

Appendix

Appendix A1 Study characteristics: Torgesen, Wagner, Rashotte, & Herron, 2003 (randomized controlled 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 Web site: 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>Auditory Discrimination in Depth</i> [®] (ADD) and 16 to <i>Read, Write and Type</i> [™] (RWT). At three schools, 38 students were randomly assigned to ADD, 38 to RWT, and 42 to a control group (J.K. Torgesen, personal communication, September 7, 2006). Two students left the ADD and RWT groups, and 1 student left the control group. The final sample for the analysis comparing ADD to RWT included 52 ADD students and 53 RWT students across five schools. The final sample for the analysis comparing ADD to control students included 36 ADD 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 lunch, but students ranged in their socio-economic status.
Setting	Five elementary schools (locations unknown).
Intervention	Students assigned to the ADD 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	RWT students received the same format and time of instruction as the ADD students, but the type of activities students engaged in differed. RWT teachers began their sessions with warm up activities, and students then spent the remaining time working on computers with the teachers lending support, particularly when children ran into specific difficulties. The computer program emphasizes phonological awareness, letter sound correspondence, and phonemic decoding as children express themselves in written language. The control group continued using the 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 children 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 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 non-word efficiency). (See Appendices A2.1 and A2.2 for more detailed descriptions of outcome measures.)
Teacher training	No information was provided on teacher training.

Appendix A2.1 Outcome measures in the alphabetic domain

Outcome measure	Description
Phonological awareness	
Comprehensive Test of Phonological Processes (CTOPP): Phoneme Blending Subtest	The phoneme blending subtest measures the child's ability to blend separately presented sounds together to form words. This is a standardized test (as cited in Torgesen et al., 2003).
Comprehensive Test of Phonological Processes (CTOPP): Phoneme Elision Subtest	The phoneme elision subtest measures the child's ability to manipulate sounds in words. This is a standardized test (as cited in Torgesen et al., 2003).
Comprehensive Test of Phonological Processes (CTOPP): Phoneme Segmenting Subtest	The phoneme segmenting subtest measures the child's ability to isolate and pronounce the sounds in words. This is a standardized test (as cited in Torgesen et al., 2003).
Phonics	
Woodcock Reading Mastery Test: Word Identification Subtest	The word identification subtest is a measure of word reading vocabulary in which the child 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 child reads non-words. This is a standardized test (as cited in Torgesen et al., 2003).

Appendix A2.2 Outcome measures in the comprehension domain

Outcome measure	Description
Comprehension	
Woodcock Reading Mastery Test: Passage Comprehension Subtest	The passage comprehension subtest measures the child's ability to comprehend the meaning of short passages. This is a standardized test (as cited in Torgesen et al., 2003).
Vocabulary	
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 child'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 Alphabetic domain: Summary of findings by construct¹

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (ADD – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Auditory Discrimination in Depth [®] group	Comparison group				
Construct: Phonological awareness								
Torgesen et al., 2003 (randomized controlled trial)⁷								
Comparison #1: Auditory Discrimination in Depth[®] vs. Read, Write and Type![™] group								
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 Segmenting Subtest	Grade 1	104	16.20 (6.60)	15.30 (5.30)	0.9	0.15	ns	+6
Comparison #2: Auditory Discrimination in Depth[®] vs. regular instruction/support group								
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 Segmenting Subtest	Grade 1	77	15.60 (3.70)	11.70 (4.50)	3.9	0.93	Statistically significant	+32
Construct: Phonics								
Torgesen et al., 2003 (randomized controlled trial)								
Comparison #1: Auditory Discrimination in Depth[®] vs. Read, Write and Type![™] group								
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 **Alphabetic domain: Summary of findings by construct** *(continued)*

Outcome measure	Study sample	Sample size (students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (ADD – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Auditory Discrimination in Depth [®] group	Comparison group				
Comparison #2: Auditory Discrimination in Depth[®] vs. regular instruction/support group								
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
Averages⁸								
Average for alphabetics, Comparison #1 (Torgesen et al., 2003)						0.14	ns	+6
Average for alphabetics, Comparison #2 (Torgesen et al., 2003)						0.76	Statistically significant	+28
Domain average for alphabetic domain across comparisons (Torgesen et al., 2003)						0.45	Statistically significant	+17

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 the 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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.
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 Torgesen et al. (2003), corrections for multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
8. The WWC-computed average effect sizes for each comparison and for the domain across comparisons 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			
			Mean outcome (standard deviation ²)		Mean difference ³ (ADD – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			Auditory Discrimination in Depth [®] group	Comparison group				
Construct: Reading comprehension								
Torgesen et al., 2003 (randomized controlled trial)⁷								
Comparison #1: Auditory Discrimination in Depth[®] vs. Read, Write and Type![™] group								
Woodcock Reading Mastery Test: Passage Comprehension Subtest	Grade 1	104	99.90 (12.50)	99.30 (10.50)	0.60	0.05	ns	+2
Comparison #2: Auditory Discrimination in Depth[®] vs. regular instruction/support group								
Woodcock Reading Mastery Test: Passage Comprehension Subtest	Grade 1	77	102.20 (10.00)	95.40 (14.40)	6.8	0.54	ns	+20
Construct: Vocabulary								
Torgesen et al., 2003 (randomized controlled trial)								
Comparison #1: Auditory Discrimination in Depth[®] vs. Read, Write and Type![™] group								
Stanford Binet Intelligence Scale: Vocabulary Subtest	Grade 1	104	95.50 ⁸	95.50 ⁸	0.0	0.0	ns	0
Comparison #2: Auditory Discrimination in Depth[®] vs. regular instruction/support group								
Stanford Binet Intelligence Scale: Vocabulary Subtest	Grade 1	77	96.10 (12.50)	95.90 (11.30)	0.2	0.02	ns	+1
Averages⁹								
Average for comprehension, Comparison #1 (Torgesen et al., 2003)						0.03	ns	+1
Average for comprehension, Comparison #2 (Torgesen et al., 2003)						0.28	ns	+11
Domain average for comprehension across comparisons (Torgesen et al., 2003)						0.15	ns	+6

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices.

(continued)

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

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 the 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 versus the percentile rank of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting results favorable to the intervention group.
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 Torgesen et al. (2003), corrections for multiple comparisons were needed, so the significance levels may differ from those reported in the original study.
8. The authors did not present standard deviations for this outcome; however, assuming that there were positive standard deviations, the zero mean difference between comparison and treatment groups generates a zero effect size.
9. The WWC-computed average effect sizes for each comparison and for the domain across comparisons are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect sizes.

Appendix A4.1 Auditory Discrimination in Depth® 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 *Auditory Discrimination in Depth*® as having potentially positive effects. It did not meet the criteria for positive effects because only one study met WWC evidence standards. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *Auditory Discrimination in Depth*® 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.

- 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.

- 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. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A4.2 *Auditory Discrimination in Depth*[®] rating for the comprehension 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 comprehension, the WWC rated *Auditory Discrimination in Depth*[®] as having no discernible effects. It did not meet the criteria for other ratings (positive effects, potentially positive effects, mixed effects, potentially negative effects, and negative effects) because the one study that met WWC standards did not show statistically significant or substantively important effects.

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.

- 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.

- 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 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 effect, either positive or negative.

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

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

(continued)

Appendix A4.2 *Auditory Discrimination in Depth*[®] rating for the 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 showed a statistically significant or substantively important negative effect.

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

Met. No study showed a statistically significant or substantively important positive effect.

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 showed a statistically significant or substantively important negative effect.

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

Met. No study showed statistically significant or substantively important positive 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. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence ¹
		Schools	Students	
Alphabetics	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 “moderate to large” requires at least two studies and two schools across studies in one domain and a total sample size across studies of at least 350 students or 14 classrooms. Otherwise, the rating is “small.”