

# Appendix

## Appendix A1.1 Study characteristics: Gettinger, 1986 (randomized controlled trial)

Characteristic	Description
<b>Study citation</b>	Gettinger, M. (1986). Prereading skills and achievement under three approaches to teaching word recognition. <i>Journal of Research and Development in Education</i> , 19(2), 1–9.
<b>Participants</b>	The study began with 122 four- and five-year-old children who were pretested on four measures. Based on pretest scores, 26 children were eliminated because they scored with greater than 75% accuracy on at least one of the pretest measures. The remaining children were blocked on pretest scores, gender, and age and randomly assigned to either the intervention or comparison conditions. The matching procedure resulted in a loss of 24 children, resulting in a final sample of 72 children. Sixty-six percent of the children were white, 22% were black, and 12% were Hispanic or other backgrounds. Forty-four percent of the children were female, and a range of socioeconomic status levels were represented (11% upper middle, 38% middle, 31% lower middle, and 20% lower).
<b>Setting</b>	The study took place in preschools in two Midwestern communities.
<b>Intervention</b>	The children in the intervention group participated in phonological awareness and letter knowledge ( <i>PAT + LK</i> ) skills training. <i>PAT + LK</i> skills training was delivered to children in instructional subgroups, to which they had been randomly assigned, during three 30-minute lessons a week over three consecutive weeks. During each skills training session children were taught to recognize, name, produce the sound for, and blend (initial and final positions) two consonants each day, so that by the end of the intervention all children had been taught 18 consonants. Each lesson followed a similar structure and ended with a cumulative review of all syllables learned. Next, children participated in nine additional 30-minute reading instruction lessons over three consecutive weeks (i.e., three lessons a week for three weeks) to learn how to read using either a sight word approach, a linguistic approach, or a phonetic approach. The reading instruction lessons were designed to assess the effect of <i>PAT + LK</i> skills training on early reading/writing outcomes. The WWC does not report the results of the separate reading instruction conditions in this report because they are not <i>Phonological Awareness Training plus Letter Knowledge Training</i> interventions. However, the WWC does report the results for the early reading/writing outcomes combined across reading instruction conditions because effects on those measures can be attributed to the initial skills training.
<b>Comparison</b>	The children in the comparison group participated in training in skills other than those related to phonological awareness and letter knowledge. The other skills training was delivered to children in instructional subgroups, to which they had been randomly assigned, during three 30-minute lessons a week over three consecutive weeks. During the other skills training, children participated in activities such as practice in color and number naming and picture identification. Next, children participated in nine additional 30-minute reading instruction lessons over three consecutive weeks (i.e., three lessons a week for three weeks) to learn how to read using either a sight word approach, a linguistic approach, or a phonetic approach. The WWC does not report the results of the separate reading instruction conditions in this report because they are not <i>Phonological Awareness Training plus Letter Knowledge Training</i> interventions. However, the WWC does report the results for the early reading/writing outcomes combined across reading instruction conditions because effects on those measures can be attributed to the initial skills training.
<b>Primary outcomes and measurement</b>	The primary outcome domains assessed were print knowledge and phonological processing (measured immediately after the skills training), and early reading/writing (measured after completion of the skills training but during reading instruction training). Print knowledge was assessed with one nonstandardized measure—consonant names. Phonological processing was assessed with three nonstandardized measures—consonant sounds, sound memory, and sound blending. Early reading/writing was assessed by five nonstandardized measures—training words, transfer words, short vowel in transfer words, trials-to-criterion, and discrimination. (See Appendices A2.2–A2.4 for more detailed descriptions of outcome measures.)
<b>Teacher training</b>	The intervention and comparison conditions were implemented by 11 preschool teachers or aides who had at least two years experience with preschool-age children and who received two hours of training involving the reading of prepared scripts and simulated activities for the lessons. All teachers were familiar with the children in the groups they taught.

## Appendix A1.2 Study characteristics: Roberts & Neal, 2004 (randomized controlled trial with attrition problems)

Characteristic	Description
<b>Study citation</b>	Roberts, T., & Neal, H. (2004). Relationships among preschool English language learners' oral proficiency in English, instructional experience and literacy development. <i>Contemporary Educational Psychology, 29</i> (3), 283–311.  <i>Additional source:</i> Roberts, T. A. (2003). Effects of alphabet-letter instruction on young children's word recognition. <i>Journal of Educational Psychology, 95</i> (1), 41–51.
<b>Participants<sup>1</sup></b>	The study began with 43 three- to four-year-old low-income children. During the course of the study, four children moved, one child was excluded because of a high level of missing data, and five children were excluded because English was their primary language. The final sample included 33 Hmong- or Spanish-speaking children. The children ranged in age from 42 to 58 months (mean age = 52.8 months), and 64% were female. The children were blocked by primary language and randomly assigned across morning and afternoon classrooms to either the intervention or comparison conditions.
<b>Setting</b>	The study took place in a half-day, state-funded preschool program.
<b>Intervention</b>	The WWC designated the letter-rhyme group as the intervention condition for this review. The children in this group participated in a total of 48 lessons lasting 20–25 minutes each (three lessons a week for 16 weeks) in small groups that focused on improving children's phonological awareness skills and letter knowledge. Each week, the children were introduced to a new letter in the alphabet, learned to name and write the letter, and used the letter to participate in rhyming activities (e.g., distinguishing rhyming words from nonrhyming words, recognizing rhyme, generating rhyme).
<b>Comparison</b>	The WWC designated the language comprehension group as the comparison condition for this review. The children in the language comprehension condition participated in a total of 48 lessons lasting 20–25 minutes each (three lessons a week for 16 weeks) in small groups. Each week, the children watched a video of a book followed by pretend reading of the book with teacher support (e.g., the teacher responded to children's story-related language and pointing). During subsequent weekly sessions, the children engaged in activities to learn key vocabulary from the text, fingertip reading of the text to promote print awareness, and activities such as acting out the events from the story and putting in order pictures representing events in the story.
<b>Primary outcomes and measurement<sup>2</sup></b>	The primary outcome domains assessed were oral language, print knowledge, phonological processing, and early reading/writing. Oral language was measured by two non-standardized tests—vocabulary and story event sequencing—and a standardized test of English oral language proficiency—the Pre-Idea Proficiency Test. The WWC does not include the Pre-Idea Proficiency Test in this review because it was not intended to assess the effects of the intervention. Print knowledge was assessed by one nonstandardized measure—letter names. Phonological processing was measured by a nonstandardized test of rhyming. Early reading/writing was assessed by one nonstandardized measure of writing. The study also used a nonstandardized test of print concepts; however, it measured elements of both oral language and print knowledge and cannot be appropriately placed in either domain. So, the WWC does not include this measure in the review. (See Appendices A2.1–A2.4 for more detailed descriptions of outcome measures.)
<b>Teacher training</b>	The intervention and comparison conditions were conducted by two undergraduate students who alternated between the letter/rhyme and comprehension conditions every two weeks for 16 weeks. They received about four hours of training and ongoing feedback from the researcher. Initial training included reviewing and discussing the lesson scripts and goals, observing two lessons, and practicing two lessons.

1. The overlap in samples between the two studies was substantial enough to treat this as one study. The children in Roberts (2003) were a subsample from the Roberts and Neal (2004) study. Specifically, Roberts (2003) included 29 children from Roberts and Neal (2004) and four of the five native English speakers from the original sample.
2. The immediate posttests for Roberts (2003) are the same as those used in Roberts and Neal (2004) and are not included in this review. The WWC does not include the measures used following the word-learning training because the purpose of these measures is to test the parameters of learnability (i.e., teaching letters helps children learn the phonetics involved in reading).

## Appendix A1.3 Study characteristics: Pietrangelo, 1999 (quasi-experimental design)

Characteristic	Description
<b>Study citation</b>	Pietrangelo, D. J. (1999). Outcomes of an enhanced literacy curriculum on the emergent literacy skills of Head Start preschoolers. <i>Dissertation Abstracts International</i> , 60(4), 1014A. (UMI No. 9927614).
<b>Participants</b>	The study began with 139 four-year-old low-income children. During the course of the study, 10 children left the study, resulting in a final sample of 129 children from 10 classes. Eighty-three percent of the treatment and comparison children came from English-speaking families, and 17% of the children resided with non-English-speaking families. Twenty-nine percent of the children were black, 22% were Hispanic or Latino, 42% were white, and 7% were Asian (primarily Afghan). Forty-eight percent of the children were female. Classrooms were first matched on half-day or full-day status, and nine of the classrooms were randomly assigned to either the intervention or comparison conditions. The 10th classroom was placed in the comparison group because the intervention materials were not accessible to the teacher. Because the 10th classroom was assigned by convenience, the design for this study is quasi-experimental.
<b>Setting</b>	The study took place in a Head Start program in upstate New York.
<b>Intervention</b>	The children in the intervention group participated in 14 weeks of early literacy instruction designed to supplement the existing Head Start curriculum. <sup>1</sup> The early literacy instruction focused on teaching phonological awareness skills and letter knowledge, such as letter names, letter sounds, and phonemic composition of words. Twenty preschool books were introduced in large and small groups (about six children per small group), and children were encouraged to participate in discussions and to read. Children were also exposed to explicit instruction in letter names and letter sounds, and each lesson plan included phoneme awareness training using game-like activities.
<b>Comparison</b>	The children in the comparison group participated in their regular Head Start curriculum that consisted of few emergent literacy activities and varied book reading activities.
<b>Primary outcomes and measurement</b>	The primary outcome domains assessed were oral language, print knowledge, phonological processing, early reading/writing, and cognition. Oral language was measured with the Peabody Picture Vocabulary Test (PPVT-III), a standardized measure. Print knowledge was assessed with four nonstandardized measures—alphabet knowledge, letter identification, letter-sound correspondence, and print conventions. Phonological processing was assessed with three nonstandardized measures—alliteration, rhyming, and phoneme blending. Early reading/writing was assessed with two nonstandardized measures—invented spelling and word identification. The cognition domain was assessed with two nonstandardized measures—sentence memory and word memory. (See Appendices A2.1–A2.5 for more detailed descriptions of outcome measures.)
<b>Teacher training</b>	Teachers received an orientation packet, participated in one session of training prior to the intervention, and received weekly training once the intervention began. During the weekly training, teachers reviewed lesson plans, and trainers addressed teacher concerns and suggestions and answered teacher questions. Because teachers implemented the intervention in their respective program and groups, they were familiar with the children in the intervention and comparison conditions.

1. Pietrangelo (1999) reported contradictory intervention lengths (14 weeks and 12 weeks). Some of the supplementary materials were drawn from the Ready Readers program.

## Appendix A2.1 Outcome measures in the oral language domain

Outcome measure	Description
<b>Vocabulary</b>	A researcher-developed measure designed to assess children's vocabulary by showing children a series of cards, each with four pictures, and asking them to point to the picture representing the target vocabulary word (as cited in Roberts and Neal, 2004).
<b>Story event sequencing</b>	A researcher-developed measure that requires children to place pictures from a story in the proper sequential order (as cited in Roberts and Neal, 2004).
<b>Peabody Picture Vocabulary Test, 3rd Edition (PPVT-III)</b>	A standardized measure of children's receptive vocabulary that requires them to point to one of four pictures that represents a word that the examiner says (as cited in Pietrangelo, 1999).

## Appendix A2.2 Outcome measures in the print knowledge domain

Outcome measure	Description
<b>Consonant names</b>	A researcher-developed measure that assesses children's knowledge of consonants by requiring them to name each of 18 consonants presented by the researcher (as cited in Gettinger, 1986).
<b>Letter names</b>	A researcher-developed measure designed to assess children's knowledge of letters by presenting children with 16 cards in random order, each with a letter of the alphabet (A through P) printed on it, and asking them to name the letter (as cited in Roberts and Neal, 2004).
<b>Alphabet knowledge</b>	A researcher-developed measure designed to assess children's knowledge of the alphabet by asking children to sing or recite the alphabet (as cited in Pietrangelo, 1999).
<b>Letter identification</b>	A researcher-developed measure designed to assess children's ability to identify letters by presenting each of the 26 letters of the alphabet in a random order, with the exception of the first letter of the child's name, which was presented first to promote success and understanding of the task (as cited in Pietrangelo, 1999).
<b>Letter-sound correspondence</b>	A researcher-developed measure administered at the same time as the letter identification measure in which children were asked to make the sound corresponding with each letter (as cited in Pietrangelo, 1999).
<b>Print conventions</b>	A researcher-developed measure based on Clay's Print Concepts Test designed to test children's knowledge of print conventions by asking them about concepts such as the book cover, from where to start reading, and pointing at words (as cited in Pietrangelo, 1999).

## Appendix A2.3 Outcome measures in the phonological processing domain

Outcome measure	Description
<b>Consonant sounds</b>	A researcher-developed measure that tested the children's ability to produce the sound of 18 consonants presented (as cited in Gettinger, 1986).
<b>Sound memory</b>	A researcher-developed measure that tested the children's ability to produce two sounds (e.g., vowel + consonant) in the same sequence as they had been presented by the examiner (as cited in Gettinger, 1986).
<b>Sound blending</b>	A researcher-developed measure that tested the children's ability to blend together two sounds (e.g., vowel + consonant) that were presented by the examiner to produce the syllable correctly (as cited in Gettinger, 1986).

(continued)

### Appendix A2.3 Outcome measures in the phonological processing domain *(continued)*

Outcome measure	Description
<b>Rhyme</b>	A researcher-developed measure that requires children to listen to 10 monosyllabic words and to produce the rhyming word (as cited in Roberts and Neal, 2004). <sup>1</sup>
<b>Alliteration</b>	A researcher-developed measure that requires children to identify two of three words that start with the same sound (as cited in Pietrangelo, 1999).
<b>Rhyming</b>	A researcher-developed measure that requires children to identify two words that rhyme from a series of three words presented in pictures (as cited in Pietrangelo, 1999).
<b>Phoneme blending</b>	A researcher-developed measure during which the child listens to segments of words and is asked to blend them together (as cited in Pietrangelo, 1999).

1. Roberts and Neal (2004) stated that there was a floor effect for this measure.

### Appendix A2.4 Outcome measures in the early reading/writing domain

Outcome measure	Description
<b>Training words</b>	A researcher-developed measure that requires children to read one of 10 words that they were trained on during the weekly lessons (as cited in Gettinger, 1986).
<b>Transfer words</b>	A researcher-developed measure that requires children to read one of 10 transfer words where six had the same initial bigrams as the training words and four had the same media vowel used in the training words (as cited in Gettinger, 1986).
<b>Short vowel in transfer words</b>	A researcher-developed measure that assesses the number of times the child produced the short vowel sound accurately in the transfer words, regardless of whether the child said the entire word correctly (as cited in Gettinger, 1986).
<b>Trials-to-criterion</b>	A researcher-developed measure that assesses children's learning rate by determining how many trials were needed to learn a word (as cited in Gettinger, 1986). For the purposes of this review and to make effect size estimates consistent across measures, the WWC reversed the direction of the effect so that a higher score reflected a better outcome.
<b>Discrimination</b>	A researcher-developed measure that tested children's ability to accurately identify training words from nonsense miscues (as cited in Gettinger, 1986).
<b>Writing</b>	A researcher-developed measure that requires children to write their names, seven letters of the alphabet, three words that are dictated to them, and any additional words that they wish to write (as cited in Roberts and Neal, 2004).
<b>Letter writing portion of the writing task</b>	A portion of the writing measure described above during which children are asked to write seven letters of the alphabet (as cited in Roberts and Neal, 2004).
<b>Invented spelling</b>	A researcher-developed measure that requires children to write five words that are said to them individually and in the context of a sentence (as cited in Pietrangelo, 1999).
<b>Word identification</b>	A researcher-developed measure that requires children to pronounce each of 20 isolated words that were presented in print (as cited in Pietrangelo, 1999).

## Appendix A2.5 Outcome measures in the cognition domain

Outcome measure	Description
<b>Sentence memory</b>	A researcher-developed measure that assesses children's memory by asking children to recall verbatim sentences ranging from three to 12 words each (as cited in Pietrangelo, 1999).
<b>Word memory</b>	A researcher-developed measure that assess children's memory by presenting children with the same words from the sentence memory test (described above) and asking children to recall in the original order presented the words in each string (as cited in Pietrangelo, 1999).

## Appendix A3.1 Summary of study findings included in the rating for the oral language domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> (PAT + LK – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			PAT + LK group	Comparison group				
<b>Roberts &amp; Neal, 2004 (randomized controlled trial with attrition problems)<sup>7</sup></b>								
Vocabulary	3–4 year olds	33	18.93 (5.32)	24.21 (5.06)	–5.28	–1.00	Statistically significant	–34
Story event sequencing	3–4 year olds	33	4.07 (3.83)	5.89 (3.84)	–1.82	–0.46	ns	–18
<b>Average<sup>8</sup> for oral language (Roberts &amp; Neal, 2004)</b>						–0.73	Statistically significant	–27
<b>Pietrangelo, 1999 (quasi-experimental design)<sup>9</sup></b>								
PPVT-III	4 year olds	10/129	93.78 (12.99)	92.50 (14.28)	1.28	0.09	ns	+4
<b>Average<sup>8</sup> for oral language (Pietrangelo, 1999)</b>						0.09	ns	+4
<b>Domain average<sup>8</sup> for oral language across all studies</b>						–0.32	na	–12

PAT + LK = Phonological Awareness Training plus Letter Knowledge Training

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 A4.1.
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 the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that 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 Roberts and Neal (2004), a correction for multiple comparisons was needed, but the significance levels do not 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 sizes.
9. In the case of Pietrangelo (1999), a correction for clustering was needed, but the significance level does not differ from that reported in the original study.

## Appendix A3.2 Summary of study findings included in the rating for the print knowledge domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (PAT + LK – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			PAT + LK group <sup>3</sup>	Comparison group <sup>3</sup>				
<b>Gettinger, 1986 (randomized controlled trial)<sup>8</sup></b>								
Consonant names	4–5 year olds	72	14.03 (2.87)	9.75 (3.72)	4.28	1.27	Statistically significant	+40
<b>Average<sup>9</sup> for print knowledge (Gettinger, 1986)</b>						1.27	Statistically significant	+40
<b>Roberts &amp; Neal, 2004 (randomized controlled trial with attrition problems)<sup>10</sup></b>								
Letter names	3–4 year olds	33	11.14 (5.30)	6.74 (5.24)	4.40	0.82	Statistically significant	+29
<b>Average<sup>9</sup> for print knowledge (Roberts &amp; Neal, 2004)</b>						0.82	Statistically significant	+29
<b>Pietrangelo, 1999 (quasi-experimental design)<sup>11</sup></b>								
Alphabet knowledge	4 year olds	10/129	24.29 (5.75)	23.68 (4.55)	0.61	0.12	ns	+5
Letter identification	4 year olds	10/129	12.25 (8.25)	11.45 (9.83)	0.80	0.09	ns	+4
Letter-sound correspondence	4 year olds	10/129	1.97 (2.87)	1.61 (4.34)	0.36	0.10	ns	+4
Print conventions	4 year olds	10/129	5.96 (2.64)	5.45 (2.61)	0.51	0.19	ns	+8
<b>Average<sup>9</sup> for print knowledge (Pietrangelo, 1999)</b>						0.12	ns	+5
<b>Domain average<sup>9</sup> for print knowledge across all studies</b>						0.74	na	+27

PAT + LK = Phonological Awareness Training plus Letter Knowledge Training

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.
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. For Pietrangelo (1999), the intervention group mean equals the comparison group mean plus the mean difference.

(continued)

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

4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Pietrangelo (1999), the mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effect when the intervention group had higher pretest scores than 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 and that 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.
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 Gettinger (1986), no corrections for clustering or multiple comparisons were needed.
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 sizes.
10. In the case of Roberts and Neal (2004), no corrections for clustering or multiple comparisons were needed.
11. In the case of Pietrangelo (1999), a correction for clustering was needed, but the significance levels do not differ from those reported in the original study.

## Appendix A3.3 Summary of study findings included in the rating for the phonological processing domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (PAT + LK – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			PAT + LK group <sup>3</sup>	Comparison group <sup>3</sup>				
<b>Gettinger, 1986 (randomized controlled trial)<sup>8</sup></b>								
Consonant sounds	4–5 year olds	72	11.31 (2.88)	3.53 (3.12)	7.78	2.56	Statistically significant	+49
Sound memory	4–5 year olds	72	28.93 (7.20)	23.66 (7.83)	5.27	0.69	Statistically significant	+26
Sound blending	4–5 year olds	72	20.12 (3.89)	7.53 (3.98)	12.59	3.16	Statistically significant	+50
<b>Average<sup>9</sup> for phonological processing (Gettinger, 1986)</b>						1.92	Statistically significant	+47
<b>Roberts &amp; Neal, 2004 (randomized controlled trial with attrition problems)<sup>10</sup></b>								
Rhyme	3–4 year olds	33	0.64 (0.84)	0.63 (0.54)	0.01	0.01	ns	+1
<b>Average<sup>9</sup> for phonological processing (Roberts &amp; Neal, 2004)</b>						0.01	ns	+1
<b>Pietrangelo, 1999 (quasi-experimental design)<sup>11</sup></b>								
Alliteration	4 year olds	10/129	5.02 (1.91)	4.31 (1.78)	0.71	0.38	ns	+15
Rhyming	4 year olds	10/129	7.10 (3.06)	5.94 (3.03)	1.16	0.38	ns	+15
Phoneme blending	4 year olds	10/129	6.56 (3.15)	5.23 (2.45)	1.33	0.47	ns	+18
<b>Average<sup>9</sup> for phonological processing (Pietrangelo, 1999)</b>						0.41	ns	+16
<b>Domain average<sup>9</sup> for phonological processing across all studies</b>						0.78	na	+28

PAT + LK = Phonological Awareness Training plus Letter Knowledge Training

ns = not statistically significant

na = not applicable

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. For Pietrangelo (1999), the intervention group mean equals the comparison group mean plus the mean difference.

(continued)

### Appendix A3.3 Summary of study findings included in the rating for the phonological processing domain *(continued)*

4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Pietrangelo (1999), the mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effect when the intervention group had higher pretest scores than 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 and that 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.
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 Gettinger (1986), a correction for multiple comparisons was needed, but the significance levels do not 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 sizes.
10. In the case of Roberts and Neal (2004), no corrections for clustering or multiple comparisons were needed.
11. In the case of Pietrangelo (1999), a correction for clustering was needed, so the significance levels may differ from those reported in the original study.

## Appendix A3.4 Summary of study findings included in the rating for the early reading/writing domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Author's findings from the study			WWC calculations		
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (PAT + LK – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			PAT + LK group <sup>3</sup>	Comparison group <sup>3</sup>				
<b>Gettinger, 1986 (randomized controlled trial)<sup>8</sup></b>								
Training words	4–5 year olds	72	23.67 (3.91)	18.48 (4.31)	5.19	1.25	Statistically significant	+39
Transfer words	4–5 year olds	72	15.51 (3.60)	11.73 (4.39)	3.78	0.93	Statistically significant	+32
Short vowel in transfer words	4–5 year olds	72	16.77 (3.92)	12.60 (3.86)	4.17	1.06	Statistically significant	+36
Trials-to-criterion	4–5 year olds	72	1.27 (0.63)	1.63 (0.70)	0.36	0.53	Statistically significant	+20
Discrimination	4–5 year olds	72	67.95 (10.88)	60.87 (10.34)	7.08	0.66	ns <sup>9</sup>	+25
<b>Average<sup>10</sup> for early reading/writing (Gettinger, 1986)</b>						0.89	Statistically significant	+31
<b>Roberts &amp; Neal, 2004 (randomized controlled trial with attrition problems)<sup>11</sup></b>								
Writing	3–4 year olds	33	13.10 (4.31)	10.81 (4.27)	2.29	0.52	ns	+20
<b>Average<sup>8</sup> for early reading/writing (Roberts &amp; Neal, 2004)</b>						0.52	ns	+20
<b>Pietrangelo, 1999 (quasi-experimental design)<sup>12</sup></b>								
Invented spelling	4 year olds	10/129	0.45 (0.98)	0.28 (0.86)	0.17	0.18	ns	+7
Word identification	4 year olds	10/129	0.63 (0.72)	0.69 (1.18)	–0.06	–0.06	ns	–2
<b>Average<sup>8</sup> for early reading/writing (Pietrangelo, 1999)</b>						0.06	ns	+2
<b>Domain average<sup>8</sup> for early reading/writing across all studies</b>						0.49	na	+19

PAT + LK = Phonological Awareness Training plus Letter Knowledge Training

ns = not statistically significant

na = not applicable

(continued)

## Appendix A3.4 Summary of study findings included in the rating for the early reading/writing domain *(continued)*

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices. Subgroup and subscale findings from the same studies are not included in these ratings, but are reported in Appendices A4.3 and A4.4.
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. For Pietrangelo (1999), the intervention group mean equals the comparison group mean plus the mean difference.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group. For Pietrangelo (1999), the mean differences were computed by the WWC and took into account the pretest difference between the study groups. The resulting effect sizes may overestimate the intervention's effect when the intervention group had lower pretest scores than the comparison group and underestimate the intervention's effect when the intervention group had higher pretest scores than the comparison group. For Gettinger (1986), the scores represent the posttest means across the three teaching approaches (sight, linguistic, and phonetic) that children were exposed to during phase two training to assess the effects of *PAT + LK* skills training during completion of reading instruction. To provide consistency in direction of effect size estimates across measures, the WWC reversed the direction of the effect for the trials-to-criterion measure so that a higher score reflected a better outcome.
5. For an explanation of the effect size calculation, see [Technical Details of WWC-Conducted Computations](#).
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that 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.
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 Gettinger (1986), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
9. Although the WWC calculations for this measure were statistically significant, the author reported nonstatistically significant findings for this measure.
10. 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 sizes.
11. In the case of Roberts and Neal (2004), no corrections for clustering or multiple comparisons were needed.
12. In the case of Pietrangelo (1999), a correction for clustering was needed, but the significance levels do not differ from those reported in the original study.

## Appendix A3.5 Summary of study findings included in the rating for the cognition domain<sup>1</sup>

Outcome measure	Study sample	Sample size (classrooms/ children)	Author's findings from the study			WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> (PAT + LK – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>	
			PAT + LK group	Comparison group					
<b>Pietrangelo, 1999 (quasi-experimental design)<sup>7</sup></b>									
Sentence memory	4 year olds	10/129	10.24 (2.67)	10.11 (2.94)	0.13	0.05	ns	+2	
Word memory	4 year olds	10/129	4.37 (2.28)	3.98 (2.32)	0.39	0.17	ns	+7	
<b>Domain average<sup>8</sup> for cognition</b>						0.11	ns	+4	

*PAT + LK = Phonological Awareness Training plus Letter Knowledge Training*

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. 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 and that 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 Pietrangelo (1999), a correction for clustering was needed, but the significance levels do not differ from those reported in the original study.
8. 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 A4.1 Summary of subgroup findings for the oral language domain<sup>1</sup>

Outcome measure	Study sample	Sample size (children) <sup>3</sup>	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (PAT + LK – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			PAT + LK group	Comparison group				
<b>Roberts &amp; Neal, 2004 (randomized controlled trial with attrition problems; level A English proficiency children)<sup>8</sup></b>								
Vocabulary	3–4 year olds	13	14.17 (3.37)	20.83 (3.97)	–6.66	–1.67	Statistically significant	–45

*PAT + LK = Phonological Awareness Training plus Letter Knowledge Training*

1. This appendix presents subgroup findings for measures that fall in the oral language domain. Total group scores were used for rating purposes and are presented in Appendix A3.1. Level A English proficiency (i.e., being able to understand and communicate in English) refers to those children who began the intervention with the lowest possible level of English proficiency.
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 number of children in the subgroup was provided by the study author.
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 and that 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.
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 Roberts and Neal (2004), no correction for clustering was needed.

## Appendix A4.2 Summary of subgroup findings for the print knowledge domain<sup>1</sup>

Outcome measure	Study sample	Sample size (children) <sup>3</sup>	Author's findings from the study					
			Mean outcome (standard deviation <sup>2</sup> )		WWC calculations			
			PAT + LK group	Comparison group	Mean difference <sup>4</sup> (PAT + LK – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
<b>Roberts &amp; Neal, 2004 (randomized controlled trial with attrition problems; level A English proficiency children)<sup>8</sup></b>								
Letter names	3–4 year olds	13	9.17 (5.85)	2.43 (1.51)	6.74	1.53	Statistically significant	+44

*PAT + LK = Phonological Awareness Training plus Letter Knowledge Training*

1. This appendix presents subgroup findings for measures that fall in the print knowledge domain. Total group scores were used for rating purposes and are presented in Appendix A3.2.
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 number of children in the subgroup was provided by the study author.
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 and that 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.
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 Roberts and Neal (2004), no correction for clustering was needed.

## Appendix A4.3 Summary of subgroup findings for the early reading/writing domain<sup>1</sup>

Outcome measure	Study sample	Sample size (children) <sup>3</sup>	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>4</sup> (PAT + LK – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			PAT + LK group	Comparison group				
<b>Roberts &amp; Neal, 2004 (randomized controlled trial with attrition problems; level A English proficiency children)<sup>8</sup></b>								
Writing	3–4 year olds	13	10.83 (4.55)	7.29 (3.30)	3.54	0.84	ns	+30

*PAT + LK = Phonological Awareness Training plus Letter Knowledge Training*

ns = not statistically significant

1. This appendix presents subgroup findings for measures that fall in the early reading/writing domain. Total group scores were used for rating purposes and are presented in Appendix A3.4.
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 number of children in the subgroup was provided by the study author.
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 and that 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.
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 Roberts and Neal (2004), no correction for clustering was needed.

## Appendix A4.4 Summary of subscale findings for the early reading/writing domain<sup>1</sup>

Outcome measure	Study sample	Sample size (children)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation <sup>2</sup> )		Mean difference <sup>3</sup> (PAT + LK – comparison)	Effect size <sup>4</sup>	Statistical significance <sup>5</sup> (at $\alpha = 0.05$ )	Improvement index <sup>6</sup>
			PAT + LK group	Comparison group				
<b>Roberts &amp; Neal, 2004 (randomized controlled trial with attrition problems)<sup>7</sup></b>								
Letter writing	3–4 year olds	33	4.89 (1.96)	3.10 (2.69)	1.79	0.72	Statistically significant	+27

*PAT + LK = Phonological Awareness Training plus Letter Knowledge Training*

1. This appendix presents subscale findings for a measure that fell in the early reading/writing domain. Total scale scores were used for rating purposes and are presented in Appendix A3.4.
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 and that 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 Roberts and Neal (2004), no correction for clustering was needed.

## Appendix A5.1 Phonological Awareness Training plus Letter Knowledge Training 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.<sup>1</sup>

For the outcome domain of oral language, the WWC rated *Phonological Awareness Training plus Letter Knowledge Training* as having potentially negative effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, or no discernible effects because one study showed a statistically significant and negative effect and one study showed an indeterminate effect. The rating of negative effects was not considered because *Phonological Awareness Training plus Letter Knowledge Training* was assigned the highest applicable rating.

### Rating received

**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.

**Met.** One of the two studies produced a negative effect that was statistically significant based on the authors' report and the WWC re-analysis.

- 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 substantively important positive effect in this domain.

### 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.** None of the studies produced a statistically significant positive effect in this domain.

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

**Not met.** One study had a statistically significant negative effect in this domain.

**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.** None of the studies produced statistically significant or substantively important positive effects in this domain.

- 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 a statistically significant negative effect, and one study had an indeterminate effect in this domain.

(continued)

## Appendix A5.1 Phonological Awareness Training plus Letter Knowledge Training rating for the oral language domain (continued)

**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 positive effect, and one study showed a statistically significant negative effect in this domain.

- 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 showed a statistically significant negative effect, and one study showed an indeterminate effect in this domain.

**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*.

**Not met.** One study showed a statistically significant negative effect in this domain.

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 A5.2 Phonological Awareness Training plus Letter Knowledge Training 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.<sup>1</sup>

For the outcome domain of print knowledge, the WWC rated *Phonological Awareness Training plus Letter Knowledge Training* as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Phonological Awareness Training plus Letter Knowledge Training* was assigned the highest applicable rating.

### Rating received

**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.

**Met.** Two of the three studies produced positive effects that were statistically significant based on the author's report. One of these two studies met WWC evidence standards for a strong design.

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

**Met.** No studies had statistically significant or substantively important negative effects in this domain.

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 A5.3 Phonological Awareness Training plus Letter Knowledge Training rating for the phonological processing 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.<sup>1</sup>

For the outcome domain of phonological processing, the WWC rated *Phonological Awareness Training plus Letter Knowledge Training* as having potentially positive effects. It did not meet the criteria for positive effects because only one study showed statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Phonological Awareness Training plus Letter Knowledge Training* 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 of the three studies showed a statistically significant positive effect, one showed a substantively important positive effect, and one showed an indeterminate effect based on the authors' reports and the WWC re-analysis.

- 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, and the number of studies showing indeterminate effects (one) is not greater than the number of studies showing statistically significant or substantively important positive effects (two) in this domain.

### 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 of the three studies showed a statistically significant positive effect in this domain.

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

**Met.** No studies showed statistically significant or substantively important negative effects in this domain.

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 A5.4 Phonological Awareness Training plus Letter Knowledge Training 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.<sup>1</sup>

For the outcome domain of early reading/writing, the WWC rated *Phonological Awareness Training plus Letter Knowledge Training* as having potentially positive effects. It did not meet the criteria for positive effects because only one study showed statistically significant positive effects. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Phonological Awareness Training plus Letter Knowledge Training* 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 a statistically significant positive effect, and one study showed a substantively important positive effect in this domain.

- 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 or substantively important negative effects, and fewer studies showed indeterminate effects (one) than statistically significant or substantively important positive effects (two) in this domain.

### 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 significantly positive effects in this domain.

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

**Met.** No studies showed statistically significant or substantively important negative effects in this domain.

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 A5.5 Phonological Awareness Training plus Letter Knowledge Training rating for the cognition 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.<sup>1</sup>

For the outcome domain of cognition, the WWC rated *Phonological Awareness Training plus Letter Knowledge Training* 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 because 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 statistically significant or substantively important effects, either positive or negative. One study showed an indeterminate effect in this domain.

### 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 significantly positive effects in this domain.

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

**Met.** No studies had statistically significant or substantively important negative effects in this domain.

**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 statistically significant or substantively important positive effects in this domain.

- 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 had statistically significant or substantively important positive effects and one study showed indeterminate effects in this domain.

**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 statistically significant or substantively important effects, either positive or negative, in this domain.

- 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 had statistically significant or substantively important effects. One study showed indeterminate effects in this domain.

(continued)

## Appendix A5.5 Phonological Awareness Training plus Letter Knowledge Training rating for the cognition 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 statistically significant or substantively important negative effects in this domain.

- 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 statistically significant or substantively important positive effects in this domain.

**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 negative effects in this domain.

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

**Met.** No studies showed statistically significant or substantively important positive effects in this domain.

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.