

Appendix

Appendix A1 Study Characteristics: Torgeson, Wagner, Rashotte, & Herron, 2003 (randomized clinical trial)

Characteristic	Description
Study citation	Torgesen, J., Wagner, R., Rashotte, C., & Herron, J. (2003). <i>Summary of outcomes from first grade study with Read, Write and Type and Auditory Discrimination in Depth instruction and software with at-risk children</i> (FCRR Tech. Rep. No. 2). Retrieved from Florida Center for Reading Research website: http://www.fcrr.org/TechnicalReports/RWTfullrept.pdf .
Participants	The study included 150 first-grade students in five elementary schools. All students scored in the lowest 35% on a letter-sound knowledge measure. At two of the schools, 16 students were randomly assigned to <i>LiPS</i> [®] and 16 to <i>Read, Write and Type!</i> [™] (<i>RWT</i>). At three schools, 38 students were randomly assigned to <i>LiPS</i> [®] , 38 to <i>RWT</i> , and 42 to a control group (J. K. Torgesen, personal communication, September 7, 2006). Two students left the <i>LiPS</i> [®] and <i>RWT</i> groups, and one student left the control group. The final sample for the analysis comparing <i>LiPS</i> [®] to <i>RWT</i> included 52 <i>LiPS</i> [®] students and 53 <i>RWT</i> students across five schools. The final sample for the analysis comparing <i>LiPS</i> [®] to control students included 36 <i>LiPS</i> [®] students and 41 control students across three schools. Approximately 34% of the sample were minority children (primarily African-American). Approximately 35% of the sample received free/reduced-price lunch, but the socioeconomic status of the students varied.
Setting	Five elementary schools (locations unknown).
Intervention	Students assigned to the <i>LiPS</i> [®] program were divided into groups of three children and received four 50-minute sessions a week from October through May. A trained teacher devoted half of each session to direct instruction. The remainder of the time the students worked individually on the computer practicing the same skills with the teacher in a support role.
Comparison	<i>RWT</i> students had the same format and time of instruction as did the <i>LiPS</i> [®] students, but they had different activities. <i>RWT</i> teachers began their sessions with warm-up activities, and students then worked on computers, with teachers lending support (particularly when students had difficulties). The computer component emphasized phonological awareness, letter sound correspondence, and phonemic decoding as students expressed themselves in written language. The control group had classroom instruction and support typically available to them (J. K. Torgesen, personal communication, September 7, 2006). Two of the three schools with regular instruction comparison groups used Open Court's <i>Collections for Young Scholars</i> as the whole-class reading curriculum.
Primary outcomes and measurement	The authors assessed students at the end of the study period using a battery of tests. All students in the sample were given the Phoneme Blending, Phoneme Elision, and Phoneme Segmenting subtests of the Comprehensive Test of Phonological Processes and the Word Attack, Word Identification, and Passage Comprehension subtests of the Woodcock Reading Mastery Test. Students in the study were also given the Vocabulary subtest of the Stanford Binet Intelligence Scale, which the authors used as a proxy for verbal IQ. Other outcomes were reported in the study but were not included in this review either because they were outside the scope of the beginning reading review (developmental spelling and probability of reading disability) or because sufficient information on the measure name, description, or validity and reliability was not reported (word efficiency and nonword efficiency). For a more detailed description of these outcome measures, see Appendices A2.1 and A2.2.
Staff/teacher training	No information was provided on teacher training.

Appendix A2.1 Outcome measures for the alphabetic domain

Outcome measure	Description
<i>Phonological awareness</i>	
Comprehensive Test of Phonological Processes (CTOPP): Phoneme Blending subtest	The Phoneme Blending subtest measures the student's ability to blend separately presented sounds together to form words. This is a standardized test (as cited in Torgesen et al., 2003).
CTOPP: Phoneme Elision subtest	The Phoneme Elision subtest measures the student's ability to manipulate sounds in words. This is a standardized test (as cited in Torgesen et al., 2003).
CTOPP: Phoneme Segmenting subtest	The Phoneme Segmenting subtest measures the student's ability to isolate and pronounce the sounds in words. This is a standardized test (as cited in Torgesen et al., 2003).
<i>Phonics</i>	
Woodcock Reading Mastery Test: Word Identification subtest	The Word Identification subtest is a measure of word reading vocabulary in which the student reads list of words of increasing difficulty. This is a standardized test (as cited in Torgesen et al., 2003).
Woodcock Reading Mastery Test: Word Attack subtest	The Word Attack subtest is a measure of phonemic reading ability in which the student reads nonwords. This is a standardized test (as cited in Torgesen et al., 2003).

Appendix A2.2 Outcome measures for the comprehension domain

Outcome measure	Description
<i>Comprehension</i>	
Woodcock Reading Mastery Test: Passage Comprehension subtest	The Passage Comprehension subtest measures the student's ability to comprehend the meaning of short passages. This is a standardized test (as cited in Torgesen et al., 2003).
<i>Vocabulary</i>	
Stanford Binet Intelligence Scale: Vocabulary subtest	The measure is based on the Vocabulary subtest of the Stanford Binet Intelligence Scale. The Vocabulary subtest measures the student's ability to provide names of pictures and definitions of words. This is a standardized test (as cited in Torgesen et al., 2003).

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

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) ²		Mean difference ³ (LiPS [®] - comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
		LiPS [®] group	Comparison group					
Torgeson, Wagner, Rashotte, & Herron, 2003 (randomized controlled trial)⁷								
Phonological Awareness								
Comparison #1: LiPS[®] vs. Read, Write, and Type								
CTOPP: Phoneme Blending subtest	Grade 1	104	18.80 (5.30)	18.90 (4.90)	-0.1	-0.02	ns	-1
CTOPP: Phoneme Elision subtest	Grade 1	104	14.30 (4.50)	13.50 (4.50)	0.8	0.18	ns	+7
CTOPP: Phoneme Segmentation subtest	Grade 1	104	16.20 (6.60)	15.30 (5.30)	0.8	0.15	ns	+6
Comparison #2: LiPS[®] vs. regular instruction/support								
CTOPP: Phoneme Blending subtest	Grade 1	77	20.60 (4.50)	18.20 (5.40)	2.4	0.48	ns	+18
CTOPP: Phoneme Elision subtest	Grade 1	77	15.30 (4.20)	12.50 (4.60)	2.8	0.63	Statistically significant	+23
CTOPP: Phoneme Segmentation subtest	Grade 1	77	15.60 (3.70)	11.70 (4.50)	3.9	0.93	Statistically significant	+32
Phonics								
Comparison #1: LiPS[®] vs. Read, Write, and Type								
Woodcock Reading Mastery Test: Word Attack subtest	Grade 1	104	109.70 (14.00)	106.30 (13.60)	3.4	0.24	ns	+10
Woodcock Reading Mastery Test: Word Identification subtest	Grade 1	104	107.10 (14.30)	105.10 (13.40)	2.0	0.14	ns	+6

(continued)

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

Outcome measure	Authors' findings from the study							
	Study sample	Sample size (students)	Mean outcome (standard deviation) ²		WWC calculations			
			LiPS [®] group	Comparison group	Mean difference ³ (LiPS [®] - comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
Torgeson, Wagner, Rashotte, & Herron, 2003 (randomized controlled trial)⁷								
Comparison #2: LiPS[®] vs. regular instruction/support								
Woodcock Reading Mastery Test: Word Attack subtest	Grade 1	77	113.70 (12.20)	99.50 (14.50)	14.2	1.04	Statistically significant	+35
Woodcock Reading Mastery Test: Word Identification subtest	Grade 1	77	110.60 (12.20)	100.10 (15.60)	10.5	0.74	Statistically significant	+27
Average for alphabetics, Comparison #1 (Torgeson et al., 2003)⁸						0.14	ns	+6
Average for alphabetics, Comparison #2 (Torgeson et al., 2003)⁸						0.76	Statistically significant	+28
Domain average for alphabetics across comparisons (Torgeson et al., 2003)⁸						0.45	Statistically significant	+17

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the alphabetic domain.
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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Torgeson et al. (2003), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

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

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			<i>LiPS</i> [®] group	Comparison group	Mean outcome (standard deviation) ²	Mean difference ³ (<i>LiPS</i> [®] -comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)
Torgeson, Wagner, Rashotte, & Herron, 2003 (randomized controlled trial)⁷								
Reading comprehension								
Comparison #1: <i>LiPS</i>[®] vs. Read, Write, and Type								
Woodcock Reading Mastery Test: Passage Comprehension subtest	Grade 1	104	99.90 (12.50)	99.30 (10.50)	0.6	0.05	ns	+2
Comparison #2: <i>LiPS</i>[®] vs. regular instruction/support								
Woodcock Reading Mastery Test: Passage Comprehension subtest	Grade 1	77	102.20 (10.00)	95.40 (14.40)	6.8	0.54	ns	+20
Vocabulary								
Comparison #1: <i>LiPS</i>[®] vs. Read, Write, Type								
Stanford Binet Intelligence Scale: Vocabulary subtest	Grade 1	104	95.50	95.50	0.0	0.00	ns	0
Comparison #2: <i>LiPS</i>[®] vs. regular instruction/support								
Stanford Binet Intelligence Scale: Vocabulary subtest	Grade 1	77	96.10 (12.50)	95.90 (11.30)	0.2	0.02	ns	+1
Average for comprehension, Comparison #1 (Torgeson et al., 2003)⁸						0.03	ns	+1
Average for comprehension, Comparison #2 (Torgeson et al., 2003)⁸						0.28	ns	+11
Domain average for comprehension across comparisons (Torgeson et al., 2003)⁸						0.15	ns	+6

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the comprehension domain.
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. For the formulas the WWC used to calculate statistical significance, see Technical Details of WWC-Conducted Computations. In the case of Torgeson et al. (2003), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
8. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4.1 *LiPS*[®] rating for the alphabets domain

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

For the outcome domain of alphabets, the WWC rated *LiPS*[®] as having potentially positive effects. It did not meet the criteria for positive effects because only one study met WWC evidence standards. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and 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 comparison within one study showed statistically significant positive effects.

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 studies showed statistically significant or substantively important negative effects or indeterminate effects.

Other ratings considered

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

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

Not Met. Only one study met WWC evidence standards for a strong design.

AND

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

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

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

Appendix A4.2 *LiPS*[®] rating for the 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 comprehension, the WWC rated *LiPS*[®] as having no discernible effects. The remaining ratings (potentially negative effects and negative effects) were not considered, as *LiPS*[®] was assigned the highest applicable rating.

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 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. Only one study met the WWC evidence standards for a strong design.

AND

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

Met. No study showed statistically significant or substantively important negative effects.

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

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

Not met. No study 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 showed a statistically significant or substantively important negative effect, but one study showed indeterminate effects.

Mixed effects: Evidence of inconsistent effects as demonstrated through EITHER of the following.

- 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 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 showed a statistically significant or substantively important effect, while one study showed indeterminate effects.

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

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Alphabetics	1	5	146	Small
Fluency	0	0	0	na
Comprehension	1	5	146	Small
General reading achievement	0	0	0	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.”