

Appendix

Appendix A1.1 Study characteristics: Bramlett, 1994 (quasi-experimental design)

Characteristic	Description
Study citation	Bramlett, R. K. (1994). Implementing cooperative learning: A field study evaluating issues for school-based consultants. <i>Journal of School Psychology, 32</i> (1), 67–84.
Participants	Eighteen third-grade teachers volunteered to participate in the study. They were matched on the basis of geographic representativeness (school district) and years of teaching experience. In the analysis sample, the <i>CIRC</i> [®] group included 198 students in nine classrooms, and the comparison group included 194 students in nine classrooms. Each of the two groups of children were divided into three ability levels (lowest 33%, middle 33%, and upper 34%) based on the students' percentile rankings of the pretest California Achievement Test (CAT) total reading scores. ¹
Setting	The study took place in eight school districts in rural southern Ohio. The number of participating schools was not provided in the study.
Intervention	Students in the nine intervention classes were given only the reading components of the <i>CIRC</i> [®] program: basal related activities, partner reading, story structure, words out loud, word meaning, story retelling, spelling, direct instruction in reading comprehension, and independent reading.
Comparison	Students in the comparison group received their regular reading curriculum, which was not indicated in the study. Teachers in the comparison group were promised that they could receive <i>CIRC</i> [®] training at the completion of the study, and six of them were subsequently trained.
Primary outcomes and measurement	The author administered four California Achievement Test (CAT) subtests: Reading Vocabulary, Reading Comprehension, Total Reading, and Word Analysis. The Total Reading subtest has not been included in this review because it is based on the results from the two subtests reported separately: Reading Vocabulary and Reading Comprehension (see Appendix A2 for more detailed descriptions of outcome measures).
Teacher training	The teachers received a one-day (6 hour) training in <i>CIRC</i> [®] by a certified trainer, as well as the project supplemental materials. Following training, the teachers were given assistance via observation and behavioral consultation sessions (approximately 15–30 minutes). Teachers also attended three half-day meetings during the study year to discuss implementation issues. The teachers in the comparison group were promised training and materials upon completion of the first year of the project.

1. For the lowest 33% subgroup, the study did not establish that the comparison group was equivalent to the intervention group at baseline. Analyses of the other two subgroups (middle ability and higher ability) are presented in Appendix A4.

Appendix A1.2 Study characteristics: Skeans, 1991 (quasi-experimental design)

Characteristic	Description
Study citation	Skeans, Sharon E. S. (1991). The effects of cooperative integrated reading and composition, fidelity of implementation, and teacher concerns on student achievement. <i>Dissertation Abstracts International</i> , 53(02), 0455A. (UMI No. 9217026)
Participants	The study used intact classroom groups from the same school district, with twenty-four third-grade teachers matched on students' achievement, demographic factors, and hours of training. ¹ No attrition was reported. In the analysis sample, 169 students were in the 12 intervention classrooms and 141 students were in the 12 comparison classrooms.
Setting	The study took place in a suburban district north of Houston, Texas.
Intervention	The intervention group was taught <i>CIRC</i> [®] integrated language arts and cooperative learning for 18 weeks. The program used daily 90-minute lessons to focus on story-related activities, direct instruction in reading comprehension, and integrated reading and language arts activities. Within four-member cooperative teams, pairs of mixed-ability students worked together to read. Within their teams, students re-read the story and discussed to clarify unknown vocabulary, understand the main idea, and comprehend stories. They worked through the writing process together, drafting, revising, and editing each other's writing.
Comparison	Students in the comparison classes continued using the district's integrated language arts program. The children's pretest scores on the Metropolitan Achievement Test were used as a covariant in the analyses to account for the differences between groups.
Primary outcomes and measurement	The author administered two subtests of the Metropolitan Achievement Test (MAT): Vocabulary and Reading Comprehension. Third graders received the elementary level of the test. The MAT Language subtest was also used in the study, but has not been included in this review because it was outside the scope of the Beginning Reading review (see Appendix A2 for more detailed descriptions of outcome measures).
Teacher training	The intervention teachers received at least 12 hours of <i>CIRC</i> [®] training. The teachers in both conditions had received at least 18 hours of training in the district's integrated language arts program. In order to continue teaching with the program, teachers needed to return a Stages of Concern questionnaire and a formative evaluation of <i>CIRC</i> [®] , indicating that they were implementing the program.

1. The fifth-grade sample that was 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](#).

Appendix A2 Outcome measures in the comprehension domain by construct

Outcome measure	Description
<i>Reading comprehension</i>	
California Achievement Test (CAT): Reading Comprehension (Form E)	A group-administered, standardized assessment of reading comprehension (as cited in Bramlett, 1994).
CAT: Word Analysis (Form E)	A group-administered, standardized assessment of word analysis (as cited in Bramlett, 1994).
Metropolitan Achievement Test-6 (MAT-6): Reading Comprehension (Form L)	A group-administered, standardized assessment of reading comprehension (as cited in Skeans, 1991).
<i>Vocabulary development</i>	
CAT: Reading Vocabulary (Form E)	A group-administered, standardized assessment of vocabulary (as cited in Bramlett, 1994).
MAT-6: Vocabulary subtest (Form L)	A group-administered, standardized assessment of vocabulary (as cited in Skeans, 1991).

Appendix A3 Summary of study findings included in the rating for the comprehension domain by construct¹

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
			CIRC [®] group ³	Comparison group	Mean difference ⁴ (CIRC [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
Construct: Reading comprehension								
Bramlett, 1994 (quasi-experimental design)⁸								
CAT: Reading Comprehension	Third grade	18/392	687.0 (56.4)	681.0 (61.1)	6.0	0.10	ns	+4
CAT: Word Analysis	Third grade	18/392	667.0 (43.3)	662.0 (49.7)	5.0	0.11	ns	+4
Skeans, 1991 (quasi-experimental design)⁸								
MAT-6: Reading Comprehension	Third grade	24/305	58.6 (20.7)	57.0 (20.6)	1.6	0.08	ns	+3
Construct: Vocabulary development								
Bramlett, 1994 (quasi-experimental design)⁸								
CAT: Reading Vocabulary	Third grade	18/392	684.0 (48.7)	682.0 (59.5)	2.0	0.04	ns	+1
Skeans, 1991 (quasi-experimental design)⁸								
MAT-6: Vocabulary	Third grade	24/310	61.2 (17.9)	57.5 (18.3)	3.7	0.20	ns	+8
Average⁹ for comprehension domain (Bramlett, 1994)						0.08	ns	+3
Average⁹ for comprehension domain (Skeans, 1991)						0.14	ns	+6
Domain average⁹ for comprehension domain						0.11	na	+4

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the improvement index. Subgroup findings for high ability students (defined as upper 34% of the sample) and medium ability students (middle 33%) are presented in Appendix A4.
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. The CIRC[®] group means were adjusted for pretest. Pretest reading scores were used as a covariant.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).

(continued)

Appendix A3 Summary of study findings included in the rating for the comprehension domain by construct¹ *(continued)*

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 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.
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 (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 Bramlett (1994), a correction for clustering and multiple comparisons was needed, so the significance levels may differ from those reported in the original study. In the case of Skeans (1991), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.
9. 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 size.

Appendix A4 Summary of subgroup findings for the comprehension domain by construct¹

Outcome measure	Study sample	Sample size (classrooms/ students)	Authors' findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
			CIRC [®] group ³	Comparison group	Mean difference ⁴ (CIRC [®] – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
Construct: Reading comprehension								
Bramlett, 1994 (quasi-experimental design)⁸								
CAT: Reading Comprehension	Third grade/ medium	18/151	698.0 (45.6)	695.0 (35.7)	3.0	0.07	ns	+3
CAT: Word Analysis	Third grade/ medium	18/151	670.0 (29.9)	673.0 (38.3)	-3.0	-0.09	ns	-3
CAT: Reading Comprehension	Third grade/ high	18/92	744.0 (32.7)	735.0 (35.5)	9.0	0.26	ns	+10
CAT: Word Analysis	Third grade/ high	18/92	712.0 (38.2)	704.0 (37.1)	8.0	0.21	ns	+8
Construct: Vocabulary development								
Bramlett, 1994 (quasi-experimental design)⁸								
CAT: Reading Vocabulary	Third grade/ medium	18/151	694.0 (36.7)	693.0 (30.0)	1.0	0.03	ns	+1
CAT: Reading Vocabulary	Third grade/ high	18/92	736.0 (33.1)	738.0 (31.6)	-2.0	-0.06	ns	-2

ns = not statistically significant

1. This appendix presents subgroup findings for high ability students (defined as upper 34% of the sample) and medium ability students (middle 33%) for measures that fall in the comprehension domain. Total group scores were used for rating purposes and are presented in Appendix A3.
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. The CIRC[®] group means were adjusted for pretest. Pretest reading scores were used as a covariant.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
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 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.
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 (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 Bramlett (1994), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

Appendix A5 CIRC® 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 CIRC® as having no discernible effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, and negative effects) because the one study that met WWC standards did not show statistically significant or substantively important effects.

Rating received

No discernible effects: No affirmative evidence of effects.

- Criterion 1: None of the studies shows a statistically significant or substantively important effect, either *positive* or *negative*.

Met. No study showed a statistically significant or substantively important effect, either positive or negative.

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 study showed statistically significant positive effects or met the WWC evidence standards for a strong design.

AND

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

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

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.

Not met. No study showed a statistically significant or 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.

Not met. No study showed a statistically significant or substantively important negative effect, but two studies showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through either of the following criteria.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect, and at least one study showing a statistically significant or substantively important *negative* effect, but no more such studies than the number showing a statistically significant or substantively important *positive* effect.

Not met. No study showed a statistically significant or substantively important effect, either positive or negative.

OR

- Criterion 2: At least one study showing a statistically significant or substantively important effect, and more studies showing an *indeterminate* effect than showing a statistically significant or substantively important effect.

Not met. No study showed a statistically significant or substantively important effect, while two studies showed indeterminate effects.

Appendix A5 CIRC® rating for the comprehension domain (continued)

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

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

Not met. No study showed a statistically significant or substantively important negative effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *positive* effect, or more studies showing statistically significant or substantively important *negative* effects than showing statistically significant or substantively important *positive* effects.

Met. No study showed a statistically significant or substantively important positive effect.

Negative effects: Strong evidence of a negative effect with no overriding contrary evidence.

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

Not met. No study showed a statistically significant or substantively important negative effect.

AND

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

Met. No study showed statistically significant or substantively important positive 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 ²
		Schools ¹	Students	
Alphabetics	0	0	0	na
Fluency	0	0	0	na
Comprehension	2	over 8	702	Moderate to large
General reading achievement	0	0	0	na

na = not applicable/not studied

1. No information is provided about the number of schools. Bramlett (1994) study took place in eight school districts (and 18 classrooms), while 24 teachers participated in Skeans (1991) study.
2. 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.”