

Appendix

Appendix A1 Study Characteristics: Torgesen et al., 2001 (randomized controlled trial)

Characteristic	Description
Study citation	Torgesen, J. K., Alexander, A. W., Wagner, R. K., Rashotte, C. A., Voeller, K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. <i>Journal of Learning Disabilities, 34</i> (1), 33.
Participants	The sample for this study included a total of 60 students between eight and ten years of age. All participants had been identified by school staff as having learning disabilities based on a discrepancy of at least one standard deviation between their scores on a standardized test of reading and their full-scale score on an intelligence test. Each year for three years, the researchers worked with staff from the three elementary schools to select a sample of students with learning disabilities (20 students per year) who met the following criteria: (1) they were identified by their teachers as having serious difficulty in acquiring word-level reading skills, (2) their average standard score on two measures of word-level reading was at least 1.5 standard deviations below average, (3) their estimated verbal intelligence was above 75, and (4) they performed below minimum required levels for their grade on a measure of phonological awareness. The 60 children eligible for the study were randomly assigned to one of two groups (30 per group): <i>LiPS</i> [®] or <i>Embedded Phonics</i> (an instructional program developed by the authors). The interventions were provided to students in two eight-week phases: an intensive phase and an application/generalization phase. Pretest data were collected two to three weeks prior to the start of the interventions, posttest data were collected two to three weeks after completion of the eight-week intensive intervention period, and follow-up data were collected one year and two years following the posttest. Ten children were not available for second year follow-up data collection, leaving 50 students in the analysis sample—26 students that received <i>LiPS</i> [®] and 24 students that received <i>Embedded Phonics</i> . Children in the study had the following characteristics: all were 4th graders, 72% were male, 65% were Caucasian and 35% were African-American, and the average full-scale IQ for students in the study was about 96. Additional findings reflecting students' outcomes at the first and second year follow-ups can be found in Appendices A4.1–4.10.
Setting	The study was conducted with students with learning disabilities from three public elementary schools in Florida. Students were between eight and ten years of age and enrolled in the 4th grade.
Intervention	This study involved a comparison of the effectiveness of two interventions: (1) <i>LiPS</i> [®] and (2) <i>Embedded Phonics</i> . Both of the interventions provided explicit instruction in word-level reading skills; they differed in method of teaching and in relative amount of time spent on instructional activities. <i>LiPS</i> [®] was provided to students in two phases. In the first phase of the intervention, intensive instruction was delivered on a 1:1 basis for two 50-minute sessions, five days a week, for eight weeks, until a total of 67.5 hours of instruction had been provided. During this time of intensive instruction, <i>LiPS</i> [®] substituted for time the students would normally have spent in their learning disability resource room. In the second, less intensive phase of <i>LiPS</i> [®] , students received one 50-minute lesson per week for eight additional weeks in their learning disability resource room, applying skills they had learned during the intensive phase to regular classroom materials. <i>LiPS</i> [®] placed primary emphasis on building skills in phonemic awareness and phonemic decoding with individual words. <i>LiPS</i> [®] has three goals: (1) to provide a basis for accurate discriminations among phonemes by teaching the distinctive kinesthetic, auditory, and visual (mouth form pictures) features associated with all the common phonemes of the English language; kinesthetic and visual features are taught to help make the phoneme more concrete, and to allow children to both hear and feel phonemic contrasts and identities in spoken patterns; (2) to teach children to use their knowledge of the distinctive features of phonemes to monitor and represent sequences of sounds in spoken syllables; and (3) to use problem-solving activities to teach children self-monitoring skills.

(continued)

Appendix A1 Study Characteristics: Torgesen et al., 2001 (randomized controlled trial) (continued)

Characteristic	Description
Comparison	Students in the comparison group participated in a competing intervention, developed by the study authors, called <i>Embedded Phonics</i> . This intervention was delivered for the same amount of time as <i>LiPS</i> [®] and also taught phonemic awareness and phonemic decoding. However, the <i>Embedded Phonics</i> instruction emphasized application through reading meaningful text and recognizing, practicing, and spelling high-frequency sight words.
Primary outcomes and measurement	The authors assessed students with a battery of tests at the pretest, posttest, one-year follow-up, and two-year follow-up time points. In the domain of alphabets, phonological awareness was measured by administration of the Lindamood Auditory Conceptualization Test and the Elision subtest, the Non-word Repetition subtest, and the Rapid Letter Naming subtest from the Comprehensive Test of Phonological Processes (CTOPP); word attack was measured by administration of the Word Attack subtest of the Woodcock Reading Mastery Test–Revised (WRMT–R) and the Phonemic Decoding and Sight Word Efficiency subtests of the Test of the Word Reading Efficiency (TOWRE); and letter-word identification was measured by administration of the Word Identification subtest of the WRMT–R. The domain of reading fluency was measured by administration of the Reading Accuracy and Reading Rate subtests of the Gray Oral Reading Test–III (GORT–III). The domain of reading comprehension was measured by administration of the Passage Comprehension subtest of the WRMT–R and the Reading Comprehension subtest of the GORT–III. The domain of writing was measured by administration of the Spelling subtest of the Kaufman Test of Educational Achievement. The domain of math was measured by administration of the Calculation subtest from the Woodcock-Johnson Psycho-Educational Battery–Revised (WJ–R). Other outcomes were reported in the study but were not included in this report because they were outside the scope of the Students with Learning Disabilities review. For a more detailed description of these outcome measures, see Appendices A2.1–2.5.
Staff/teacher training	The teachers who administered each program all had at least one year’s experience teaching children with reading disabilities using that method or one very similar to it. The teachers who taught <i>LiPS</i> [®] were all drawn from those working at a clinic where the program had been used for the previous five years.

Appendix A2.1 Outcome measures for the alphabetic domain

Outcome measure	Description
<i>Phonological awareness construct</i>	
Lindamood Auditory Conceptualization Test (LAC)	The LAC is an individually administered, norm-referenced assessment that measures a child's ability to perceive and conceptualize speech sounds using a visual medium (as cited in Torgesen et al., 2001).
Elision subtest from the Comprehensive Test of Phonological Processes (CTOPP)	The CTOPP Elision subtest measures the child's ability to manipulate sounds in words (as cited in Torgesen et al., 2001).
Non-word Repetition subtest from the CTOPP	The CTOPP Non-word Repetition subtest measures the ability of a child to combine sounds that are presented orally and to put the separate sounds together to form a nonsense word (as cited in Torgesen et al., 2001).
Rapid Letter Naming subtest from the CTOPP	The CTOPP Rapid Letter Naming subtest measures rate of access to phonological information in long-term memory (as cited in Torgesen et al., 2001).
<i>Word attack construct</i>	
Word Attack subtest from the Woodcock Reading Mastery Test-Revised (WRMT-R)	The WRMT-R Word Attack subtest measures phonemic decoding skills by asking students to read pseudowords. Students are aware that the words are not real (as cited in Torgesen et al., 2001).
Phonemic Decoding subtest from the Test of Word Reading Efficiency (TOWRE)	The TOWRE Phonetic Efficiency Decoding subtest measures the number of pronounceable non-printed words that can be accurately decoded within 45 seconds (as cited in Torgesen et al., 2001).
Sight Word Efficiency subtest from the TOWRE	The TOWRE Sight Word Efficiency subtest assesses the number of real printed words that can be accurately identified within 45 seconds (as cited in Torgesen et al., 2001).
<i>Letter-word identification construct</i>	
Word Identification subtest from the WRMT-R	The WRMT-R Word Identification subtest measures basic word reading skills and requires the child to read aloud isolated words that range in frequency and difficulty (as cited in Torgesen et al., 2001).

Appendix A2.2 Outcome measures for the reading fluency domain

Outcome measure	Description
Reading Accuracy subtest from the Gray Oral Reading Test-III (GORT-III)	The GORT-III Reading Accuracy subtest measures the number of word reading errors that occurred while reading a series of short paragraphs that increase in difficulty (as cited in Torgesen et al., 2001).
Reading Rate subtest from the GORT-III	The GORT-III Reading Rate subtest measures the amount of time taken to read short paragraphs that increase in difficulty (as cited in Torgesen et al., 2001).

Appendix A2.3 Outcome measures for the reading comprehension domain

Outcome measure	Description
Passage Comprehension subtest from the WRMT–R	The WRMT–R Passage Comprehension subtest asks students to read silently and fill in missing words in a short paragraph (as cited in Torgesen et al., 2001).
Reading Comprehension subtest from the GORT–III	The GORT–III Reading Comprehension subtest asks students to read paragraphs and answer five comprehension questions for each paragraph. The questions are read to students by the tester (as cited in Torgesen et al., 2001).

Appendix A2.4 Outcome measures for the writing domain

Outcome measure	Description
Spelling subtest from the Kaufman Test of Educational Achievement (KTEA)	The KTEA Spelling subtest is a norm-referenced test consisting of 50 items (as cited in Torgesen et al., 2001).

Appendix A2.5 Outcome measures for the math domain

Outcome measure	Description
Calculation subtest from the Woodcock-Johnson Psycho-Educational Battery–Revised (WJ–R)	The WJ–R Calculation subtest measures the subject’s skill in performing mathematical calculations including addition, subtraction, multiplication, and division. The calculations involve decimals, fractions, and whole numbers (as cited in Torgesen et al., 2001).

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

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS [®] –comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS [®] group ³	Comparison group				
Torgesen et al., 2001⁸								
LAC	Grade 4	50	82.3 (10.1)	69.0 (17.3)	13.3	0.93	Statistically significant	+32
CTOPP Elision subtest	Grade 4	50	96.4 (14.3)	97.9 (12.8)	-1.5	-0.11	ns	-4
CTOPP Non-word Repetition subtest	Grade 4	50	102.2 (16.1)	103.0 (19.0)	-0.8	-0.04	ns	-2
CTOPP Rapid Letter Naming subtest	Grade 4	50	93.3 (10.3)	93.1 (11.5)	0.2	0.02	ns	+1
WRMT–R Word Attack subtest	Grade 4	50	98.0 (7.0)	90.3 (8.3)	7.7	0.99	Statistically significant	+34
TOWRE Phonemic Decoding subtest	Grade 4	50	84.7 (4.8)	83.7 (5.8)	1.0	0.19	ns	+7
TOWRE Sight Word Efficiency subtest	Grade 4	50	72.1 (5.8)	72.7 (7.8)	-0.6	-0.09	ns	-3
WRMT–R Word Identification subtest	Grade 4	50	79.9 (11.2)	80.5 (9.6)	-0.6	-0.06	ns	-2
Domain average for alphabets⁹						0.23	ns	+9

ns = not statistically significant

LAC = Lindamood Auditory Conceptualization Test

CTOPP = Comprehensive Test of Phonological Processing

WRMT–R = Woodcock Reading Mastery Test–Revised

TOWRE = Test of Word Reading Efficiency

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the alphabets domain. One- and two-year follow-up findings from Torgesen et al. (2001) are not included in these ratings, but are reported in Appendices A4.1 and A4.6.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.

(continued)

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain¹ *(continued)*

5. For an explanation of the effect size calculation, see WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
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 average improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®-comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
		LiPS® group ³	Comparison group					
Torgesen et al., 2001⁸								
GORT-III Reading Accuracy subtest	Grade 4	50	93.1 (12.4)	87.5 (13.4)	5.6	0.43	ns	+17
GORT-III Reading Rate subtest	Grade 4	50	75.6 (8.2)	72.1 (7.9)	3.5	0.43	ns	+17
Domain average for reading fluency⁹						0.43	ns	+17

ns = not statistically significant

GORT-III = Gray Oral Reading Test-III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the reading fluency domain. One- and two-year follow-up findings from Torgesen et al. (2001) are not included in these ratings, but are reported in Appendices A4.2 and A4.7.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.
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 average improvement index is calculated from the average effect size.

Appendix A3.3 Summary of study findings included in the rating for the reading comprehension domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®-comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
		LiPS® group ³	Comparison group					
Torgesen et al., 2001⁸								
WRMT-R Passage Comprehension subtest	Grade 4	50	90.2 (9.0)	92.0 (19.8)	-1.8	-0.12	ns	-5
GORT-III Reading Comprehension subtest	Grade 4	50	91.7 (10.0)	86.0 (10.4)	5.7	0.55	ns	+21
Domain average for reading comprehension⁹						0.21	ns	+8

ns = not statistically significant

WRMT-R = Woodcock Reading Mastery Test-Revised

GORT-III = Gray Oral Reading Test-III

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the reading comprehension domain. One- and two-year follow-up findings from Torgesen et al. (2001) are not included in these ratings, but are reported in Appendices A4.3 and A4.8.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.
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 average improvement index is calculated from the average effect size.

Appendix A3.4 Summary of study findings included in the rating for the writing domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®-comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS® group ³	Comparison group				
Torgesen et al., 2001⁸								
KTEA Spelling subtest	Grade 4	50	76.3 (5.0)	80.0 (7.4)	-3.7	-0.58	Statistically significant	-22
Domain average for writing						-0.58	Statistically significant	-22

KTEA = Kaufman Test of Educational Achievement

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the writing domain. One- and two-year follow-up findings from Torgesen et al. (2001) are not included in these ratings, but are reported in Appendices A4.4 and A4.9.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A3.5 Summary of study findings included in the rating for the math domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®–comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS® group ³	Comparison group				
Torgesen et al., 2001⁸								
WJ–R Calculation subtest	Grade 4	50	90.9 (14.1)	86.9 (11.5)	4.0	0.30	ns	+12
Domain average for math						0.30	ns	+12

ns = not statistically significant

WJ–R = Woodcock–Johnson Psycho-Educational Battery–Revised

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the math domain. One- and two-year follow-up findings from Torgesen et al. (2001) are not included in these ratings, but are reported in Appendices A4.5 and A4.10.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A4.1 Summary of one-year follow-up findings for the alphabetic domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations				
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®–comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷	
			LiPS® group ³	Comparison group					
Torgesen et al., 2001⁸									
LAC	Grade 4	50	75.4 (14.0)	72.0 (13.9)	3.4	0.24	ns	+9	
CTOPP Elision subtest	Grade 4	50	93.3 (13.7)	94.4 (14.3)	–1.1	–0.08	ns	–3	
CTOPP Non-word Repetition subtest	Grade 4	50	109.8 (15.6)	108.2 (17.4)	1.6	0.10	ns	+4	
CTOPP Rapid Letter Naming subtest	Grade 4	50	90.0 (10.6)	92.4 (10.2)	–2.4	–0.23	ns	–9	
WRMT–R Word Attack subtest	Grade 4	50	92.3 (9.3)	87.0 (8.9)	5.3	0.57	ns	+22	
TOWRE Phonemic Decoding subtest	Grade 4	50	83.0 (7.4)	80.6 (9.1)	2.4	0.29	ns	+11	
TOWRE Sight Word Efficiency subtest	Grade 4	50	76.9 (6.4)	74.4 (9.6)	2.5	0.30	ns	+12	
WRMT–R Word Identification subtest	Grade 4	50	80.2 (9.6)	78.2 (11.3)	2.0	0.19	ns	+7	

ns = not statistically significant

LAC = Lindamood Auditory Conceptualization Test

CTOPP = Comprehensive Test of Phonological Processes

WRMT–R = Woodcock Reading Mastery Test–Revised

TOWRE = Test of Word Reading Efficiency

1. This appendix presents one-year follow-up findings for measures that fall in the alphabetic domain. Posttest 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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.

(continued)

Appendix A4.1 Summary of one-year follow-up findings for the alphabetic domain¹ *(continued)*

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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.

Appendix A4.2 Summary of one-year follow-up findings for the reading fluency domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS [®] –comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS [®] group ³	Comparison group				
Torgesen et al., 2001⁸								
GORT–III Reading Accuracy subtest	Grade 4	50	97.4 (12.3)	90.8 (14.8)	6.6	0.48	ns	+18
GORT–III Reading Rate subtest	Grade 4	50	75.2 (9.3)	72.1 (13.2)	3.1	0.27	ns	+11

ns = not statistically significant

GORT–III = Gray Oral Reading Test–III

1. This appendix presents one-year follow-up findings for measures that fall in the reading fluency domain. Posttest 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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A4.3 Summary of one-year follow-up findings for the reading comprehension domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS [®] –comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
		LiPS [®] group ³	Comparison group					
Torgesen et al., 2001⁸								
WRMT–R Passage Comprehension subtest	Grade 4	50	92.0 (8.0)	91.5 (10.8)	0.5	0.05	ns	+2
GORT–III Reading Comprehension subtest	Grade 4	50	96.3 (10.0)	88.1 (12.2)	8.2	0.73	Statistically significant	+27

ns = not statistically significant

WRMT–R = Woodcock Reading Mastery Test–Revised

GORT–III = Gray Oral Reading Test–III

1. This appendix presents one-year follow-up findings for measures that fall in the reading comprehension domain. Posttest scores were used for rating purposes and are presented in Appendix A3.3.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.

Appendix A4.4 Summary of one-year follow-up findings for the writing domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®–comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS® group ³	Comparison group				
Torgesen et al., 2001⁸								
KTEA Spelling subtest	Grade 4	50	75.5 (5.8)	74.0 (6.6)	1.5	0.24	ns	+9

ns = not statistically significant

KTEA = Kaufman Test of Educational Achievement

1. This appendix presents one-year follow-up findings for measures that fall in the writing domain. Posttest 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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A4.5 Summary of one-year follow-up findings for the math domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®–comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS® group ³	Comparison group				
Torgesen et al., 2001⁸								
WJ–R Calculation subtest	Grade 4	50	83.6 (13.5)	86.5 (10.1)	–2.9	–0.24	ns	–9

ns = not statistically significant

WJ–R = Woodcock–Johnson Psycho–Educational Battery–Revised

1. This appendix presents one-year follow-up findings for measures that fall in the math domain. Posttest scores were used for rating purposes and are presented in Appendix A3.5.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A4.6

Summary of two-year follow-up findings for the alphabetic domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations				
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS [®] –comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷	
			LiPS [®] group ³	Comparison group					
Torgesen et al., 2001⁸									
LAC	Grade 4	50	75.3 (14.1)	76.2 (13.4)	–0.9	–0.06	ns	–3	
CTOPP Elision subtest	Grade 4	50	93.3 (13.9)	98.8 (13.6)	–5.5	–0.39	ns	–15	
CTOPP Non-word Repetition subtest	Grade 4	50	114.3 (13.7)	112.8 (14.0)	1.5	0.11	ns	+4	
CTOPP Rapid Letter Naming subtest	Grade 4	50	93.5 (11.3)	93.3 (12.4)	0.2	0.02	ns	+1	
WRMT–R Word Attack subtest	Grade 4	50	93.4 (12.5)	89.9 (10.4)	3.5	0.30	ns	+12	
TOWRE Phonemic Decoding subtest	Grade 4	50	85.7 (7.5)	82.7 (10.7)	3.0	0.32	ns	+13	
TOWRE Sight Word Efficiency subtest	Grade 4	50	79.7 (6.5)	77.8 (9.5)	1.9	0.23	ns	+9	
WRMT–R Word Identification subtest	Grade 4	50	84.5 (12.1)	83.9 (12.2)	0.6	0.05	ns	+2	

ns = not statistically significant

LAC = Lindamood Auditory Conceptualization Test

CTOPP = Comprehensive Test of Phonological Processes

WRMT–R = Woodcock Reading Mastery Test–Revised

TOWRE = Test of Word Reading Efficiency

1. This appendix presents two-year follow-up findings for measures that fall in the alphabetic domain. Posttest 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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.

(continued)

Appendix A4.6 Summary of two-year follow-up findings for the alphabetics domain¹ *(continued)*

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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A4.7 Summary of two-year follow-up findings for the reading fluency domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS [®] –comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS [®] group ³	Comparison group				
Torgesen et al., 2001⁸								
GORT–III Reading Accuracy subtest	Grade 4	50	95.0 (15.5)	90.4 (14.7)	4.6	0.30	ns	+12
GORT–III Reading Rate subtest	Grade 4	50	72.9 (9.5)	70.7 (12.9)	2.2	0.19	ns	+8

ns = not statistically significant

GORT–III = Gray Oral Reading Test–III

1. This appendix presents two-year follow-up findings for measures that fall in the reading fluency domain. Posttest 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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A4.8 Summary of two-year follow-up findings for the reading comprehension domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®–comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS® group ³	Comparison group				
Torgesen et al., 2001⁸								
WRMT–R Passage Comprehension subtest	Grade 4	50	93.9 (8.9)	96.9 (11.5)	–3.0	–0.29	ns	–11
GORT–III Reading Comprehension subtest	Grade 4	50	94.0 (11.8)	87.2 (15.1)	6.8	0.50	ns	+19

ns = not statistically significant

WRMT–R = Woodcock Reading Mastery Test–Revised

GORT–III = Gray Oral Reading Test–III

1. This appendix presents two-year follow-up findings for measures that fall in the reading comprehension domain. Posttest scores were used for rating purposes and are presented in Appendix A3.3.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A4.9 Summary of two-year follow-up findings for the writing domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS [®] –comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS [®] group ³	Comparison group				
Torgesen et al., 2001⁸								
KTEA Spelling subtest	Grade 4	50	75.0 (6.7)	75.3 (6.0)	–0.3	–0.05	ns	–2

ns = not statistically significant

KTEA = Kaufman Test of Educational Achievement

1. This appendix presents two-year follow-up findings for measures that fall in the writing domain. Posttest 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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A4.10 Summary of two-year follow-up findings for the math domain¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ⁴ (LiPS®–comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			LiPS® group ³	Comparison group				
Torgesen et al., 2001⁸								
WJ–R Calculation subtest	Grade 4	50	85.5 (12.4)	89.4 (11.1)	–3.9	–0.33	ns	–13

ns = not statistically significant

WJ–R = Woodcock–Johnson Psycho–Educational Battery–Revised

1. This appendix presents two-year follow-up findings for measures that fall in the math domain. Posttest scores were used for rating purposes and are presented in Appendix A3.5.
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. Each intervention group mean is calculated as the unadjusted control mean plus the WWC-adjusted mean difference. Standard deviations are unadjusted.
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 WWC Procedures and Standards Handbook, Appendix B.
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 results favorable to the intervention group.
8. The level of statistical significance was reported by the study authors or, when 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. For the formulas the WWC used to calculate the statistical significance, see WWC Procedures and Standards Handbook, Appendix C for clustering and WWC Procedures and Standards Handbook, Appendix D for multiple comparisons. In the case of Torgesen et al. (2001), no corrections for clustering or multiple comparisons were needed.

Appendix A5.1 *LiPS*[®] rating for the alphabets domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of alphabets, the WWC rated *LiPS*[®] as having potentially positive effects for students with learning disabilities. The remaining ratings (mixed effects, indeterminate effects, potentially negative effects, negative effects) were not considered, as *LiPS*[®] 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 of *LiPS*[®] showed a statistically significant positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No study of *LiPS*[®] showed a statistically significant or substantively important negative effect, nor did any study show an indeterminate effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

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

Not met. One study of *LiPS*[®] had a strong design and showed a statistically significant positive effect.

AND

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

Met. No study of *LiPS*[®] showed a statistically significant or substantively important negative effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A5.2 *LiPS*[®] rating for the reading fluency domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of reading fluency, the WWC rated *LiPS*[®] as having potentially positive effects on students with learning disabilities. The remaining ratings (mixed effects, indeterminate effects, potentially negative effects, negative effects) were not considered, as *LiPS*[®] 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 of *LiPS*[®] showed a substantively important positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No study of *LiPS*[®] showed a statistically significant or substantively important negative effect, nor did any study show an indeterminate effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

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

Not met. No study of *LiPS*[®] showed a statistically significant positive effect.

AND

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

Met. No study of *LiPS*[®] showed a statistically significant or substantively important negative effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A5.3 **LiPS®** rating for the reading comprehension domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹ For the outcome domain of reading comprehension, the WWC rated *LiPS®* as having no discernible effects on students with learning disabilities.

Rating received

No discernible effects: No affirmative evidence of effects.

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

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

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

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

Not met. No study of *LiPS®* showed a statistically significant positive effect.

AND

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

Met. No study of *LiPS®* showed a statistically significant or substantively important negative effect.

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

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

Not met. No study of *LiPS®* showed a statistically significant or substantively important positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Not met. No study of *LiPS®* showed a statistically significant or substantively important negative effect, but one study showed an indeterminate effect.

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 of *LiPS®* showed a statistically significant or substantively important effect, either positive or negative.

OR

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

Not met. No study of *LiPS®* showed a statistically significant or substantively important effect, either positive or negative, but one study showed an indeterminate effect.

(continued)

Appendix A5.3 *LiPS*[®] rating for the reading comprehension domain (continued)

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

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

Not met. No study of *LiPS*[®] showed a statistically significant or substantively important negative effect.

AND

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

Met. No study of *LiPS*[®] showed a statistically significant or 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 study of *LiPS*[®] showed a statistically significant negative effect.

AND

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

Not met. No study of *LiPS*[®] showed a statistically significant or substantively important positive effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A5.4 *LiPS*[®] rating for the writing domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of writing, the WWC rated *LiPS*[®] as having potentially negative effects on students with learning disabilities. The remaining rating (negative effects) was not considered, as *LiPS*[®] 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 study of *LiPS*[®] showed a statistically significant negative effect.

AND

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

Met. No study of *LiPS*[®] showed a statistically significant or substantively important positive effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

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

Not met. No study of *LiPS*[®] showed a statistically significant positive effect.

AND

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

Not met. One study of *LiPS*[®] showed a statistically significant negative effect.

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

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

Not met. No study of *LiPS*[®] showed a statistically significant or substantively important positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Not met. One study of *LiPS*[®] showed a statistically significant negative effect. No study of *LiPS*[®] showed an indeterminate effect.

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.

(continued)

Appendix A5.4 *LiPS*[®] rating for the writing domain (continued)

Not met. One study of *LiPS*[®] showed a statistically significant negative effect. No study of *LiPS*[®] showed a statistically significant or substantively important positive effect.

OR

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

Not met. One study of *LiPS*[®] showed a statistically significant negative effect. No study of *LiPS*[®] showed an indeterminate effect.

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 of *LiPS*[®] showed a statistically significant negative effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A5.5 *LiPS*[®] rating for the math domain

The WWC rates an intervention's effects for a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of math, the WWC rated *LiPS*[®] as having potentially positive effects on students with learning disabilities. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *LiPS*[®] 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 of *LiPS*[®] showed a substantively important positive effect.

AND

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. No study of *LiPS*[®] showed a statistically significant or substantively important negative effect, nor did any study show an indeterminate effect.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

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

Not met. No study of *LiPS*[®] showed a statistically significant positive effect.

AND

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

Met. No study of *LiPS*[®] showed a statistically significant or substantively important negative effect, nor did any study show an indeterminate effect.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain-level effect. The WWC also considers the size of the domain-level effect for ratings of potentially positive or potentially negative effects. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

Appendix A6 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Alphabetics	1	3	50	Small
Reading fluency	1	3	50	Small
Reading comprehension	1	3	50	Small
Writing	1	3	50	Small
Math	1	3	50	Small
General reading achievement	0	na	na	na
Science	0	na	na	na
Social studies	0	na	na	na
Progressing in school	0	na	na	na

na = not applicable/not studied

1. A rating of “medium to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.” For more details on the extent of evidence categorization, see the WWC Procedures and Standards Handbook, Appendix G.