What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

Beginning Reading July 16, 2007

Peer-Assisted Learning Strategies

Program description[1]

Peer-Assisted Learning Strategies (PALS) is a peer-tutoring program. According to the developer's web site, it is designed to be incorporated into the existing curriculum with the goal of improving the academic performance of children with diverse academic needs. Teachers train students to use *PALS* procedures. Students partner with peers, alternating the role of tutor while reading aloud, listening, and providing feedback in various structured activities. *PALS* is typically implemented three times a week for 30 to 35 minutes. Although *PALS* can be used in different subject areas and grade levels, this intervention report focuses on the use of *PALS* to improve reading skills of students in kindergarten through third grade.[2]

Research

Four studies of *Peer-Assisted Learning Strategies* met the What Works Clearinghouse (WWC) evidence standards with reservations. The four studies included more than 360 students from first to third grades in the United States.[3] The WWC considers the extent of evidence for *PALS* to small for alphabetics, fluency, and comprehension. No studies that met WWC evidence standards with or without reservations addressed general reading achievement.

Effectiveness

PALS was found to have potentially positive effects on alphabetics, fluency, and comprehension.

	Alphabetics	*Fluency*	*Comprehension*	*General reading achievement*
Rating of effectiveness	Potentially positive effects	Potentially positive effects	Potentially positive effects	na

(continued)

1. The descriptive information for this program was obtained from publicly available sources: the program's web site (http://kc.vanderbilt.edu/pals/, retrieved March 2007) and the research literature (Fuchs, Fuchs, Kazdan, & Allen, 1999; Mathes & Babyak, 2001; Mathes, Howard, Allen, & Fuchs, 1998; and Mathes, Torgesen, Clancy-Menchetti, Santi, Nicholas, Robinson, & Grek, 2003). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
2. The What Works Clearinghouse (WWC) also reviewed the effects of *PALS* on the reading achievement of English language learners with learning disabilities. The findings are reported in a separate WWC intervention report.
3. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

Effectiveness *(continued)*

	Alphabetics	*Fluency*	*Comprehension*	*General reading achievement*
Improvement index[4]	Average: +19 percentile points Range: –15 to +45 percentile points	Average: +13 percentile points Range: –8 to +31 percentile points	Average: +13 percentile points Range: –17 +28 percentile points	na

na = not applicable

Additional program information[2]

Developer and contact

Developed by Lynn and Doug Fuchs, *Peer-Assisted Learning Strategies* is distributed by Vanderbilt Kennedy Center for Research on Human Development. Address: Vanderbilt University, Attn: Flora Murray/PALS Orders, Box 328 Peabody, Nashville, TN 37203-5701. Email: flora.murray@vanderbilt.edu. Web: http://kc.vanderbilt.edu/pals/. Telephone: (615) 343-4782.

Scope of use

Peer-Assisted Learning Strategies was developed more than ten years ago to be used with students in kindergarten through high school. It has been implemented in Tennessee, and teacher trainings have been conducted in Iowa, Minnesota, Illinois, Arizona, and Ohio. The program has been used with students with diverse ability levels, including English language learners and students with learning disabilities.

Teaching

PALS is designed to supplement the existing reading curriculum. It includes separate versions for kindergarten (called *K-PALS*), grade 1 (*First-Grade PALS*), and grades 2–6. In each version students engage in peer-tutoring routines through a series of structured interactions. *K-PALS Reading* and *First-Grade PALS* include a set of 70 student lesson sheets, and teachers choose appropriate reading material for partner reading. In higher grades *PALS* does not provide any reading material; teachers select appropriate reading materials.

PALS sessions usually last 30 to 35 minutes three times a week. A typical lesson for the first-grade students begins with 15 minutes of Sounds and Words, which focuses on learning to hear and identify sounds, sounding out words, learning sight words, and practicing passage reading. The next 15 minutes of Story Sharing focuses on predicting story plots, oral reading, and retelling stories. A typical lesson for students in grades 2 to 6 includes specific activities to improve reading accuracy, fluency, and reading comprehension.

PALS offers teacher training in an all-day workshop where teachers learn to implement the program through modeling and role playing. Teachers are also provided with a manual describing the program.

Cost

The manual for each grade-level reading version of *PALS* costs $35. It includes teaching scripts and master copies of necessary student materials. Video materials that provide an overview of the grades 2 to 6 program are available for $15. Information on the cost of *PALS* training workshops is not available.

4. These numbers show the average and range of improvement indices for all findings across the studies.

Research

Eleven studies reviewed by the WWC investigated the effects of *PALS*. Four studies met WWC evidence standards with reservations. Two studies (Fuchs, Fuchs, Kazdan, & Allen, 1999; Mathes & Babyak, 2001) were randomized control trials with randomization problems, and two studies (Mathes, Howard, Allen, & Fuchs, 1998; Mathes, Torgesen, Clancy-Menchetti, Santi, Nicholas, Robinson, & Grek, 2003) were quasi-experimental designs. The remaining seven studies did not meet WWC evidence screens.

Met evidence standards with reservations

Fuchs et al. (1999) included 45 second- and third-grade students from 15 general education classrooms. Fuchs et al. compared two interventions—*Peer-Assisted Learning Strategies* and *Peer-Assisted Learning Strategies plus Help Giving*—to a comparison group that used the same curriculum as the intervention group but did not implement collaborative learning. Teachers were randomly assigned to the intervention or comparison group, but after random assignment, teachers selected three students with different achievement levels within each participating classroom to be part of the study. The WWC review of this study focused on the comparison of *PALS* and the comparison group with a total of 10 at-risk students in the second and third grades.[5]

Mathes and Babyak (2001) included 110 first-grade students from five schools in a medium-sized school district in Florida. Mathes and Babyak compared two interventions—*Peer-Assisted Learning Strategies* and *Peer-Assisted Learning Strategies plus Mini-Lessons*—to a comparison group that used a typical reading curriculum with no supplement. Teachers were matched on demographic characteristics to form a stratified sample and randomly assigned to the intervention or comparison group, but after random assignment, teachers selected five students with different achievement levels within each participating classroom to be part of the study.[6]

Mathes et al. (1998) included 96 first-grade students from six schools in an urban school district in the southeastern United States. Some teachers were randomly assigned to the treatment or comparison condition, but some were matched based on teaching profiles, generating a quasi-experimental study design. After teacher-level assignment, study authors selected five students with different achievement levels per classroom to be part of the study. The study compared *PALS* to a comparison group that used a typical reading curriculum with no supplement.

Mathes et al. (2003) included 89 low-achieving first-grade students taught by 22 teachers from six schools in a medium-sized southeastern school district. Some teachers were randomly assigned to the treatment or comparison condition, but some were matched based on teaching profiles, generating a quasi-experimental study design. After teacher-level assignment, study authors selected up to five low-achieving students per classroom to participate in the study. Mathes et al. compared *PALS* to a program similar to *PALS* but with teacher-directed instruction, and to a comparison group that participated in their usual reading curriculum.

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or moderate to large (see the What Works Clearinghouse Extent of Evidence Categorization Scheme). The extent of evidence takes into account the number of studies and the total sample size across the studies that met WWC evidence standards with or without reservations.[7]

5. Findings on the comparison of *PALS plus HG* and the comparison groups are included in Appendix A4.3 but do not factor into the intervention rating.
6. Findings for the comparison between *PALS plus ML* and the comparison groups are included in Appendices A4.1 and 4.2 but do not factor into the intervention rating.
7. The Extent of Evidence Categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept, external validity, such as the students' demographics and the types of settings in which studies took place, are not taken into account for the categorization.

Research *(continued)*

The WWC considers the extent of evidence for *PALS* to be small for alphabetics, fluency, and comprehension. No studies that met WWC evidence standards with or without reservations addressed general reading achievement.

Effectiveness

Findings

The WWC review of interventions for beginning reading addresses student outcomes in four domains: alphabetics, including phonemic awareness and phonics constructs; fluency; comprehension; and general reading achievement.[8] The four studies reviewed in this intervention report address student outcomes in the alphabetics, fluency, and comprehension domains. The findings below present the authors' and the WWC-calculated estimates of the size and statistical significance of the effects of *PALS* on student performance.[9]

Alphabetics. Three studies examined the effects of *PALS* on two constructs in the alphabetics domain: phonological awareness and phonics.

For phonological awareness, Mathes and Babyak (2001) found that *PALS* students had greater growth than comparison students on one measure (Continuous Progress Monitoring (CPM) Phonological Awareness Augmentation subtest). The WWC confirmed the statistically significant positive effect.

Mathes et al. (1998) found statistically significant positive growth in phonological awareness for low-achieving students but no statistically significant effect for average- and high-achieving students on one measure (CPM phonological awareness segmentation subtest). The WWC did not find a statistically significant effect of *PALS* for any single group, but found a statistically significant positive effect of *PALS* across all three ability groups combined.

Mathes et al. (2003) compared *PALS* students to two other groups.

- When *PALS* was compared with the usual curriculum group, the authors reported statistically significant positive effects on two measures of phonological awareness (the Comprehensive Test of Phonological Processes (CTOPP) Phonemic Segmentation subtest and the CPM Phoneme Segmentation subtest). The WWC confirmed the first but not the second finding.

  For phonics, the study authors found statistically significant positive effects on two of three measures (the Test of Word Reading Efficiency (TOWRE) Phonemic Decoding subtest and the Woodcock Reading Mastery Tests–Revised (WRMT–R) Word Attack subtest). The WWC confirmed the statistically significant effect on the second.
- When *PALS* was compared with the teacher-directed instruction group, the authors and the WWC did not find any statistically significant differences between the groups on either phonological awareness test or any of the three phonics outcomes.

The average effect size across all comparisons and outcomes in the alphabetics domain in Mathes et al. (2003) was statistically significant and positive.

Fluency. Three studies examined outcomes in the fluency domain. Mathes and Babyak (2001) reported that low- and average-achieving students, but not high-achieving students, made greater gains than comparison students on one fluency measure (CPM Oral Reading Fluency subtest). The WWC found that there were no statistically significant differences for any of the groups, but the average effect across all groups was large enough to be considered substantively important according to WWC criteria (that is, at least 0.25).

Mathes et al. (1998) reported a statistically significant positive effect on the low-achieving group and no statistically significant

8. For definitions of the domains, see the Beginning Reading Protocol.
9. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of *Peer-Assisted Learning Strategies,* corrections for clustering and multiple comparisons were needed.

Effectiveness *(continued)*

differences for the average- and high-achieving groups on one fluency measure (CPM Oral Reading Fluency subtest). The WWC found no statistically significant differences for any of the groups, but the average effect across all groups was large enough to be considered substantively important.

Mathes et al. (2003) compared *PALS* students with the two groups described in the research section on two fluency measures (the WRMT–R Word Identification subtest and the CPM Oral Reading subtest). The study authors and the WWC did not find any statistically significant differences between any of the groups and the average effect size was not large enough to be considered substantively important.

Comprehension. Two studies examined outcomes in the comprehension domain. Fuchs et al. (1999) reported and the WWC confirmed a statistically significant positive effect on one comprehension measure (Stanford Diagnostic Reading Test-III Reading Comprehension subtest).

Mathes et al. (2003) compared *PALS* students with two comparison groups on one comprehension outcome (the WRMT–R Passage Comprehension subtest). For both comparisons, the study authors and the WWC found no statistically significant effect of *PALS*. In addition, across comparisons, the average effect size was not large enough to be considered substantively important.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings, the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Intervention Rating Scheme).

The WWC found *Peer-Assisted Learning Strategies* to have potentially positive effects on alphabetics, fluency, and comprehension

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see Technical Details of WWC-Conducted Computations). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is based entirely on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analyses. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.

The average improvement index for alphabetics is +19 percentile points across three studies with a range of –15 to +45 percentile points across findings. The average improvement index for fluency is +13 percentile points across three studies with a range of –8 to +31 percentile points across findings. The average improvement index for comprehension is +13 percentile points across two studies with a range of –17 to +28 percentile points across findings.

Summary

The WWC reviewed 11 studies on *PALS*.[10] Four of these studies met WWC evidence standards with reservations; the remaining studies did not meet WWC evidence screens. Based on these four studies, the WWC found potentially positive effects in alphabetics, fluency, and comprehension. The evidence presented in this report may change as new research emerges.

10. A single-case design study was identified but is not included in this review because the WWC does not have standards yet for reviewing single-case design studies.

References

Met WWC evidence standards with reservations

Fuchs, L. S., Fuchs, D., Kazdan, S., & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *The Elementary School Journal*, *99*(3), 201–219.

Mathes, P. G., & Babyak, A. E. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional mini-skills lessons. *Learning Disabilities Research & Practice*, *16*(1), 28–44.

Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly*, *33*(1), 62–94.

Mathes, P. G., Torgesen, J. K., Clancy-Menchetti, J., Santi, K., Nicholas, K., Robinson, C., et al. (2003). A comparison of teacher-directed versus peer-assisted instruction to struggling first-grade readers. *The Elementary School Journal*, *103*(5), 459–479.

Did not meet WWC evidence screens

Bergeron, J. (1998). A comparison of classwide cross-age and same-age peer tutoring for second-grade students at risk for reading failure. *Dissertation Abstracts International*, *59*(09), 3390A. (UMI No. 9905010)[11]

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*(1), 174–206.[12]

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), 309–318.[13]

Hudson, K. G. (2004). The effects of Peer-Assisted Learning Strategies on the reading achievement of elementary students with and without decoding weaknesses. *Dissertation Abstracts International*, *65*(10), 3754A. (UMI No. 3149163)[13]

Pearson, J. J. M. (2004). The effect of peer-assisted literacy strategies on the social standing of first-grade readers. *Dissertation Abstracts International, 65*(02), 412A. (UMI No. 3122359)[14]

Sáenz, L. M., Fuchs, L. S., & Fuchs, D. (2005). Peer-Assisted Learning Strategies for English language learners with learning disabilities. *Exceptional Children, 71*(3), 231–247.[12]

Wehby, J. H., Falk, K. B., Barton-Arwood, S., Lane, K. L., & Cooley, C. (2003). The impact of comprehensive reading instruction on the academic and social behavior of students with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 11*(4), 225.[15]

Disposition Pending

Falk, K. B., & Wehby, J. H. (2001). The effects of peer-assisted learning strategies on the beginning reading skills of young children with emotional or behavioral disorders. *Behavioral Disorders*, 26(4), 344–359.[16]

For more information about specific studies and WWC calculations, please see the WWC *Peer-Assisted Learning Strategies* Technical Appendices.

11. Confound: there was only one classroom in each study condition, so the effects of the intervention could not be separated from the effects of the teacher.
12. The sample is not appropriate for this review: the parameters for this WWC review specified that students should be in grades kindergarten through 3; this study did not disaggregate students in the eligible range from those outside the range.
13. The sample is not appropriate to this review: the parameters for this WWC review specified that students should be in grades kindergarten through 3 at the time of the intervention; this study did not focus on the targeted grades.
14. The outcome measures are not relevant to this review: the parameters for this WWC review specified student outcome measures, but this study did not focus on students.
15. Confound: this study included *PALS* but combined it with another intervention, so the analysis could not separate the effects of the intervention from other factors.
16. The disposition is pending development of WWC evidence standards for single subject designs.

Appendix

Appendix A1.1 Study characteristics: Fuchs, Fuchs, Kazdan, & Allen, 1999 (randomized controlled trial with randomization problems[1])

Characteristic	Description
Study citation	Fuchs, L. S., Fuchs, D., Kazdan, S. & Allen, S. (1999). Effects of peer-assisted learning strategies in reading with and without training in elaborated help giving. *The Elementary School Journal, 99*(3), 201–219.
Participants	Fifteen second- and third-grade teachers and nine fourth-grade teachers were randomly assigned within each grade to intervention and comparison groups.[2] Teachers in the intervention group were then randomly assigned to two intervention conditions: *Peer-Assisted Learning Strategies (PALS)* or *Peer-Assisted Learning Strategies plus Help-Giving (PALS plus HG).* Each teacher identified three students within their classroom to be included in the analysis sample: one at-risk student (defined by having pervasive social behavior problems), one student with average reading achievement, and one student with high reading achievement. The study authors analyzed and reported findings for all achievement groups. However, the WWC reports findings only for the at-risk intervention and comparison groups, who had similar characteristics at baseline.[3] No attrition occurred during the course of intervention. In the grade 2–3 sample, 31% were female, 38% were African-American, and 24% eligible for free/reduce-priced lunch.
Setting	The study took place in 15 general education classrooms in one district in the United States. Information about the location of the district was not provided.
Intervention	This study included two intervention groups: *PALS* and *PALS plus HG*. In the *PALS* condition, teachers implemented *PALS* with their entire class for 21 weeks in 35-minute sessions three times a week during language art time. In each lesson, a stronger and a weaker reader were paired to engage in the *PALS* activities to improve reading accuracy, fluency, and comprehension. The content and structure of *PALS plus HG* condition is the same as *PALS,* except that teachers in *PALS plus HG* taught students strategies to determine correct responses rather than receiving the correct response from a peer as in *PALS.*
Comparison	The comparison group teachers used the same reading curriculum and same books as the intervention groups. However, the comparison group teachers did not implement a collaborative learning program.
Primary outcomes and measurement	The reading comprehension subtest of the third edition of the Stanford Diagnostic Reading Test was used for both the pre- and posttest. The red level of Form G was used for the second graders. The green level of Form G was used for the third graders (see Appendix A2.3 for more detailed descriptions of outcome measures).[4]
Teacher training	*PALS* and *PALS plus HG* teachers participated in a separate full-day workshop where they learned to implement the techniques through modeling and role playing. Each intervention teacher was assigned a research assistant who met with the teacher for five to 10 minutes every one to two weeks, observed teachers gave feedback, and provided support needed to implement the program. Teachers were also provided with *PALS* scripts.

1. The study design was based on random assignment of teachers to the intervention and comparison conditions. However, teacher selection of students for the analysis sample was not random and the WWC could not confirm that the selection was unrelated to treatment status, so the study met evidence standards with reservations.
2. The fourth-grade sample included in this study is not reviewed in this report because it is outside the scope of the review. For sample relevancy criteria, please see the Beginning Reading Protocol.
3. Findings for the comparison between *PALS plus HG* and the comparison groups for at-risk and high-achieving students are included in Appendices A4.1 and 4.2 but do not factor into the intervention rating.
4. The study authors also included measures of help-giving strategies that were not included in this review because they did not measure WWC Beginning Reading outcomes.

Appendix A1.2 Study characteristics: Mathes & Babyak, 2001 (randomized controlled trial with randomization problems[1])

Characteristic	Description
Study citation	Mathes, P. G. & Babyak, A. E. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional mini-skills lessons. *Learning Disabilities Research & Practice*, *16*(1), 28–44.
Participants	Thirty first-grade teachers from five schools matched on demographic characteristics were selected to form a stratified sample and were randomly assigned to one of three conditions: *Peer-Assisted Learning Strategies (PALS*; 10 teachers), *Peer-Assisted Learning Strategies plus Mini-Skills Lesson* (*PALS plus ML*; 10 teachers), or a comparison group (10 teachers). After rank-ordering students by their reading ability within the classroom, each teacher identified five students to be included in the analysis sample: one high-achieving student, one average-achieving student, and three low-achieving students. High- and average-achieving students from the *PALS plus ML* group did not participate in the ML component of the intervention and thus sample sizes varied across groups. The study began with 150 first-grade students. After attrition, the final analysis sample was 130 students (61 students in *PALS*, 20 in *PALS plus ML*, and 49 in the comparison group) and 28 teachers.[2] The WWC intervention rating focused on the comparison of *PALS* and the comparison group with a total of 110 students across different ability groups.[3] The mean age of the participating students was 6.9 years old. Forty-seven percent of the students were female, 39% African-American, 59% Caucasian, and 32% special needs.
Setting	The study took place in five schools in a medium-sized school district in Florida.
Intervention	This study included two intervention conditions: *PALS* and *PALS plus ML*. Teachers implemented *PALS* with their entire class for 14 weeks in 35-minute sessions three times a week. In each lesson, a stronger and a weaker reader were paired. In Sounds and Words activities, students practiced phonemic segmentation, applied alphabetic knowledge to decoding novel words, and read connected text built on previously mastered phonological elements. During Story Sharing time, students made predictions about a book prior to reading it, shared the experience of reading the book with a peer, had repeated exposure to the text, and summarized the text through verbal retelling. In the *PALS plus ML* condition, a 15–20 minute mini-lesson was given to small groups of low-achieving students in each classroom three times a week during the last six weeks of the *PALS* intervention. Teachers taught the mini-lessons before the *PALS* sessions. The content of the mini-lessons was the same as the Words and Sounds portion of *PALS*.
Comparison	Teachers used their regular reading curriculum. *PALS* staff collected student data weekly using the Continuous Progress Monitoring (CPM) measure. Teachers were given a graph showing students' progress every month. Teachers did not receive any recommendations or feedback about instruction from the researchers.
Primary outcomes and measurement	The primary outcome measure in the alphabetics domain was the CPM Phonological Awareness measure. The primary outcome measure in the fluency domain was CPM Oral Reading Fluency (see Appendices A2.1–2.2 for more detailed descriptions of outcome measures).[4]
Teacher training	Intervention teachers participated in an all-day in-service workshop prior to the intervention. They were provided with a manual describing *PALS* and practiced using the intervention. During training, *PALS* project staff were available to provide support needed to implement the program.

1. The study design was based on random assignment of teachers to the intervention and comparison conditions. However, teacher selection of students for the analysis sample was not random and the WWC could not confirm that the selection was unrelated to treatment status, so the study met evidence standards with reservations.
2. The postattrition samples were checked for equivalence at pretest by the WWC and found to be comparable.
3. Findings for the comparison between *PALS plus ML* and the comparison groups are included in Appendices A4.1 and 4.2, but not the intervention rating.
4. The authors also reported findings on the Word Identification, Word Attack, and Passage Comprehension subtests of the Woodcock Johnson Mastery Test as well as the measure of basic skills, a composite of the Woodcock Johnson Word Attack and Word Identification subtests. However, not enough information was reported to calculate an effect size for these measures.

Appendix A1.3 Study characteristics: Mathes, Howard, Allen, & Fuchs, 1998 (quasi-experimental design[1])

Characteristic	Description
Study citation	Mathes, P. G., Howard, J. K., Allen, S. H., & Fuchs, D. (1998). Peer-assisted learning strategies for first-grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 33(1)*, 62–94.
Participants	Twenty teachers from six schools were recruited to participate in the study to form a representative stratified sample reflecting the district teaching population. Teachers were assigned to intervention and comparison groups either through random assignment or matching on teaching profiles. The research team selected five students within each intervention and comparison class to be included in the analysis sample: three low-achieving students, one average-achieving student, and one high-achieving student per class. Achievement status was based on a curriculum-based measurement of phonological segmentation. The study began with 100 first-grade students from 20 classrooms. Four low-achieving students left the sample. The final analysis sample includes 96 students and 20 teachers. Among the student participants, 50% were male and 45% were African-American.
Setting	Six schools were drawn from one urban district in the southeastern United States.
Intervention	Teachers implemented *PALS* with their entire class for 16 weeks in 35-minute sessions three times a week during language art time. Other typical reading instruction was not changed. A stronger and a weaker reader were paired. Each lesson had two phases: Sounds and Words used the direct instruction model for 10 minutes for learning letter-sound correspondence, phonemic awareness, phonological recoding, and integration. The next 20 minutes used Partner Read-Aloud with emphasis on prediction, oral reading, and retelling. *PALS* staff collected student data weekly and assisted teachers with implementation.
Comparison	Comparison teachers used their usual reading curriculum. The teachers used whole language, phonics, or a mix of instructional techniques.
Primary outcomes and measurement	The primary outcome measure in the alphabetics domain was CPM Phonological Awareness. The primary outcome measure in the fluency domain was CPM Oral Reading Fluency (see Appendices A2.1–2.2 for more detailed descriptions of outcome measures).[2]
Teacher training	Teachers received training in an all-day in-service session. Teachers role-played each of the routines and were guided through the first-grade *PALS* manual. Four of the 10 treatment teachers had participated in the pilot for this study and returned with greater program experience than the other treatment teachers.

1. The study design was based on partial random assignment of teachers to the intervention and comparison conditions. Other teachers were assigned to condition after matching them on teaching profiles. Thus the study was classified as a quasi-experimental design and meet WWC evidence standards with reservation.
2. The authors also reported findings on the Word Attack and Passage Mastery subtests of Woodcock Reading Mastery Test-Revised, Test of Early Reading Ability-2, and the Comprehensive Reading Assessment Battery-Revised. However, not enough information was reported to calculate an effect size for these measures.

Appendix A1.4 Study characteristics: Mathes et al., 2003 (quasi-experimental design[1])

Characteristic	Description
Study citation	Mathes, P. G., Torgesen, J. K., Clancy-Menchetti, J. Santi, K., Nicholas, K., Robinson, C., & Grek, M. (2003). A comparison of teacher-directed versus peer-assisted instruction to struggling first-grade readers. *Elementary School Journal, 103*(5), 459–479.
Participants	Twenty-two first-grade teachers from six schools participated in the study. Teachers were assigned to three conditions either through random assignment or matching on teaching and school profiles: the full *PALS* program, teacher-directed instruction, and the comparison condition. One hundred students were initially identified for the study after they were determined to be at risk based on two reading measures: oral reading and phonemic segmentation fluency. Students were ranked based on their oral reading fluency in each class and up to five students with the lowest segmentation scores in each class were selected for the study after teachers confirmed their academic performance. During the course of intervention, 11 students left the participating classrooms. The final analysis sample included seven teachers and 31 students in the *PALS* group, seven teachers and 30 students in the teacher-directed instruction group, and eight teachers and 28 students in the comparison group.[2] Among the student participants, 33% were female, 42% African-American, and 29% qualified for free/reduced price lunch.
Setting	The six participating schools were drawn from one medium-sized southeastern school district.
Intervention	Teachers in the *PALS* group used the program in three 35-minute sessions a week for 16 weeks. Implementation of the program was integrated with the regular reading program, replacing either individual seat work or silent reading time. In each lesson, students worked in pairs to perform two 15-minute routines. Although each pair usually consisted of a stronger and a weaker reader, each child performed both the tutor and tutee roles for part of the lesson. Each lesson began with 15 minutes of Sounds and Words focusing on letter sounds, hearing sounds, sounding out words, reading sight words, and passage reading practice. Then students engaged in 15 minutes of Story Sharing which emphasized story prediction, oral reading, and retelling of stories.
Comparison	In the teacher-directed instruction condition, teachers conducted a 30-minute lesson with a small group of 4–5 students three times a week teaching the same content as in PALS. Teachers worked toward mastery of the Sound and Word lesson content for each student. In the comparison condition, teachers used their usual reading curriculum. Assessments were conducted in these classes every second week. Teachers were given performance data every six weeks.
Primary outcomes and measurement	The authors used the Word Identification, Word Attack, and Passage Comprehension subtests of Woodcock Reading Mastery Tests–Revised (WRMT–R), and the Phonemic Decoding and Sight Word Efficiency subtests of Test of Word Reading Efficiency (TOWRE). Additional tests included the Comprehensive Test of Phonological Processes (CTOPP) Phonemic Segmentation subtest and the Continuous Progress Monitoring Oral Fluency and Phonemic Segmentation measures (see Appendices A2.1–2.3 for more detailed descriptions of outcome measures).
Teacher training	*PALS* and teacher-directed instruction teachers attended an all-day workshop designed for their interventions. Teachers were provided with a manual describing the programs and practiced using the interventions. During training for *PALS* teachers, trainers were available to provided needed support. During training for teacher-directed instruction teachers, *PALS* staff observed the first week of implementation and offered assistance where necessary.

1. The study design was based on partial random assignment of teachers to the intervention and comparison conditions. Other teachers were assigned to condition after matching on teaching and school profiles. Thus the study was classified as a quasi-experimental design and meet WWC evidence standards with reservation.
2. The postattrition samples were checked for equivalence at pretest by the WWC and found to be comparable.

Appendix A2.1 Outcome measures in the alphabetics domain

Outcome measure	Description
Phonological awareness	
Comprehensive Test of Phonological Processes (CTOPP): Phoneme Segmenting subtest	The phoneme blending subtest measures the child's ability to blend separately presented sounds together to form words (as cited in Mathes, Torgesen, Clancy-Minchetti et al, 2003).
Continuous Progress Monitoring (CPM):[1] Phonological Awareness	A researcher-developed measure designed to measure phonological awareness segmentation skills following the model of curriculum-based measurement (as cited in Mathes & Babyak, 2001; Mathes et al., 1998).
CPM1: Phonemic Segmentation	Students' oral segmentation of the words was timed to generate a score of segments per minute (as cited in Mathes et al., 2003).
Phonics	
Woodcock Reading Mastery Tests–Revised (WRMT–R): Word Attack subtest	The word attack subtest is a measure of phonemic reading ability in which the child reads non-words. This is a standardized test with 51 items (as cited in Mathes et al., 2003).
WRMT–R: Word Identification subtest	The word identification subtest is a measure of word reading vocabulary in which the child reads list of words of increasing difficulty. This is a standardized test (as cited in Mathes et al., 2003).
Test of Word Reading Efficiency (TOWRE): Phonemic Decoding	This test measures the number of non-words read/decoded in 45 seconds (as cited in Mathes et al., 2003).
TOWRE: Sight Word Efficiency	This test measures the number of high-frequency words read in 45 seconds (as cited in Mathes et al., 2003).

1. The terms Curriculum-Based Measurement (CBM) and Continuous Progress Monitoring (CPM) were used across the different studies. For consistency, these measures have all been called Continuous Progress Monitoring (CPM) in the WWC report and appendixes.

Appendix A2.2 Outcome measure in the fluency domain

Outcome measure	Description
CPM: Oral Reading Fluency	A researcher-developed measure designed to measure words per minute. Students were tested on oral reading fluency using ten 400 word stories at the first-grade level (as cited in Mathes et al., 2003).

Appendix A2.3 Outcome measures in the comprehension domain

Outcome measure	Description
Reading comprehension	
Stanford Diagnostic Reading Test (SDRT): Reading Comprehension subtest	A standardized measure of children's reading comprehension (as cited in Fuchs et al., 1999).
WRMT–R: Passage Comprehension subtest	In this standardized test, comprehension is measured by having students read silently and fill in missing words in a short paragraph (as cited in Mathes et al., 2003).

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Phonological awareness								
Mathes & Babyak, 2001 (randomized controlled trial with randomization problems)[7]								
CPM: Phonological Awareness	Grade 1 (low ability)	20/56	44.50 (10.39)	33.79 (16.65)	10.71	0.75	ns	+27
CPM: Phonological Awareness	Grade 1 (average ability)	20/27	48.45 (9.31)	45.34 (17.32)	3.11	0.24	ns	+9
CPM: Phonological Awareness	Grade 1 (high ability)	20/27	51.90 (9.29)	46.04 (9.98)	5.86	0.64	ns	+24
Mathes et al., 1998 (quasi-experimental design)[8]								
CPM: Phonological Awareness	Grade 1 (low ability)	20/56	31.96 (9.58)	25.39 (9.65)	6.57	0.67	ns	+25
CPM: Phonological Awareness	Grade 1 (average ability)	20/20	38.64 (7.15)	35.20 (5.55)	3.44	0.51	ns	+20
CPM: Phonological Awareness	Grade 1 (high ability)	20/20	43.32 (8.16)	38.18 (8.03)	5.14	0.61	ns	+23
Mathes et al., 2003 (quasi-experimental design)[9]								
Comparison #1: PALS vs. usual reading curriculum group								
CTOPP: Phonemic Segmentation	Grade 1	15/59	14.84 (3.37)	7.75 (4.84)	7.09	1.69	Statistically significant	+45
CPM: Phoneme Segmentation	Grade 1	15/59	50.01 (14.45)	38.29 (16.66)	11.72	0.74	ns	+27
Comparison #2: PALS vs. teacher-directed instruction group								
CTOPP: Phonemic Segmentation	Grade 1	14/61	16.75 (3.37)	15.1 (4.03)	1.65	0.44	ns	+17
CPM: Phoneme Segmentation	Grade 1	14/61	50.99 (14.45)	50.56 (13.13)	–0.43	0.03	ns	+1

(continued)

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain *(continued)*

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Phonics								
Mathes et al., 2003 (quasi-experimental design)[10]								
Comparison #1: PALS vs. usual reading curriculum group								
TOWRE: Phonemic Decoding	Grade 1	15/59	9.4 (3.84)	6.59 (5.92)	2.81	0.56	ns	+21
TOWRE: Sight Word Efficiency	Grade 1	15/59	22.47 (8.90)	20.82 (9.33)	1.65	0.18	ns	+7
WRMT: Word Attack	Grade 1	15/59	13.48 (7.43)	6.14 (8.19)	7.34	0.93	Statistically significant	+32
Comparison #2: PALS vs. teacher-directed instruction								
TOWRE: Phonemic Decoding	Grade 1	14/61	9.38 (3.84)	11.05 (5.10)	–1.67	–0.37	ns	–14
TOWRE: Sight Word Efficiency	Grade 1	14/61	21.76 (8.90)	25.77 (11.77)	–4.01	–0.38	ns	–15
WRMT: Word Attack	Grade 1	14/61	14.01 (7.43)	15.90 (8.51)	–1.89	–0.23	ns	–9
Average[8] for alphabetics, Comparison #1 (Mathes et al., 2003)						0.82	Statistically significant	+29
Average[8] for alphabetics, Comparison #2 (Mathes et al., 2003)						–0.10	ns	–4
Average[10] for alphabetics (Mathes & Babyak, 2001)						0.54	Statistically significant	+21
Average[7] for alphabetics (Mathes et al., 1998)						0.60	Statistically significant	+23
Average[8] for alphabetics (Mathes et al., 2003)						0.36	ns	+14
Domain average[8] for alphabetics across all studies						0.50	na	+19

ns = not statistically significant

na = not applicable

1. This appendix reports findings that are considered for the effectiveness rating and the average improvement indices in the alphabetics domain. Additional findings on variations of *PALS, PALS plus Mini-Lessons* from Mathes & Babyak (2001) and *PALS plus Help-Giving* from Fuchs et al. (1999), are reported in Appendix A4.1.

(continued)

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain *(continued)*

2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean equals the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Mathes & Babyak (2001), a correction for clustering was needed, so the significance levels may differ from those reported in the original study. In addition, study authors reported findings on the CPM based on a growth curve model with four assessments over 14 weeks, while the WWC reported on the findings based on the final assessment.
8. In the case of Mathes et al. (1998), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
9. In the case of Mathes et al. (2003), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A3.2 Summary of study findings included in the rating for the fluency domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Mathes & Babyak, 2001 (randomized controlled trial with randomization problems)[7]								
CPM: Oral Reading Fluency	Grade 1 (low ability)	20/56	18.10 (9.98)	13.68 (14.32)	4.42	0.35	ns	+14
CPM: Oral Reading Fluency	Grade 1 (average ability)	20/27	38.49 (17.26)	24.40 (12.67)	14.09	0.87	ns	+31
CPM: Oral Reading Fluency	Grade 1 (high ability)	20/27	94.70 (44.36)	81.03 (35.91)	13.67	0.32	ns	+13
Mathes et al., 1998 (quasi-experimental design)[8]								
CPM: Oral Reading Fluency	Grade 1 (low ability)	20/56	25.48 (24.19)	14.97 (10.23)	10.51	0.56	ns	+21
CPM: Oral Reading Fluency	Grade 1 (average ability)	20/20	58.33 (30.92)	47.45 (19.70)	10.88	0.40	ns	+16
CPM: Oral Reading Fluency	Grade 1 (high ability)	20/20	102.15 (46.25)	96.48 (26.58)	5.67	0.14	ns	+6
Mathes et al., 2003 (quasi-experimental design)[9]								
Comparison #1: PALS vs. usual reading curriculum group								
WRMT–R: Word Identification	Grade 1	15/59	30.08 (9.01)	25.89 (9.90)	4.19	0.44	ns	+17
CPM: Oral Reading Fluency	Grade 1	15/59	28.93 (14.20)	22.20 (15.19)	6.73	0.45	ns	+17
Comparison #2: PALS vs. teacher-directed instruction								
WRMT–R: Word Identification	Grade 1	14/61	30.59 (9.01)	31.87 (9.22)	–1.28	–0.14	ns	–6
CPM: Oral Reading Fluency	Grade 1	14/61	28.53 (14.20)	31.42 (15.30)	–2.89	–0.19	ns	–8
Average for fluency, Comparison #1 (Mathes et al., 2003)						0.45	ns	+17
Average for fluency, Comparison #2 (Mathes et al., 2003)						–0.17	ns	–7

(continued)

Appendix A3.2 Summary of study findings included in the rating for the fluency domain *(continued)*

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Average for fluency across all ability groups (Mathes & Babyak, 2001)						0.51	ns	+20
Average for fluency across all ability groups (Mathes et al., 1998)						0.37	ns	+14
Average[8] for fluency (Mathes et al., 2003)						0.14	ns	+6
Domain average[8] for fluency across all studies						0.34	na	+13

ns = not statistically significant

na = not applicable

1. This appendix reports findings that are considered for the effectiveness rating and the average improvement indices in the fluency domain. Additional findings on variations of *PALS* and *PALS plus Mini-Lessons* from Mathes & Babyak (2001) are reported in Appendix A4.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean equals the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Mathes & Babyak (2001), a correction for clustering was needed, so the significance levels may differ from those reported in the original study. In addition, study authors reported findings on the CPM based on a growth curve model with four assessments over 14 weeks, while the WWC reported on the findings based on the final assessment.
8. In the case of Mathes et al. (1998), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. In the case of Mathes et al. (2003), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.

Appendix A3.3 Summary of study findings included in the rating for the comprehension domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Fuchs et al., 1999 (randomized controlled trial with randomization problems)[7]								
Stanford Diagnostic Reading Test III: Reading Comprehension	Grade 2–3 (at-risk students)	10/10	41.60 (13.15)	36 (10.65)	5.60	0.76	Statistically significant	+28
Average[8] for alphabetics (Fuchs et al., 1999)						0.76	Statistically significant	+28
Mathes et al., 2003 (quasi-experimental design)[9]								
Comparison #1 – PALS vs. usual reading curriculum group								
WRMT: Passage Comprehension	Grade 1	15/59	12.04 (6.21)	10.21 (7.86)	1.83	0.26	ns	+10
Comparison #2 – PALS vs. teacher-directed instruction group								
WRMT: Passage Comprehension	Grade 1	14/61	12.50 (6.21)	15.47 (6.81)	–2.97	–0.45	ns	–17
Average[8] for comprehension (Mathes et al., 2003)						–0.10	ns	–4
Domain average[8] for comprehension across all studies						0.33	na	+13

ns = not statistically significant
na = not applicable

1. This appendix reports findings that are considered for the effectiveness rating and the average improvement indices in the comprehension domain. Additional findings on variations of *PALS* and *PALS plus Help-Giving* from Fuchs et al. (1999) are reported in Appendix A4.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean equals the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations. For Fuchs et al. (1999), the effect size was reported by the study authors.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Fuchs et al. (1999), no corrections for clustering or multiple comparisons were needed.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.
9. In the case of Mathes et al. (2003), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.

Appendix A4.1 Summary of alternative intervention group findings for the alphabetics domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Mathes & Babyak, 2001 (randomized controlled trial with randomization problems)[7] for *PALS plus Mini-Lessons* vs. comparison group								
CPM: Phonological Awareness	Grade 1 (low ability)	18/49	44.37 (12.13)	33.79 (16.65)	10.58	0.69	Statistically significant	+26

1. This appendix presents findings for an alternative intervention group, *PALS plus Mini-Lessons*, on alphabetics measures. Only primary group findings were used for rating purposes and are presented in Appendix A3.1.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean equals the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Mathes & Babyak (2001), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

Appendix A4.2 Summary of alternative intervention group findings for the fluency domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Mathes & Babyak, 2001 (randomized controlled trial with randomization problems)[7] for *PALS plus Mini-Lessons* vs. comparison group								
CPM: Oral Reading Fluency	Grade 1 (low ability)	18/49	22.28 (11.37)	13.68 (14.32)	8.60	0.64	ns	+24

1. This appendix presents findings for an alternative intervention group, *PALS plus Mini-Lessons*, on fluency measures. Only primary group findings were used for rating purposes and are presented in Appendix A3.2.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean equals the comparison group mean plus the mean difference.
4. For an explanation of effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Mathes & Babyak (2001), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

Appendix A4.3 Summary of alternative intervention group findings for the comprehension domain[1]

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *PALS* group	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[3] (*PALS* – comparison)	WWC calculations: Effect size[4]	WWC calculations: Statistical significance[5] (at $\alpha = 0.05$)	WWC calculations: Improvement index[6]
Fuchs et al., 1999 (Randomized controlled trial with randomization problems)[7] for *PALS plus Help-Giving* vs. comparison group								
Stanford Diagnostic Reading Test: Reading Comprehension subtest	Grade 2–3 (at-risk)	10/10	32.00 (11.90)	36.00 (10.65)	–4.00	–0.168[8]	ns	–13
Fuchs et al., 1999 (Randomized controlled trial with randomization problems)[9] for *PALS plus Help-Giving* vs. comparison group								
Stanford Diagnostic Reading Test: Reading Comprehension subtest	Grade 2–3 (high achieving)	10/10	46.00 (1.14)	45.60 (1.30)	0.40	0.30	ns	+12

1. This appendix presents findings for an alternative intervention group, *PALS plus Help-Giving*, on comprehension measures. Only primary group findings were used for rating purposes and are presented in Appendix A3.3.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The intervention group mean equals the comparison group mean plus the mean difference.
4. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools (corrections for multiple comparisons were not done for findings not included in the overall intervention rating). For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate statistical significance. In the case of Fuchs et al. (1999), no correction for clustering was needed.
8. Effect size was reported by the study authors and accounted for pretest scores.
9. In the case of Fuchs et al. (1999), no correction for clustering was needed.

Appendix A5.1 *Peer-Assisted Learning Strategies* rating for the alphabetics domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of alphabetics, the WWC rated *Peer-Assisted Learning Strategies* as having potentially positive effects. It did not meet the criteria for positive effects as none of the four studies reviewed met WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Peer-Assisted Learning Strategies* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. Three studies examined the effect of *Peer-Assisted Learning Strategies* on alphabetics and all showed statistically significant positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies showed statistically significant or substantively important negative effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. No studies reviewed met WWC evidence standards for a strong design.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.2 *Peer-Assisted Learning Strategies* rating for the fluency domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of fluency, the WWC rated *Peer-Assisted Learning Strategies* as having potentially positive effects. It did not meet the criteria for positive effects as none of the four studies reviewed met WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Peer-Assisted Learning Strategies* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. Two studies examined the effect of Peer-Assisted Learning Strategies on fluency and showed substantively important positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies showed statistically significant or substantively important negative effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. No studies reviewed met WWC evidence standards for a strong design.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A5.3 *Peer-Assisted Learning Strategies* rating for the comprehension domain

The WWC rates an intervention's effects in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.[1]

For the outcome domain of comprehension, the WWC rated *Peer-Assisted Learning Strategies* as having potentially positive effects. It did not meet the criteria for positive effects as none of the four studies reviewed met WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Peer-Assisted Learning Strategies* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

  Met. One study examined the effect of Peer-Assisted Learning Strategies on comprehension and showed statistically significant positive effects.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

  Met. No studies showed statistically significant or substantively important negative effects. One study showed statistically significant positive effects and one study showed indeterminate effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a strong design.

  Not met. No studies reviewed met WWC evidence standards for a strong design.

AND

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

  Met. No studies showed statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. See the WWC Intervention Rating Scheme for a complete description.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	3	17	295	Small
Fluency	3	5	295	Small
Comprehension	2	6	99	Small
General reading achievement	0	0	0	na

na = not applicable/not studied

1. A rating of "moderate to large" requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is "small."