T. S. (2007). Improving Mathematics Achievement of Exceptional Learners through Differentiated and Peer-Mediated Instruction. *Nova Southeastern University*.

*Jitendra, A. K., Dupuis, D. N., & Rodriguez, M. C. (2012). Effectiveness of Small-Group Tutoring Interventions for Improving the Mathematical Problem-Solving Performance of Third-Grade Students with Mathematics Difficulties: A Randomized Experiment. *Society for Research on Educational Effectiveness*.

*Kane, B. J., & Alley, G. R. (1980). A peer-tutored, instructional management program in computational mathematics for incarcerated, learning disabled juvenile delinquents. *Journal of Learning Disabilities*, *13*(3), 148-151. doi:10.1177/002221948001300308

Kroesbergen, E. H., & Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs a meta-analysis. *Remedial and special education*, *24*(2), 97-114. doi:10.1177/07419325030240020501

Kunsch, C. A., Jitendra, A. K., & Sood, S. (2007). The Effects of Peer-Mediated Instruction in Mathematics for Students with Learning Problems: A Research Synthesis. *Learning Disabilities Research & Practice*, *22*(1), 1-12. doi:10.1111/j.1540-5826.2007.00226.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in psychology, 4, 863. doi:10.3389/fpsyg.2013.00863

Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of educational research*, *76*(2), 275-313. doi:10.3102/00346543076002275

*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. *African Journal for the Psychological Studies of Social Issues, 17(1), 96-105.* doi:10.1080/0020739X.2017.1342284

Ledgerwood, A., & Sherman, J. W. (2012). Short, sweet, and problematic? The rise of the short report in psychological science. *Perspectives on Psychological Science, 7(1), 60-66.* doi:10.1177/1745691611427304

Leung, K. C., Marsh, H. W., & Craven, R. G. (2005). Are peer tutoring programs effective in promoting academic achievement and self-concept in educational settings: A meta-analytical review. In *Annual Conference, Cairns*.

Leung, K. C. (2015). Preliminary empirical model of crucial determinants of best practice for peer tutoring on academic achievement. *Journal of Educational Psychology*, *107*(2), 558. doi:10.1037/a0037698

Lippit, P. (1976). Learning through close-age peer tutoring: Why and how. *Children as Teachers: Theory and Research on Tutoring, edited by VL Allen. New York: Academic Press.*

Luca, J., & Clarkson, B. (2002). Promoting Student Learning through Peer Tutoring--A Case Study. *ED-MEDIA 2002 World Conference on Educational Multimedia, Hypermedia & Telecommunications.* Proceedings (14th, Denver, Colorado, June 24-29, 2002).

*Mahan, A. (2010). Effects of Peer Tutoring on the Mathematics Achievement of Fifth Grade Students with Academic Learning Disabilities. *Valdosta State University.*

Marín-Martínez, F., & Sánchez-Meca, J. (2010). Weighting by inverse variance or by sample size in random-effects meta-analysis. *Educational and Psychological Measurement*, *70*(1), 56-73. doi:10.1177/0013164409344534

*McKethan, L. D. (1982). An attitudinal and achievement comparison of mathematics deficient Lincoln University freshmen resulting from structured peer tutoring versus no peer tutoring in mathematics. *Temple University.*

Meca, J. S. (2010). Cómo realizar una revisión sistemática y un meta-análisis. *Aula abierta*, *38*(2), 53-64.

Menesses, K. F., & Gresham, F. M. (2009). Relative efficacy of reciprocal and nonreciprocal peer tutoring for students at-risk for academic failure. *School Psychology Quarterly*, *24*(4), 266. doi:10.1037/a0018174

Mohan, M. (1972). Peer tutoring as a technique for teaching the unmotivated. Fredonia, NY: State University of New York. *Teacher Education Research Center*.

Morris, S. B. (2007). Estimating effect sizes from the pretest-posttest-control group designs. *Organizational Research Methods*. doi:10.1177/1094428106291059

Morris, S. B. & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological methods*, *7*(1), 105. doi:10.1037/1082-989X.7.1.105

Murphy, K. R., Myors, B., & Wolach, A. (2014). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests*. Routledge.

*Murugan, S. (2015). Effectiveness of peer tutoring with student to student counselling in learning Mathematics at Standard IX level. *Alagappa University.*

*Nazzal, A. (2002). Peer tutoring and at-risk students: An exploratory study. *Action in teacher Education*, *24*(1), 68-80. doi:10.1080/01626620.2002.10463268

*Nesselrodt, P. S., & Alger, C. L. (2005). Extending opportunity to learn for students placed at risk. *Journal of Education for Students Placed at Risk*, *10*(2), 207-224. doi:10.1207/s15327671espr1002_7

*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. *University of Pennsylvania.*

Nuijten, M. B., van Assen, M. A., Veldkamp, C. L., & Wicherts, J. M. (2015). The replication paradox: Combining studies can decrease accuracy of effect size estimates. *Review of General Psychology, 19(2), 172.* http://dx.doi.org/10.1037/gpr0000034

*Obidoa, M. A., Eskay, M., & Onwubolu, C. O. (2013). Remedial Help in Inclusive Classrooms: Gender Differences in the Enhancement of Mathematics Achievement of Students through PAL (Peer-Assisted Learning). *US-China Education Review A, ISSN 2161-623X. March 2013, Vol. 3, No. 3, 172-180.*

Osguthorpe, R. T., & Scruggs, T. E. (1986). Special education students as tutors: A review and analysis. *Remedial and Special Education*, *7*(4), 15-25. doi:10.1177/074193258600700405

*Phillips, N. B., Hamlett, C. L., Fuchs, L. S., & Fuchs, D. (1993). Combining classwide curriculum-based measurement and peer tutoring to help general educators provide adaptive education. *Learning Disabilities Research & Practice*.

*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*.

Robinson, D. R., Schofield, J. W., & Steers-Wentzell, K. L. (2005). Peer and cross-age tutoring in math: Outcomes and their design implications. *Educational Psychology Review*, *17*(4), 327-362.

Rohrbeck, C. A., Ginsburg-Block, M. D., Fantuzzo, J. W., & Miller, T. R. (2003). Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, *95*(2), 240. doi:10.1037/0022-0663.95.2.240

Rosenthal, R., & Rubin, D. B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin*, *92*(2), 500. doi:10.1037/0033-2909.92.2.500

Rosenthal, R. (1984). Meta-analytic procedures for social science research. Beverley Hills, CA: Sage.

*Schneck, A. J. (2010). *The effects of after-school peer tutoring programs in mathematics* (Doctoral dissertation, University of Minnesota).

Scruggs, T. E., & Osguthorpe, R. T. (1985). Tutoring Interventions within Special Education Settings: A Comparison of Cross-Age and Peer Tutoring.

*Shamir, A., Tzuriel, D., & Rozen, M. (2006). Peer Mediation The Effects of Program Intervention, Maths Level, and Verbal Ability on Mediation Style and Improvement in Maths Problem Solving. *School Psychology International*, *27*(2), 209-231. doi:10.1177/0143034306064548

*Sharpley, A. M., Irvine, J. W., & Sharpley, C. F. (1983). An examination of the effectiveness of a cross-age tutoring program in mathematics for elementary school children. *American Educational Research Journal*, *20*(1), 103-111. doi:10.3102/00028312020001103

*Sprinthall, N. A., & Scott, J. R. (1989). Promoting psychological development, math achievement, and success attribution of female students through deliberate psychological education. *Journal of Counseling Psychology*, *36*(4), 440. doi:10.1037/0022-0167.36.4.440

Swanson, H. L., Carson, C., & Sachse-Lee, C. M. (1996). A Selective Synthesis of Intervention Research for Students with Learning Disabilities. *School Psychology Review*, *25*(3), 370-91.

*Tella, A. (2013). The effect of peer tutoring and explicit instructional strategies on primary school pupils learning outcomes in mathematics. *Bulgarian Journal of Science and Education Policy*, *7*(1), 5.

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research: A simplified methodology. *Work-Learning Research*, 1-9.

Topping, K. (2000). *Tutoring*. Genf, Switzerland: International Academy of Education.

Topping, K., & Ehly, S. (Eds.). (1998). *Peer-assisted learning*. Routledge.

Topping, K. J. (1996). The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher education*, *32*(3), 321-345. doi:10.1007/BF00138870

*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. doi:10.1080/00131881.2011.552239

*Tsuei, M. (2012). Using synchronous peer tutoring system to promote elementary students' learning in mathematics. *Computers & Education*, *58*(4), 1171-1182. doi:10.1016/j.compedu.2011.11.025

*Vazquez, I. R. (1995). Effect of Peer Tutoring on the Arithmetic Skills of Severely Emotionally Disturbed Sixth Grade Students. *Metlife Fellow Teachers Network Policy Institute. Miami, Florida.*

White, H. D. (2009). Scientific communication and literature retrieval. *The handbook of research synthesis and meta-analysis*, *2*, 51-71.

*Xu, Y., Hartman, S., Uribe, G., & Mencke, R. (2001). The effects of peer tutoring on undergraduate students' final examination scores in mathematics. *Journal of College Reading and Learning*, *32*(1), 22-31. doi:10.1080/10790195.2001.10850123