Appendix

Appendix A1.1   Study characteristics: Cognitive Concepts, 2003 (randomized controlled trial)

Characteristic	Description
Study citation	Cognitive Concepts, Inc. (2003). *Outcomes report: Los Angeles Unified School District, California.* Retrieved from http://www.cogcon.com/research/proven/LAUSD.pdf
Participants	Nineteen teachers identified students in Kindergarten through third grade with reading difficulties. More than 80% of students were English language learners. The study author administered pretests (ORAL-J and Test of Memory and Learning (TOMAL)) to students to divide them into two similar groups.[1] The groups were then randomly assigned to be either the intervention or comparison group. Each group originally had 43 students, but there was some attrition due to poor attendance.[2] In the analysis sample, 39 students were in the intervention group and 35 students were in the comparison group.
Setting	The study took place in one elementary school in Los Angeles, California.
Intervention	Students in the intervention group were given directions on how to use *Earobics*® software. They received instruction with *Earobics*® for 30 minutes a day, five days a week over three months. In addition, the intervention group received the regular whole class reading instruction with the *Open Court Reading* curriculum.
Comparison	Students in the comparison classes received the regular whole class reading instruction with the *Open Court Reading* curriculum during the language arts period.
Primary outcomes and measurement	For both pre- and posttests, the authors administered six subtests of the ORAL-J Early Literacy Achievement test: Blending into Words, Segmenting into Sounds, Rhyming Words, Letter Naming, and Sound of Letters subtests as well as three administrations of the Words per Minute subtest.[3] The Test of Memory and Learning (TOMAL) was also used in the study, but was not included in this review because it was outside the scope of the Beginning Reading review (see Appendices A2.1–2.2 for more detailed descriptions of outcome measures).
Teacher training	No information on teaching training is provided. The *Earobics*® group worked with minimal teacher supervision in a computer lab.

1. Equivalence of the two groups at pretest was confirmed through data sent by the author, M. Poblanz.
2. Some information about attrition was provided through personal communication with the author.
3. Some of the test data was not in the published report and was provided directly by the author.

Appendix A1.2   Study characteristics: Valliath, 2002 (quasi-experimental design)

Characteristic	Description
Study citation	Valliath, S. (2002). An evaluation of a computer-based phonological awareness training program: Effects on phonological awareness, reading and spelling. *Dissertation Abstracts International, 63*(04), 1291A. (UMI No. 3050601)
Participants	Ten teachers each identified three children with the lowest reading ability within their first-grade classrooms. The study author administered a pretest (the Woodcock Reading Mastery Test: Word Identification subtest) to students to divide them into two similar groups.[1] All children came from English-speaking monolingual homes; none received any special education or speech and language services. The sample consisted of 16 boys and 14 girls, ranging in age from 6.5 to 7.5 years. In the analysis sample, 15 students were in the intervention group and 15 students were in the comparison group.
Setting	The study took place in three public elementary schools from a high-achieving school district in a northwest suburb of Chicago, Illinois.
Intervention	Students in the intervention group spent 20 minutes each day playing one of the six *Earobics*® games. *Earobics*® delivered phonological awareness training in the auditory mode and provided minimal sound-to-print training. The children played individually and were provided headsets. They started at the lowest skill level for each game and progressed at their own pace. The games were rotated systematically on a daily basis during the ten-week training program. The average number of days attended by the students in intervention group was 46.47 of a possible 50 days.
Comparison	Students in the comparison classes received comparable amounts of daily exposure (approximately 20 minutes) to math training software, *Knowledge Adventure's Jump Start Math for First Graders*. The software has no linguistic training component and consists of eight math games appropriate for children in the first grade. The average number of days attended by the students in comparison group was 45.8 of a possible 50 days.
Primary outcomes and measurement	For both pre- and posttests, the authors administered the Comprehensive Test of Phonological Processing (CTOPP): Blending Words, Blending Non-words, Elision, and Sound Matching subtests and the Woodcock Reading Mastery Test: Word Identification and Word Attack subtests. The CTOPP Memory for Digits subtest and the Spelling subtest of the Wide Range Achievement Test were also used in the study but have not been included because they were outside the scope of the Beginning Reading review (see Appendices A2.1–2.2 for more detailed descriptions of outcome measures).
Teacher training	The experimenter trained the computer lab technicians in each of the three schools on how to use the software. Detailed instructions, attendance sheets, and appropriate rotations of the *Earobics*® games were discussed. No other information on teaching training is provided.

1. The pretest also confirmed that students' performance was low-average.

Appendix A2.1   Outcome measures in the alphabetics domain by construct

Outcome measure	Description
Phonological awareness	
Comprehensive Test of Phonological Processing (CTOPP): Elision subtest	This subtest measures the ability of a child to manipulate the components of a word. The examinee is prompted to say a compound word (for example, "cowboy") and then to say the word without the first part "cow", or to say the word without a specific sound (such as "f" in "farm" – "arm") (as cited in Valliath, 2002).
CTOPP: Blending Words subtest	This subtest measures the ability of a child to combine separately spoken sounds and put them together to form a real word (as cited in Valliath, 2002).
CTOPP: Sound Matching subtest	This subtest measures the ability of a child to choose the word that contains a target sound. Words are presented orally and the subject is shown a card containing pictures of the four words. The child must indicate which word contains the sound. The target sound is tested both in the initial and final position in the word (as cited in Valliath, 2002).
ORAL-J: Early Literacy Achievement Blending into Words subtest	This task requires children to combine or blend the separate sounds of a word to say the word. For example, child is given sounds like /k/ /a/ /t/ and has to say "cat" (as cited in author communication[1]).
ORAL-J: Early Literacy Achievement Segmenting into Sounds subtest	This task requires students to segment words into sounds. The child is given a word and has to give individual sounds (as cited in author's communication[1]).
ORAL-J: Early Literacy Achievement: Rhyming Words subtest	This task requires students to generate words that rhyme. The child is given a word and has to supply a word that rhymes (as cited in author's communication[1]).
CTOPP: Blending Non-Words subtest	This subtest measures the ability of a child to combine sounds that are presented orally and put the separate sounds together to form a nonsense word (as cited in Valliath, 2002).
Letter knowledge	
ORAL-J: Early Literacy Achievement: Letter Naming subtest	Children get a card with 100 letters and are to name each one. The subtest score is determined by how many letters they name in a minute (as cited in author's communication[1]).
Phonics	
ORAL-J: Early Literacy Achievement: Sound of Letters subtest	Children name the sound of letters on a card with 59 letters. The subtest score is determined by how many sounds they name in a minute (as cited in author's communication[1]).
Woodcock Reading Mastery Test (WRMT): Word Identification subtest	This subtest of the standardized WRMT measures basic word reading skills and requires the child to read aloud isolated words that range in frequency and difficulty (as cited in Valliath, 2002).
WRMT: Word Attack subtest	The Word Attack subtest of the WRMT measures the child's ability to apply phonic and structural analysis skills to pronounce unfamiliar words. Subjects cannot read the pseudowords by sight and must rely on phonological processes to decode them (as cited in Valliath, 2002).

1.   The information was received from M. Poblanz, author of Cognitive Concepts (2003).

Appendix A2.2   Outcome measure in the fluency domain by construct

Outcome measure	Description
ORAL-J: Early Literacy Achievement Words per Minute	As students read a story, the teacher times their word recognition for one minute. There are three different stories on each ORAL-J test from which three different WPM scores are derived (as cited in author's communication[1]).

1. The information was received from M. Poblanz, author of Cognitive Concepts (2003).

Summary of study findings included in the rating for the alphabetics domain by construct[1]

Outcome measure	Study sample	Sample size (students)	Mean outcome (standard deviation[2])		WWC calculations			
			Earobics® group[3]	Comparison group	Mean difference[4] (Earobics® – comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
Construct: Phonological awareness								
			Cognitive Concepts, 2003 (randomized controlled trial)[8]					
ORAL-J: Blending into Words subtest[9]	Grades K–3	74	17.31 (3.54)	14.86 (4.10)	2.45	0.64	Statistically significant	+24
ORAL-J: Segmenting into Sounds[9]	Grades K–3	74	45.31 (14.31)	35.80 (15.82)	9.51	0.63	Statistically significant	+23
ORAL-J: Rhyming Words[9]	Grades K–3	74	7.16 (5.31)	4.26 (4.36)	2.90	0.59	Statistically significant	+22
			Valliath, 2002 (quasi-experimental design)[8]					
CTOPP: Elision	First grade	30	104.00 (11.98)	97.67 (7.04)	6.33	0.63	ns	+23
CTOPP: Blending Words	First grade	30	105.66 (4.88)	103.33 (9.76)	2.33	0.29	ns	+12
CTOPP: Sound Matching	First grade	30	103.63 (4.58)	95.00 (9.45)	8.63	1.13	Statistically significant	+37
CTOPP: Blending Non-words	First grade	30	111.00 (9.02)	105.33 (10.26)	5.67	0.57	ns	+22
Construct: Letter knowledge								
			Cognitive Concepts, 2003 (randomized controlled trial)[8]					
ORAL-J: Letter Naming[9]	Grades K–3	74	57.49 (18.78)	57.26 (20.63)	0.23	0.01	ns	0
Construct: Phonics								
			Cognitive Concepts, 2003 (randomized controlled trial)[8]					
ORAL-J: Sound of Letters[9]	Grades K–3	74	27.80 (6.89)	26.17 (7.72)	1.63	0.22	ns	+9

(continued)

| | | | Authors' findings from the study | | WWC calculations | | |
| | | | Mean outcome (standard deviation[2]) | | | | |
Outcome measure	Study sample	Sample size (students)	*Earobics*® group[3]	Comparison group	Mean difference[4] (*Earobics*® − comparison)	Effect size[5]	Statistical significance[6] (at $\alpha = 0.05$)	Improvement index[7]
			Valliath, 2002 (quasi-experimental design)[8]					
WRMT: Word Identification	First grade	30	104.07 (5.16)	100.87 (3.76)	3.2	0.69	ns	+25
WRMT: Word Attack	First grade	30	103.33 (6.18)	101.33 (6.90)	2.0	0.30	ns	+12
Average[10] for alphabetics domain (Cognitive Concepts, 2003)						0.42	ns	+16
Average[10] for alphabetics domain (Valliath, 2002)						0.60	ns	+23
Domain average[10] for alphabetics						0.51	na	+19

ns = not statistically significant

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement index.
2. The standard deviation across all students in each group shows how dispersed the participants' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. The *Earobics*® group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
4. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
5. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
6. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
7. The improvement index represents the difference between the percentile rank of the average student in the intervention condition 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.
8. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the WWC Tutorial on Mismatch. See Technical Details of WWC-Conducted Computations for the formula the WWC used to calculate statistical significance. In the case of Cognitive Concepts (2003) and Valliath (2002), corrections for multiple comparisons were needed, so the significance levels may differ from those reported in the original studies.
9. Means and standard deviations were received through communication with the author.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size.

| Outcome measure | Study sample | Sample size (students) | Authors' findings from the study[2] | | WWC calculations | | | |
| | | | Mean outcome (standard deviation[3]) | | | | | |
			Earobics® group[4]	Comparison group	Mean difference[5] (Earobics® – comparison)	Effect size[6]	Statistical significance[7] (at $\alpha = 0.05$)	Improvement index[8]
			Cognitive Concepts, 2003 (randomized control trial)[9]					
ORAL-J: Words per Minute 1	Grades K–3	74	39.21 (22.95)	35.49 (26.32)	3.72	0.15	ns	+6
ORAL-J: Words per Minute 2	Grades K–3	74	34.11 (25.91)	31.63 (33.64)	2.48	0.08	ns	+3
ORAL-J: Words per Minute 3	Grades K–3	74	36.70 (27.35)	33.86 (32.02)	2.84	0.10	ns	+4
Domain average[10] for fluency (Cognitive Concepts, 2003)						0.11	ns	+4

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
2. Means and standard deviations were received through communication with the author.
3. 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.
4. The Earobics® group mean equals the comparison group mean plus the mean difference. The computation of the mean difference took into account the pretest difference between the study groups.
5. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
6. For an explanation of the effect size calculation, see Technical Details of WWC-Conducted Computations.
7. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
8. 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.
9. 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 Cognitive Concepts (2003), a correction for multiple comparisons was needed, so the significance levels may differ from those reported in the original study.
10. The WWC-computed average effect sizes for each study and for the domain across studies are simple averages rounded to two decimal places. The average improvement indices are calculated from the average effect size. For a single study included in the fluency domain, the study average is equal to domain average.

Appendix A4.1 *Earobics*® rating for the alphabetics 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 alphabetics, the WWC rated *Earobics*® as having positive effects. The remaining ratings (potentially positive effects, mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *Earobics*® was assigned the highest applicable rating.

Rating received

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

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

  Met. Two studies showed statistically significant positive effects, and one study had 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.

---

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   *Earobics*® rating for the fluency 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 fluency, the WWC rated *Earobics*® 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.** No studies show statistically significant positive effects.

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 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.

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.

(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.

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 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.

AND

- 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	2	4	104	Small
Fluency	1	1	74	Small
Comprehension	0	0	0	na
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."