

Appendix

Appendix A1 Study characteristics: Taylor, Frye, Short, & Shearer, 1991 (randomized controlled trial)

Characteristic	Description
Study citation	Taylor, B. M., Frye, B. J., Short, R., & Shearer, B. (1991). <i>Early Intervention in Reading: Preventing reading failure among low-achieving first grade students</i> . Minneapolis, MN: University of Minnesota, Center for Urban and Regional Affairs and Office of the Vice President of Academic Affairs.
Participants	<p>Twelve first grade teachers from two schools were randomly assigned either to the intervention or to a control group (six teachers were assigned to <i>EIR</i>[®] and six teachers were assigned to the comparison group). In each classroom, five or six of the lowest scoring students participated in the study. Students were identified initially by teacher recommendations based on reading test scores and confirmed through testing by study assistants using knowledge of consonant sounds, reading of sight words on the Dolch preprimer list, and the Burns-Roe Informal Reading Inventory, an auditory phonemic segmentation and blending test. Thirty-one low-achieving students from six <i>EIR</i>[®] classes and 28 students from six comparison classes participated in the study (there were five or six students in each class, but only three low-achieving students in one of the comparison classrooms). The district reports 20 percent of students receive free or reduced price lunch and 10 percent are minority students, but no specific demographic information was given about the study participants.</p> <p>Twenty-nine of the original 31 students in the treatment group remained throughout the study. All of the 28 comparison students remained in the study.¹</p>
Setting	The study took place in one suburban district in a metropolitan area in the Midwest.
Intervention	The program involved pulling aside the lowest-achieving students in each class to work as a group with the teacher. The program was implemented in three-day cycles from October to April of the school year. On day one, the teacher read a picture book (this part of the intervention occurred with the entire class). The teacher then taught the intervention students to segment words and blend phonemes into words. On days two and three, the intervention students read a story summary with minimal assistance. They also wrote one sentence a day that was related to the story with the teacher's help. In addition to the 15–20 minutes that students worked with teachers each day, children worked individually (for 5 minutes) or in pairs (for 10 minutes) with a trained aide or project assistant. Running records were taken by the teacher or aide weekly to assess students' progress. In this study, the project assistants, who were graduate students from a local university, spent time listening to intervention students read individually and providing teachers with feedback on the program.
Comparison	Students in the comparison classes participated in their regular reading instruction, supplemented with additional instruction from teachers and reading specialists. Some students received 30-minute pull-out sessions, while others were aided by special reading teachers within their own classes.
Primary outcomes and measurement	For both pre- and posttests, the authors administered a vowel sounds test, a test of segmentation and blending, and the Gates-MacGinitie reading test. Two additional tests, the Burns-Roe Informal Reading Inventory and the percent of children reading a 150-word selection at the first grade level, were used in the study, but have not been included in this review. ² (See Appendix A2.1–2.2 for more detailed descriptions of outcome measures.)
Teacher training	Intervention teachers attended an all-day workshop the summer before implementation. Three afternoon meetings were also held to support implementation. Project assistants (graduate students) observed and assisted (listening to program students read aloud) in intervention classes. These assistants were in program classes about 90 minutes per week. Assistants gave feedback and suggestions for improvement to program teachers.

1. Outcome tests were conducted over two days, thus the total number of students in the analysis samples varies depending on the measure assessed and student absences each day.
2. The administration of the tests involved substantial reading and interaction between students and testers, who served as assistants in the intervention classrooms. The WWC eliminated this test from consideration in the review because students in the intervention group had a pre-existing relationship with testers, which created unequal testing conditions across the intervention and comparison groups.

Appendix A2.1 Outcome measures in the alphabetics domain

Outcome measure	Description
Phonological awareness	
Segmentation and blending	An 18-item version of a segmentation (6-items) and blending (6-items) test adapted from Taylor and Pearson (as cited in Taylor, Frye, Short, & Shearer, 1991). At posttest, children were asked to say each sound in a three- to four-letter word, then blend the sounds together. Twelve of the 18 words were on the pretest, and the other six were new.
Phonics	
Vowel sounds	A test measuring students' knowledge of letter sounds for 15 pairs of vowels. The same test was given at pre- and posttest (as cited in Taylor, Frye, Short, & Shearer, 1991).

Appendix A2.2 Outcome measures in the comprehension domain

Outcome measure	Description
Gates-MacGinitie reading test	A standardized test of reading readiness; form R was given as the pretest and Level A as the posttest (as cited in Taylor, Frye, Short, & Shearer, 1991).

Appendix A3.1 Summary of study findings included in the rating for the alphabets domain¹

Outcome measure	Study sample	Sample size (classes/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (EIR [®] – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Early Intervention in Reading [®] group	Comparison group				
Taylor, Frye, Short, & Shearer, 1991 (randomized controlled trial)⁷								
Construct: Phonological awareness								
Segmentation and blending	Grade 1	12/56	14.30 (4.09)	10.41 (5.41)	3.89	0.80	Statistically significant	+29
Construct: Phonics								
Vowel sounds	Grade 1	12/56	10.62 (3.18)	6.44 (2.72)	4.18	1.39	Statistically significant	+42
Average⁸ for alphabets domain (Taylor, Frye, Short, & Shearer, 1991)						1.10	Statistically significant	+36

1. This appendix reports findings considered for the effectiveness rating and the average improvement index.
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.
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 formula the WWC used to calculate statistical significance. In the case of Taylor, Frye, Short, & Shearer (1991), corrections for clustering and multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
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 size.

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain¹

Outcome measure	Study sample	Sample size (classes/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (EIR [®] – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Early Intervention in Reading [®] group	Comparison group				
Taylor, Frye, Short, & Shearer, 1991 (randomized control trial)⁷								
Gates-MacGinitie reading test	Grade 1	12/57	20.76 (8.03)	17.14 (6.97)	3.62	0.47	ns	+18
Average for comprehension domain (Taylor, Frye, Short, & Shearer, 1991)						0.47	ns	+18

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
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.
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 Taylor, Frye, Short, & Shearer (1991), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

Appendix A4.1 *Early Intervention in Reading*[®] rating for the alphabetic 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 alphabetic, the WWC rated *Early Intervention in Reading*[®] as having potentially positive effects. It did not meet the criteria for positive effects because only one study met WWC evidence standards. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Early Intervention in Reading*[®] 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. The one study showed a statistically significant positive effect and had a strong design.

- 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. There were no studies showing 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. There was only one study that met WWC standards.

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

Met. There were no studies showing 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 A4.2 Early Intervention in Reading® 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 *Early Intervention in Reading*® as having potentially positive effects. It did not meet the criteria for positive effects because only one study met WWC evidence standards. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered because *Early Intervention in Reading*® 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. The one study showed a substantively important positive effect and had a strong design.

- 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. There were no studies showing 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. There was only one study that met WWC evidence standards.

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

Met. There were no studies showing 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 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Alphabetics	1	2	56	Small
Fluency	0	0	0	na
Comprehension	1	2	57	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.”