Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Learning

Do peer-tutored reading interventions improve reading comprehension in secondary age pupils attending mainstream schools?

Section 1: Summary

This review focuses on "peer tutoring" reading interventions. Peer tutoring typically involves one pupil taking on the role of "tutor" and one pupil taking on the role of "tutee" in order to learn a specific skill, such as reading. This review sought to look at *all* types of peer tutoring.

This review concludes that the weighted evidence from the reviewed studies shows that peer tutoring does appear to improve reading comprehension for secondary school pupils with reading difficulties. However, there should be caution regarding its generalisability to all pupils as these studies specifically looked at improving reading skills in pupils with ascertained reading difficulties.

In terms of practice, a peer tutoring reading approach could be an effective, simple and low cost intervention to introduce to a group of secondary pupils with reading difficulties.

Section 2. Introduction

Peer Tutoring

Peer tutoring, in its broadest sense, involves a more able child working cooperatively, on a particular skill, with a less able child, under close supervision of a teacher. The more able child ("tutor") may be the same age as the less able child ("tutee") or they can be from different year groups, termed as "cross age" peer tutoring. Peer tutoring can also be "reciprocal", whereby the role of tutor and tutee rotates so that both members of the dyad practise the same skill. Programs have also been developed that involve small group peer

tutoring. Over the years many different programs have been developed based on this concept of children teaching children. Details are provided below of the different approaches identified in the review papers.

It is not a new concept and has been seen informally for centuries, its efficacy, however, has only been evidenced in the past 35 years (Veerkamp, Kamps & Cooper, 2007). There is evidence that both the tutee *and* the tutor improve skills, as well as self-esteem (Miller, Topping & Thurston, 2010; Topping, 1987, 1989), which will be of particular importance to schools.

Numerous studies have demonstrated the benefits of peer tutoring on pupils' academic achievement (Maheady, Harper & Malette, 2001; Topping & Ehly, 1998). In particular, studies have found positive effects of peer tutoring in relation to reading skills, such as accuracy, comprehension and fluency (Dufrene et al., 2006; Mathes & Fuchs, 1994). Importantly, a recent meta- analysis reported that peer tutoring can produce long term gains (Suggate, 2014).

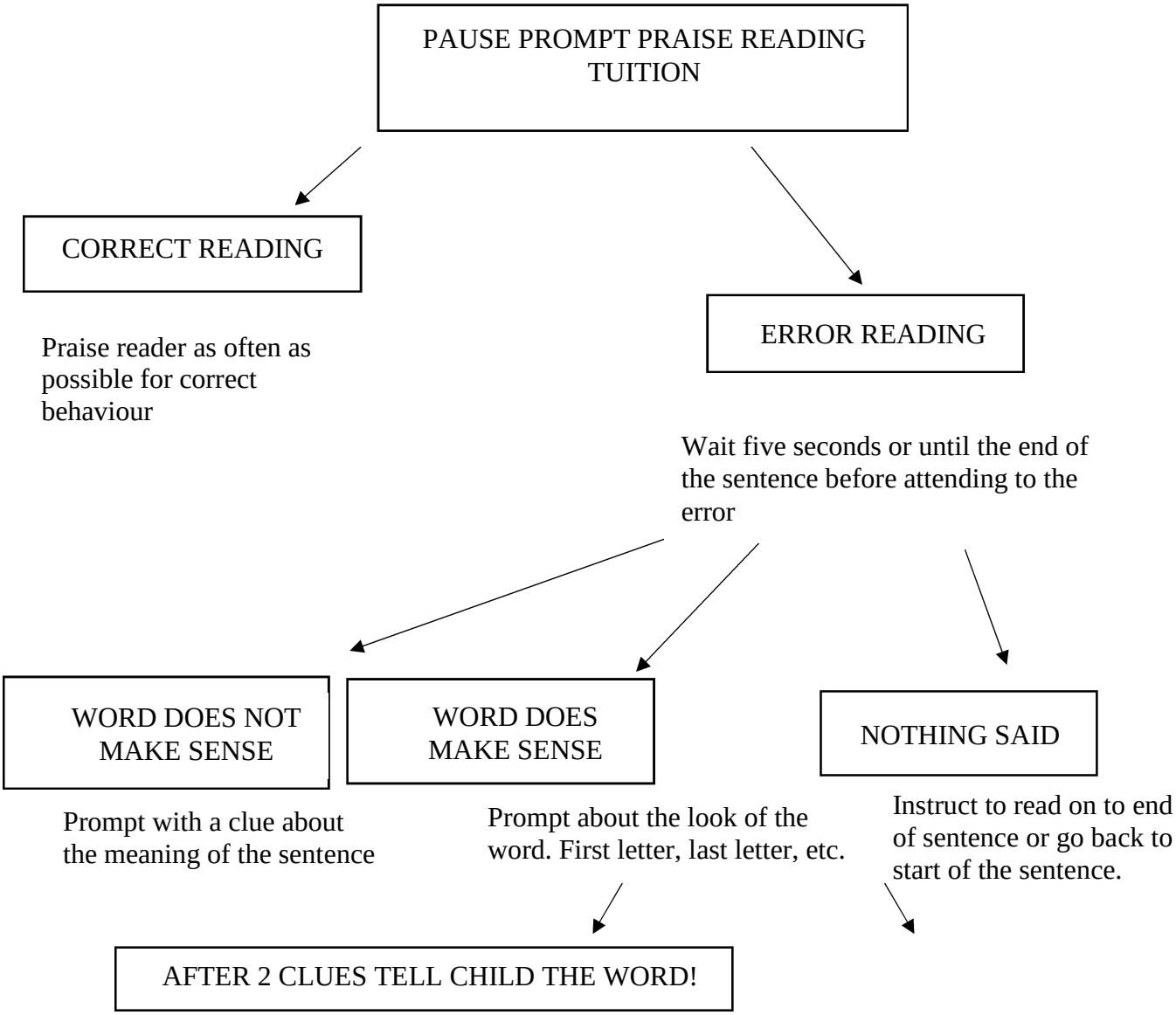
The review searches revealed several different reading peer tutoring approaches. These are explained below.

Intraclass peer tutoring

The earliest study reviewed in this systematic review (Carlton, Litton & Zinkgraf, 1985) is based on intraclass peer tutoring, demonstrated by Brown, Fenrick and Klemme (1971), in which tutors successfully taught selected words to peers in the same class. In this instance, the peer tutoring is a simple form of flash card drill in order to teach basic sight words.

Pause, Prompt and Praise (PPP)

Pause, Prompt and Praise was first developed by McNaughton and Glynn in 1981. This is a behavioural method for tutoring oral reading which is based on research that suggests self-correction of errors is a natural part of reading, with good readers showing increased rates of self-correction (Clay, 1972). Self-correction should therefore be encouraged, hence the "pause" aspect. The researchers found that second year readers who received immediate correction from their teacher made more errors and produced fewer self-corrections than second year readers who received delayed error feedback. See diagram below for details of the procedure.



Classwide Peer Tutoring

Classwide Peer tutoring (CWPT) was first developed by Delquadri, Greenwood, Stretton and Hall in 1983 in order to improve third-grade pupils' spelling attainment. . In this program all the pupils have increased opportunity to respond to the learning materials, for example, more time to practise the spellings. Lower ability and higher ability pairs are typically formed, with the lower ability pupil completing reading activities; such as passage reading; and the higher ability pupil correcting errors and then reading the passage themselves. The higher achieving pupil can then ask a series of questions, thus tapping into comprehension skills as well. Greenwood, Delquadri and Hall (1989) showed that Grades 1 to 4, low socio-economic status (SES) pupils, following CWPT, significantly improved their long term reading achievement compared with low SES and high SES control groups. The CWPT also uses an element of contingency reinforcement with the use of points, dyad and team rewards. It is also noted that the tutor's behaviour is viewed as imperative to the intervention and thus points are also awarded to the tutors (see Table 1).

Table 1: An example of a CWPT reading procedure.

Action	Details
Part 1: First 10 minutes	"Tutee" reads passage to "Tutor" The tutor observes the child read, awards points, and corrects errors. To correct the error the tutor pronounces the correct word and the child rereads the sentence until it is correct. If it is correct, the tutor tells the child it is correct and to give him or herself two points. The teacher supervises the tutoring, providing assistance and awarding bonus points to tutors for correct tutoring behaviours. Tutees are also given bonus points for responding immediately and for working cooperatively with their tutor.
Part 2: Second 10 minutes	The "Tutor" becomes the "Tutee" and the "Tutee" becomes the "Tutor". The same procedure is followed as above.
Part 3	Students' points are written on a large team chart that produces the team totals. The winning team is applauded for winning, as is the losing team for making a good effort.

Note. Adapted from Delquadri, Greenwood, Whorton, Carta and Hall (1986).

Peer Assisted Learning Strategies (PALs)

PALs is an extension of CWPT which incorporates practising strategic reading behaviours. It was first developed by Fuchs in 1995 and studies have found that elementary school pupils with learning disabilities, as well as low and average achieving pupils, improve their reading development (Fuchs, Fuchs, Mathes & Simmons, 1997). Calhoon (2005) chose to design their own phonological program based on CWPT, with the aim of complementing Fuchs' PALs program.

Basis in Psychological Theory

The main theories underpinning peer tutoring, including reading interventions, are borne out of the behaviourist approach. For example, "Pause, Prompt and Praise" is based on basic reward theories (Skinner, 1938) and Bandura's "social learning theory" (1977). According to these theories the tutee will improve their reading skills, for example, due to "modelling" the behaviour of the more advanced tutor. The tutee will also aim to read accurately in order to receive a reward (praise or sometimes contingency rewards).

Peer tutoring can simply focus on simple drill and practice but there are more elaborate and cognitively demanding forms of peer tutoring, such as peer tutoring in thinking skills (McKinstery & Topping, 2003). The tutoring of thinking skills relies on "scaffolding" the learning, thus tapping into the tutee's "zone of proximal development" (Vygotsky, 1980).

Finally, the tutor will often have a clear structure to follow that provides the tutee with many more opportunities to practise, for example, reading a passage, compared with usual class teaching. As it has been established that active learning time is necessary to achievement in classroom settings, this is another suggested reason why peer tutoring improves reading skills (Bloom, 1974; Rosenshine & Berliner, 1978).

Relevance to Educational Psychology Practice

Reading is a vital life skill as well as it being necessary to gain educational attainment. Ofsted state that, "*Too many pupils still emerge from our schools without the confident and secure literacy skills they need to thrive as adults*" (Ofsted, 2013) (p.6).

The Literacy Trust (2012) found that one in every six adults struggles with literacy, with a literacy level below that expected of an 11-year-old. It is known through research that "ordinary teaching" ('no intervention') does not enable children with literacy difficulties to catch up" (Brooks, 2007) (p.18).

Fortunately there is a wealth of research on the effectiveness of reading peer tutoring with primary school pupils (Delquadri, Greenwood, Whorton, Carta & Hall, 1986; Topping, 1987). In fact Brooks (2007) (p.18) said, "Where reading partners are available and can be given appropriate training and support, partnership approaches can be very effective. Children's comprehension skills can be improved if directly targeted". However, there is much less research on the effectiveness of peer tutoring reading interventions at secondary school (Mastropieri, Scruggs, Mohler, Beranek, Spencer, Boon & Talbott, 2001).

Reading difficulties at secondary level may be more difficult to remedy as the pupils have a longer history of failure and have shown low motivation and academic self-confidence (Phelan, Yu & Davidson, 1994). The setting can prove more difficult for interventions as the pupils have to move from class to class and teaching for exams is important.

In terms of reading comprehension, secondary school pupils are expected to comprehend complex and diverse texts to be able to meet National Curriculum standards and pass formal

qualifications. However, the readability level of some texts may be too high for struggling readers, resulting in comprehension challenges (Mastropieri, Scruggs, & Graetz, 2003). Evidence suggests that those with reading difficulties at primary school are likely to still need support at secondary school (Hulme & Snowling, 2013).

Appropriate evidence based secondary school reading comprehension interventions are much scarcer than primary interventions. In fact, the vast majority of reading interventions identified as effective in *"What works for children and young people with literacy difficulties?"* (Brooks, 2007) were based on primary school studies. Establishing whether peer tutoring is effective in improving comprehension in secondary school pupils is of importance to educational psychologists as it would be an evidenced based intervention that can then be recommended to schools. This links to the recent range of policy initiatives that have stressed the importance of basing EP practice on firm evidence (Harper, Gannon & Robinson, 2012).

This all suggests the need to find effective secondary school reading interventions that teachers can easily implement at little cost. As previous research has given an indication that peer tutoring could be effective this is the intervention that will be reviewed. There are various aspects of reading to focus on but as comprehending difficult and complex texts is a prerequisite for secondary school pupils and vital for achieving formal qualifications the review will focus on comprehension skills.

Review Question

The review question for the present systematic literature review is therefore:

Do peer-tutored reading interventions improve reading comprehension in secondary age pupils attending mainstream schools?

Section 3. Critical Review of the Evidence Base

Searches

A search was carried out on PsycINFO, ERIC ProQuest and Web of Science on 30th January 2016. Those databases were chosen to yield the best range of appropriate articles possible. This database search produced 282 records (see Figure 2: Database Search below for details). Next all the duplicates were removed, leaving 225 articles. The inclusion/exclusion criteria were applied to the titles (see Table 3 below for details) which then left 181 articles. The abstracts were then screened against the inclusion/exclusion criteria, leaving 42 full articles to be checked against the inclusion/exclusion criteria (see Appendix A for full references and rationale for inclusion/exclusion). The references of these full articles were also checked as an ancestral search. A search was also carried out in the three databases (ERIC ProQuest, PscyINFO and Web of Science) for Douglas Fuchs, an author/researcher who has published widely in this area. '. No new studies were identified that met the inclusion/exclusion criteria, therefore they have not been included in Figure 2. This procedure resulted in six studies to be included in the review.

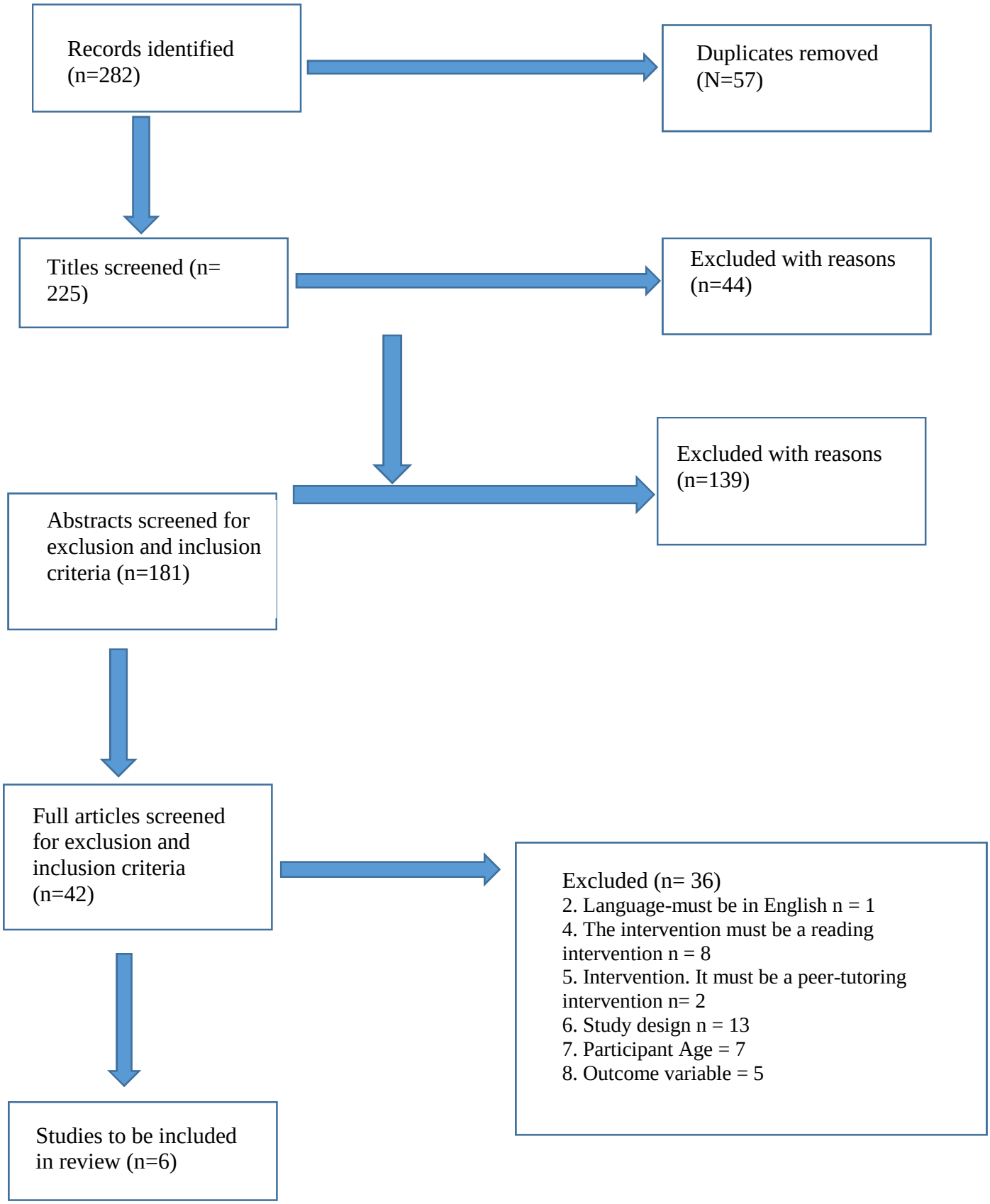
Table 2
Database search terms

Databases	Search terms	Other search criteria
ERIC ProQuest	"peer tut*," school and reading	Peer reviewed
PsycINFO	peer tutored/peer tutoring, school/schools and reading	Peer reviewed
Web of Science	peer tut*, school and reading	None (this was checked manually at a later stage)

Table 3
Inclusion and exclusion criteria used for literature search with rationale

Criteria	Inclusion	Exclusion	Rationale
1. Type of publication	Peer reviewed journal	Not a peer reviewed journal	Peer reviewed journals ensure a level of quality checks have been completed.
2. Language	The study is written in English	The study is not written in English	In order that the paper can be read and understood it needs to be in English so that it can then be appraised. Translators were not available as part of this assignment but translations can still lose their meaning and make appraisals difficult
3. Setting	The pupils must be attending mainstream school	The pupils must not be attending a non-mainstream school	The review question is specifically assessing only reading interventions in mainstream schools.
4. Intervention	It is an empirical paper on an English “reading” intervention	It is not an empirical paper on an English “reading” intervention	The review question is specifically only assessing empirical English “reading” interventions.
5. Intervention	The intervention uses peer-tutoring	The intervention does not use peer tutoring	The review question is specifically on “peer tutoring”
6. Type of study	The study design must have an “active” control	The study design does not have an “active” control	Control designed studies are more robust as they aim to eliminate the effects of other variables on the outcomes.
7. Participants	Participants must be age 11+ attending a middle school/high school or secondary school	Any participant who is not aged 11+ and at a middle school/high school or secondary school	The review question seeks to review the effects of peer tutoring comprehension interventions on secondary pupils as there is evidence that comprehension levels are poor at this age.
8. Outcome variables	The outcomes must look at reading comprehension	Outcomes other than reading comprehension	The review question is specifically around reading comprehension.

Figure 2.
Literature Search and Screening Strategy: Database search



Weight of Evidence (WoE)

The six studies which met the inclusion criteria are summarised in Appendix B and they were evaluated using Gough's 2007, Weight of Evidence Framework (WoE). This framework was used as it is a well-known framework and is suitable for assessing studies against a specific question. Studies are rated for methodological quality (WoE A), methodological relevance to the review question (WoE B), relevance of evidence to the review question (WoE C) and the overall weight of evidence (WoE D). See Table 4 for the summary weightings of the six reviewed studies. Appendix C provides a detailed explanation of the WoE criteria and ratings used on each of the six studies.

Table 4
Summary of weight of evidence ratings (based on Gough, 2007)

Author	WoE A: **Quality of Methodology**	WoE B: **Relevance of Methodology**	WoE C: **Relevance of evidence to the review question**	WoE D: **Overall weight of evidence**
Wheldall & Mettem (1985)	Medium (1.75)	Low (1.5)	Low (1.5)	Low (1.59)
Mastropieri et al. (2001)	Low (1.5)	High (2.5)	High (2.5)	Medium (2.12)
Fuchs et al. (1999)	Medium (1.75)	Low (1.5)	Low (1.5)	Low (1.58)
Carlton et al. (1985)	Low (1.0)	Medium (2.00)	Low (1.00)	Low (1.33)
Calhoon (2005)	Low (1.25)	Low (1.5)	Medium (2.00)	Low (1.58)
Mastropieri et al. (2003)	Medium (2.25)	High (2.5)	High (2.5)	High (2.42)

Participants

In total there were at least 162 pupils who received peer tutoring (i.e. they were the tutee at some point) and were assessed for changes in reading comprehension. Fuchs, Fuchs and Kazdan (1999) did not specify the number involved in the *assessment* but there were 18 classes involved in the study, containing 102 pupils in total. Participant ages ranged from 11 – 15 years and there were generally more boys in each study (ranging from 65% to 81%). Out of the studies that recorded ethnicity (78 pupils in total), 40% were Hispanic American, 35% were Caucasian (21 pupils were from one study, Mastropieri et al., 2001), 14% were European American (only in Carlton, Litton & Zinkgraf, 1985), 9% were African American and 2% were Asian American (only from Mastropieri, Scruggs, Spencer & Fontana, 2003). These demographic differences between studies may be a factor in differing outcomes but this review did not seek to investigate this.

In four studies participants attended both mainstream classes (between 20% and 50% of their day) and a specialist resource. The participants in the fifth study (Wheldall & Mettem, 1985) were in streamed “remedial” classes. Fuchs et al. (1999) did not make it clear. In all of the studies the participants were categorised as having a “learning disability” with difficulties in reading. Some were also deemed “mentally retarded”. As the review question asks about secondary “mainstream” pupils those studies in which pupils participated in some mainstream classes scored more highly.

Design

Randomness

There was variation across the studies on the "randomness" of assignment to experiment and control group. A fully randomised control is most effective for reducing selection bias (Barker, Pistrang & Eilliot, 2002; Kratochwill, 2003) so studies were weighted as "high" under the WoE B if they were completely randomised. Only Carlton et al. (1985), was rated as high, as six classes were randomly selected to receive the intervention and six classes were randomly selected to receive the usual instruction. Wheldall and Mettem (1985) mention a "pool of 24 selected" but it does not say on what basis and Fuchs et al. (1999) chose an unknown subset of pupils to assess. These studies therefore were only weighted as "low" under the WoE B.

Control intervention

As the review question asks about *effectiveness* it was deemed appropriate for all the review studies to contain an "active" control group, as identified in the inclusion/exclusion criteria. This enables a comparison of treatment and non-treatment group, reducing the chances of other factors influencing changes to the pupils' comprehension ability following a peer tutored reading intervention. There are, however, different types of "active" control, such as a placebo, wait list or comparison group.

It is also important that the control group intervention is as similar to the peer group intervention, but without the peer tutoring element (this is the "active ingredient"), as possible to try and control for any other variables that may influence the reading outcomes. Studies, therefore, in which both control and peer tutoring interventions were delivered over the same duration and used the same materials were weighted as "high" under WoE B. Two studies, Mastropieri et al. (2003) and Mastropieri et al. (2001) were in

fact weighted as “high” as they used the same reading materials, with the same lesson plans, in both conditions and both conditions were delivered during the same lessons for the same duration.

Although Calhoon (2005) used a contrast treatment which had many similarities to the phonological part of the peer tutoring condition, it was not similar for the comprehension part. The control group also received more than 10 hours extra tuition so it was not deemed a fair comparison and thus scored a “low” on WoE B. Carlton et al. (1985) also received a low weighting as there were minimal details on the control intervention.

Equivalence:

Equivalence was sought using different group characteristics. Equivalence is important in order to allow for differences that might have been present in the first place. Most studies found no significant differences in pre testing measures.

Calhoon (2005) found differences in age between the groups which may be a factor in any changes post intervention and this study therefore only received a WoE A Comparison rating of 1 out of 3 (see Appendix C, Weight of Evidence A: Methodological quality for details). Wheldall and Mettem (1985) only mention that the mean reading ages of the groups were “similar” and does not mention significance. They showed the three groups were not significantly different in reading accuracy but did not check for reading comprehension. This study was rated a 2 on the WoE A Comparison section as it did not provide clear evidence of equivalence.

Fidelity

"Data on the program integrity will be critical to determine if the intervention was responsible for the positive outcomes reported by the researchers" (Kratochwill, 2003) (p.37). Studies scored higher on WoE A if the study provided evidence of ongoing supervision/consultation, coding sessions, or audio/video tapes and use of a manual. All the researchers tried to ensure fidelity of the study in similar ways by the use of observations, videotapes, checklists, regular contact with the Research Assistant, training, and manuals. All studies were weighted as a "high"(3) under the WoE A Fidelity section, apart from Fuchs et al. (1999), who scored a medium (2) as they did not provide lessons plans.

Measures

In terms of answering the review question, it was deemed important to weight the studies in terms of how focused they were on reading comprehension so this was one of two criteria set in order to measure WoE C. Studies where reading comprehension was the main focus received a "high" under WoE C criteria. Carlton et al. (1985) was the only study to receive a low rating (1) as the focus of the intervention was on enhancing vocabulary, although comprehension was still measured.

In order to score a "high" under the WoE A Measurement section, the data should have been collected using multiple methods, collected from multiple sources and measures should report a reliability coefficient of at least .85. Only Mastropieri et al. (2003) was weighted a "medium" and the others were all weighted as "low". It was mainly due to the reliability not being reported..

Findings

All the studies had pre and post measures for control and experiment groups. The studies, however, analysed their data in differing ways. As this review is related to Reading Comprehension only these are the only outcome measures that will be discussed. In terms of trying to ascertain a sense of comparison for the purposes of this review, Cohen's d standardised effect sizes for each study were calculated by comparing the post score means of the intervention group with the control group. The results were then interpreted using Cohen's (1992) interpretation of small (.2), medium (.5) and large (.8). Details of these effect sizes can be seen in Table 5 below.

Fuchs et al. (1999) did not provide sample sizes, Mastropieri et al. (2001) did not report on sample sizes for the post tests and Wheldall and Mettem (1985) did not find a significant difference between the groups at post comprehension testing, so effect sizes are not available for these studies. Carlton et al. (1985) showed a small effect size, Calhoon (1985) and Mastropieri et al. (2003) showed large effect sizes of .86 and 2.18 respectively (see Table 5).

Calhoon (2005) found statistically significant differences between groups with LST/PALs outperforming the contrast condition on Passage Comprehension, $F (1, 37) = 11.35$, $p = .01$. This significance is supported with a large effect size (0.86) however the overall weighting for this study was "low" so these findings need to be taken with caution with reference to the review question.

Mastropieri et al. (2003) found significant differences on pre-test versus post-test, condition, $F (1, 11) = 9.46$, $p = 0.01$ and pre-test versus post-test by condition interaction, $F (1, 11) = 7.91$, $p = 0.017$. . The effect size was very large at 2.18 and as the study was

weighted as “high” these results can be viewed as more robust than the other studies in terms of answering the review question.

Mastropieri et al. (2001) found a significant difference between control and tutoring group of post test scores, T= (19) = 2.72, p= 0.013 but as three pupils were unavailable for post testing, without specifying from which intervention, it was not possible to calculate the standardised effect size

Another way that the effects were analysed was on the “gains” the groups made, taking into account their pre and post measures. Wheldall and Mettem (1985) found a mean gains in excess of 6 months for comprehension which was better than the two control groups. Carlton et al. (1985) found gain scores higher for the intervention group than for the control group on the reading subtest. There was a significant interaction between group membership and time of testing for comprehension subtest ($F = 3.30$, $df = 2, 87$, $p \leq .05$). Simple main effects using the Scheffe procedure showed both tutors and tutees made significant gains from pre to post whereas the controls did not.

Calhoon (2005) found significant growth for LST/PALs over the contrast group for Passage Comprehension, F (1, 37 = 11.35, P = .01). They calculated an effect size of .94 for growth on Passage Comprehension, however, it does not report on what effect size is being used though.

Finally, Fuchs et al. (1999) found that on a “number of questions answered correctly” (measure of comprehension) the growth of PALs’ students exceeded that of contrast students. They report an effect size of.34. Differences on final status (post scores were not significant for either measures).

Table 5
Effect sizes of peer tutoring vs. control

Author	Measures	Intervention				Control 1 Non Tutored				Control 2				Effect size (Cohen's d)	Effect size descriptor (High, Medium or Low)	Overall WoE
		N	Pre M (S.D)	Post M (S.D)	Follow up (S.D)	N	Pre M (S.D)	Post M (S.D)	Follow up M (S.D)	N	Pre M (S.D)	Post M (S.D)	Follow up M (S.D)			
Wheldall,& Mettem (1985)	Neale Analysis of Reading Ability (Form A, B and C) Comprehension part	8	10.38 (2.45)	**15.25 (4.71)**	**14.25 (4.59)**	8	10.38 (4.63)	**13.88 (3.18)**	**13.38 (6.02)**	8	9.75 (2.60)	**12.50 (4.17)**	**13.50 (5.83)**	*NS	n/a	LOW
Mastropieri et al. (2001)	Criterion referenced comprehension tests	12		**81% correct (13.9)**	n/a	12	63% (17.2)		n/a	n/a	n/a	n/a	n/a	Unable to report on this effect size as post-test sample size not known	n/a	MEDIUM
Carlton et al. (1985)	Gates-MacGinitie Reading Test	30 Tutees	284.7 (67.	**315.3 (73.6)**	n/a	30	276.0 (94.7)	**278.8 (98.7**	n/a	n/a	n/a	n/a	n/a	.4193	LOW	LOW

	(Comprehension part)		9)					**)**								
Author	Measures	Intervention				Control 1 Non Tutored				Control 2				Effect size (Cohen's d)	Effect size descriptor (High, Medium or Low)	Overall WoE
		N	Pre M (S.D)	Post M (S.D)	Follow up (S.D)	N	Pre M (S.D)	Post M (S.D)	Follow up M (S.D)	N	Pre M (S.D)	Post M (S.D)	Follow up M (S.D)			
Calhoon, (2005)	WJ- III Comprehension sub-set	18	78.88 (10.19)	**85.44 (8.45)**	n/a	20	77.80 (11.19)	**76.60 (11.66)**	n/a	n/a	n/a	n/a	n/a	.8608	HIGH	LOW
Mastropieri et al. (2003)	Comprehension Strategy	7	1.43 (0.53)	**2.71 (0.76)**	n/a	9	1.33 (0.52)	**1.33 (0.52)**	n/a	n/a	n/a	n/a	n/a	2.1764	HIGH	HIGH
Fuchs et al. (1999)	Comprehensive Reading Assessment Battery (CRAB) Measure of comprehension	unknown	5.88 (2.56)	**7.22 (2.23)**	n/a	unknown	6.10 (2.49)	**6.64 (2.44)**	n/a	n/a	n/a	n/a	n/a	Unable to compute as sample size not known	n/a	LOW

*NS means "not significant"

Section 4. Conclusions and Recommendations

This review aimed to ascertain whether peer-tutored reading interventions improve reading comprehension in secondary age pupils attending mainstream schools. The main findings were that the peer tutored interventions used in these studies showed improvements between pre and post reading comprehension scores in all but one study, Wheldall and Mettem (1985). This study received an overall weighting of “low” so in terms of answering the review question the results need to be viewed tentatively

Three of the studies also found significant differences between the peer tutored and control group post measures (Calhoon, 2005; Mastropieri et al., 2001 & Mastropieri et al., 2003). Mastropieri et al. (2003), as mentioned, was given an overall “high” weighting, Mastropieri (2001) received an overall “medium” weighting and Calhoon (2005) achieved an overall “low” weighting. This means that more significance can be placed on the Mastropieri findings, and based on previous research, it is likely that peer tutoring to improve secondary school reading comprehension is an effective intervention.

In terms of implications for EP practice, this review provides an evidence base for recommending peer tutored comprehension interventions to secondary schools. Improving comprehension skills can help pupils improve their accessibility to all subjects and exams, as comprehending texts and questions are fundamental skills within secondary education.

Although these studies did focus only on pupils with identified reading difficulties there is good evidence that both the tutor and tutee improve their learning (Miller, Topping &

Thurston, 2010; Topping, 1987, 1989). This was in fact supported in the Carlton et al. (1985) study, who tested both the tutor and tutee comprehension skills. It would therefore be plausible to recommend developing a class wide approach, as opposed to identifying a group of children with reading difficulties, which could make the delivery easier as the whole class would be receiving the same instruction. If working with a specific group of children with reading difficulties then a reciprocal form of peer tutoring, could be recommended as all the pupils involved will practise the same skills.

Schools may also like peer tutoring for a number of other reasons. For example, the allocation of extra reading time in itself can result in improved reading skills (Bloom, 1974; Rosenshine & Berliner, 1978). Recommending a reading intervention to schools responds to the research suggesting that ordinary teaching will not improve the literacy skills of those with difficulties and that those with difficulties will need continued support (Brooks, 2007; Hulme & Snowling, 2013). This review suggests that it could be effective after just 5 weeks, a fairly short term intervention. Peer tutored reading gives the added benefit of being delivered by the peers, alongside close teaching supervision, which might make this a more attractive intervention for schools who may be concerned about the teaching time needed to implement an intervention.

This review did not differentiate between the different types of peer reading interventions, however, based on practicalities it would probably be best to recommend same age tutoring as opposed to cross age due to the differing timetables.

Finally, in terms of delivery, some of the peer tutoring approaches are manualised, for example, “Pause, Prompt and Praise” but could also be monitored by an EP to ensure fidelity. Costs will therefore vary according to the type of peer tutoring. It should be

remembered, though, that interventions that are relatively easy and cost effective increase the chances of a school implementing them (O'Keefe & Medway, 1997).

Further research could look at the processes needed in order for the peer tutoring to be effective. Winter (1987) made some attempts to look at the processes that might be involved in paired reading. It would be advantageous if there are some really simple processes that could easily be implemented by teachers in most classrooms.

The interventions were of varying duration, with the longest one lasting 31 weeks (Calhoon, 2005) and the shortest one only lasting for five weeks (Mastropieri et al., 2001). Interestingly Calhoon et al. (2005) was weighted overall as “low” and Mastropieri et al. (2001) rated overall as “medium”, with both showing significant improvements in the peer tutoring intervention. This suggests that the intervention does not need to be that long to be effective. It would be beneficial to research this further in order to make implementation as simple as possible in secondary schools.

Limitations of the current review

This review could have looked at the “gains’” effect sizes made by the different groups. The mean pre-post change in the treatment condition minus the mean pre-post change in the comparison condition, divided by the pooled pre-test standard deviation. This method of calculating effect sizes that has been found to be favourable on measures of bias, precision and robustness (Morris, 2007). However as group equivalence had been shown for the studies it was decided that calculating the effect size using the mean differences between the post measures would suffice.

References

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.

Barker, C., Pistrang, N., & Elliot, R. (2002).*Research Methods in Clinical Psychology: An Introduction for Students and Practitioners, Second Edition.* John Wiley & Sons, Ltd.

Bloom, B. S. (1974). Time and learning. *American Psychologist*, *29*(9), 682.

Brooks, G. (2007). *What Works for Children with Literacy Difficulties?* The effectiveness of Intervention Schemes. DCFS Publications.

Brown, L., Fenrick, N., & Klemme, H. (1971). Trainable pupils learn to teach each other. *Teaching Exceptional Children*, *4*(1), 18.

Calhoon, M. B. (2005). Effects of a peer-mediated phonological skill and reading comprehension program on reading skill acquisition for middle school students with reading disabilities. *Journal of Learning Disabilities*, 38(5), 424-433.

Carlton, M. B., Litton, F. W., & Zinkgraf, S. A. (1985). The effects of an intraclass peer tutoring program on the sight-word recognition ability of students who are mildly mentally retarded. *Mental Retardation*, (23)2, 74-78.

Clay, M. M. (1977). Reading: The patterning of complex behaviour (2nd ed.). London: Heinemann Educational

Confederation for British Industry (2011). *Building for growth*. www.cbi.org.uk/media-centre/news-articles/2011/05/building-for-growth.

Crits-Christoph, P., & Mintz, J. (1991). Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *Journal of Consulting and Clinical Psychology*, *59*(1), 20.

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

Delquadri, J. C., Greenwood, C.R., Stretton, K., & Hall, R.V. (1983). The peer tutoring spelling game: A classroom procedure for increasing opportunity to respond and spelling performance. *Education and Treatment of Children, 6*(3), 225-239.

Department for Education and Skills (2002). *What Works for Children with Literacy Difficulties? The Effectiveness of Intervention Scheme.* DfES publications.

Dufrene, B.A., Hennington, C., & Townsend, A.E. (2006). Peer tutoring for reading fluency: Student implementation and effects on reading fluency. *Journal of Evidence Based Practice for Schools*, 7(2), 118-137.

Final report of the EU High Level Group of experts on literacy, 2012; ec.europa.eu/education/literacy/resources/final-report/index_en.htm.

Folio, M. R., & Norman, A. (1981). Toward more success in mainstreaming: A peer teacher approach to physical education. *Teaching Exceptional Children*, *13*(3), 110.

Fuchs, L. S., Fuchs, D., & Kazdan, S. (1999). Effects of peer-assisted learning strategies on high school students with serious reading problems. *Remedial and Special Education*, 20(5), 309.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*(1), 174-206.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213-228.

Greenwood, C. R., Delquadri, J. C., & Hall, R. V. (1989). Longitudinal effects of classwide peer tutoring. *Journal of Educational Psychology*, *81*(3), 371.

Harper, D. J., Gannon, K. N. & Robinson, M. (2012). Beyond evidence-based practice: Rethinking the relationship between research, theory and practice. In R. Bayne & G. Jinks (Eds.). *Applied psychology: practice, training and new directions* (2nd ed). London: SAGE.

Hulme, C. & Snowling, M. (2013). Learning to read: What we know and what we need to understand better. Child Development Perspectives, 7(1), 1-5.

Juel, C. (1996). What makes literacy tutoring effective? *Reading Research Quarterly*, *31*(3), 268-289.

Kamps, D. M., Veerkamp, M. B., & Cooper, L. (2007). The effects of classwide peer tutoring on the reading achievement of urban middle school students. *Education and Treatment of Children*, *30*(2), 21-51.

Kratochwill, T. R. (2014). Task force on evidence-based interventions in school psychology.

National Literacy Trust. (2012). *Literacy: state of the nation* www.literacytrust.org.uk/research/nlt_research/2364_literacy_state_of_the_nation

Maheady, L., Harper, G. F., & Mallette, B. (2001). Peer-mediated instruction and interventions and students with mild disabilities. *Remedial and Special Education*, *22*(1), 4-14.

Mastropieri, M. A., Scruggs, T., Mohler, L., Beranek, M., Spencer, V., Boon, R. T., & Talbott, E. (2001). Can middle school students with serious reading difficulties help each other and learn anything? *Learning Disabilities Research & Practice*, 16(1), 18-27.

Mastropieri, M. A., Scruggs, T. E., Spencer, V., & Fontana, J. (2003). Promoting success in high school world history: Peer tutoring versus guided notes. *Learning Disabilities Research & Practice*, 18(1), 52-65.

Mastropieri, M. A., Scruggs, T. E., & Graetz, J. E. (2003). Reading comprehension instruction for secondary students: Challenges for struggling students and teachers. *Learning Disability Quarterly*, *26*(2), 103-116.

Mathes, P. G., & Fuchs, L. S. (1994). The efficacy of peer tutoring in reading for students with mild disabilities: A best-evidence synthesis. *School Psychology Review*, 23(1), 59-80.

McKinstery, J., & Topping, K. J. (2003). Cross-age peer tutoring of thinking skills in the high school. *Educational Psychology in Practice*, *19*(3), 199-217.

McNaughton, S., & Glynn, T. (1981). Delayed versus immediate attention to oral reading errors: effects on accuracy and self-correction. *Educational Psychology*, *1*(1), 57-65.

Miller, D., Topping, K., & Thurston, A. (2010). Peer tutoring in reading: The effects of role and organization on two dimensions of self-esteem. *British Journal of Educational Psychology*, *80*(3), 417-433.

Ofsted (April 2013). *Improving literacy in secondary schools: a shared responsibility*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/413182/Improving_literacy_in_secondary_schools.pdf

O'Keefe, D. J., & Medway, F. J. (1997). The application of persuasion research to consultation in school psychology. *Journal of School Psychology*, *35*(2), 173-193.

Phelan, P., Yu, H. C., & Davidson, A. L. (1994). Navigating the psychosocial pressures of adolescence: The voices and experiences of high school youth. *American Educational Research Journal*, *31*(2), 415-447.

Rosenshine, B. V., & Berliner, D. C. (1978). Academic engaged time. *British Journal of Teacher Education*, *4*(1), 3-16.

Skinner, B. F. (1938). The behavior of organisms: an experimental analysis.

Spooner, A. L., Baddeley, A. D., & Gathercole, S. E. (2004). Can reading accuracy and comprehension be separated in the Neale Analysis of Reading Ability? *British Journal of Educational Psychology*, *74*(2), 187-204.

Suggate, S. P. (2014). A Meta-Analysis of the Long-Term Effects of Phonemic Awareness, Phonics, Fluency, and Reading Comprehension Interventions. *Journal of Learning Disabilities* 49(1), 77–96

Topping, K. (1989). Peer tutoring and paired reading: Combining two powerful techniques. *The Reading Teacher*, *42*(7), 488-494.

Topping, K. (1987). Peer tutored paired reading: Outcome data from ten projects. *Educational Psychology*, *7*(2), 133-145.

Vygotsky, L. S. (1980). Mind in society: The development of higher psychological processes. Harvard university press.

Wheldall, K., & Mettem, P. (1985). Behavioural peer tutoring: Training 16-year-old tutors to employ the "Pause, Prompt and Praise" method with 12-year-old remedial readers. *Educational Psychology*, 5(1), 27-44

Winter, S. (1996). Paired reading: Three questions. *Educational Psychology in Practice*, *12*(3), 182-190.

Appendix A: Excluded studies with rationale and full references (excluded at full article)

Excluded Study	**Rationale for exclusion**
Arblaster, G. R., Butler, C., Taylor, A. L., Arnold, C., & Pitchford, M. (1991). Same-age tutoring, mastery learning and the mixed ability teaching of reading. *School Psychology International*, 12(1-2), 111-118.	7. Participants must be age 11+ attending a middle school/high school or secondary school. In this study they are attending a primary school.
Bagley, C., & Mallick, K. (1996). Towards achievement of reading skill potential through peer tutoring in mainstreamed 13-year-olds. *Disability & Society*, 11(1), 83-90.	8. Outcome variable. The study must have an outcome variable that measures reading comprehension. This study does not measure reading comprehension.
Dufrene, B. A., Reisener, C. D., Olmi, D. J., Zoder-Martell, K., McNutt, M. R., & Horn, D. R. (2010). Peer tutoring for reading fluency as a feasible and effective alternative in response to intervention systems. *Journal of Behavioral Education*, 19(3), 239-256.	8. Outcome variable. The study must have an outcome variable that measures reading comprehension. This study measures reading fluency and accuracy only.
Burns, E. (2006). Pause, Prompt and Praise--Peer Tutored Reading for Pupils with Learning Difficulties. *British Journal of Special Education*. Volume 33, Number 2, pp 62-67.	6. The study design must have an “active” control. This study is pre-post.
Collins, B. C., Branson, T. A., & Hall, M. (1995). Teaching generalized reading of cooking product labels to adolescents with mental disabilities through the use of key words taught by peer tutors. *Education and Training in Mental Retardation and Developmental Disabilities*, 65-75.	6. The study design must have an “active” control. This study is a small N design.
Ehly, S., Keith, T. Z., & Bratton, B. (1987). The benefits of tutoring: An exploration of expectancy and outcomes. *Contemporary Educational Psychology*, 12(2), 131-134.	4. The intervention must be an empirical paper on an English “reading” intervention. . This study is around peer tutoring in general.
Eiserman, W. D. (1988). Three Types of Peer Tutoring Effects on the Attitudes of Students with Learning Disabilities and Their Regular Class Peers. *Journal of Learning Disabilities,* 21(4), 249-252.	7. Participants must be age 11+ attending a middle school/high school or secondary school. In this study they are attending an elementary school.
Fuchs, D., Fuchs, L. S., & Burish, P. (2000). Peer-assisted learning strategies: An evidence-based practice to promote reading	7. Participants must be age 11+ attending a middle school/high school or secondary school. In this study they are

Excluded Study	Rationale for exclusion
achievement. *Learning Disabilities Research & Practice*, 15(2), 85-91.	up to age 11 and attending an elementary school.
Greer, R. D., & Polirstok, S. R. (1982). Collateral gains and short term maintenance in reading and on task responses by inner city adolescents as a function of their use of social reinforcement while tutoring. *Journal of Applied Behavior Analysis*, 15(1), 123-139.	6. The study design must have an "active" control. This study is small N design.
Greenwood, C. R., Dinwiddie, G., Terry, B., Wade, L., Stanley, S. O., Thibadeau, S., & Delquadri, J. C. (1984). Teacher-versus peer-mediated instruction: An ecobehavioral analysis of achievement outcomes. *Journal of applied behavior analysis,* 17(4), 521-538.	7. Participants must be age 11+ attending a middle school/high school or secondary school. In this study they are up to age 11 and attending an elementary school.
Hassett, J. J. (1974). *Peer tutoring in new york city high schools.*	4. The intervention must be an empirical paper on a "reading" intervention This is not a new study but a review
Houghton, S., & Bain, A. (1993). Peer tutoring with ESL and below-average readers. Journal of Behavioral Education, 3(2), 125-142.	6. The study design must have an "active" control. This study is small N design.
Howard, K. (1977). A peer tutoring program in a technical school. *Journal of Reading*, 21(2), 115-120	6. The study design must have an "active" control. This study is qualitative.
Huling, L. L. (1981). Interactive Research and Development: A Promising Strategy for Teacher Educators. *Journal of Teacher Education,* 32(6), 13-14.	8. Outcome variable. The study must have an outcome variable that measures reading comprehension. This study does not measure reading comprehension.
James, J., Charlton, T., Leo, E., & Indoe, D. (1991). A peer to listen. *Support for learning,* 6(4), 165-169.	4. The intervention must be an empirical paper on an English "reading" intervention. This is not a new study.
Kamps, D. M., Veerkamp, M. B., & Cooper, L. (2007). The effects of classwide peer tutoring on the reading achievement of urban middle school students. *Education and Treatment of Children*, 30(2), 21-51.	6. The study design must have an "active" control. This study is small N design.
Kroeger, S. D., Burton, C., & Preston, C. (2009). Integrating evidence-based practices in middle science reading. *Teaching Exceptional Children*, 41(3), 6.	6. The study design must have an "active" control. This is a single case study
LaGue, K. M., & Wilson, K. (2010). Using peer tutors to improve reading comprehension. *Kappa Delta Pi Record*, 46(4), 182-186.	4. The intervention must be an empirical paper on an English "reading" intervention.
Lingo, A. S. (2014). Tutoring middle school students with disabilities by high school students: effects on oral reading fluency. *Education*	8. Outcome variable. The study must have an outcome variable that measures reading comprehension.

Excluded Study	Rationale for exclusion
and Treatment of Children, 37(1), 53-76.	This study measures reading fluency and accuracy only.
Maheady, L., Sacca, M. K., & Harper, G. F. (1988). Classwide peer tutoring with mildly handicapped high school students. *Exceptional Children,* 55(1), 52-59.	6. The study design must have an "active" control. This study is small N design.
Martino, L. R. (1994). Peer tutoring classes for young adolescents: A cost-effective strategy. *Middle School Journal*, *25*(4), 55-58.	4. The intervention must be an empirical paper on an English "reading" intervention. This is not a new study but an article on useful strategies.
Mathes, P. G., Fuchs, D., & Fuchs, L. S. (1995). Accommodating diversity through Peabody classwide peer tutoring. *Intervention in School and Clinic*, 31(1), 46-50.	4. The intervention must be an empirical paper on an English "reading" intervention. This is not a new study but a review
Miciano, R. Z. (2006). Piloting a peer literacy program: Implications for teacher education. Asia Pacific Education Review, 7(1), 76-84.	6. The study design must have an "active" control. This study is pre-post.
Miracle, S. A., Collins, B. C., Schuster, J. W., & Grisham-Brown, J. (2001). Peer-versus teacher-delivered instruction: Effects on acquisition and maintenance. *Education and Training in Mental Retardation and Developmental Disabilities*, 373-385.	6. The study design must have an "active" control. This study is small N design.
Neddenriep, C. E., Skinner, C. H., Wallace, M. A., & McCallum, E. (2009). Classwide peer tutoring: Two experiments investigating the generalized relationship between increased oral reading fluency and reading comprehension. *Journal of applied school psychology*, 25(3), 244-269.	6. The study design must have an "active" control. This study is a Small N Design.
Oakland, T., & Williams, F. C. (1975). An evaluation of two methods of peer tutoring. *Psychology in the Schools*, 12(2), 166-171.	7. Participants must be age 11+ attending a middle school/high school or secondary school. In this study they are attending an elementary school
Palincsar, A. S., Brown, A. L., & Martin, S. M. (1987). Peer interaction in reading comprehension instruction. *Educational psychologist,* 22(3-4), 231-253.	5. Intervention. It must be a peer-tutoring intervention This is a tutoring intervention, not a peer tutoring intervention.
Pickens, J., & McNaughton, S. (1988). Peer tutoring of comprehension strategies. *Educational Psychology*, 8(1-2), 67-80.	6. The study design must have an "active" control. This study is a Small N Design.
Polirstok, S. R., & Greer, R. D. (1986). A replication of collateral effects and a component analysis of a successful tutoring package	6. The study design must have an "active" control. This study is a Small N Design.

Excluded Study	Rationale for exclusion
for inner-city adolescents. *Education and Treatment of Children*, 101-121.	
Shaver, J. P., & Nuhn, D. (1971). The effectiveness of tutoring underachievers in reading and writing. *The Journal of Educational Research*, 65(3), 107-112.	8. Outcome variables. The outcomes must look at reading comprehension. This study only has a whole Reading score.
Spörer, N; Brunstein, J.C. (2009). Fostering the reading comprehension of secondary school students through peer-assisted learning: Effects on strategy knowledge, strategy use, and task performance. *Contemporary Educational Psychology 34*, pp. 289–297	2. The study must be written in English. This is a study written in German.
Topping, K; Miller, D; Thurston, A; McCavock, K; Conlin, N. (2011). Peer tutoring in reading in Scotland: Thinking big. *Literacy* Volume 45 Number 1, pp, 3-9	4. The intervention must be an empirical paper on an English “reading” intervention. This is not a new study but a review.
Topping, K. (1991). Achieving more with less: Raising reading standards via parental involvement and peer tutoring. *Support for Learning*, 6(3), 112-115.	4. The intervention must be an empirical paper on an English “reading” intervention. This is not a new study but a review
Trovato, J., & Bucher, B. (1980). Peer tutoring with or without home-based reinforcement, for reading remediation. *Journal of Applied Behavior Analysis*, 13(1), 129-141.	7. Participants must be age 11+ attending a middle school/high school or secondary school. In this study they are an elementary school.
Van Keer, H., & Vanderlinde, R. (2010). The impact of cross-age peer tutoring on third and sixth graders' reading strategy awareness, reading strategy use, and reading comprehension. *Middle Grades Research Journal*, 5(1), 33-46.	7. Participants must be age 11+ attending a middle school/high school or secondary school. In this study they are up to age 11 and attending an elementary school
Willis, J., & Crowder, J. (1974). Does tutoring enhance the tutor's academic learning? *Psychology in the Schools*.	5. Intervention. It must be a peer-tutoring intervention This is a tutoring intervention, not a peer tutoring intervention.

Appendix B Details of included studies

Author	Sample	Study Design	Age of participants	Other relevant characteristics	Country	Intervention	Deliverers	Outcome variables	Outcome	Follow up
Mastropieri, Scruggs, Mohler, Beranek, Spencer, Boon & Talbott. (2001)	24 pupils were tutees	RCT	12 and 13 years	17 were male, 21 Caucasian, 2 Hispanic-American and 1 African American. 20 with learning disabilities and 4 with mild retardation	US	Tutoring instruction based on Classwide Peer Tutoring (CWPT) and Peer Assisted Learning (PALs) 50 minutes daily for 5 weeks	2 special education teachers and undergraduate student	Criterion referenced comprehension test. Students' views of tutoring	There was a significant difference between the two conditions Tutoring condition students scored 81.8% correct (SD = 13.9) and control condition scored 63.3% (SD=12). 83% liked the peer tutoring.	n/a
Carlton, Litton, & Zinkgraf. (1985)	60 pupils were tutees and assessed	RCT	11-13 years	Classified as "mildly mentally retarded" and attended self-contained classes	US	Flash card drill to investigate the effects of an intraclass peer tutoring program on sight-word recognition skills	Teachers	Reading Test. Vocab test	Tutors and tutees had significantly higher gain scores than did the controls	n/a
Calhoon. (2005)	38	Intervention and contrast	32 6th, 5 7th and 1 8th grades	Students were from four special	US	LST/PALS based on Fuchs et al.(1993).	4 teachers	Diagnostic achievement test WJ-III:	Statistically significant differences	n/a

Author	Sample	Study Design	Age of participants	Other relevant characteristics	Country	Intervention	Deliverers	Outcome variables	Outcome	Follow up
				education self-contained language arts classrooms and all had a reading disability		Mediated verbal rehearsal, step by step positive feedback, frequent verbal and written interaction between tutor and tutee and reciprocity. LST was a phonological skill program.		Letter word identification, word attack, reading fluency and passage comprehension	berween groups LST/PALS outperforming contrast condition on letter word, word attack and passage comprehension. No significant differences found between reading fluency.	
Wheldall & Mettem. (1985)	24 pupils were tutees	RCT	11-12 years old	Tutees: The pupils were selected from the remedial classes of a large mixed comprehensive school. Tutors: 16 fifth year pupils	UK	Pause, Prompt and Praise intervention with a standard control, an untrained peer tutored control group as well as the trained peer tutored group. Three sessions a week for eight weeks.	Older peers (16 year olds) were the deliverers of the intervention. They were trained first.	Reading accuracy and comprehension using Neale Analysis of Reading Ability. Tutoring skills	Intervention group tutees made a mean gain of 6 months (maintained at follow up), Control group 1 (untrained) made a mean gain of 2.4 months and a mean gain of 0.9 months for the Control	Form C of Neale Analysis of Reading Ability was given two months later.

Author	Sample	Study Design	Age of participants	Other relevant characteristics	Country	Intervention	Deliverers	Outcome variables	Outcome	Follow up
									group 2 (no peer tutoring). The comprehension scores were not statistically different at post-test stage between the three groups although there were higher mean gains in the intervention group than the other groups.	
Fuchs, Fuchs & Kazdan. (1999)	Unknown But 18 classes with 102 in total- a selection were chosen to be assessed	RCT	Average age was 15	All classes were taught special education remedial reading and those chosen to be assessed had to have Grade 2 – 6 reading levels.	US	Peer Assisted Learning Strategy (PALs)	18 special education teachers teach and monitor the intervention that the peers deliver.	CRAB reading test. Student belief questionnaire.	The number of correct questions – the growth of PALs students exceeded that of contrast students (ES = .34).	n/a

Author	Sample	Study Design	Age of participants	Other relevant characteristics	Country	Intervention	Deliverers	Outcome variables	Outcome	Follow up
				Information about the pupils assessed is not known.						
Mastropieri, Scruggs, Spencer and Fontana (2003)	16 pupils were tutees	RCT	Average age was 15 years	They had mild disabilities. 14 with a learning disability and one with mild mental retardation and one with a learning disability and emotional disability. 81% were males. 6 Caucasian, 4 Hispanic American, 4 African-American and 2 Asian-American	US	Tutoring instruction based on Classwide Peer Tutoring (CWPT) and Peer Assisted Learning (PALs)	One special education teacher and an assistant co-teacher	Content tests (Chapter test, Unit tests and End of ear final exam). Reading fluency Reading comprehension strategy Surveys of student views of instruction	The tutoring condition resulted in significantly higher gains on content-area tests. No significant differences were found between conditions for fluency. Pupils in the tutoring condition had better comprehension that the control group	n/a

Appendix C: Weight of Evidence

Weight of Evidence A: Methodological quality

In order to assess the methodological quality of the studies an adapted version of Kratochwill's (2003) coding protocol was used for all the studies in order to fit the review question. The protocols and a list of amendments can be found in Appendix D. There were a few protocols available but this protocol was chosen as it was designed by a "Task force on interventions in school psychology" to review group based interventions and this reflected the types of studies being reviewed. The protocol enabled each study to be rated and thus generate Weight of Evidence (WoE) A. The table below provides information on each rated section of the Kratochwill's protocol and its criteria.

Note that the Identifiable Components have not been included as all the studies scored a 0 on this section and it was not deemed necessary to score this for the review question.

The scoring criteria for WoE A is highlighted below:

High = 2.40 – 3.00
Medium = 1.70 – 2.30
Low = 1.00 – 1.60

Weight of Evidence A: Methodological quality

Area	Weight of evidence (Strong=3, Promising = 2, Weak = 1, limited/no = 0)	Criteria
Measurement	Strong	A rating of **3** requires that, instruments produce a reliability coefficient of .85 or higher. The reliability information should be provided including the type of reliability statistic used. Data should have been collected using multiple methods, and collected from multiple sources, when appropriate. A case needs to be presented for using those measures.
	Promising	A rating of **2** requires that, instruments produce a reliability coefficient of .70 or higher. In addition, data should have been collected either (1) using multiple methods and/or (2) from multiple sources, when appropriate.
	Weak	A rating of **1** requires that, instruments produce a reliability coefficient of .50 or higher.
	Limited/no	A rating of **0** indicates that the measures did not produce reliable scores or produced scores with low reliability (<.50), AND/OR data were not collected using multiple methods.
Comparison Group	Strong	To score a **3** there must be at least one type of "active" comparison group. Initial group equivalency must be established, preferably through random assignment of participants to intervention conditions. There must be evidence that change agents were counterbalanced, as well as the study must meet the criteria for equivalent mortality and low attrition at post, and if applicable, at follow-up

Area	Weight of evidence (Strong=3, Promising = 2, Weak = 1, limited/no = 0)	Criteria
Comparison group	Promising	To score a **2** there must be, at least a "no intervention group". In addition, there must be evidence for at least two of the following: (1) counterbalancing of change agents, (2) group equivalence established, or (3) equivalent mortality with low attrition.
	Weak	A rating of **1** would require a comparison group and at least one of the following: (1) counterbalancing of change agents, (2) group equivalence established, or (3) equivalent mortality with low attrition.
	Limited/no	A rating of **0** indicates that no efforts were made to ensure group equivalence.
Fidelity	Strong	To score a **3** fidelity should be measured through at least two of the following: ongoing supervision/consultation, coding sessions, or audio/video tapes, and use of a manual (written materials or a formal training). The study must also demonstrate evidence of acceptable adherence.
	Promising	To score a **2** fidelity should be measured through at least one of the following: ongoing supervision/consultation, coding sessions, or audio/video tapes, and use of a manual (written materials or a formal training). The study must also demonstrate evidence of acceptable adherence.
	Weak	To score a **1** fidelity should be measured through at least one of the following: ongoing supervision/consultation, coding sessions, or audio/video tapes, OR use of a manual (written materials or a formal training). The study must also demonstrate evidence of acceptable adherence.
		To score a 0 the study would have done

Area	Weight of evidence (Strong=3, Promising = 2, Weak = 1, limited/no = 0)	Criteria
	Limited/no	nothing to ensure implementation fidelity or evidence indicates unacceptable adherence.
Follow up	Strong	To receive a rating of **3,** the study must have conducted follow up assessments over multiple intervals (e.g., 6 months, 1 year), with all participants that were included in the original sample, using similar measures used to analyse data from primary or secondary outcomes.
Follow up	Promising	To receive a rating of **2,** the study must have conducted follow up assessments at least once (e.g., 6 months), with the majority of participants that were included in the original sample, using similar measures used to analyse data from primary or secondary outcomes.
	Weak	To receive a rating of 1, there must been a follow up at least once (e.g., 6 months), with some participants from the original sample.
	Limited/no	A rating of **0**, would indicate that no follow up assessment was built into the study.

Weight of Evidence A: Methodological quality

Author	Measurement	Comparison Group	Fidelity	Follow up	Overall WoE A
Wheldall,& Mettem (1985)	1	2	3	1	Medium (1.75)
Mastropieri et al. (2001)	1	2	3	0	Low (1.5)
Fuchs et al. (1999)	1	2	2	2	Medium (1.75)

Carlton et al. (1985)	0	2	2	0	Low (1.0)
Calhoon (2005)	1	1	3	0	Low (1.25)
Mastropieri et al. (2003)	2	2	3	2	Medium(2.25)

Weight of Evidence B: Methodological relevance

Weight of Evidence B refers to the appropriateness of the type of evidence/design of the study to be able to answer this review question.

As this review question is specifically looking at reading comprehension only studies with an outcome variable measuring reading comprehension, as per the inclusion/exclusion criteria, are included. Therefore, reading comprehension measures will not be covered in WoE B.

The following were identified as areas to weight according to the review question:

Randomness

Fully randomised control is considered the most robust study design to measure effectiveness (Kratochwill, 2003) as it aims to minimise the risk of confounding variables and provide the most reliable evidence of the effectiveness of an intervention. However within each study there were differences in how random they allocated interventions. It was thus decided that the level of "randomness" would be a WoE B criteria.

Weight of Evidence B: Randomness

Weighting	Descriptive

High (3)	Fully randomised control
Medium (2)	Randomised Block control
Low (1)	Non-randomised control/unknown

Control group intervention

As the review question asks about *effectiveness* it was deemed appropriate for all the review studies to contain an “active” control group, as identified in the inclusion/exclusion criteria. This enables a comparison of treatment and non-treatment group, reducing the chances of other factors influencing changes to the pupils’ comprehension ability following a peer tutored Reading intervention. There are however different types of “active” control, such as a placebo, wait list or comparison group.

It is also important that the control group intervention is as similar to the peer group intervention, without the peer tutoring element (this is the “active ingredient), as possible to try and control for any other variables that may influence the reading outcomes, other than the peer tutoring.

Weight of Evidence B: Type of control group

Weighting	Descriptive
High (3)	Same time allocated to both groups AND same materials used in both groups
Medium (2)	Same time allocated to both groups OR same materials used in both groups

Low (1)	Unknown aspects of control group

Overall WoE B scoring criteria:

High = 2.40 – 3.00
Medium = 1.70 – 2.30
Low = 1.00 – 1.60

Weight of Evidence B: Methodological relevance

Author	Randomness	Control group	Overall WoE B
Wheldall,& Mettem (1985)	1	2	Low (1.5)
Mastropieri et al. (2001)	2	3	High (2.5)
Fuchs et al. (1999)	1	2	Low (1.5)
Carlton et al. (1985)	3	1	Medium (2.00)
Calhoon (2005)	2	1	Low (1.5)
Mastropieri et al. (2003)	2	3	High (2.5)

Weight of Evidence C (WoE C)

Weight of Evidence C rates the relevance of focus of study to the review question i.e. how close if this topic to my question?

The inclusion/exclusion criteria for this review only looked at secondary age pupils so that will not be a factor in the WoE C criteria. This review looks at the effectiveness of peer tutoring and the inclusion/exclusion criteria already only allows for studies comparing a peer tutoring intervention with a control group so this will not be a factor in the WoE C weighting.

Reading Comprehension

As this review looks specifically at reading comprehension and although it was part of the inclusion/exclusion criteria the studies were different on what aspects of reading they focused on. Therefore, it was deemed important to weight the studies in terms how focused they were on reading comprehension and thus how much could the study answer the review question.

The ratings for comprehension are shown below.

Weight of Evidence C: Reading comprehension

Weighting	Descriptive
High (3)	Comprehension main focus
Medium (2)	Comprehension and reading fluency main focus
Low (1)	Reading fluency main focus

Mainstream

The review question is around mainstream school so the studies that use mainstream pupils will score a higher weighting than those that focus on SEN pupils. See below for details.

Weight of Evidence C: Participants' education setting

Weighting	Descriptive
High (3)	All mainstream lessons
Medium (2)	Mainstream lessons for some of the time
Low (1)	Unit or resource in mainstream school for all of the time

Overall WoE C scoring criteria:

High = 2.40 – 3.00
Medium = 1.70 – 2.30
Low = 1.00 – 1.60

Weight of Evidence C: Relevance of focus of study to the Review Question

Author	Reading comprehension	Mainstream	Overall WoE C
Wheldall,& Mettem (1985)	2	1	Low (1.5)
Mastropieri et al. (2001)	3	2	High (2.5)
Fuchs et al. (1999)	2	1	Low (1.5)
Carlton et al. (1985)	1	1	Low

			(1.00)
Calhoon (2005)	2	2	Medium (2.00)
Mastropieri et al. (2003)	3	2	High (2.5)

Weight of Evidence D

Using the criteria explained above, each study was given a weighting of between 1 and 3 for A, B and C. The mean average of these three scores was calculated and corresponds to an overall weight (WoE D) for each study.

Overall WoE D scoring criteria:

High = 2.40 – 3.00
Medium = 1.70 – 2.30
Low = 1.00 – 1.60

Weight of Evidence D: Overall Weight of Evidence

Author	WoE A	WoE B	WoE C	Overall WoE D
Wheldall,& Mettem (1985)	Medium (1.75)	Low (1.5)	Low (1.5)	Low (1.59)

Mastropieri et al. (2001)	Low (1.5)	High (2.5)	High (2.5)	Medium (2.12)
Fuchs et al. (1999)	Medium (1.75)	Low (1.5)	Low (1.5)	Low (1.58)
Carlton et al. (1985)	Low (1.0)	Medium (2.00)	Low (1.00)	Low (1.33)
Calhoon (2005)	Low (1.25)	Low (1.5)	Medium (2.00)	Low (1.58)
Mastropieri et al. (2003)	Medium (2.25)	High (2.5)	High (2.5)	High (2.42)

Appendix D.1 Coding protocol amendments

Section Excluded	Rationale
Domain (apart from “Academic intervention programs”)	Review question focuses only on academic intervention programs
B7 Coding- For studies using qualitative research methods	Inclusion criteria means it has to be quantitative data
C. Primary/Secondary Outcomes Are Statistically Significant	Only outcomes relating to reading comprehension were being analysed and effect sizes were being calculated in another section of the review
D. Educational/Clinical Significance	Education significance is being assessed as part of the main review
G. Replication (answer G1, G2, G3, and G4)	Not relevant to current review
H. Site of Implementation	The review was only looking at state funded mainstream school based interventions
A 2. Participant Characteristics Specified for Treatment and Control group	Information already gathered and provided in summary of study table
A4. Receptivity/acceptance by target participant population (treatment group)	Not relevant to the review

Coding Protocol: Group-Based Design

Domain:

- ☐ ~~School- and community-based intervention programs for social and behavioral problems~~
- ☐ Academic intervention programs
- ☐ ~~Family and parent intervention programs~~
- ☐ ~~School-wide and classroom-based programs~~
- ☐ ~~Comprehensive and coordinated school health services~~

Name of Coder(s): XXXXXX

Date: 05.02.2016
M / D / Y

Full Study Reference in APA format: Wheldall, K; Mettem, P. (1985). Behavioural peer tutoring: Training 16-year-old tutors to employ the "Pause, Prompt and Praise" method with 12-year-old remedial readers. *Educational Psychology*, Vol. S, No. 1. pp. 27-44

Intervention Name (description from study): Peer tutored intervention using “pause, prompt and praise” Reading intervention

Study ID Number (Unique Identifier): 1

Type of Publication: (Check one)

- ☐ Book/Monograph
- ☐ Journal article
- ☐ Book chapter
- ☐ Other (specify):

I. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)

A1.1 ☐ Completely randomized design
A1.2 ☐ Randomized block design (between-subjects variation)
A1.3 ☐ Randomized block design (within-subjects variation)
A1.4 ☐ Randomized hierarchical design

All participants were taken from "remedial" reading classes

A2. Nonrandomized designs (if nonrandom assignment design, select one of the following)

A2.1 ☐ Nonrandomized design
A2.2 ☐ Nonrandomized block design (between-participants variation)
A2.3 ☐ Nonrandomized block design (within-participants variation)
A2.4 ☐ Nonrandomized hierarchical design
A2.5 ☐ Optional coding of Quasi-experimental designs (see Appendix C)

A3. Overall confidence of judgment on how participants were assigned (select one of the following)

A3.1 ☐ Very low (little basis)
A3.2 ☐ Low (guess)
A3.3 ☐ Moderate (weak inference)
A3.4 ☐ High (strong inference)
A3.5 ☐ Very high (explicitly stated)
A3.6 ☐ N/A
A3.7 ☐ Unknown/unable to code

B. Statistical Treatment/Data Analysis (answer B1 through B6)

B1. Appropriate unit of analysis	yes	
B2. Familywise error rate controlled	Yes	No n/a
B3. Sufficiently large *N*	yes	No

Statistical Test: **ANCOVA 3 groups**
_ level: **0.05**
ES: **Medium effect size anticipated**
N required: **74 in total**

B4. Total size of sample (start of the study): **24- not sure if this is the case at post testing though**
N

B5. Intervention group sample size: **8**
N

B6. Control group sample size: **8 in Control 1 group and 8 in Control 2 group**
N

~~**For studies using qualitative research methods, code B7 and B8**~~

~~B7. Coding~~

~~B7.1 Coding scheme linked to study's theoretical-empirical basis (select one) ☐ yes ☐ no~~

~~B7.2 Procedures for ensuring consistency of coding are used (select one) ☐ yes ☐ no~~
~~Describe procedures: _____~~

~~B7.3 Progression from abstract concepts to empirical exemplars is clearly articulated (select one) ☐ yes ☐ no~~

~~B8. Interactive process followed (select one) ☐ yes ☐ no~~

~~Describe process: _____~~

C. Type of Program (select one)

C1. ☐ Universal prevention program
C2. ☐ Selective prevention program
C3. ☐ Targeted prevention program
C4. ☐ Intervention/Treatment
C5. ☐ Unknown

It is Pause, Prompt and Praise reading intervention using cross age peers as the tutors.

D. Stage of the Program (select one)

D1. ☐ Model/demonstration programs
D2. ☐ Early stage programs
D3. ☐ Established/institutionalized programs
D4. ☐ Unknown

Although Pause, Prompt and Praise is established it was used with parents and not peers as tutors

E. Concurrent or Historical Intervention Exposure (select one)

E1. ☐ Current exposure
E2. ☐ Prior exposure
E3. ☐ Unknown

The participants are all currently taught in remedial classes

II. Key Features for Coding Studies and Rating Level of Evidence/ Support

(3=Strong Evidence 2=Promising Evidence 1=Weak Evidence 0=No Evidence)

A. Measurement (answer A1 through A4)

A1. Use of outcome measures that produce reliable scores for the majority of primary outcomes. The table for Primary/Secondary Outcomes Statistically Significant allows for listing separate outcomes and will facilitate decision making regarding measurement (select one of the following)

A1.1☐ Yes
A1.2☐ No
A1.3☐ Unknown/unable to code

The measure used for looking at Comprehension was the Neale Analysis of Reading Ability Forms A-C. Test retest reliability is increased with 3 different forms. This does not matter in this case as the scores are being compared to the other group not the population as a whole. I was unable to find a reliability coefficient in the article or on the internet.

A2. Multi-method (select one of the following)

A2.1☐ Yes
A2.2☐ No
A2.3☐ N/A
A2.4☐ Unknown/unable to code

A3. Multi-source (select one of the following)

A3.1☐ Yes
A3.2☐ No
A3.3☐ N/A
A3.4☐ Unknown/unable to code

A4. Validity of measures reported (select one of the following)

A5.1☐ Yes validated with specific target group
A5.2☐ In part, validated for general population only
A5.3☐ No
A5.4☐ Unknown/unable to code

There is some question over the validity of the measure as it was designed to measure Reading ability as a whole. It is standardized up to 13 year olds but for the general population. There has been some research that suggests you cannot separate the accuracy and comprehension parts and that the NARA can underestimate the comprehension ability of children with weak decoding skills and children who have some difficulty with open-ended questions. If there are differences between the groups in terms of decoding skills then this will be an issue, however an ANOVA on pre test scores of accuracy showed the groups were not significant.

Rating for Measurement (select 0, 1, 2, or 3)**:** ☐ **3** ☐ **2** ☐ **1** ☐ **0**

The NARA is somewhat reliable but it was the only source and method of measuring the comprehension of the participants.

B. Comparison Group

B1. Type of Comparison Group (select one of the following)

B1.1 ☐ Typical contact
B1.2 ☐ Typical contact (other) specify:
B1.3 ☐ Attention placebo
B1.4 ☐ Intervention elements placebo
B1.5 ☐ Alternative intervention
B1.6 ☐ PharmacotherapyB1.1
B1.7 ☐ No intervention
B1.8 ☐ Wait list/delayed intervention
B1.9 ☐ Minimal contact
B1.10 ☐ Unable to identify comparison group

There were two control groups- Control Group 1 – there were peer tutors who delivered the PPP programme but they were not trained in it and Control Group 2 in which the group received their usual reading activity of silent reading.

Rating for Comparison Group (select 0, 1, 2, or 3)**:** ☐ **3** ☐ **2** ☐ **1** ☐ **0**

There is at least 1 active control group and they were randomly chosen to be in either of the three groups. They were equivalent on their reading accuracy scores at pre-test. There was not any counterbalancing of the change agents (i.e. tutors or teachers swapping) but attrition was low.

B2. Overall confidence rating in judgment of type of comparison group (select one of the following)

B2.1 ☐ Very low (little basis)
B2.2 ☐ Low (guess)
B2.3 ☐ Moderate (weak inference)
B2.4 ☐ High (strong inference)
B2.5 ☐ Very high (explicitly stated)
B2.6 ☐ Unknown/Unable to code

B3. Counterbalancing of Change Agents (answer B3.1 to B3.3)

B3.1 ☐ By change agent
B3.2 ☐ Statistical
B3.3 ☐ Other

Unknown

B4. Group Equivalence Established (select one of the following)

B4.1 ☐ Random assignment
B4.2 ☐ Posthoc matched set
B4.3 ☐ Statistical matching
B4.4 ☐ Post hoc test for group equivalence

The participants were all in remedial classes and matched into dyads and then the dyads were randomly selected to either group.

B5. Equivalent Mortality (answer B5.1 through B5.3)

B5.1☐ Low Attrition (less than 20% for Post)

B5.2☐ Low Attrition (less than 30% for follow-up)

B5.3☐ Intent to intervene analysis carried out

Findings_____

~~**C. Primary/Secondary Outcomes Are Statistically Significant**~~

~~C1. Evidence of appropriate statistical analysis for **primary outcomes** (answer C1.1 through C1.3)~~

~~C1.1☐ Appropriate unit of analysis (rate from previous code)~~

~~C1.2☐ Familywise/experimenterwise error rate controlled when applicable (rate from previous code)~~

~~C1.3☐ Sufficiently large *N* (rate from previous code)~~

~~C2. Percentage of **primary outcomes** that are significant (select one of the following)~~

~~C2.1☐ Significant primary outcomes for at least 75% of the total primary outcome measures for each key construct~~

~~C2.2☐ Significant primary outcomes for between 50% and 74% of the total primary outcome measures for each key construct~~

~~C2.3☐ Significant primary outcomes for between 25% and 49% of the total primary outcome measures for any key construct~~

~~**Rating for Primary Outcomes Statistically Significant** (select 0, 1, 2, or 3)**:** ☐**3** ☐**2** ☐**1** ☐**0**~~

~~C3. Evidence of appropriate statistical analysis for **secondary outcomes** (answer C3.1 through C3.3)~~

~~C3.1☐ Appropriate unit of analysis~~

~~C3.2☐ Familywise/experimenterwise error rate controlled when applicable (rate from previous code)~~

~~C3.3☐ Sufficiently large *N* (rate from previous code)~~

~~C4. Percentage of **secondary outcomes** that are significant (select one of the following)~~

~~C4.1☐ Significant secondary outcomes for at least 75% of the total secondary outcome measures for each key construct~~

~~C4.2☐ Significant secondary outcomes for between 50% and 74% of the total secondary outcome measures for each key construct~~

~~C4.3☐ Significant secondary outcomes for between 25% and 49% of the total secondary outcome measures for any key construct~~

~~**Rating for Secondary Outcomes Statistically Significant** (select 0, 1, 2, or 3)**:** ☐**3** ☐**2** ☐**1** ☐**0**~~

~~C5. Overall Summary of Questions Investigated~~

~~C5.1 Main effect analyses conducted~~	~~(select one)~~	~~yes~~	~~no~~
~~C5.2 Moderator effect analyses conducted~~	~~(select one)~~	~~yes~~	~~no~~

~~Specify results:~~ _____

~~C5.3. Mediator analyses conducted (select one) ☐ yes ☐ no~~

~~Specify results:~~ _____

~~C. Primary/Secondary Outcomes Statistically Significant (only list *p* ≤ .05)~~

~~(list primary outcomes first in alphabetical order, followed by secondary outcomes in alphabetical order)~~

~~Outcomes~~	~~Primary vs. Secondary~~	~~Who Changed~~		~~What Changed~~	~~Source~~	~~Treatment Information~~	~~Outcome Measure Used~~	~~Reliability~~	~~ES~~	~~(1-_)~~
~~Outcome #1:~~	☐ ~~Primary~~ ☐ ~~Secondary~~ ☐ ~~Unknown~~	☐ ~~Child~~ ☐ ~~Teacher~~ ☐ ~~Parent/sign. adult~~ ☐ ~~Ecology~~ ☐ ~~Other~~ ~~Unknown~~		☐ ~~Behavior~~ ☐ ~~Attitude~~ ☐ ~~Knowledge~~ ☐ ~~Other~~ ☐ ~~Unknown~~	~~Self Report~~ ☐ ~~Parent Report~~ ☐ ~~Teacher Report~~ ☐ ~~Observation~~ ☐ ~~Test~~ ☐ ~~Other~~ ~~Unknown~~					
~~Outcome #2~~	☐ ~~Primary~~ ☐ ~~Secondary~~ ☐ ~~Unknown~~	☐ ~~Child~~ ☐ ~~Teacher~~ ☐ ~~Parent/sign. Adult~~ ☐ ~~Ecology~~ ☐ ~~Other~~ ~~Unknown~~		☐ ~~Behavior~~ ☐ ~~Attitude~~ ☐ ~~Knowledge~~ ☐ ~~Other~~ ☐ ~~Unknown~~	~~Self Report~~ ☐ ~~Parent Report~~ ☐ ~~Teacher Report~~ ☐ ~~Observation~~ ☐ ~~Test~~ ☐ ~~Other~~ ~~Unknown~~					
~~Outcome #3:~~	☐ ~~Primary~~ ☐ ~~Secondary~~ ☐ ~~Unknown~~	☐ ~~Child~~ ☐ ~~Teacher~~ ☐ ~~Parent/sign. Adult~~ ☐ ~~Ecology~~ ☐ ~~Other~~ ☐ ~~Unknown~~		☐ ~~Behavior~~ ☐ ~~Attitude~~ ☐ ~~Knowledge~~ ☐ ~~Other~~ ☐ ~~Unknown~~	~~Self Report~~ ☐ ~~Parent Report~~ ☐ ~~Teacher Report~~ ☐ ~~Observation~~ ☐ ~~Test~~ ☐ ~~Other~~ ☐ ~~Unknown~~					
~~Outcome #4:~~	☐ ~~Primary~~ ☐ ~~Secondary~~ ☐ ~~Unknown~~	☐ ~~Child~~ ☐ ~~Teacher~~ ☐ ~~Parent/sign. Adult~~ ☐ ~~Ecology~~ ☐ ~~Other~~ ~~Unknown~~		☐ ~~Behavior~~ ☐ ~~Attitude~~ ☐ ~~Knowledge~~ ☐ ~~Other~~ ☐ ~~Unknown~~	~~Self Report~~ ☐ ~~Parent Report~~ ☐ ~~Teacher Report~~ ☐ ~~Observation~~ ☐ ~~Test~~ ☐ ~~Other~~ ~~Unknown~~					
~~Outcome #5:~~	☐ ~~Primary~~ ☐ ~~Secondary~~ ☐ ~~Unknown~~	☐ ~~Child~~ ☐ ~~Teacher~~ ☐ ~~Parent/sign. Adult~~ ☐ ~~Ecology~~ ☐ ~~Other~~ ~~Unknown~~		☐ ~~Behavior~~ ☐ ~~Attitude~~ ☐ ~~Knowledge~~ ☐ ~~Other~~ ☐ ~~Unknown~~	~~Self Report~~ ☐ ~~Parent Report~~ ☐ ~~Teacher Report~~ ☐ ~~Observation~~ ☐ ~~Test~~ ☐ ~~Other~~ ~~Unknown~~					

Null Findings/Negative Outcomes Associated with the Intervention (listed alphabetically by outcome)

Outcomes	Primary vs. Secondary	Who Was Targeted for Change		What Was Targeted for Change	Source	Note null/negative outcomes	Outcome Measure Used	Reliability	ES
Outcome #1:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other Unknown		☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other Unknown				
Outcome #2	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other Unknown		☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other Unknown				
Outcome #3:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other Unknown		☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other Unknown				
Outcome #4:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other ☐ Unknown		☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	☐ ☐ ☐ ☐ ☐ ☐	Self Report Parent Report Teacher Report Observation Test Other Unknown			
Outcome #5:	☐ Primary ☐ Secondary ☐ Unknown	☐ Child ☐ Teacher ☐ Parent/sign. Adult ☐ Ecology ☐ Other Unknown		☐ Behavior ☐ Attitude ☐ Knowledge ☐ Other ☐ Unknown	Self Report ☐ Parent Report ☐ Teacher Report ☐ Observation ☐ Test ☐ Other Unknown				

Type of Data Effect Size is Based On	Confidence Rating in ES Computation
(check all that apply) ☐ Means and SDs ☐ *t* - value or *F* – value ☐ Chi-square (*df* = 1) ☐ Frequencies or proportions (dichotomous) ☐ Frequencies or proportions (polytomous) ☐ Other (specify): ☐ Unknown	(select one of the following) ☐ Highly estimated (e.g., only have *N* p value) ☐ Moderate estimation (e.g., have complex but complete statistics) ☐ Some estimation (e.g., unconventional statistics that require conversion) ☐ Slight estimation (e.g., use significance testing statistics rather than descriptives) ☐ No estimation (e.g., all descriptive data is present)

D. Educational/Clinical Significance

Outcome Variables:	Pretest	Posttest	Follow Up
D1. Categorical Diagnosis Data	Diagnostic information regarding inclusion into the study presented: ☐ Yes ☐ No ☐ Unknown	Positive change in diagnostic criteria from pre to posttest: ☐ Yes ☐ No ☐ Unknown	Positive change in diagnostic criteria from posttest to follow up: ☐ Yes ☐ No ☐ Unknown
D2. Outcome Assessed via continuous Variables		Positive change in percentage of participants showing clinical improvement from pre to posttest: ☐ Yes ☐ No ☐ Unknown	Positive change in percentage of participants showing clinical improvement from posttest to follow up: ☐ Yes ☐ No ☐ Unknown
D3. Subjective Evaluation: The importance of behavior change is evaluated by individuals in direct contact with the participant.	Importance of behavior change is evaluated: ☐ Yes ☐ No ☐ Unknown	Importance of behavior change from pre to posttest is evaluated positively by individuals in direct contact with the participant: ☐ Yes ☐ No ☐ Unknown	Importance of behavior change from posttest to follow up is evaluated positively by individuals in direct contact with the participant: ☐ Yes ☐ No ☐ Unknown
D4. Social Comparison: Behavior of participant at pre, post, and follow up is compared to normative data (e.g., a typical peer).	Participant's behavior is compared to normative data ☐ Yes ☐ No ☐ Unknown	Participant's behavior has improved from pre to posttest when compared to normative data: ☐ Yes ☐ No ☐ Unknown	Participant's behavior has improved from posttest to follow up when compared to normative data: ☐ Yes ☐ No ☐ Unknown

Rating for Educational/Clinical Significance (select 0, 1, 2, or 3)**:** ☐ **3** ☐ **2** ☐ **1** ☐ **0**

E. Identifiable Components (answer E1 through E7)

E1. Evidence for primary outcomes (rate from previous code): ☐ 3 ☐ 2 ☐ 1 ☐ 0 n/a

E2. Design allows for analysis of identifiable components (select one) ☐ yes ☐ no

Although the Control Group 1 does not have the trained element, the intervention itself is not broken down into identifiable components that are analysed separately against the outcomes.

E3. Total number of components: **Unknown**

N

E4. Number of components linked to primary outcomes: **Treated as a whole**

N

Additional criteria to code descriptively:

E5. Clear documentation of essential components (select one) ☐ yes ☐ no

E6. Procedures for adapting the intervention are described in detail (select one) ☐ yes ☐ no

E7. Contextual features of the intervention are documented (select one) ☐ yes ☐ no

Rating for Identifiable Components (select 0, 1, 2, or 3)**:** ☐ **3** ☐ **2** ☐ **1** ☐ **0**

F. Implementation Fidelity

F1. Evidence of Acceptable Adherence (answer F1.1 through

F1.3) ☐ F1.1 Ongoing supervision/consultation
F1.2 ☐ Coding intervention sessions/lessons or procedures
F1.3 ☐ Audio/video tape implementation (select F1.3.1 or F1.3.2):

F1.3.1 ☐ Entire intervention
F1.3.2 ☐ Part of intervention

Alternate tutoring sessions were taped to monitor tutor ad tutee performance.

F2. Manualization (select all that apply)

F2.1 ☐ Written material involving a detailed account of the exact procedures and the sequence in which they are to be used

F2.2 ☐ Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used

F2.3 ☐ Written material involving an overview of broad principles and a description of the intervention phases

F2.4 ☐ Formal or informal training session involving an overview of broad principles and a description of the intervention phases

F3. Adaptation procedures are specified (select one) ☐ yes ☐ no ☐ unknown

Rating for Implementation Fidelity (select 0, 1, 2, or 3)**:** ☐ **3** ☐ **2** ☐ **1** ☐ **0**

There is ongoing consultation, feedback and a manual

~~**G. Replication** (answer G1, G2, G3, and~~

~~G4) ☐ G1. Same Intervention~~
~~G2. ☐ Same Target Problem~~
~~G3. ☐ Independent evaluation~~

~~**Rating for Replication** (select 0, 1, 2, or 3)**:** ☐ **3** ☐ **2** ☐ **1** ☐ **0**~~

~~**H. Site of Implementation**~~

~~H1. School (if school is the site, select one of the following~~

~~options)~~ ☐ ~~H1.1 Public~~

~~H1.2~~ ☐ ~~Private~~
~~H1.3~~ ☐ ~~Charter~~
~~H1.4~~ ☐ ~~University Affiliated~~
~~H1.5~~ ☐ ~~Alternative~~
~~H1.6~~ ☐ ~~Not specified/unknown~~

~~H2. Non School Site (if it is a non school site, select one of the following options)~~

~~H2.1~~ ☐ ~~Home~~
~~H2.2~~ ☐ ~~University Clinic~~
~~H2.3~~ ☐ ~~Summer Program~~
~~H2.4~~ ☐ ~~Outpatient Hospital~~
~~H2.5~~ ☐ ~~Partial inpatient/day Intervention Program~~
~~H2.6~~ ☐ ~~Inpatient Hospital~~
~~H2.7~~ ☐ ~~Private Practice~~
~~H2.8~~ ☐ ~~Mental Health Center~~
~~H2.9~~ ☐ ~~Residential Treatment Facility~~
~~H2.10~~ ☐ ~~Other (specify):_____~~
~~H2.11~~ ☐ ~~Unknown/insufficient information provided~~

~~**Rating for Site of Implementation** (select 0, 1, 2, or 3)**:** ☐ **3** ☐ **2** ☐ **1** ☐ **0**~~

I. Follow Up Assessment

☐ Timing of follow up assessment: specify **2 MONTHS**

☐ Number of participants included in the follow up assessment: specify **UNKNOWN**

☐ Consistency of assessment method used: specify **Version C of Neal Analysis of Reading**

Rating for Follow Up Assessment (select 0, 1, 2, or 3)**:** ☐ **3** ☐ **2** ☐ **1** ☐ **0**

It can only be rated a 1 as it is unclear how many participants were in the follow up

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

A1. Sampling procedures described in detail ☐ yes ☐ no

Specify rationale for selection: **Remedial readers in bottom classes as they want to improve reading ability in remedial readers. Not sure about rationale for the number of participants though**

Specify rationale for sample size: **not known**

A1.1 Inclusion/exclusion criteria specified ☐ yes ☐ no

A1.2 Inclusion/exclusion criteria similar to school practice ☐ yes ☐ no

A1.3 Specified criteria related to concern ☐ yes ☐ no

~~A2. Participant Characteristics Specified for Treatment and Control Group~~

~~Participants from Treatment Group~~	~~Grade/age~~	~~Gender~~	~~Ethnicity or Multi-ethnic~~	~~Ethnic Identity~~	~~Race(s)~~	~~Acculturation~~	~~Primary Language~~	~~SES~~	~~Family Structure~~	~~Locale~~	~~Disability~~	~~Functional Descriptors~~
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~												
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~												
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~												
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~												

~~Participants from Control Group~~	~~Grade/age~~	~~Gender~~	~~Ethnicity or Multi-ethnic~~	~~Ethnic Identity~~	~~Race(s)~~	~~Acculturation~~	~~Primary Language~~	~~SES~~	~~Family Structure~~	~~Locale~~	~~Disability~~	~~Functional Descriptors~~
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~												
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~												
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~												
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~												

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes ☐ yes ☐ no

Specify: ______________________________

A3.2 Have relevance to inclusion criteria ☐yes ☐no

Specify: **Remedial readers with the aim of improving their reading**

~~A4. Receptivity/acceptance by target participant population (treatment group)~~

~~**Participants from Treatment Group**~~	~~**Results**~~ ~~**(What person reported to have gained from participation in program)**~~	~~**General Rating**~~
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~		☐ ~~Participants reported benefiting overall from the intervention~~ ☐ ~~Participants reported not benefiting overall from the intervention~~
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~		☐ ~~Participants reported benefiting overall from the intervention~~ ☐ ~~Participants reported not benefiting overall from the intervention~~
☐ ~~Child/Student~~ ☐ ~~Parent/caregiver~~ ☐ ~~Teacher~~ ☐ ~~School~~ ☐ ~~Other~~		☐ ~~Participants reported benefiting overall from the intervention~~ ☐ ~~Participants reported not benefiting overall from the intervention~~

A5. Generalization of Effects:

A5.1 Generalization over time

A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated ☐yes ☐no

Specify:______________________________________

A5.1.2 Procedures for maintaining outcomes are specified ☐yes ☐no

Specify: _____________________________________

A5.2 Generalization across settings

A5.2.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context ☐yes ☐no

Specify: ____________________________________

A5.2.2 Documentation of efforts to ensure application of intervention to other settings ☐yes ☐no

Specify: _____

A5.2.3 Impact on implementers or context is sustained ☐ yes ☐ no

Specify: _____

A5.3 Generalization across persons

Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated
☐ yes ☐ no

Specify: _____

B. Length of Intervention (select B1 or B2)

B1. ☐ Unknown/insufficient information provided

B2. ☐ Information provided (if information is provided, specify one of the following:) B2.1 weeks **8**
N

B2.2 months _____
N

B2.3 years _____
N

B2.4 other _____
N

C. Intensity/dosage of Intervention (select C1 or

C2) ☐ C1. Unknown/insufficient information

provided

C2. ☐ Information provided (if information is provided, specify both of the following:)

C2.1 length of intervention session **unknown**
N

C2.2 frequency of intervention session **three times a week**
N

D. Dosage Response (select D1 or D2)

D1. ☐ Unknown/insufficient information provided

D2. ☐ Information provided (if information is provided, answer D2.1)

D2.1 Describe positive outcomes associated with higher dosage: _____

E. Program Implementer (select all that apply)

E1. ☐ Research Staff
E2. ☐ School Specialty Staff
E3. ☐ Teachers
E4. ☐ Educational Assistants
E5. ☐ Parents
E6. ☐ College Students
E7. Peers
E8. ☐ Other
E9. ☐ Unknown/insufficient information provided

F. Characteristics of the Intervener

F1. ☐ Highly similar to target participants on key variables (e.g., race, gender, SES)
F2. ☐ Somewhat similar to target participants on key variables
F3. ☐ Different from target participants on key variables

The tutors are older and have higher reading ability. The dyads are matched for same gender.

G. Intervention Style or Orientation (select all that apply)

G1. ☐ Behavioral
G2. ☐ Cognitive-behavioral
G3. ☐ Experiential
G4. ☐ Humanistic/interpersonal
G5. ☐ Psychodynamic/insight oriented
G6. ☐ other (specify):_____
G7. ☐ Unknown/insufficient information provided

H. Cost Analysis Data (select G1 or G2)

H1. ☐ Unknown/insufficient information provided
H2. ☐ Information provided (if information is provided, answer H2.1)

H2.1 Estimated Cost of Implementation:_____

I. Training and Support Resources (select all that apply)

I1. ☐ Simple orientation given to change agents
I2. ☐ Training workshops conducted

# of Workshops provided **Two**

Average length of training **30 minutes**_____

Who conducted training (select all that apply)

I2.1 ☐ Project Director
I2.2 ☐ Graduate/project assistants

I2.3 ☐ Other (please specify):
I2.3 ☐ Unknown

I3. ☐ Ongoing technical support
I4. ☐ Program materials obtained
I5. ☐ Special Facilities
I6. ☐ Other (specify):

J. Feasibility

J1. Level of difficulty in training intervention agents (select one of the following)

J1.1 ☐ High
J1.2 ☐ Moderate
J1.3 ☐ Low
J1.4 ☐ Unknown

J2. Cost to train intervention agents (specify if known): _____

J3. Rating of cost to train intervention agents (select one of the following)

J3.1 ☐ High
J3.2 ☐ Moderate
J3.3 ☐ Low
J3.4 ☐ Unknown

Summary of Evidence for Group-Based Design Studies

Indicator	Overall Evidence Rating NNR = No numerical rating or 0 - 3	Description of Evidence Strong Promising Weak No/limited evidence or Descriptive ratings
General Characteristics		
General Design Characteristics	NNR	All participants were taken from "remedial" reading classes
Statistical Treatment	NNR	Sample size of 24 but needed 74 to get 80% power for a medium effect.
Type of Program	NNR	Pause, Prompt and Praise reading Intervention using cross age peer to tutor.
Stage of Program	NNR	Although Pause, Prompt and Praise is established it was used with parents and not peers as tutors so it is Early Stages.
Concurrent/Historical Intervention Exposure	NNR	Although Pause, Prompt and Praise is established it was used with parents and not peers as tutors
Key Features		
Measurement	1	Weak Evidence
Comparison Group	2	Promising Evidence
~~Primary/Secondary Outcomes are Statistically Significant~~	N/A	N/A
~~Educational/clinical significance~~	N/A	N/A
Identifiable Components	0	Limited or no evidence
Implementation Fidelity	3	Strong evidence
~~Replication~~	N/A	N/A
~~Site of Implementation~~	N/A	N/A
Follow Up Assessment Conducted	1	Weak evidence

Descriptive or Supplemental Criteria		
External validity indicators	NNR	Remedial readers in bottom classes as they want to improve reading ability in remedial readers. Not sure about rationale for the number of participants though
Length of Intervention	NNR	8 weeks
Intensity/dosage	NNR	3 times a week- it is not known the length of each lesson.
Dosage Response	NNR	Not known
Program Implementer	NNR	The implementers were the older peers in the school selected to act as tutors.
Characteristics of the Intervener	NNR	Some demographics are known of the peers. Eg they were matched in same sex dyads.
Intervention Style/Orientation	NNR	It is a behaviourist intervention
Cost Analysis Data Provided	NNR	The information is unknown
Training and Support Resources	NNR	There were two 30 minute training sessions for the tutors
Feasibility	NNR	This is unknown