

Appendix

Appendix A1.1 Study characteristics: Fuchs, Fuchs, Kazdan, & Allen, 1999 (randomized controlled trial with randomization problems¹)

| 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. <i>The Elementary School Journal</i> , 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: <i>Peer-Assisted Learning Strategies (PALS)</i> or <i>Peer-Assisted Learning Strategies plus Help-Giving (PALS plus HG)</i> . 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: <i>PALS</i> and <i>PALS plus HG</i> . In the <i>PALS</i> condition, teachers implemented <i>PALS</i> 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 <i>PALS</i> activities to improve reading accuracy, fluency, and comprehension. The content and structure of <i>PALS plus HG</i> condition is the same as <i>PALS</i> , except that teachers in <i>PALS plus HG</i> taught students strategies to determine correct responses rather than receiving the correct response from a peer as in <i>PALS</i> . |
| 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 | <i>PALS</i> and <i>PALS plus HG</i> 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 <i>PALS</i> 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 problems¹)

| 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. <i>Learning Disabilities Research & Practice, 16</i> (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: <i>Peer-Assisted Learning Strategies (PALS)</i> ; 10 teachers), <i>Peer-Assisted Learning Strategies plus Mini-Skills Lesson (PALS plus ML)</i> ; 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 <i>PALS plus ML</i> 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 <i>PALS</i> , 20 in <i>PALS plus ML</i> , and 49 in the comparison group) and 28 teachers. ² The WWC intervention rating focused on the comparison of <i>PALS</i> 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: <i>PALS</i> and <i>PALS plus ML</i> . Teachers implemented <i>PALS</i> 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 <i>PALS plus ML</i> 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 <i>PALS</i> intervention. Teachers taught the mini-lessons before the <i>PALS</i> sessions. The content of the mini-lessons was the same as the Words and Sounds portion of <i>PALS</i> . |
| Comparison | Teachers used their regular reading curriculum. <i>PALS</i> 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 alphabetic 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 <i>PALS</i> and practiced using the intervention. During training, <i>PALS</i> 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 design¹)

| 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. <i>Reading Research Quarterly, 33</i> (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 <i>PALS</i> 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. <i>PALS</i> 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 alphabetic 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 <i>PALS</i> 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 design¹)

| 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. <i>Elementary School Journal</i> , 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 <i>PALS</i> 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 <i>PALS</i> 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 <i>PALS</i> 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 <i>PALS</i> . 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 | <i>PALS</i> 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 <i>PALS</i> teachers, trainers were available to provide needed support. During training for teacher-directed instruction teachers, <i>PALS</i> 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 alphabetic 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 |
|---|---|
| <i>Reading comprehension</i> | |
| 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 alphabetic domain¹

| Outcome measure | Study sample | Sample size (classrooms/ students) | Authors' findings from the study | | | | | |
|--|---------------------------|------------------------------------|---|------------------|--|--------------------------|---|--------------------------------|
| | | | Mean outcome (standard deviation ²) | | WWC calculations | | | |
| | | | PALS group | Comparison group | Mean difference ³ (PALS – comparison) | Effect size ⁴ | Statistical significance ⁵ (at $\alpha = 0.05$) | 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 | | | | | | | | |
| <i>Comparison #1: PALS vs. usual reading curriculum group</i> | | | | | | | | |
| 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 |
| <i>Comparison #2: PALS vs. teacher-directed instruction group</i> | | | | | | | | |
| 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 alphabets domain (continued)

| Outcome measure | Study sample | Sample size (classrooms/ students) | Authors' findings from the study | | | | | |
|---|--------------|------------------------------------|---|------------------|--|--------------------------|---|--------------------------------|
| | | | Mean outcome (standard deviation ²) | | WWC calculations | | | |
| | | | PALS group | Comparison group | Mean difference ³ (PALS – comparison) | Effect size ⁴ | Statistical significance ⁵ (at $\alpha = 0.05$) | Improvement index ⁶ |
| Phonics | | | | | | | | |
| Mathes et al., 2003 (quasi-experimental design)¹⁰ | | | | | | | | |
| <i>Comparison #1: PALS vs. usual reading curriculum group</i> | | | | | | | | |
| 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 |
| <i>Comparison #2: PALS vs. teacher-directed instruction</i> | | | | | | | | |
| 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 |
| Average⁸ for alphabets, Comparison #1 (Mathes et al., 2003) | | | | | | 0.82 | Statistically significant | +29 |
| Average⁸ for alphabets, Comparison #2 (Mathes et al., 2003) | | | | | | -0.10 | ns | -4 |
| Average¹⁰ for alphabets (Mathes & Babyak, 2001) | | | | | | 0.54 | Statistically significant | +21 |
| Average⁷ for alphabets (Mathes et al., 1998) | | | | | | 0.60 | Statistically significant | +23 |
| Average⁸ for alphabets (Mathes et al., 2003) | | | | | | 0.36 | ns | +14 |
| Domain average⁸ for alphabets 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 alphabets 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 alphabetic 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 domain¹

| Outcome measure | Study sample | Sample size (classrooms/ students) | Authors' findings from the study | | WWC calculations | | | |
|--|---------------------------|------------------------------------|---|------------------|--|--------------------------|---|--------------------------------|
| | | | Mean outcome (standard deviation ²) | | Mean difference ³ (PALS – comparison) | Effect size ⁴ | Statistical significance ⁵ (at $\alpha = 0.05$) | Improvement index ⁶ |
| | | | PALS group | Comparison group | | | | |
| 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 | | | | | | | | |
| <i>Comparison #1: PALS vs. usual reading curriculum group</i> | | | | | | | | |
| 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 |
| <i>Comparison #2: PALS vs. teacher-directed instruction</i> | | | | | | | | |
| 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 | | WWC calculations | | | |
|--|--------------|------------------------------------|---|------------------|--|--------------------------|---|--------------------------------|
| | | | Mean outcome (standard deviation ²) | | Mean difference ³ (PALS – comparison) | Effect size ⁴ | Statistical significance ⁵ (at $\alpha = 0.05$) | Improvement index ⁶ |
| | | | PALS group | Comparison group | | | | |
| 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 |
| Average⁸ for fluency (Mathes et al., 2003) | | | | | | 0.14 | ns | +6 |
| Domain average⁸ 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 domain¹

| Outcome measure | Study sample | Sample size (classrooms/students) | Authors' findings from the study | | WWC calculations | | | |
|---|------------------------------|-----------------------------------|---|------------------|--|--------------------------|---|--------------------------------|
| | | | Mean outcome (standard deviation ²) | | Mean difference ³ (PALS – comparison) | Effect size ⁴ | Statistical significance ⁵ (at $\alpha = 0.05$) | Improvement index ⁶ |
| | | | PALS group | Comparison group | | | | |
| 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 |
| Average⁸ for alphabets (Fuchs et al., 1999) | | | | | | 0.76 | Statistically significant | +28 |
| Mathes et al., 2003 (quasi-experimental design)⁹ | | | | | | | | |
| <i>Comparison #1 – PALS vs. usual reading curriculum group</i> | | | | | | | | |
| WRMT: Passage Comprehension | Grade 1 | 15/59 | 12.04 (6.21) | 10.21 (7.86) | 1.83 | 0.26 | ns | +10 |
| <i>Comparison #2 – PALS vs. teacher-directed instruction group</i> | | | | | | | | |
| WRMT: Passage Comprehension | Grade 1 | 14/61 | 12.50 (6.21) | 15.47 (6.81) | –2.97 | –0.45 | ns | –17 |
| Average⁸ for comprehension (Mathes et al., 2003) | | | | | | –0.10 | ns | –4 |
| Domain average⁸ 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 alphabetic domain¹

| Outcome measure | Study sample | Sample size (classrooms/ students) | Authors' findings from the study | | | | | |
|--|-----------------------|------------------------------------|---|------------------|--|--------------------------|---|--------------------------------|
| | | | Mean outcome (standard deviation ²) | | WWC calculations | | | |
| | | | PALS group | Comparison group | Mean difference ³ (PALS – comparison) | Effect size ⁴ | Statistical significance ⁵ (at $\alpha = 0.05$) | 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 alphabetic 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 domain¹

| Outcome measure | Study sample | Sample size (classrooms/ students) | Authors' findings from the study | | | | | |
|--|-----------------------|------------------------------------|---|------------------|--|--------------------------|---|--------------------------------|
| | | | Mean outcome (standard deviation ²) | | WWC calculations | | | |
| | | | PALS group | Comparison group | Mean difference ³ (PALS – comparison) | Effect size ⁴ | Statistical significance ⁵ (at $\alpha = 0.05$) | 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 domain¹

| Outcome measure | Study sample | Sample size (classrooms/ students) | Authors' findings from the study | | | | | |
|--|----------------------------|------------------------------------|---|------------------|--|--------------------------|---|--------------------------------|
| | | | Mean outcome (standard deviation ²) | | WWC calculations | | | |
| | | | PALS group | Comparison group | Mean difference ³ (PALS – comparison) | Effect size ⁴ | Statistical significance ⁵ (at $\alpha = 0.05$) | 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 alphabets 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 alphabets, 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 alphabets 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.”