What Works Clearinghouse

INSTITUTE OF EDUCATION SCIENCES

Beginning Reading August 13, 2007

Voyager Universal Literacy System®

Program description[1]

The *Voyager Universal Literacy System*® is a core reading program designed to help students learn to read at or above grade level by the end of the third grade. This program uses strategies such as individual reading instruction, higher-level comprehension activities, problem solving, and writing. Students are also exposed to computer-based practice and reinforcement in phonological skills, comprehension, fluency, language development, and writing. The program uses whole classroom, small group, and independent group settings. *Voyager Universal Literacy System*® emphasizes regular assessments, with biweekly reviews for struggling students and quarterly assessments for all students.

Research

Two studies of *Voyager Universal Literacy System*® met WWC evidence standards with reservations. The two studies included over 600 kindergarten students from Florida, Ohio, and Washington, DC.[2] The WWC considers the extent of evidence for *Voyager Universal Literacy System*® to be moderate to large for alphabetics and small for comprehension. No studies that met WWC evidence standards with or without reservations addressed fluency or general reading achievement.

Effectiveness

Voyager Universal Literacy System® was found to have potentially positive effects on alphabetics and potentially negative effects on comprehension.

	Alphabetics	*Fluency*	*Comprehension*	*General reading achievement*
Rating of effectiveness	Potentially positive	na	Potentially negative	na
Improvement index[3]	Average: +11 percentile points Range: –8 to +27 percentile points	na	Average: –25 percentile points	na

na = not applicable

1. The descriptive information for this program was obtained from publicly available sources: the program's website (www.voyagerlearning.com; downloaded April 2007) and the research literature (Frechtling, Zhang, & Silverstein, 2006). 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 based on available research. Findings and conclusions may change as new research becomes available.
3. These numbers show the average and range of student-level improvement indices for all findings across the studies.

Additional program information[1]

Developer and contact

Voyager Universal Literacy System® was developed by Sharon Vaughn, Ed Kame'ennui, Deborah Simmons, Roland Good, and Jeri Nowakowski. *Voyager Universal Literacy System*® is a program of *Voyager Expanded Learning* and is owned and distributed by Pro Quest Education. Address: Voyager Expanded Learning, One Hickory Centre, 1800 Valley View Suite 400, Dallas, TX 75234-8923. Web: www.voyagerlearning.com. Telephone: (888) 399-1995.

Scope of use

The program was first published in 2000. According to the developer, *Voyager Universal Literacy System*® has been implemented with students reading at all levels, including students who receive special education services. Since 2002, *Voyager Universal Literacy System*® has been used in 360 districts in 22 states across the US. Almost 17, 500 teachers and over 331,000 students have used the program.

Teaching

Sequenced lessons provide the teachers with tools and directions for instruction and assessment. Classroom activities include read-alouds directed by the teacher and students reading at different levels in whole group, small group, and independent settings. The program also involves computer-based practice in phonological skills, comprehension, fluency, writing, and language development. *Voyager Universal Literacy System*® involves use of a progress monitoring system four times a year to determine if any students are struggling. Struggling readers are provided with 10–20 minutes of supplemental in-school instruction and, if they continue to struggle, have the option of enrolling in an 80-hour summer reading intervention program. The program has a home study curriculum with 15-minute activities to use with parents. In addition, each child receives a take-home library to initiate the child's own book collection. The *Voyager Universal Literacy System*® program utilizes ongoing professional development and school-based reading coaches.

Cost

The cost for *Voyager Universal Literacy System*® is $244 per student for the first year and $160 for subsequent years. This includes curriculum materials—student books, home study guides, and assessment record sheets for each grade level, as well as daily lesson plans and teacher training materials, teacher's guides for reading intervention and enrichment activities, a classroom management packet, a literature library, and a teacher supply pack with manipulatives, CDs, puppets, games, and additional materials. Other elements of the program include a progress monitoring system with an online data management system, a *Struggling Reader Intervention*, and summer *Advanced Reader Modules* programs. The cost also includes initial teacher and reading coach training, done on-site. Costs can vary based on which elements are selected. Further training kits (using videos and tutorials) are also available.

Research

Seven studies reviewed by the WWC investigated the effects of the *Voyager Universal Literacy System*®. Two studies (Frechtling, Zhang, & Silverstein, 2006; Hecht, 2003) were quasi-experimental designs that met WWC evidence standards with reservations. The remaining five studies did not meet WWC evidence screens.

Frechtling, Zhang, & Silverstein (2006) included 447 kindergarten students in eight schools. Students in the intervention schools used *Voyager Universal Literacy System*® for two hours a day and students in the comparison schools used only their schools' existing curriculum. In the final analysis sample 202 intervention students were compared with 196 comparison students. The two groups scored similarly on achievement pretests after attrition.

Research *(continued)*

Hecht (2003) included 213 students in four low-income schools.[4] Students in the intervention schools used *Voyager Universal Literacy System*® as their daily reading program. Students in the comparison schools used their schools' existing curriculum. The two groups scored similarly on achievement pretests after attrition.

Extent of evidence

The WWC categorizes the extent of evidence in each domain as small or moderate to large (see the What Works Clearinghouse Extent of Evidence Categorization Scheme). The extent of evidence takes into account the number of studies and the total sample size across the studies that met WWC evidence standards with or without reservations.[5]

The WWC considers the extent of evidence for *Voyager Universal Literacy System*® to be moderate to large for alphabetics and small for comprehension. No studies that met WWC evidence standards with or without reservations addressed fluency or general reading achievement.

Effectiveness

Findings

The WWC review of interventions for beginning reading addresses student outcomes in four domains: alphabetics, fluency, comprehension, and general reading achievement.[6] The studies included here cover outcomes in alphabetics and comprehension. Within alphabetics, results for four constructs are reported: phonological awareness, print awareness, letter knowledge, and phonics. The findings below present the authors' estimates and WWC-calculated estimates of the size and statistical significance of the effects of *Voyager Universal Literacy System*® on students.[7]

Alphabetics

Phonological awareness. Frechtling, Zhang, and Silverstein (2006) reported positive, but not statistically significant, effects on the four phonological awareness measures (Comprehensive Test of Phonological Processing (CTOPP) Elision, Blending Words, Blending Nonwords, and Segmentation subtests).

Hecht (2003) examined effects for three phonological awareness measures (Blending, CTOPP Elision, and CTOPP Segmentation) and found a positive and statistically significant effect on the CTOPP Segmentation subtest. None of these effects were statistically significant according to the WWC analysis.

Letter Knowledge. Frechtling, Zhang, and Silverstein (2006) found a positive, but not statistically significant effect on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Letter Naming Fluency subtest.

Hecht (2003) reported a positive and statistically significant effect on using a researcher-designed measure of letter naming

4. The study originally included 429 students and was designed to examine outcomes for intervention and comparison students within and between schools. However, data on the within-school comparisons was not reported in the study due to what the study authors called poor implementation of the intervention at the schools used for the within-school comparisons. The WWC typically considers the success of implementation of the intervention to be part of the effect of the intervention and reports on study findings regardless of implementation. However, data for the within-schools comparisons were not presented and the WWC cannot report on the effectiveness of the intervention for this portion of the study.
5. 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.
6. For definitions of the domains, see the Beginning Reading Protocol.
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 all studies of the *Voyager Universal Literacy System*®, corrections for clustering and multiple comparisons were needed, so the significance levels differ from those reported in the original studies.

Effectiveness *(continued)*

fluency. According to WWC calculations, the effect was not statistically significant.

Print Awareness. Hecht (2003) reported a negative, but not statistically significant, effect on the Concepts about Print subtest.

Phonics. Frechtling, Zhang, & Silverstein (2006) reported positive, but not statistically significant, effects for the two Woodcock Reading Mastery (WRMT) subtests: Word Identification and Word Attack.

Hecht (2003) reported a statistically significant positive effect for the Letter Sounds test. However, this outcome was not statistically significant according to the WWC analysis. The study found negative effects on the DIBELS Nonsense Word Fluency and Woodcock Word Identification subtests and a positive effect on the Woodcock Reading Mastery Test-Revised (WRMT-R) Word Attack subtest, but none of the effects were statistically significant.

Across all constructs in the alphabetics domain, the average effect size in Frechtling, Zhang, & Silverstein (2006) was positive and large enough to be considered substantively important according to the WWC criteria (that is, at least 0.25). The average effect size for Hecht (2003) was positive, but not large enough to be considered substantively important.

Comprehension

Hecht (2003) reported a negative and statistically significant effect on the Stanford-Binet Intelligence Expressive Vocabulary subtest. This outcome was not statistically significant according to the WWC analysis but the effect size was large enough to be substantively important (that is, smaller than –0.25).

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,[8] the size of the difference between participants in the intervention and the comparison conditions, and the consistency in findings across studies (see the WWC Intervention Rating Scheme).

The WWC found *Voyager Universal Literacy System®* to have potentially positive effects on alphabetics and potentially negative effects on comprehension

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 Technical Details of WWC-Conducted Computations). 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. Unlike the rating of effectiveness, the improvement index is based entirely on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analyses. The improvement index can take on values between –50 and +50, with positive numbers denoting results favorable to the intervention group.

The average improvement index for alphabetics is +11 percentile points across the two studies, with a range of –8 to +27 percentile points across findings. The improvement index for comprehension is –25 for the one outcome studied.

Summary

The WWC reviewed seven studies on *Voyager Universal Literacy System®*. Two of these studies met WWC evidence standards with reservations; the remaining studies did not meet WWC evidence screens. Based on these two studies, the WWC found potentially positive effects on alphabetics and potentially negative effects on comprehension. Evidence presented in this report may change as new research emerges.

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, see the WWC Tutorial on Mismatch. See the Technical Details of WWC-Conducted Computations for the formulas the WWC used to calculate the statistical significance. In the case of *Voyager Universal Literacy System®,* corrections for clustering and multiple comparisons were needed.

References

Met WWC evidence standards with reservations

Frechtling, J. A., Zhang, X., Silverstein, G. (2006). The Voyager Universal Literacy System: Results from a study of kindergarten students in inner-city schools. *Journal of Education for Students Placed At-Risk*, *11*(1), 75–95.

Additional sources:

Frechtling, J., Silverstein, G., & Zhang, X. (2003). *Evaluation of the Voyager Universal Literacy System*. Retrieved from Voyager Expanded Learning© Web site: http://www.voyagerlearning.com/docs/difference/report_studies/Westat.pdf

Frechtling, J., Zhang, X., & Wang, L. W. (2004). *Evaluation of the Voyager Universal Literacy System: Year 2*. Retrieved from Voyager Expanded Learning© Web site: http://www.voyagerlearning.com/docs/difference/report_studies/WESTAT_Voyager_2004_3.pdf

Hecht, S. A. (2003). *A study between Voyager and control schools in Orange County, Florida 2002–2003*. Retrieved from Voyager Expanded Learning© Web site: http://www.voyagerlearning.com/docs/difference/report_studies/ocps_2002_03.pdf

Did not meet WWC evidence screens

Hecht, S. A., & Torgesen, J. K. (2002). *Within school treatment and control study: Voyager Universal Literacy System: Orange County, Florida 2001–2002*. Retrieved from Voyager Expanded Learning© Web site: http://www.voyagerlearning.com/ResearchStudyDocuments/OrangeCounty_FL_Treatment_Control_Study_2001-2002.pdf[9]

Roberts, G. (2002, June). *Evaluation report on the impact of the Voyager Universal Literacy System in Birmingham City Schools.* Retrieved April 19, 2007, from http://www.voyagerlearning.com/docs/difference/report_studies/Birmingham.pdf[10]

Roberts, G. (2003). *Longitudinal study of the effect of universal literacy: A hierarchical linear modeling analysis of curriculum-based measurement data.* Austin, TX: Evaluation Research Services.[10]

Roberts, G., & Allen, A. S. (2003). *Impact of the Voyager Universal Literacy System as measured by PALS in Virginia*. Retrieved from Voyager Expanded Learning© Web site: http://www.voyagerlearning.com/ResearchStudyDocuments/ULS_measuredby_PALS_Richmond_VA.pdf[11]

Starnes, D., Taylor, D., & Betourne, M. (2004). *Voyager Universal Literacy System second year evaluation report: Fulton County Schools.* Atlanta, GA: EMSTAR Research, Inc.[12]

For more information about specific studies and WWC calculations, please see the WWC *Voyager Universal Literacy System*® Technical Appendices.

9. Incomparable groups: this study was a quasi-experimental design with substantial differences in student and teacher characteristics prior to the start of the intervention.
10. Does not use a strong causal design: this study did not use a comparison group.
11. Incomparable groups: this study was a quasi-experimental design that used achievement pretests, but it did not establish that the comparison group was comparable to the treatment group prior to the start of the intervention.
12. This study was a quasi-experimental design but did not use achievement pretests to establish that the comparison group was equivalent to the intervention group at baseline.

Appendix

Appendix A1.1 Study characteristics: Frechtling, Zhang, and Silverstein, 2006 (quasi-experimental design)

Characteristic	Description
Study citation	Frechtling, J. A., Zhang, X., Silverstein, G. (2006). The Voyager Universal Literacy System: Results from a study of Kindergarten students in inner-city schools. *Journal of Education for Students Placed At-Risk, 11*(1), 75–95.
Participants	The study included 447 Kindergarten students. The final analysis sample included 398 students (202 intervention and 196 comparison students).[1] Over 95% of students were African-American and almost 90% of students qualified for free or reduced price lunch.
Setting	Eight schools from Cleveland, Ohio, and Washington, DC, were included in the study.
Intervention	Students received two hours of the *Voyager Universal Literacy System®* program daily, which included whole group instruction (20 minutes); differentiated, small group instruction, including two student-led independent stations and one teacher-led station (70 minutes); and a teacher-facilitated writing activity (30 minutes). According to study authors, 9 of 11 teachers demonstrated high or moderate fidelity to the intervention and 2 demonstrated low fidelity.
Comparison	The comparison condition used the schools' existing reading program and the teachers were already familiar with the curriculum. The study authors noted that comparison schools used reading activities that explicitly addressed phonemic awareness, phonics, and sight words and that literacy skills were also integrated into other lessons. Small groups were routinely used in literacy instruction. One comparison school had large numbers of students who resided in a homeless shelter or domestic violence center, and another accepted students from out of the typical school boundaries through a lottery. According to study authors, these characteristics may have led to lower and higher parental involvement, respectively.
Primary outcomes and measurement	Measures used for both pretests and posttests include the Comprehensive Test of Phonological Processing (CTOPP) Elision, Blending Words, Blending Nonwords, and Segmenting Words subtests; the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test of Letter Naming Fluency; and the Woodcock Reading Mastery Test Revised (WRMT-R) Word Identification and Word Attack subtests.[2] (See Appendix A2.1–2.2 for more detailed descriptions of outcome measures.)
Teacher training	*Voyager Universal Literacy System®* training includes an initial two-day session for district and campus coaches and a three-day training session for teachers. There were also eight 3-hour professional development modules throughout the school year. In addition, *Voyager Universal Literacy System®* staff periodically observed teachers during the reading block to assess implementation fidelity.

1. This WWC review focuses on the first year of the study, which included findings from Kindergarten. Findings from the second year included 255 students from the original sample who were tested at the end first grade were not included because the study authors did not establish the pretest equivalence of the intervention and comparison groups for this sample.
2. The authors reported other measures that are not included here. The DIBELS Oral Reading Fluency subtest was not given as a pretest, so baseline equivalence could not be established. The Wide Range Achievement Test Letter Writing and Spelling subtests were also administered but are not reported here because they do not fall within the domains of interest to the WWC Beginning Reading topic.

Appendix A1.2 Study characteristics: Hecht, 2003 (quasi-experimental design)

Characteristic	Description
Study citation	Hecht, S. A. (2003). A study between Voyager and control schools in Orange County, Florida 2002–2003. Retrieved from Voyager Expanded Learning Web site: http://www.voyagerlearning.com/docs/difference/report_studies/ocps_2002_03.pdf
Participants	The study included 429 economically disadvantaged Kindergarten students at two intervention and two comparison schools. The initial study design called for analysis of outcomes for intervention and comparison classrooms within schools and across the four schools. However, the study authors did not report findings on the within school comparisons due to poor implementation of the intervention.[1] The analysis sample for the between school comparisons included 213 students. This left 213 students in the between schools study: 101 students in the intervention group and 112 students in the comparison group.4 Over 80% of students were African-American, and approximately 80% qualified for free or reduced price lunches.
Setting	Four schools in Orange County, Florida.
Intervention	The *Voyager Universal Literacy System®* program was used as the core reading program in intervention classrooms for five months. No other information about implementation of the program is given.
Comparison	The two schools in the comparison group used their school's existing curriculum, either *Houghton Mifflin* or *Success for All.* No other information about instruction for the comparison group was given.
Primary outcomes and measurement	Hecht (2003) used the Comprehensive Test of Phonological Awareness (CTOPP) Elision, Segmenting, and Blending subtests as well as the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) test of Nonsense Word Fluency. Letter Name Knowledge, Letter Sound Knowledge, and Concepts about Print measures were also used. In addition, the Woodcock Reading Mastery Test-Revised (WRMT-R) Word Identification and Word Analysis subtests were used as well as the Stanford-Binet Intelligence Scale (4th Edition) Vocabulary subtest. Spelling subtests of the Wide Range Achievement Test were administered, but are beyond the scope of this review. (See Appendix A2.1–2.2 for more detailed descriptions of outcome measures.)
Teacher training	No information was given about teacher training in this study.

1. The WWC typically considers the success of implementation of the intervention as part of the effect of the intervention and reports on study findings regardless of implementation. However, data for the within-schools comparisons were not presented and the WWC cannot report on the effectiveness of the intervention within schools.
2. Post-attrition equivalence on all pretest measures was established by data provided in author communication.

Appendix A2.1 Outcome measures in the alphabetics domain

Outcome measure	Description
Phonological awareness	
Blending	In this researcher-developed test, students combined phonemes to form words. Sounds were given separately and the student was asked to blend them together and identify the word the sounds made. There were 20 items on this test (as cited in Hecht, 2003).
Comprehensive Test of Phonological Processing (CTOPP): Blending Words subtest	This standardized assessment includes 20 items that measured the extent to which the child could combine separately spoken sounds and blend together to form a real word (as cited in Frechtling, Zhang, & Silverstein, 2006; Hecht, 2003).
CTOPP: Blending Nonwords subtest	This standardized assessment includes 18 items that measured the extent to which the child could combine separately spoken sounds and blend together to form a nonsense word (as cited in Frechtling, Zhang, & Silverstein, 2006; Hecht, 2003).
CTOPP: Elision subtest	This is a standardized measure of children's phonological awareness skills. Children were asked to say a word. Then, children were asked what the word would be if a specific phoneme in the word were deleted. The remaining phonemes were used to form a word. There are 20 items on the test (as cited in Frechtling, Zhang, & Silverstein, 2006; Hecht, 2003).
CTOPP: Segmenting Words subtest	This standardized 20-item subtest requires that the student repeat a word and then say the word one sound at a time (as cited in Frechtling, Zhang, & Silverstein, 2006; Hecht, 2003).
Letter knowledge	
Dynamic Indicators of Basic Literacy Skills (DIBELS): Letter Naming Fluency	This is a subtest of a standardized measure in which students are presented with a page of upper- and lower-case letters arranged in a random order and are asked to name as many letters as they can. The score is the number of letters named correctly in one minute (as cited in Frechtling, Zhang, & Silverstein, 2006).
District Letter Name Knowledge	A district measure given to all students designed to measure the total number of randomly placed upper and lower case letter names correctly pronounced (as cited in Hecht, 2003).
Letter Name Knowledge	In this researcher-developed measure, students gave the names of the 26 letters of the alphabet (as cited in Hecht, 2003).
Print awareness	
Concepts about Print test	Students perform tasks related to printed language concepts (for example, directionality and word concepts) while reading a book. This assessment, developed by Clay, is not standardized and is based on 18 questions (as cited in Hecht, 2003).

(continued)

Appendix A2.1 Outcome measures in the alphabetics domain *(continued)*

Outcome measure	Description
Phonics	
DIBELS: Nonsense Word Fluency subtest	This standardized subtest measures children's word reading ability, including letter-sound correspondence and the ability to blend letter sounds into words (as cited in Hecht, 2003).
Letter Sound Knowledge	In this researcher developed test, students indicated the sounds individual letters make in words. Score were out of a possible 38 (as cited in Hecht, 2003).
Woodcock Reading Mastery Test (WRMT): Word Identification subtest	This standardized test measures decoding skills by requiring children to read aloud isolated real words that range in frequency and difficulty (as cited in Frechtling, Zhang, & Silverstein, 2006; Hecht, 2003).
WRMT: Word Attack subtest	This standardized test measures phonemic decoding skills by asking students to read pseudo-words. Students were aware that the words are not real (as cited in Frechtling, Zhang, & Silverstein, 2006; Hecht, 2003).

Appendix A2.2 Outcome measure in the comprehension domain

Outcome measure	Description
Vocabulary	
Stanford Binet Intelligence Scale: Expressive Vocabulary subtest	This standardized subtest measured children's ability to provide names of pictures and definitions of words (as cited in Hecht, 2003).

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain[1]

Outcome measure	Study sample	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) *Voyager* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) Comparison group	WWC calculations: Mean difference[4] (*Voyager* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Phonological Awareness								
Frechtling, Zhang, and Silverstein, 2006 (quasi-experimental design)[8]								
CTOPP: Elision	Kindergarten	8/398	3.47 (3.05)	2.76 (2.83)	0.71	0.24	ns	+10
CTOPP: Blending Words	Kindergarten	8/398	4.89 (3.77)	3.14 (3.43)	1.75	0.48	ns	+19
CTOPP: Blending Nonwords	Kindergarten	8/398	2.67 (2.47)	1.33 (1.97)	1.34	0.60	ns	+22
CTOPP: Segmenting Words	Kindergarten	8/398	3.66 (3.96)	1.35 (2.38)	2.31	0.66	ns	+24
Hecht, 2003 (quasi-experimental design)[8]								
Blending	Kindergarten	4/213	9.90 (5.30)	9.20 (5.50)	0.70	0.13	ns	+5
CTOPP: Elision	Kindergarten	4/213	3.20 (3.20)	3.80 (2.90)	–0.60	–0.20	ns	–8
CTOPP: Segmenting Words	Kindergarten	4/213	7.20 (5.10)	4.60 (3.90)	2.60	0.57	ns	+22
Letter Knowledge								
Frechtling, Zhang, and Silverstein, 2006 (quasi-experimental design)[8]								
DIEBELS: Letter Naming Fluency	Kindergarten	8/398	39.39 (14.20)	35.05 (18.34)	4.34	0.26	ns	+10
Hecht, 2003 (quasi-experimental design)[8]								
Letter Name Knowledge	Kindergarten	4/213	26.20 (2.40)	25.20 (4.90)	1.0	0.25	ns	+10

(continued)

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain[1] *(continued)*

Outcome measure	Study sample	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Voyager* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Voyager* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Print Awareness								
Hecht, 2003 (quasi-experimental design)[8]								
Concepts About Print test	Kindergarten	4/213	12.80 (3.70)	13.50 (5.20)	–0.70	–0.15	ns	–6
Phonics								
Frechtling, Zhang, and Silverstein, 2006 (quasi-experimental design)[8]								
WRMT: Word Identification	Kindergarten	8/398	9.83 (9.83)	8.31 (10.12)	1.52	0.15	ns	+6
WRMT: Word Attack	Kindergarten	8/398	4.73 (5.51)	1.34 (3.27)	3.39	0.74	ns	+27
Hecht, 2003 (quasi-experimental design)[8]								
DIBELS Nonsense Word Fluency	Kindergarten	4/213	29.30 (15.30)	30.60 (19.00)	–1.30	–0.07	ns	–3
Letter Sound Knowledge	Kindergarten	4/213	26.00 (4.50)	23.80 (6.60)	2.2	0.38	ns	+15
WRMT: Word Identification	Kindergarten	4/213	9.40 (10.40)	10.40 (10.30)	–1.00	–0.10	ns	–4
WRMT: Word Attack	Kindergarten	4/213	5.30 (5.50)	4.80 (4.60)	0.5	0.10	ns	+4
Average[9] for alphabetics (Frechtling, Zhang, and Silverstein, 2006)						0.45	ns	+17
Average[9] for alphabetics (Hecht, 2003)						0.10	ns	+4
Domain average[9] for alphabetics						0.28	na	+11

ns = not statistically significant
na = not applicable

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. Standard deviations for Frechtling, Zhang, & Silverstein (2006) and Hecht (2003) were provided in author communications.
3. The intervention group values for mean outcome performance are the control scores plus the difference in mean gains between the *Voyager* and comparison groups. For Hecht (2003), raw scores were provided by the author.

(continued)

Appendix A3.1 Summary of study findings included in the rating for the alphabetics domain[1] *(continued)*

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 formulas the WWC used to calculate statistical significance. In the case of all studies of the *Voyager Universal Literacy System®*, corrections for clustering and multiple comparisons were needed, so the significance levels differ from those reported in the original studies.
9. 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.

Appendix A3.2 Summary of study findings included in the rating for the comprehension domain[1]

Outcome measure	Study sample	Sample size (schools/ students)	Authors' findings from the study: Mean outcome (standard deviation[2]) — *Voyager* group[3]	Authors' findings from the study: Mean outcome (standard deviation[2]) — Comparison group	WWC calculations: Mean difference[4] (*Voyager* – comparison)	WWC calculations: Effect size[5]	WWC calculations: Statistical significance[6] (at $\alpha = 0.05$)	WWC calculations: Improvement index[7]
Vocabulary								
Hecht, 2003 (quasi-experimental design)[8]								
Stanford Binet: Expressive Vocabulary	Kindergarten	4/213	14.30 (3.60)	17.00 (4.40)	–2.70	–0.67	ns	–25
Domain average[9] for comprehension						–0.67	na	–25

ns = not statistically significant
na = not applicable

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. Standard deviations for Hecht (2003) were provided in author communications.
3. The intervention group values for mean outcome performance are the control scores plus the difference in mean gains between the *Voyager* and comparison 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 formulas the WWC used to calculate statistical significance. In the case of Hecht (2003), corrections for clustering were needed, so the significance levels differ from those reported in the original studies.
9. 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.

Appendix A4.1 *Voyager Universal Literacy System®* 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 *Voyager Universal Literacy System®* as having potentially positive effects. It did not meet the criteria for the positive effects because none of the studies showed statistically positive significant effects or met WWC evidence standards for a strong design. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, negative effects) were not considered, as *Voyager Universal Literacy System®* 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 substantively important 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 a statistically significant or substantively important negative effect and more studies showed positive effects than 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. No studies showed statistically significant positive effects or met WWC evidence standards for a strong design.

AND

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

  Met. There were no statistically significant or substantively important negative effects in the alphabetics domain.

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 *Voyager Universal Literacy System*® 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 *Voyager Universal Literacy System*® as having potentially negative effects. It did not meet the criteria for positive effects, potentially positive effects, mixed effects, or no discernible effects as the one study showed a substantively important negative effect. The remaining rating (negative effects) was not considered, as *Voyager Universal Literacy System*® was assigned the highest applicable rating.

Rating received

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.

  Met. One study showed substantively important negative effects.

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 studies showed statistically significant or substantively important positive effects; one study showed substantively important negative 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. The study did not show statistically significant positive effects and did not meet WWC standards for a strong design.

AND

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

  Not met. One study showed 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. The study did not show statistically significant or substantively important 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.

  Not met. One study showed substantively important negative effects.

(continued)

Appendix A4.2 *Voyager Universal Literacy System*® rating for the comprehension domain *(continued)*

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. One study showed substantively important negative effects.

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. One study showed 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 A5 Extent of evidence by domain

Outcome domain	Number of studies	Sample size		Extent of evidence[1]
		Schools	Students	
Alphabetics	2	12	611	Moderate to large
Fluency	0	0	0	na
Comprehension	1	4	213	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."