Appendix

Appendix A1 Study Characteristics: Greenwood et al., 1993 (randomized controlled trial)

Characteristic	Description
Study citation	Greenwood, C. R., Terry, B., Utley, C. A., Montagna, D., & Walker, D. (1993). Achievement placement and services: Middle school benefits of ClassWide Peer Tutoring used at the elementary school. *School Psychology Review, 22*(3), 497–516.
Participants	Two-hundred and ninety-three first-grade students (170 students in intervention, 123 students in comparison) participated in this longitudinal study that followed students during program implementation from first grade to fourth grade and followed up two years later in sixth grade.[1] The study assigned schools to conditions—four schools were randomly assigned to the intervention and two schools to the comparison. About 24% of the students in the intervention group and 27% of the students in the comparison group were lost to analysis at follow-up at sixth grade.[2] The study demonstrated equivalence of baseline scores of students in the intervention and comparison samples included in the analysis.
Setting	Six Chapter I elementary schools in one school district in Kansas City, Kansas.
Intervention	Most teachers were involved in the study for one year. Two teachers refused to participate in the *CWPT* program, but agreed for the assessments to take place. Teachers received either three hours of paid university credit or a monetary compensation for their participation in the study.
Comparison	Comparison group students received their regular reading instruction program and Title I services. *CWPT* training or implementation was not conducted in the comparison schools. Teachers received either three hours of paid university credit or a monetary compensation for their participation in the study.
Primary outcomes and measurement	The primary outcome measure at grade 6 was the reading subtest of the Comprehensive Test of Basic Skills–Form U (CTBS–U), 3rd edition. The reading subtest of the Metropolitan Achievement Test (MAT) was administered at grade 4, but is not reviewed in this intervention report because of severe attrition of students. Language subtests for both measures were also administered but are not included in this review because they do not reflect outcome domains that are the focus of this Beginning Reading WWC review (see Appendix A2 for more detailed descriptions of the outcome measure reviewed for rating purposes).
Teacher training	Most of the participating classrooms were taught by a different teacher each year. The participating teachers received *CWPT* training each year. Teachers first read the program manual and then discussed with their consultants the changes to be made in their classroom practices. After the initial planning and preparation, consultants helped teachers implement the intervention in their classrooms. Teachers were considered trained when they received a score of 85% or above on the consultants' observation checklists.

1. This intervention report focuses on the part of the study that randomly assigned Chapter I schools to intervention and control conditions. Greenwood et al. (1993) also compared *CWPT* students in Chapter I schools and non-*CWPT* students in non-Chapter I schools.
2. In an earlier analysis of outcomes at the end of the program implementation at fourth grade, attrition was more severe and differential by treatment status, 68.2% of the students in the intervention group and 44.2% of the students in the comparison group were not included in the analysis at the end of the fourth grade. However, study authors were able to locate some students in sixth grade when they fed back into a common middle school. Because of these high and differential rates of attrition, outcomes at grade 4 are not reviewed in this WWC intervention report.

Appendix A2 Outcome measure in the general reading achievement domain

Outcome measure	Description
CTBS-U Reading subtest	This is the reading subtest of the Comprehensive Test of Basic Skills–Form U (CTBS-U), 3rd edition. The sub-skills measured by this standardized subtest include visual discrimination, letter recognition, auditory discrimination, sight vocabulary, phoneme/grapheme consonants and vowels, vocabulary in context, word part clues, and comprehension (as cited by Greenwood et al., 1993).

Appendix A3 Summary of study findings included in the rating for the general reading achievement domain[1]

Outcome measure	Study sample	Sample size[3] (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]): *CWPT* group	Authors' findings from the study: Mean outcome (standard deviation[2]): Comparison group	WWC calculations: Mean difference[4] (*CWPT* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Greenwood & Terry, 1993 (randomized controlled trial)[8] – four years of intervention, after two years of follow-up								
CTBS-U Reading subtest	Grade 6[9]	6/218	46.17 (15.78)	40.77 (14.99)	5.40	0.35	ns	+14
Domain average[10] for general reading achievement						0.35	ns	+14

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
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. Sample sizes were requested by the WWC and received by the study author.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. The means included here were adjusted for reading achievement at baseline and IQ scores.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and 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.
8. 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 Greenwood et al. (1993), corrections for clustering during the time of intervention delivery were needed, so the significance levels differ from those reported in the original study.
9. Students started the study when they were in first grade; intervention students participated in *CWPT* for four years (first through fourth grade).
10. This row provides the domain average, which in this instance is also the single finding from one study.

Appendix A4 *CWPT* rating for the general reading achievement 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 general reading achievement, the WWC rated *CWPT* as having potentially positive effects. It did not meet the criteria for positive effects because only one study met WWC standards, and that study did not show statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *CWPT* 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 of *CWPT* showed a substantively important positive effect.

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 of *CWPT* showed indeterminate effects. In addition, no studies showed a statistically significant or substantively important negative effect.

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 showed a statistically significant positive effect. In addition, only one study of *CWPT* met the WWC standards with or without reservations.

AND

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

  Met. No studies of *CWPT* showed a statistically significant or substantively important negative effect.

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 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	0	0	0	na
Fluency	0	0	0	na
Comprehension	0	0	0	na
General reading achievement	1	6	218	Small

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."