

Appendix

Appendix A1.1 Study characteristics: Justice & Ezell, 2002 (randomized controlled trial)

Characteristic	Description
Study citation	<p>Justice, L. M., & Ezell, H. K. (2002). Use of storybook reading to increase print awareness in at-risk children. <i>American Journal of Speech-Language Pathology</i>, 11(1), 17–29.</p> <p><i>Additional source:</i> Justice, L. M. (2000). An experimental evaluation of an intervention to stimulate written language awareness in preschool children from low-income households. <i>Dissertation Abstracts International</i>, 61(07), 2587A. (UMI No. 9980417).</p> <p>This study, including the additional source, is not included in the overall effectiveness rating because it compared variations of <i>Interactive Shared Book Reading</i> to each other, which does not allow the effects of <i>Interactive Shared Book Reading</i> to be determined.</p>
Participants	<p>The study began with 38 three- to five-year-old children from low-income households who met the researchers' eligibility requirements. The four children who did not complete all project activities and the four age-matched children were dropped from the study, leaving a final sample of 30 children. The mean age of the children at the start of the study was 53 months. Half of the children were female, 90% were Caucasian, 7% were Asian, and 3% were African-American. All children in the study attended preschool at a Head Start center, were considered at-risk, and were from households with incomes at or below the 133% poverty line (\$22,211 annual income for a family of four). The children were divided into 15 pairs matched on age; one from each pair was then randomly assigned to the print focus group and the other was assigned to the picture focus group.</p>
Setting	<p>The study took place at a Head Start center in a rural Appalachian region of southeastern Ohio.</p>
Print focus group	<p>The children in this group participated in <i>Interactive Shared Book Reading</i> with a print focus to enhance their familiarity with print during the reading of target books. The intervention lasted eight weeks and children participated in groups of three to five for a total of 24 sessions. Adult readers posed nine prompts. Each prompt was one of three general types: print conventions, concept of word, or alphabet knowledge. Scripts were developed for each book with three prompts of each type. During each reading session, the reader called on each child to respond to a prompt at least once and the reader also commented on enjoyment of the book reading activity.</p>
Picture focus group	<p>The children in this group participated in <i>Interactive Shared Book Reading</i> with a picture focus. The intervention lasted eight weeks and children participated in groups of three to five for a total of 24 sessions. Adult readers posed nine prompts focused on the pictures in the book. Each prompt was one of three general types: character focus, perceptual focus, or action focus. Each type of prompt occurred three times per book reading.</p>
Primary outcomes and measurement	<p>The primary outcome domain assessed was print knowledge, which was assessed with six non-standardized measures: letter orientation and discrimination, print concepts, print recognition, words in print, alphabet knowledge, and literacy terms. The authors also reported on a print awareness composite score that was a sum of scores across the six measures. However, the WWC does not include this composite in this report because the WWC includes the six individual measures used to develop the composite (see Appendix A2.2 for more detailed descriptions of outcome measures).</p>
Teacher training	<p>No information on teacher training was provided and the first author was the instructor for both intervention conditions. The WWC found no reasons to believe that the person implementing the intervention and comparison conditions was not equally trained and motivated to implement each condition.</p>

Appendix A1.2 Study characteristics: Lamb, 1986 (randomized controlled trial)

Characteristic	Description
Study citation	Lamb, H. A. (1986). The effects of a read-aloud program with language interaction. <i>Dissertation Abstracts International</i> , 47(5A). (UMI No. 8616894).
Participants	The study included 36 three- to five-year-old children, who were predominantly from a minority racial group and low-income families. These children were randomly assigned to three intervention groups and one no-treatment comparison group participating in regular preschool activities. Results for the 19 children in the read-aloud with language interaction group and the read-aloud only group are included in this report.
Setting	The study took place at a day care center serving minority and lower socio-economic status children in Tallahassee, Florida.
Intervention	This study included three intervention groups: read-aloud with language interaction; language interaction only; and read-aloud only. The WWC chose the read-aloud with language interaction group as the <i>Interactive Shared Book Reading</i> intervention group for this report. For this group, the researcher read pre-selected books to the children daily for ten weeks, with discussion taking place prior to, during, and following reading of the text. The language interaction only group was excluded from the review because it is not an appropriate comparison group to isolate the effects of <i>Interactive Shared Book Reading</i> .
Comparison	The study included a no-treatment comparison group that did not receive read-aloud techniques, language interaction, or researcher contact. However, this comparison group could not be used to estimate the impacts of the interventions because the researcher implemented the intervention in all three groups but had no or minimal contact with the children in the no-treatment comparison group and the effects of the individual providing the intervention cannot be separated from the effects of the intervention. The WWC chose the read-aloud only group as the comparison group for this report. For this group, the researcher read the same pre-selected books to the children daily for ten weeks, but did not initiate any interaction.
Primary outcomes and measurement	The primary outcome domains assessed were oral language and print knowledge. Oral language was measured with a non-standardized test (Record of Oral Language) and a standardized test (Peabody Picture Vocabulary Test—Revised; PPVT-R). Print knowledge was measured with a standardized test (Concepts About Print: Sand and Stones) (see Appendices A2.1–2.2 for more detailed descriptions of outcome measures).
Teacher training	The article did not provide information on teacher training and the researcher was the instructor for the three intervention conditions. The WWC found no reasons to believe that the person implementing the intervention and comparison conditions was not equally trained and motivated to implement each condition.

Appendix A1.3 Study characteristics: Mautte, 1991 (randomized controlled trial)

Characteristic	Description
Study citation	Mautte, L.A. (1991). The effects of adult-interactive behaviors within the context of repeated storybook readings upon the language development and selected prereading skills of prekindergarten at-risk students. <i>Dissertation Abstracts International</i> , 52(1), 122A. (UMI No. 9115887).
Participants	The study began with 66 four-year-old at-risk low-income children from families with a history of abuse or neglect. All children were attending pre-kindergarten At-Risk Program classes in an Early Childhood Education Center in Tampa, Florida. Three children withdrew during the pre-testing phase and 10 children withdrew during the intervention leaving a total of 53 children. Fifty-one percent of the sample was female, 9% were Caucasian, 4% were Hispanic, and 87% were African-American. Results for the 38 children who had been randomly assigned within two developmental strata (average or delayed) to the repeated reading with adult interaction group or the repeated reading without adult interaction group are included in this report. ¹
Setting	The study took place at an inner-city Early Childhood Education center located in the third largest school system in Florida, the Hillsborough County School System in Tampa, Florida. ² The center was open 12 months a year, five days a week. When the study was conducted, the center had a total enrollment of 267 children, of whom 243 were black non-Hispanics, 7 were white non-Hispanics, and 17 were Hispanics.
Intervention	There were two storybook reading intervention groups: repeated reading of Big Book storybooks with adult interaction and repeated reading of Big Book storybooks without adult interaction. The WWC chose the repeated reading with adult interaction group as the intervention group for this report. Children in this group listened to stories in groups of seven. The researcher read a story to each small group three times per week for 20 weeks. Each session lasted about 25 minutes and only one book was used per week. In the first reading session each week, the researcher read the story to the children and interacted with them before, during, and after reading of the text. In the second and third reading sessions each week, the researcher read the story to the children and interacted with them in a manner designed to generate more elaborative responses and encourage children in the storybook reading.
Comparison	The study included a no-treatment comparison group that participated in their regular pre-kindergarten curriculum activities as mandated by the Hillsborough County Department of Early Childhood Education. However, this comparison group could not be used to estimate the impacts of the intervention because the researcher implemented the intervention in both groups but had no or minimal contact with the children in the no-treatment comparison group and the effects of the individual providing the intervention cannot be separated from the effects of the intervention. The WWC chose the repeated reading without adult interaction group as the comparison group for this report. For this group, there were the same number of sessions that followed a similar format as described above; however, the researcher did not implement the story reading model or elicit discussion or interaction. The children were exposed to a reading of the storybooks without adult interaction.
Primary outcomes and measurement	The primary outcome domains assessed were oral language and print knowledge. Oral language was measured using a standardized measure (Preschool Language Scale; PLS) and print knowledge was also measured using a standardized measure (Test of Early Reading Ability; TERA) (see Appendices A2.1-2.2 for more detailed descriptions of outcome measures).
Teacher training	The article did not provide information on teacher training and the researcher was the instructor for both intervention conditions. The researcher enlisted the assistance of reading instructors from the University of Florida to listen to tape-recorded treatment sessions and check adherence of sessions to scripts in order to ensure fidelity of treatment. The WWC found no reasons to believe that the person implementing the intervention and comparison conditions was not equally trained and motivated to implement each condition.

1. The study was not downgraded by the WWC due to differential attrition because the authors demonstrated pre-test equivalence between the children remaining in the repeated reading with adult interaction group and the repeated reading without adult interaction group.

2. The other inner-city center in the Hillsborough School System was used to field-test materials.

Appendix A1.4 Study characteristics: McCormick & Mason, 1989 (quasi-experimental design)¹

Characteristic	Description
Study citation	<p>McCormick, C. E., & Mason, J. M. (1989). Fostering reading for Head Start children with Little Books. In J. Allen & J. M. Mason (Eds.), <i>Risk makers, risk takers, risk breakers: Reducing the risks for young literacy learners</i> (pp. 154–177). Portsmouth, NH: Heinemann.</p> <p><i>Additional source:</i> McCormick, C. E., & Mason, J. M. (1986). Use of little books at home: A minimal intervention strategy that fosters early reading. Center for the Study of Reading: Technical Report No. 338 (ED 314742).</p>
Participants	The study included 51 three- to five-year-old children with a mean age of 4 years, 7 months. ² Ninety-six percent were Caucasian and four percent were African-American. To form intervention and comparison groups, the authors divided in half an alphabetized list of children from four classes; the first half of the list from two classes and the last half of the list from the other two classes were placed in the intervention group and the remaining children were placed in the comparison group.
Setting	The study took place in four Head Start classrooms in a small city in the Midwest.
Intervention	The children in the intervention group participated in a 10- to 15-minute “Book Recitation” session once a week over a six-week period. Children were taught in small groups of four to six. Each group was introduced to one new book each week (Little Books were chosen from the Pint Size Print series). In this intervention, children were encouraged to make predictions about each story based upon its cover and they were encouraged to relate their own experiences relevant to the story topic. The teacher modeled reading by showing the children the pictures and text and pointing to the words as she read. Lastly, the teacher encouraged children to accompany her in reading the text. Each child received a copy of the book of the week in the mail at home.
Comparison	The children in the comparison group participated in a 10 to 15-minute “Story Discussion” session once a week over a six-week period. Children were taught in small groups of four to six. Each group was introduced to one new classic story, such as the Three Little Pigs, The Three Bears, and Little Red Riding Hood, each week. The teacher told the story while presenting illustrations from the storybook and children did not see the text (i.e., reading was based on the illustrations). The children were asked to retell the story and the illustrations served as prompts. Each child received a copy of the illustrations from the story of the week in the mail at home.
Primary outcomes and measurement	The primary outcome domains assessed were print knowledge and early reading/writing. Print knowledge was measured by two non-standardized measures, letter naming and points to print. Early reading/writing was assessed with four non-standardized measures including picture label, word label, taught book, and new book. The WWC only includes the picture label measure in this report. The word label and new book measures did not have sufficient face validity or reliability to be included. The taught book measure was excluded from this report because the Book Recitation group had been exposed to the book during the intervention but the Story Discussion group had not, so it was not a fair measure of intervention effects. The researchers also included a parent questionnaire from which two outcome measures were developed, a parent estimate of child literacy and a parent literacy support measure. These measures are not included in this report because they are not relevant to the WWC review (see Appendices A2.2-2.3 for more detailed descriptions of outcome measures).
Teacher training	The article did not provide information on teacher training and the first author implemented both conditions. The WWC found no reasons to believe that the person implementing the intervention and comparison conditions was not equally trained and motivated to implement each condition.

1. McCormick and Mason (1989) implemented a second phase of the intervention when the children were in kindergarten. This phase is not included in the WWC review because it did not meet the criterion for sample age (i.e., children aged three to five or in preschool) for the WWC review of ECE interventions.

2. The additional source for this study (McCormick and Mason, 1986) indicated that 52 children comprised the sample.

Appendix A2.1 Outcome measures in the oral language domain

Outcome measure	Description
Peabody Picture Vocabulary Test—Revised (PPVT-R)	A standardized measure of children's receptive vocabulary that requires them to identify pictures that correspond to spoken words (as cited in Lamb, 1986).
Record of Oral Language	A non-standardized measure of children's oral language proficiency that requires them to repeat sentences (as cited in Lamb, 1986).
Preschool Language Scale (PLS)	A standardized measure of children's receptive and expressive language (as cited in Mautte, 1991).

Appendix A2.2 Outcome measures in the print knowledge domain

Outcome measure	Description
Print concepts	A researcher adaptation of a test to assess children's knowledge of print- and book-reading conventions (as cited in Justice & Ezell, 2002).
Print recognition	A researcher adaptation of a test to assess children's ability to recognize print in context (as cited in Justice & Ezell, 2002).
Words in print	A researcher-developed measure of children's awareness of printed words as discrete elements of written language (as cited in Justice & Ezell, 2002).
Letter orientation/discrimination	A researcher adaptation of a measure to assess children's ability to recognize letters presented to them in different orientations (e.g. upside down, sideways) (as cited in Justice & Ezell, 2002).
Alphabet knowledge	A researcher-developed measure designed to assess children's knowledge of the letters in the alphabet (as cited in Justice & Ezell, 2002).
Literacy terms	A researcher adaptation of a measure of children's knowledge of literacy-related words such as reading and writing (as cited in Justice & Ezell, 2002).
Concepts about print: Sand and stones	A standardized test of children's knowledge of book orientation, direction of words and lines of print, and other basic print concepts (as cited in Lamb, 1986).
Test of Early Reading Ability (TERA)¹	A standardized measure of young children's early reading skills that captures the following constructs: awareness of print in environmental contexts, vocabulary, listening, comprehension, knowledge of alphabet, and concepts about printed language (as cited in Mautte, 1991).
Letter naming	A non-standardized researcher-developed measure of children's ability to identify ten uppercase letters in print (as cited in McCormick & Mason, 1989).
Points to print	A non-standardized researcher-developed measure in which children were shown pictures of common objects with a one-word label for the objects beneath the pictures and children were asked to point to the text (as cited in McCormick & Mason, 1989).

1. By name, this measure sounds like it should be captured under the early reading/writing domain; however, the description of the measure identifies constructs that are pertinent to print knowledge such as knowing the alphabet, understanding print conventions, and environmental print.

Appendix A2.3 Outcome measure in the early reading/writing domain

Outcome measure	Description
Picture label	A non-standardized measure for which children were presented with black-and-white drawings of common objects such as a box of crayons and were required to name the picture (as cited in McCormick & Mason, 1989).

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

Outcome measure	Study sample	Sample size (children)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (Interactive Shared Book Reading – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Interactive Shared Book Reading group	Comparison group				
Lamb, 1986 (randomized controlled trial)⁷								
PPVT-R	3–5 year olds	19	27.30 (14.03)	27.44 (18.23)	–0.14	–0.01	ns	–0.3
Record of oral language	3–5 year olds	19	8.75 (8.30)	13.88 (10.51)	–5.13	–0.52	ns	–20
Average⁸ for oral language (Lamb, 1986)						–0.27	ns	–10
Mautte, 1991 (randomized controlled trial)⁹								
PLS-total score	4 year olds	38	64.79 (5.62)	61.34 (9.34)	3.46	0.43	ns	+17
Average⁸ for oral language (Mautte, 1991)						0.43	ns	+17
Domain average⁸ for oral language across all studies						0.08	na	+3

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup findings from the same studies are not included in these ratings, but are reported 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. 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 groups.
6. 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 favorable results.
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 Lamb (1986), 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 size.
9. In the case of Mautte (1991), no corrections for clustering or multiple comparisons were performed. However, students received the intervention in small groups, so clustering induced by grouping may have affected study findings.

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

Outcome measure	Study sample	Sample size (children)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (Interactive Shared Book Reading – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Interactive Shared Book Reading group	Comparison group				
Lamb, 1986 (randomized controlled trial)⁷								
Concepts about print: Sand and stones	3–5 year olds	19	3.60 (2.95)	4.22 (2.05)	–0.62	–0.23	ns	–9
Average⁸ for print knowledge (Lamb, 1986)						–0.23	ns	–9
Mautte, 1991 (randomized controlled trial)⁹								
TERA	4 year olds	38	8.13 (4.69)	9.79 (8.17)	–1.66	–0.24	ns	–10
Average⁸ for print knowledge (Mautte, 1991)						–0.24	ns	–10
McCormick & Mason, 1989 (quasi-experimental design)¹⁰								
Letter naming	3–5 year olds	51	2.80 (nr) ¹¹	2.70 (nr) ¹¹	0.10	0.07	ns	+3
Points to print	3–5 year olds	51	4.20 (nr) ¹¹	3.30 (nr) ¹¹	0.90	0.27	ns	+10
Average⁸ for print knowledge (McCormick & Mason, 1989)						0.17	ns	+7
Domain average⁸ for print knowledge across all studies						–0.10	na	–4

ns = not statistically significant

na = not applicable

nr = not reported

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Findings for Justice and Ezell (2002) are reported in Appendix A5 because it compares two *Interactive Shared Book Reading* interventions to each other, which does not allow the effects of *Interactive Shared Book Reading* to be determined.
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 groups.
6. 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 favorable results.

(continued)

Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain *(continued)*

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 Lamb (1986), 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 size.
9. In the case of Mautte (1991), no corrections for clustering or multiple comparisons were performed. However, students received the intervention in small groups, so clustering induced by grouping may have affected study findings.
10. In the case of McCormick and Mason (1989), no corrections for clustering or multiple comparisons were needed.
11. McCormick and Mason (1989) reported means but not the associated standard deviations. The authors were unable to provide standard deviations, so the WWC calculated the effect size based on the sample size and results from the author reported one-way ANOVA for letter naming [$F(1, 50) = 0.07$] and for points to print [$F(1, 50) = 0.93$].

Appendix A3.3 Summary of study findings included in the rating for the early reading/writing domain¹

Outcome measure	Study sample	Sample size (children)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation ²)		Mean difference ³ (Interactive Shared Book Reading – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Interactive Shared Book Reading group	Comparison group				
McCormick & Mason, 1989 (quasi-experimental design)⁷								
Picture label	3–5 year olds	51	13.10 (nr) ⁸	11.60 (nr) ⁸	1.50	0.70	Statistically significant	+26
Domain average⁹ for early reading/writing						0.70	Statistically significant	+26

nr = not reported

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. 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 groups.
6. 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 favorable results.
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 McCormick and Mason (1989), no corrections for clustering or multiple comparisons were needed.
8. McCormick and Mason (1989) reported means but not the associated standard deviations. The authors were unable to provide standard deviations, so the WWC calculated the effect size based on the sample size and results from the author reported one-way ANOVA [$F(1, 50) = 6.52$].
9. 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 Summary of subgroup findings for the oral language domain¹

Outcome measure	Study sample	Sample size (children)	Author's findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
			Interactive Shared Book Reading group	Comparison group	Mean difference ³ (Interactive Shared Book Reading – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
Mautte, 1991 (randomized controlled trial; average developmental level)⁷								
PLS—total score	4 year olds	17	69.25 (2.60)	69.67 (3.97)	–0.42	–0.12	ns	–5
TERA	4 year olds	17	11.50 (5.53)	14.67 (10.15)	–3.17	–0.36	ns	–14
Mautte, 1991 (randomized controlled trial; delayed developmental level)⁸								
PLS—total score	4 year olds	21	60.10 (3.60)	54.45 (6.20)	5.65	1.06	Statistically significant	+35
TERA	4 year olds	21	7.00 (2.79)	6.00 (2.68)	1.00	0.35	ns	+14

ns = not statistically significant

1. This appendix presents subgroup findings for measures that fall in the oral language domain. Total scale scores 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.
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 groups.
6. 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 favorable results.
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 Mautte (1991), no correction for clustering was performed. However, students received the intervention in small groups, so clustering induced by grouping may have affected study findings.
8. In the case of Mautte (1991), no correction for clustering was needed.

Appendix A5 Summary of findings for comparisons between variations of *Interactive Shared Book Reading* for the print knowledge domain¹

Outcome measure	Study sample	Sample size (children)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ⁴ (first group – second group)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			First <i>Interactive Shared Book Reading</i> group ³	Second <i>Interactive Shared Book Reading</i> group ³				
Justice & Ezell, 2002 (randomized controlled trial)⁸								
Print concepts	3–5 year olds	30	11.92 (4.02)	10.88 (2.22)	1.04	0.31	ns	+12
Print recognition	3–5 year olds	30	5.93 (3.30)	1.33 (1.40)	4.60	1.77	Statistically significant	+46
Words in print	3–5 year olds	30	7.43 (4.31)	3.04 (3.14)	4.39	1.13	Statistically significant	+37
Letter orientation/discrimination	3–5 year olds	30	17.45 (1.93)	15.18 (3.18)	2.27	0.84	ns	+30
Alphabet knowledge	3–5 year olds	30	10.92 (6.63)	7.75 (6.37)	3.18	0.48	ns	+18
Literacy terms	3–5 year olds	30	10.03 (3.93)	8.57 (2.13)	1.45	0.45	ns	+17
Domain average⁹ for print knowledge						0.83	Statistically significant	+30

ns = not statistically significant

1. This appendix presents a summary of study findings for measures that fall in the print knowledge domain for a study that is not included in the overall effectiveness ratings.
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 posttest means are covariate-adjusted means provided by the study author. The first *Interactive Shared Book Reading* group is the print focus condition. The second *Interactive Shared Book Reading* group is the picture focus condition.
4. Positive differences and effect sizes favor the first *Interactive Shared Book Reading* group; negative differences and effect sizes favor the second *Interactive Shared Book Reading* 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 groups.
7. The improvement index represents the difference between the percentile rank of the average student in the first *Interactive Shared Book Reading* condition and the percentile rank of the average student in the second *Interactive Shared Book Reading* condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the first *Interactive Shared Book Reading* 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 Justice and Ezell (2002), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study. Clustering effects due to the grouping of children may also have been present; however, we do not have sufficient information to conduct a post-hoc clustering correction. Therefore, the statistical significance of findings may be overestimated due to clustering.
9. 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 A6.1 *Interactive Shared Book Reading* rating for the oral language domain

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of oral language, the WWC rated *Interactive Shared Book Reading* as having mixed effects. It did not meet the criteria for positive effects or potentially positive effects because no studies showed statistically significant and positive effects and one study showed a substantively important and negative effect. The remaining ratings (no discernible effects, potentially negative effects, and negative effects) were not considered, as *Interactive Shared Book Reading* was assigned the highest applicable rating.

Rating received

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.

Met. One study had substantively important and positive effects and one study had substantively important and negative effects.

- 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. One study had substantively important and positive effects, one study had substantively important and negative effects, and no studies 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 showed statistically significant and positive effects.

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

Not met. One study had substantively important and 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.

Met. One study had substantively important and positive effects.

- 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. One study had substantively important and negative effects, no studies had indeterminate effects, and one study showed substantively important and positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effects. The WWC also considers the size of the domain level effects for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A6.2 Interactive Shared Book Reading rating for the print knowledge domain

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of print knowledge, the WWC rated *Interactive Shared Book Reading* as having no discernible effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, potentially negative effects, or negative effects, as no studies showed statistically significant or substantively important effects, either positive or negative.

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 studies 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 studies showed statistically significant and positive effects.

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

Met. No studies showed statistically significant or substantively important and 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 studies showed a statistically significant or substantively important and 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.

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

- 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 studies showed a statistically significant or substantively important effect.

(continued)

Appendix A6.2 Interactive Shared Book Reading rating for the print knowledge 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 studies showed a statistically significant or substantively important and negative effect.

- 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 studies showed a statistically significant or a substantively important effect, either positive or negative.

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 studies showed statistically significant and negative effects.

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

Met. No studies showed statistically significant or substantively important and positive effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effects. The WWC also considers the size of the domain level effects for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A6.3 Interactive Shared Book Reading rating for the early reading/writing domain

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of early reading/writing, the WWC rated *Interactive Shared Book Reading* as having potentially positive effects. It did not meet the criteria for positive effects as it had only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered, as *Interactive Shared Book 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. One study showed statistically significant and positive effects.

- 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 and negative, substantively important and negative, or 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. Only one study showed statistically significant and positive effects and this study did not meet WWC evidence standards for a strong design.

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

Met. No studies showed statistically significant or substantively important and negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effects. The WWC also considers the size of the domain level effects for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.