Appendix

Appendix A1 Study characteristics: Phillips, Norris, Mason, & Kerr, 1990 (randomized controlled trial)

Characteristic	Description
Study citation	Phillips, L. M., Norris, S. P., Mason, J. M., & Kerr, B. M. (1990). *Effect of early literacy intervention on kindergarten achievement* (Technical report no. 520). Champaign, IL: Center for the Study of Reading, University of Illinois at Urbana-Champaign.
Participants	The study sample included 40 Newfoundland schools divided into three groups: rural (drawing students from one small community), rural collector (drawing students from a number of small communities), and small urban communities. From each of these groups, four schools were randomly selected to participate in the study, each randomly assigned to one of four study conditions: those that used *Little Books* at home, used *Little Books* at home and school, used *Little Books* at school only, and did not use *Little Books* (comparison group).[1] The study began with 325 students attending 18 kindergarten classes at 12 schools. The analysis sample included 314 students.[2] The area in which the study was conducted has the highest rate of basic and functional illiteracy in Canada and consistently scores below the Canadian national norm on standardized tests.
Setting	This study was conducted in kindergarten classrooms across rural and urban locations in Newfoundland, Canada.
Intervention	Three variations of the *Little Books* intervention were studied, those that used the program at home only, at school only, or both at home and at school. All three used the *Little Books*, which use high-frequency words, simple sentences, and thematic topics with which children are familiar. For the home group, school staff gave a new book to each child at the start of each week for the child to take home and read with parents. For the school group, a different book was introduced by teachers each week, and approximately 10–15 minutes each day were devoted to the materials. For the school and home group, the school-only procedures were followed and at the end of each week the teachers sent the *Little Book* home.
Comparison	The control condition used the standard language development program for Newfoundland, Canada.
Primary outcomes and measurement	This review focuses on the results for the Metropolitan Readiness Test (see Appendix A2.1 for a more detailed description of outcome measures). Results for the Emergent Literacy Concepts Test, a test designed for the study, are not included in this review because of unequal testing conditions between the interventions and comparison groups. The posttest included a section that asked students to read words that came from the *Little Books,* so students in the intervention condition had exposure to these words before the posttest was administered. It is unknown whether or not the comparison condition students had exposure to the words. The authors also used the CIRCUS Listen to the Story test, which is an assessment of oral comprehension and thus falls outside the domains included in the WWC beginning reading review.[3]
Teacher training	Parents and teachers were trained to use the *Little Books*. Parents were shown a video in which a parent and child worked with several books. Guidelines were provided by the developers, which gave suggestions about setting up a comfortable reading arrangement, discussing the main idea of the book, reading the book aloud, and eliciting the child to read. Suggestions were also made for use of particular books. Teachers attended a workshop in which they were encouraged to spend 10–15 minutes each day with the *Little Books*. They were given a specific instructional procedure that involved an opening, modeling, tryouts, and a closing. They were asked to introduce and read the book to the whole class, then work with smaller groups of children reading the book for the next three days. The last day, they were to ask each child to read the *Little Book.*

1. The study authors presented data separately for each condition by geographic setting. The WWC combined effects across the geographic settings because there was only one of each of the intervention conditions and one comparison group within each geographic setting, making it difficult to separate the effects of the intervention from other characteristics of the schools.
2. The study relied on an analysis sample of 309 students because the authors included the Circus outcome measure and some students were missing this measure. The WWC did not use the Circus measure for its analysis (see *Primary outcomes and measurement* row).
3. For further details about the scope of the beginning reading topic review, see the Beginning Reading Protocol.

Appendix A2 Outcome measures in the general reading achievement domain

Characteristic	Description
Metropolitan Readiness Test (MRT)	The MRT, a standardized test published by Harcourt, is a group-administered test. It assesses literacy skills of pre-K to first-grade students, measuring auditory memory, letter recognition, and language and listening. The test has two levels of difficulty: MRT-1, which was used as the pretest, and MRT-2, which was used as the posttest (as cited in Phillips, Norris, Mason, & Kerr, 1990).

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

Outcome measure	Study sample	Sample size (students/ schools)	Authors' findings from the study, aggregated by the WWC[2]: Mean outcome (standard deviation[3]) — *Little Books* group[4]	Authors' findings from the study, aggregated by the WWC[2]: Mean outcome (standard deviation[3]) — Comparison group	WWC calculations: Mean difference[5] (*Little Books* – comparison)	WWC calculations: Effect size[6]	WWC calculations: Statistical significance[7] (at $\alpha = 0.05$)	WWC calculations: Improvement index[8]
Phillips, Norris, Mason, & Kerr, 1990 (randomized controlled trial)[9]								
Home-only group vs. comparison group[10]								
Metropolitan Reading Readiness Test	Kindergarten students	165/6	41.93 (13.20)	37.91 (12.25)	4.02	0.31	ns	+12
Home and school group vs. comparison group[10]								
Metropolitan Reading Readiness Test	Kindergarten students	156/6	41.91 (14.71)	37.91 (12.25)	4.00	0.30	ns	+12
School-only group vs. comparison group[10]								
Metropolitan Reading Readiness Test	Kindergarten students	157/6	42.17 (14.44)	37.91 (12.25)	4.26	0.32	ns	+12
Domain average for general reading achievement[11]						0.31	ns	+12

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
2. The study authors presented data separately for each condition by geographic setting. The WWC combined effects across the geographic settings because there was only one of each of the intervention conditions and one comparison group within each geographic setting, making it difficult to separate the effects of the intervention from other characteristics of the schools.
3. The standard deviation shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
4. The *Little Books* group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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.
9. The level of statistical significance was calculated by the WWC and, where necessary, corrects 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 Phillips, Norris, Mason, and Kerr (1990), a correction for clustering was needed, so the significance levels differ from those reported in the original study.
10. The WWC combined effects across the three geographic settings. For the formula used to aggregate results across groups, see Technical Details of WWC-Conducted Computations.
11. This row provides the study average, which, in this instance, is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4 *Little Books* 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 *Little Books* 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, as Little Books 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. *Little Books* had one study showing substantively important positive effect.

- 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 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. *Little Books* has only one study and it did not show statistically significant positive effects.

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

  Met. No studies 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	314	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."