

What Works Clearinghouse



Headsprout® Early Reading

Program Description¹

*Headsprout® Early Reading*² is an Internet-based supplemental early literacy curriculum consisting of eighty 20-minute animated episodes, the first 40 of which are appropriate for prekindergarten age students. The episodes are designed to teach phonemic awareness, phonics, fluency, vocabulary, and comprehension.

The program adapts to a child’s responses, providing additional instruction and review if a child does not choose the correct answer. Teachers may use stories based on the episodes to reinforce instruction provided in the lessons.

Research³

One study of *Headsprout® Early Reading* meets What Works Clearinghouse (WWC) evidence standards and no studies meet WWC evidence standards with reservations. This study included 62 preschool children across five classrooms in two Head Start centers in Florida.⁴

Based on this study, the WWC considers the extent of evidence for *Headsprout® Early Reading* to be small for oral language and print knowledge. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Headsprout® Early Reading* in the phonological processing, early reading and writing, cognition, and math domains.

Effectiveness

Headsprout® Early Reading was found to have potentially positive effects on oral language and print knowledge.

	Oral language	Print knowledge	Phonological processing	Early reading and writing	Cognition	Math
Rating of effectiveness	Potentially positive	Potentially positive	na	na	na	na
Improvement index	+22 percentile points	+22 percentile points	na	na	na	na

na = not applicable

1. The descriptive information for this program was obtained from a publicly available source: the program’s website (www.headsprout.com/, downloaded August 2009). The WWC requests developers to review the program description sections for accuracy from their perspective. Further verification of the accuracy of the descriptive information for this program is beyond the scope of this review.
2. The evidence presented in this report is applicable only to the first 40 episodes of *Headsprout® Early Reading*.
3. The studies in this report were reviewed using WWC Evidence Standards, Version 2.0 (see the WWC Procedures and Standards Handbook, Chapter III).
4. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.

Absence of conflict of interest

Effectiveness of reading and mathematics software products: Findings from two student cohorts (Campuzano et al., 2009), which was prepared by staff of Mathematica Policy Research,

was determined ineligible for review, and thus presented no conflict of interest.

Additional program information

Developer and contact

Developed and distributed by Headsprout®. Address: 127 Broadway Ave E., Suite #200, Seattle, WA 98102. Email: schoolinfo@headsprout.com. Web: www.headsprout.com/school/products/basics.cfm. Telephone: (206) 329-3660.

Scope of use

Headsprout® Early Reading was developed in 1999 and is currently being used by school districts in almost every state across the country, as well as numerous locations overseas.

Teaching

Headsprout® Early Reading is a supplemental early literacy curriculum accessed via the Internet. The prekindergarten curriculum—which emphasizes phonemic awareness, phonics, fluency, vocabulary, and comprehension—is made up of forty 20-minute animated episodes (the first half of the 80 episode K-2 curriculum), 30 stories, and 100 printable flashcards. Animated cartoon characters guide children through interactive episodes in locations such as outer space, under the sea, or the land of the dinosaurs. Children use the mouse to navigate through the episode; for example, helping a worm get home by

identifying, from among four pairs of letters, the letters that represent a sound they learned. The worm moves closer to his hole with each correct answer. The curriculum provides individualized, adaptive instruction, and children work through the lessons at their own pace. The program responds to a child's pattern of errors with tutorials and reviews to provide extra assistance to children struggling to comprehend the material. Children must meet specific performance criteria in order to progress to the next lesson. Cumulative review is built into the curriculum to help ensure retention. Printed versions of stories in the episodes are found in six Headsprout® Readers. The stories only contain material that children have learned up to that point in the curriculum. The Readers serve to reinforce the skills taught during the series and provide children with the opportunity to practice basic reading. The program generates performance reports, allowing teachers to monitor their students' progress.

Cost

A subscription must be purchased for each student and can be carried over into the following year until the student completes the entire program. Discounts are available for larger purchases. No other cost information was provided.

Research

One study reviewed by the WWC investigated the effects of *Headsprout® Early Reading*. This study (Huffstetter, 2005) is a randomized controlled trial that meets WWC evidence standards. Eleven studies do not meet either WWC evidence standards or eligibility screens. One study uses a single subject design for which the WWC is currently developing standards and, therefore, could not be reviewed at this time.

Huffstetter (2005) conducted a randomized controlled trial with 62 children in two Head Start centers in Florida. Thirty-one children were randomly assigned to an experimental group, and the other 31 were assigned to the comparison group. The experimental group received instruction through the *Headsprout Reading Basics™* program⁵, whereas the comparison group received instruction through *Millie's Math House®*, which teaches

5. Previously, the first 40 episodes of *Headsprout® Early Reading* were called *Headsprout Reading Basics™*.

Research *(continued)*

mathematic principles. The majority of children in the study (nearly 84%) were African-American.

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or medium to large (see the WWC Procedures and Standards Handbook, Appendix G). The extent of evidence takes into account the number of studies and the total sample size across

the studies that meet WWC evidence standards with or without reservations.⁶ The WWC considers the extent of evidence for *Headsprout® Early Reading* to be small for oral language and print knowledge. No studies that meet WWC evidence standards with or without reservations examined the effectiveness of *Headsprout® Early Reading* in the domains of phonological processing, early reading and writing, cognition, and math.

Effectiveness Findings

The WWC review of interventions for early childhood education addresses children's outcomes in six domains: oral language, print knowledge, phonological processing, early reading and writing, cognition, and math. The studies included in this report cover two domains: oral language and print knowledge. The findings below present the authors' estimates and WWC-calculated estimates of the size and the statistical significance of the effects of *Headsprout® Early Reading* on children.⁷

Oral language. Huffstetter (2005) analyzes the differences between *Headsprout® Early Reading* and comparison groups for one measure in the oral language domain, the Test of Language Development–Primary: 3rd Edition (TOLD–P:3). The authors report, and the WWC confirms, a statistically significant positive effect of *Headsprout® Early Reading* on oral language of children in this study.

Print knowledge. Huffstetter (2005) analyzes the differences between *Headsprout® Early Reading* and comparison groups for one measure in the print knowledge domain, the Test of Early Reading Ability–3rd Edition (TERA–3). The authors report, and the WWC confirms, a statistically significant positive effect of *Headsprout® Early Reading* on print knowledge of children in this study.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings, the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Procedures and Standards Handbook, Appendix E).

The WWC found *Headsprout® Early Reading* to have potentially positive effects on oral language and print knowledge

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see WWC

Procedures and Standards Handbook, Appendix F). The improvement index represents the difference between the percentile rank of the average student in the intervention condition and the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement

6. The extent of evidence categorization was developed to tell readers how much evidence was used to determine the intervention rating, focusing on the number and size of studies. Additional factors associated with a related concept—external validity, such as the students' demographics and the types of settings in which studies took place—are not taken into account for the categorization. Information about how the extent of evidence rating was determined for *Headsprout® Early Reading* is in Appendix A5.
7. 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, 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.

index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results for the intervention group.

The improvement index is +22 percentile points for one finding in the oral language domain and +22 percentile points for one finding in the print knowledge domain.

Summary

The WWC identified 13 studies on *Headsprout*[®] *Early Reading* and reviewed 12 of them.⁸ One of these studies meets WWC evidence standards; the remaining 11 studies do not meet either WWC evidence standards or eligibility screens. Based on the one study, the WWC found potentially positive effects on oral language and print knowledge. The conclusions presented in this report may change as new research emerges.

References **Meets WWC evidence standards**

Huffstetter, M. (2005). *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*. Unpublished doctoral dissertation, University of South Florida, Tampa. (68813195)

Studies that fall outside the Early Childhood Education review protocol or do not meet WWC evidence standards

Budlong Elementary, Los Angeles Unified School District. (2007). *Results count: Outcome data and case studies* (pp. 3–6).

Seattle, WA: Headsprout. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Campuzano, L., Dynarski, M., Agodini, R., & Rall, K. (2009). *Effectiveness of reading and mathematics software products: Findings from two student cohorts* (NCEE 2009-4041). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Additional source:

Dynarski, M., Agodini, R., Heaviside, S., Novak, T., Carey, N., Campuzano, L., Means, B., Murphy, R., Penuel, W., Javitz, H.,

Emery, D., & Sussex, W. (2007). *Effectiveness of reading and mathematics software products: Findings from the first student cohort*. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Clarfield, J. (2006). Examining the efficacy of two computerized reading programs for kindergarten students at-risk for reading and behavior problems. (Doctoral dissertation, University of Massachusetts Amherst, 2006). *Dissertation Abstracts International*, 67, 05A. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Clarfield, J., & Stoner, G. (2005). The effects of computerized reading instruction on the academic performance of students identified with ADHD. *School Psychology Review*, 34(2), 246–254. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2003). *Headsprout Early Reading*[™]: Reliably teaching children to read. *Behavioral Technology Today*, 3, 7–20. The study is ineligible for review because it does not use a comparison group.

Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2004). Selected for success: How *Headsprout Reading Basics*[™] teaches beginning reading. In D. J. Moran, & R. W. Malott (Eds.), *Evidence-based educational methods* (pp. 171–197). San Diego, CA: Elsevier Academic Press. The study does not meet WWC

8. A single-case design study was identified but is not included in this review because the WWC does not yet have standards for reviewing regression discontinuity or single-case design studies.

References *(continued)*

evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

Learners using *Headsprout* at home. (2007). *Results count: Outcome data and case studies* (pp. 11–12). Seattle, WA: Headsprout. The study is ineligible for review because it does not examine an intervention implemented in a way that falls within the scope of the review.

NY students reading above grade level with *Headsprout*. (2007). *Results count: Outcome data and case studies* (pp. 7). Seattle, WA: Headsprout. The study does not meet WWC evidence standards because it uses a quasi-experimental design in which the analytic intervention and comparison groups are not shown to be equivalent.

PS 106, New York Public Schools. (2007). *Results count: Outcome data and case studies* (pp. 1–2). Seattle, WA: Headsprout. The study is ineligible for review because it does not use a sample within the age or grade range specified in the protocol.

Vail Unified School District, Vail, Arizona. (2007). *Results count: Outcome data and case studies* (pp. 8–9). Seattle, WA: Headsprout. The study is ineligible for review because it does not use a comparison group.

Woodcock Johnson Letter-Word Identification subtest pre- and posttest scores. (2007). *Results count: Outcome data and case studies* (pp. 10). Seattle, WA: Headsprout. The study is ineligible for review because it does not use a comparison group.

Studies with disposition pending

Layng, T. V. J., Twyman, J. S., & Stikeleather, G. (2004). Engineering discovery learning: The contingency adduction of some precursors of textual responding in a beginning reading program. *Analysis of Verbal Behavior, 20*, 99–109. The study is not included because it uses a design for which the WWC is currently developing standards.

Appendix

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

Characteristic	Description
Study citation	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)
Participants	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.
Setting	This study took place in a city on the east coast of Florida.
Intervention	The experimental group received 30 minutes of daily instruction in the <i>Headsprout Reading Basics</i> TM program for an 8-week period. The computers were housed in the mobile computer lab.
Comparison	The control group received 30 minutes of daily instruction in <i>Millie's Math House</i> [®] for an 8-week period. <i>Millie's Math House</i> [®] is software that uses cartoon characters to build math skills, such as counting, addition, and subtraction.
Primary outcomes and measurement	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.
Staff/teacher training	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> TM episodes and the <i>Millie's Math House</i> [®] 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
Test of Language Development–Primary: 3rd Edition	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
Test of Early Reading Ability–3rd Edition	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). ¹

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 domain¹

Outcome measure	Study sample	Sample size (centers/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation) ²		Mean difference ⁴ (Headsprout® Early Reading – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			Headsprout® Early Reading group ³	Comparison group ³				
Huffstetter, 2005⁸								
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
Average for oral language domain (Huffstetter, 2005)⁹						0.57	na	+22

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 domain¹

Outcome measure	Study sample	Sample size (centers/ students)	Authors' findings from the study			WWC calculations		
			Mean outcome (standard deviation) ²		Mean difference ⁴ (Headsprout® Early Reading – comparison)	Effect size ⁵	Statistical significance ⁶ (at $\alpha = 0.05$)	Improvement index ⁷
			Headsprout® Early Reading group ³	Comparison group ³				
Huffstetter, 2005⁸								
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
Average for print knowledge domain (Huffstetter, 2005)⁹						0.57	na	+22

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.¹

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.¹

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 ¹
		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.”