

# Appendix

## Appendix A1 Study characteristics: Huffstetter, 2005 (randomized controlled trial)

Characteristic	Description
<b>Study citation</b>	Huffstetter, M. (2005). <i>The effects of an Internet-based program on the early reading and oral language skills of at-risk preschool students and their teachers' perceptions of the program</i> . Unpublished doctoral dissertation, University of South Florida, Tampa. (68813195)
<b>Participants</b>	This investigation was conducted with 4-year-old children in two Head Start preschool centers in a city on the east coast of Florida. The two Head Start centers were randomly chosen from the five Head Start centers in the Florida city. Parental consent was obtained for 62 children, who were randomly assigned to either the treatment group or the comparison group using a table of random numbers. There were 31 children in the experimental group and 31 children in the control group. In the sample, 84 percent of the children were African-American, 55 percent were male, and 52 percent spoke English as a second language.
<b>Setting</b>	This study took place in a city on the east coast of Florida.
<b>Intervention</b>	The experimental group received 30 minutes of daily instruction in the <i>Headsprout Reading Basics</i> <sup>TM</sup> program for an 8-week period. The computers were housed in the mobile computer lab.
<b>Comparison</b>	The control group received 30 minutes of daily instruction in <i>Millie's Math House</i> <sup>®</sup> for an 8-week period. <i>Millie's Math House</i> <sup>®</sup> is software that uses cartoon characters to build math skills, such as counting, addition, and subtraction.
<b>Primary outcomes and measurement</b>	Children were pre- and posttested on two tests; one on oral language competency and the other on print knowledge. For a more detailed description of these outcome measures, see Appendices A2.1–2.2.
<b>Staff/teacher training</b>	Prior to the intervention, the study's principal investigator trained teachers and assistant teachers on two separate days at the two sites. The training consisted of oral explanations, modeling, and guided teacher practice. Teachers also were given access to the <i>Headsprout Reading Basics</i> <sup>TM</sup> episodes and the <i>Millie's Math House</i> <sup>®</sup> software for review prior to their students reaching each episode. Teachers were trained to respond to technology issues (e.g., volume adjustments), to access and decipher reports, and to intervene and redirect (i.e., use a minimum amount of gesturing or gentle physical guidance to return student to engagement in task) when necessary. For reference purposes, teachers and teachers' assistants also were given a copy of the implementation checklists that were used to monitor implementation integrity.

## Appendix A2.1 Outcome measure for the oral language domain

Outcome measure	Description
<b>Test of Language Development–Primary: 3rd Edition</b>	A standardized measure with six core subtests that measure semantics and syntax and three supplemental subtests that measure phonology (as cited in Huffstetter, 2005).

## Appendix A2.2 Outcome measure for the print knowledge domain

Outcome measure	Description
<b>Test of Early Reading Ability–3rd Edition</b>	A standardized measure of children’s developing reading skills with three subtests: alphabet (measuring knowledge of the alphabet and its uses), conventions (measuring knowledge and conventions of print), and meaning (measuring the construction of meaning through print) (as cited in Huffstetter, 2005). <sup>1</sup>

1. By name, this measure sounds like it should be captured under the early reading and writing domain; however, the description of the measure identifies constructs that are pertinent to print knowledge, such as knowing the alphabet, understanding print conventions, and environmental print.

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

Outcome measure	Study sample	Sample size (centers/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> (Headsprout® Early Reading – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
			Headsprout® Early Reading group <sup>3</sup>	Comparison group <sup>3</sup>				
<b>Huffstetter, 2005<sup>8</sup></b>								
Test of Language Development–Primary: 3rd Edition	4-year-olds	2/62	11.00 (15.00)	2.29 (15.00)	8.71	0.57	Statistically significant	+22
<b>Average for oral language domain (Huffstetter, 2005)<sup>9</sup></b>						<b>0.57</b>	<b>na</b>	<b>+22</b>

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the oral language 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. The reported means are average gain scores. The author only reported the standard deviations for the gain scores, and thus, the WWC used the standard deviation based on the standardized score of the nationally normed sample.
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 favorable results for 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. For the *Headsprout® Early Reading* study summarized here, 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 domain improvement index is calculated from the average effect size.

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

Outcome measure	Study sample	Sample size (centers/students)	Authors' findings from the study		WWC calculations			
			Mean outcome (standard deviation) <sup>2</sup>		Mean difference <sup>4</sup> (Headsprout® Early Reading – comparison)	Effect size <sup>5</sup>	Statistical significance <sup>6</sup> (at $\alpha = 0.05$ )	Improvement index <sup>7</sup>
		Headsprout® Early Reading group <sup>3</sup>	Comparison group <sup>3</sup>					
<b>Huffstetter, 2005<sup>8</sup></b>								
Test of Early Reading Ability—3rd edition	4-year-olds	2/62	9.55 (15.00)	0.84 (15.00)	8.71	0.57	Statistically significant	+22
<b>Average for print knowledge domain (Huffstetter, 2005)<sup>9</sup></b>						<b>0.57</b>	<b>na</b>	<b>+22</b>

na = not applicable

1. This appendix reports findings considered for the effectiveness rating and the average improvement indices for the print knowledge 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. The reported means are average gain scores. The author only reported the standard deviations for the gain scores, and thus, the WWC used the standard deviation based on the standardized score of the nationally normed sample.
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 favorable results for 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. For the *Headsprout® Early Reading* studies summarized here, 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 domain improvement index is calculated from the average effect size.

## Appendix A4.1 **Headsprout® Early Reading** rating for the oral language 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.<sup>1</sup>

For the outcome domain of oral language, the WWC rated *Headsprout® Early Reading* as potentially positive. The remaining ratings (mixed, no discernible effects, potentially negative and negative) were not considered, as *Headsprout® Early Reading* was assigned the highest applicable rating.

### Rating received

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

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

**Met.** One study showed a statistically significant 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.

**Met.** No studies showed a statistically significant or substantively important negative 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.** Only one study met evidence standards.

#### 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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A4.2 Headsprout® Early Reading rating for the print knowledge 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.<sup>1</sup>

For the outcome domain of print knowledge, the WWC rated *Headsprout® Early Reading* as potentially positive. The remaining ratings (mixed, no discernible effects, potentially negative and negative) were not considered, as *Headsprout® Early Reading* was assigned the highest applicable rating.

### Rating received

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

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

**Met.** One study showed a statistically significant 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.

**Met.** No studies showed a statistically significant or substantively important negative 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.** Only one study met evidence standards.

#### 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. For a complete description, see the WWC Procedures and Standards Handbook, Appendix E.

## Appendix A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence <sup>1</sup>
		Schools	Students	
Oral language	1	2	62	Small
Print knowledge	1	2	62	Small
Phonological processing	0	na	na	na
Early reading and writing	0	na	na	na
Cognition	0	na	na	na
Math	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.”